



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA MECÂNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA

CLAUDEMIR COSME DA SILVA

**DESEMPENHO DE MÉTODOS DE APRENDIZADO DE MÁQUINA PARA
ESTIMATIVA DE TEMPERATURA DE COLETORES SOLARES COM
NANOFLUIDOS**

FORTALEZA

2022

CLAUDEMIR COSME DA SILVA

DESEMPENHO DE MÉTODOS DE APRENDIZADO DE MÁQUINA PARA ESTIMATIVA
DE TEMPERATURA DE COLETORES SOLARES COM NANOFLUIDOS

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Mecânica da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Engenharia Mecânica. Área de concentração: Processos, Equipamentos e Sistemas para Energias Renováveis.

Orientador: Prof. Dr. Paulo Alexandre Costa Rocha.

FORTALEZA

2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- S579d Silva, Claudemir Cosme da.
Desempenho de métodos de aprendizado de máquina para estimativa de temperatura de coletores solares com nanofluidos / Claudemir Cosme da Silva. – 2022.
57 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Mecânica, Fortaleza, 2022.
Orientação: Prof. Dr. Paulo Alexandre Costa Rocha.
1. Energias Renováveis. 2. Energia Solar. 3. Coletores Solares. 4. Nanofluidos. 5. Aprendizado de Máquina. I. Título.

CDD 620.1

CLAUDEMIR COSME DA SILVA

DESEMPENHO DE MÉTODOS DE APRENDIZADO DE MÁQUINA PARA ESTIMATIVA
DE TEMPERATURA DE COLETORES SOLARES COM NANOFLUIDOS

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Mecânica da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Engenharia Mecânica. Área de concentração: Processos, Equipamentos e Sistemas para Energias Renováveis.

Aprovada em: 02/08/2022.

BANCA EXAMINADORA

Prof. Dr. Paulo Alexandre Costa Rocha (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. André Valente Bueno
Universidade Federal do Ceará (UFC)

Prof. Dr. Francisco Olimpio Moura Carneiro
Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB)

A Deus por mais esta realização.

A minha família pelo incentivo e compreensão
nas horas de ausência.

AGRADECIMENTOS

Primeiramente, agradeço a Deus, pelo dom da vida e por tornar os caminhos difíceis possíveis de serem seguidos.

O desenvolvimento deste trabalho não teria sido possível sem a ajuda e o apoio de pessoas que contribuíram, direta ou indiretamente, então venho agradecer:

A minha família pelo apoio.

Ao Prof. Dr. Paulo Alexandre Costa Rocha, por me orientar durante o trabalho de conclusão de curso, ensinado com paciência e sabedoria.

A todos os professores do Programa de Pós-graduação em Engenharia Mecânica da UFC, com quem aprendi conhecimentos e experiências.

Ao Prof. Dr. André Valente Bueno e Prof. Dr. Francisco Olimpio Moura Carneiro, que prontamente aceitaram o convite para compor a banca, e pelas reflexões, críticas e sugestões recebidas.

RESUMO

As pesquisas sobre sistemas de energias renováveis têm aumentado significativamente, sendo a finalidade encontrar novas fontes de energia que venham a complementar ou até mesmo substituir as provenientes dos combustíveis fósseis, que são consideradas não renováveis e causam impactos ambientais. Se comparada com outras fontes de energia renovável, a solar vem atraindo interesse por ser considerada de fácil acesso e abundante. Neste sentido, o objetivo deste trabalho é analisar o desempenho de modelos de aprendizagem de máquina para estimativa de temperatura em coletores solares. Os experimentos foram realizados no Laboratório de Energia Solar e Gás Natural (LESGN) avaliando o comportamento dos perfis de temperatura de quatro concentrações de nanopartículas de dióxido de titânio (12,5 ppm; 25 ppm; 50 ppm; 75 ppm) no período entre outubro e dezembro de 2018, das 06:00 às 18:00 horas, e de nanopartículas de prata (0,8125 ppm; 1,625 ppm; 3,25 ppm; 6,5 ppm) entre junho e outubro de 2019, das 5:30 às 21:30 horas, dispersas em água deionizada em um sistema de absorção direta de radiação solar. A correlação entre os preditores se baseia nas variáveis do modelo físico que se mostram mais relevantes. Os dados de entrada da modelagem numérica são: temperatura no instante de tempo posterior, temperatura ambiente, temperatura média do nanofluido, irradiância solar direta, irradiância solar global, velocidade do vento e ângulo de zênite solar, onde todos eles foram usados para estimar a temperatura final e a variação de temperatura do nanofluido dióxido de titânio dispersas em água deionizada, utilizando sete modelos numéricos: Regressão de Ridge, LASSO, *ElasticNet*, kNN, XGBoost, MARS e Modelos agregados *Ensemble*. Dentre eles, dois modelos foram utilizados como métodos de *Ensemble* (combinados): Ridge e MARS. A linguagem de programação R foi usada para a análise estatística das métricas de erro e classificar a porcentagem de importância dos preditores. Os dados foram separados em conjunto de treinamento e teste, utilizando uma validação cruzada de 10-fold para o conjunto de treinamento, sendo o RMSE utilizado para a escolha do modelo com melhor desempenho. Nas concentrações de nanopartículas de dióxido de titânio, o kNN mostrou o melhor desempenho RMSE para estimativa de variação de temperatura. O menor RMSE foi encontrado na concentração de 75 ppm, com valor de 0,146 °C. Assim, o modelo computacional apresentou uma eficiência 68,56% superior em relação ao modelo físico. O melhor desempenho RMSE nos estudos de nanopartículas de prata foi encontrado na concentração de 1,625 ppm em estimativa de variação de temperatura com o modelo *Ensemble* MARS, com cerca de 0,136 °C.

Palavras-chave: Energias renováveis; Energia solar; Coletores solares; Nanofluidos; Aprendizado de máquina; Caret R package.

ABSTRACT

Research on renewable energy systems has increased significantly, because of the need to find new sources of energy that will complement or even replace fossil fuels, which are considered non-renewable and cause environmental impacts. Compared to other renewable energy sources, solar energy has been attracting interest because it is considered easily accessible and abundant. In this sense, the objective of this work is to analyze the performance of machine learning models for temperature estimation in solar collectors. The experiments were carried out at the Solar Energy and Natural Gas Laboratory (LESGN) evaluating the behavior of the temperature profiles of four concentrations of titanium dioxide nanoparticles (12.5 ppm; 25 ppm; 50 ppm; 75 ppm) in the period between October and December 2018, from 6:00 am to 6:00 pm, and silver nanoparticles (0.8125 ppm; 1.625 ppm; 3.25 ppm; 6.5 ppm) between June and October 2019, from 5:30 am to 9:30 pm, dispersed in deionized water in a solar radiation direct absorption system. The correlation between the predictors is based on the physical model variables that are shown to be most relevant. The input data for the numerical modeling are: temperature at the latest time step, ambient temperature, mean nanofluid temperature, direct solar irradiance, global solar irradiance, wind speed, and solar zenith angle, all of which were used to estimate the final temperature and temperature variation of the titanium dioxide nanofluid dispersed in deionized water using seven numerical models: Ridge Regression, LASSO, ElasticNet, kNN, XGBoost, MARS, and Ensemble Aggregate Models. Among them, two models were used as Ensemble (combined) methods: Ridge and MARS. The R programming language was applied for statistical analysis of error metrics and to rank the importance percentage of the predictors. The data was separated into a training and test set, using a 10-fold cross-validation for the training set, and the RMSE was used to choose the best performing model. At the concentrations of titanium dioxide nanoparticles, kNN showed the best RMSE performance for estimating temperature variation. The lowest RMSE was found at a concentration of 75 ppm, with a value of 0.146 °C. Thus, the computational model showed an efficiency 68.56% higher than the physical model. The best RMSE performance in the studies of silver nanoparticles was found for the concentration of 1.625 ppm on the estimation of temperature variation with the Ensemble MARS model, of about 0.136 °C.

Keywords: Renewable energy; Solar energy; Solar collectors; Nanofluids; Machine learning; Caret R package.

LISTA DE FIGURAS

Figura 1 – Foto do aparato de teste - "parede solar". Nesta situação específica, a célula superior estava preenchida com nanofluido de TiO_2 e a inferior com água.	21
Figura 2 – Matriz de correlação entre os preditores da modelagem.	24
Figura 3 – Organização do conjunto de dados.	25
Figura 4 – Modelo agregado <i>Ensemble</i>	30

LISTA DE GRÁFICOS

Gráfico 1 – Relação de importância dos preditores em 12,5 ppm (Temperatura final).	33
Gráfico 2 – Relação de importância dos preditores em 12,5 ppm (Variação de temperatura). 34	
Gráfico 3 – Relação de importância dos preditores em 25 ppm (Temperatura final).	35
Gráfico 4 – Relação de importância dos preditores em 25 ppm (Variação de temperatura). ...	36
Gráfico 5 – Relação de importância dos preditores em 50 ppm (Temperatura final).	38
Gráfico 6 – Relação de importância dos preditores em 50 ppm (Variação de temperatura). ...	39
Gráfico 7 – Relação de importância dos preditores em 75 ppm (Temperatura final).	40
Gráfico 8 – Relação de importância dos preditores em 75 ppm (Variação de temperatura). ...	41
Gráfico 9 – Comparação de desempenho do RMSE dos modelos - TiO ₂	42
Gráfico 10 – Relação de importância dos preditores em 0,8125 ppm (Temperatura final).	43
Gráfico 11 – Relação de importância dos preditores em 0,8125 ppm (Variação de temperatura).	44
Gráfico 12 – Relação de importância dos preditores em 1,625 ppm (Temperatura final).	45
Gráfico 13 – Relação de importância dos preditores em 1,625 ppm (Variação de temperatura).	46
Gráfico 14 – Relação de importância dos preditores em 3,25 ppm (Temperatura final).	47
Gráfico 15 – Relação de importância dos preditores em 3,25 ppm (Variação de temperatura).	48
Gráfico 16 – Relação de importância dos preditores em 6,5 ppm (Temperatura final).	49
Gráfico 17 – Relação de importância dos preditores em 6,5 ppm (Variação de temperatura). 50	
Gráfico 18 – Comparação de desempenho do RMSE dos modelos - Prata	51

LISTA DE TABELAS

Tabela 1 – TiO ₂ - 12,5 ppm (Temperatura final).	33
Tabela 2 – TiO ₂ - 12,5 ppm (Variação de temperatura).	34
Tabela 3 – TiO ₂ - 25 ppm (Temperatura final).	35
Tabela 4 – TiO ₂ - 25 ppm (Variação de temperatura).	36
Tabela 5 – TiO ₂ - 50 ppm (Temperatura final).	37
Tabela 6 – TiO ₂ - 50 ppm (Variação de temperatura).	38
Tabela 7 – TiO ₂ - 75 ppm (Temperatura final).	40
Tabela 8 – TiO ₂ - 75 ppm (Variação de temperatura).	41
Tabela 9 – Prata - 0,8125 ppm (Temperatura final).	43
Tabela 10 – Prata - 0,8125 ppm (Variação de temperatura).	44
Tabela 11 – Prata - 1,625 ppm (Temperatura final).	45
Tabela 12 – Prata - 1,625 ppm (Variação de temperatura).	46
Tabela 13 – Prata - 3,25 ppm (Temperatura final).	47
Tabela 14 – Prata - 3,25 ppm (Variação de temperatura).	48
Tabela 15 – Prata - 6,5 ppm (Temperatura final).	49
Tabela 16 – Prata - 6,5 ppm (Variação de temperatura).	50

LISTA DE ABREVIATURAS E SIGLAS

DASC	<i>Direct Absorption Solar Collector</i> (Coletor solar de absorção direta)
kNN	<i>K-Nearest Neighbors</i> (K-Vizinhos mais Próximos)
LASSO	<i>Least Absolute Shrinkage and Selection Operator</i> (Menor Redução Absoluta e Operador de Seleção)
LESGN	Laboratório de Energia Solar e Gás Natural
MAE	<i>Mean Absolute Error</i> (Erro Médio Absoluto)
MARS	<i>Multivariate Adaptive Regression Splines</i> (Splines de Regressão Adaptativa Multivariada)
MBE	<i>Mean Bias Error</i> (Erro Médio por Viés)
ppm	Partículas por milhão
RMSE	<i>Root Mean Squared Error</i> (Raiz do Erro Quadrático Médio)
SAR	<i>Specific Absorption Rate</i> (Taxa de absorção específica)
XGBoost	<i>Extreme Gradient Boosting</i> (Gradiente de Aumento Extremo)

SUMÁRIO

1	INTRODUÇÃO.....	14
2	FUNDAMENTAÇÃO TEÓRICA.....	16
2.1	Energias renováveis e energia solar	16
2.2	Nanofluidos	16
2.3	Sistemas térmicos utilizando nanofluidos	18
2.4	Modelagem.....	18
2.5	Aprendizado de máquina	19
3	MATERIAIS E MÉTODOS.....	21
3.1	Descrição do aparato de testes	21
3.2	Modelo físico aplicado à parede solar.....	22
3.3	Correlação entre os preditores.....	24
3.4	Modelos Numéricos	25
3.4.1	Regressão de Ridge	25
3.4.2	LASSO.....	27
3.4.3	ElasticNet.....	27
3.4.4	kNN.....	28
3.4.5	XGBoost.....	28
3.4.6	MARS	29
3.4.7	Modelo agregado Ensemble.....	30
3.4.8	Métricas de erros	31
4	RESULTADOS E DISCUSSÕES.....	32
4.1	TiO ₂ – 12,5 ppm	32
4.2	TiO ₂ – 25 ppm	34
4.3	TiO ₂ – 50 ppm	37
4.4	TiO ₂ – 75 ppm	39
4.5	Prata – 0,8125 ppm.....	42
4.6	Prata – 1,625 ppm.....	44
4.7	Prata – 3,25 ppm.....	46
4.8	Prata – 6,5 ppm.....	48
5	CONCLUSÃO	52
	REFERÊNCIAS.....	53

1 INTRODUÇÃO

As fontes alternativas de energia têm sido um assunto de importante estudo na contemporaneidade como fator indispensável para o desenvolvimento dos países, tendo em vista as preocupações pela garantia das necessidades energéticas da sociedade. Segundo Khanlari et al. (2020), a energia solar é uma das fontes renováveis mais abundantes do mundo, e está entre as proeminentes devido a ser confiável, acessível, inesgotável e limpa, além de não produzir resíduos perigosos e gases de efeito estufa durante sua utilização.

Modelos computacionais de previsão, podem ser utilizados como componente central de redes inteligentes de sistemas solares térmicos, são usados para analisar características e desempenhos de diferentes tipos de coletores solares, como por exemplo: coletores planos e de tubo evacuado, e diferentes configurações de sistemas. Os modelos podem ser usados para alcançar a máxima eficiência, com custo mínimo em recursos computacionais (AHMAD; REYNOLDS; REZGUI, 2018).

O aprendizado de máquina é um tipo de aplicação de inteligência artificial que ajusta um modelo matemático dependendo dos dados de amostra, garante aos sistemas a capacidade de aprender e, posteriormente, prever os rótulos desconhecidos. Uma melhor previsão necessita de uma seleção de atributos relevantes e do treinamento preciso do algoritmo (AĞBULUT et al., 2020).

Neste estudo, foi realizada a análise computacional do desempenho de modelos de aprendizagem de máquina para estimativa de temperatura em coletores solares, usando as variáveis do modelo físico como preditores do experimento “aparato de teste - parede solar” estudada por Santos (2019), que avaliou o comportamento dos perfis de temperatura de quatro concentrações de nanopartículas de dióxido de titânio (12,5 ppm; 25 ppm; 50 ppm; 75 ppm), e de nanopartículas de prata (0,8125 ppm; 1,625 ppm; 3,25 ppm; 6,5 ppm) entre junho e outubro de 2019, das 5:30 às 21:30 horas, dispersas em água deionizada em um sistema de absorção direta de radiação solar.

A correlação entre os preditores se baseia nas variáveis do modelo físico que se mostram mais relevantes. Os dados de entrada da modelagem numérica são: temperatura no instante de tempo posterior, temperatura ambiente, temperatura média do nanofluido, irradiância solar direta, irradiância solar global, velocidade do vento e ângulo de zênite solar, onde todos eles foram usados para estimar a temperatura do nanofluido.

Foram utilizados no trabalho sete modelos numéricos para estimativa de temperatura e variação de temperatura: Regressão de Ridge, LASSO, *ElasticNet*, kNN,

XGBoost, MARS e Modelos agregados *Ensemble*. Dentre eles, dois modelos foram utilizados como métodos de Ensemble (combinados): Ridge e MARS. A linguagem de programação R foi usada para classificar a porcentagem de importância dos preditores e analisar estatisticamente as métricas de erro: raiz do erro quadrático médio (RMSE), coeficiente de determinação (R^2), erro absoluto médio (MAE) e erro médio por viés (MBE), sendo o RMSE utilizado para a escolha do modelo com melhor desempenho.

Perante o apresentado, o objetivo principal do presente trabalho é analisar o desempenho de modelos de aprendizagem de máquina para estimativa de temperatura em coletores solares. Como objetivos específicos se determinaram:

- Avaliar o comportamento dos perfis de temperatura das quatro concentrações de nanopartículas de dióxido de titânio e prata dispersas em água deionizada, em um sistema de absorção direta de radiação solar;
- Analisar o desempenho dos sete modelos numéricos para estimativa de temperatura e variação de temperatura;
- Comparar as variáveis preditoras com sua porcentagem de importância nos modelos de aprendizado de máquina.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Energias renováveis e energia solar

Segundo Wang et al. (2018), tanto o desenvolvimento como a utilização de novas fontes de energia se tornam imprescindíveis devido à gravidade da escassez de energia e à crise ecológica pela qual passa a sociedade.

O aquecimento global e as questões ambientais contribuíram para o aumento das preocupações acerca desta temática. Tendo em vista a busca pela garantia da manutenção das necessidades de energia da sociedade, o desenvolvimento de energia renovável está se tornando cada vez mais importante, sendo a energia solar considerada de fácil acesso e abundante (AMJAD et al., 2018).

Devido ao menor impacto ambiental, menor custo operacional de exploração e disponibilidade mundial, a energia solar é considerada uma das fontes renováveis mais populares quando se fala de energia sustentável (GUPTA; AGRAWAL; MATHUR, 2015).

Nas últimas décadas, um dos objetivos de cientistas e tecnólogos são as pesquisas para encontrar novas fontes de energia. Principalmente, isso se torna evidente ao se falar de energias renováveis como uma perspectiva desejada para o futuro. Em comparação com outras fontes de energia renovável, a solar vem atraindo interesses (BAZDIDI-TEHRANI; KHABAZIPUR; VASEFI, 2018).

2.2 Nanofluidos

Consideram-se como nanofluidos uma nova classe de fluidos projetados pela dispersão de materiais nanométricos em fluidos base, a exemplo de nanopartículas, nanofibras, nanotubos, entre outros. Mais especificamente, são suspensões coloidais em nanoescala que contêm nanomateriais condensados. Em relação aos fluidos à base de óleo ou água, os nanofluidos possuem propriedades termofísicas aperfeiçoadas (condutividade térmica, difusividade térmica, viscosidade e coeficientes de transferência de calor por convecção) (ALI; TEIXEIRA; ADDALI, 2018; YU; XIE, 2012).

A absorção solar direta baseada em nanopartículas é uma tecnologia promissora para futuros sistemas solares térmicos. São encontrados vários estudos individuais com diferentes nanomateriais para aproveitamento de energia solar, considerando que é necessário haver uma comparação abrangente das características de conversão fototérmica de vários

nanofluidos nas mesmas condições experimentais (AMJAD et al., 2018).

Em uma investigação de desempenho da conversão fototérmica de seis nanomateriais, em coletores solares de absorção direta (DASC) sob um fluxo solar simulado focado, foram usados os nanopós comerciais de seis materiais, incluindo Ag, Cu e Zn (Sigma Aldrich Co.) e Fe, Si e Al₂O₃- γ (Nanostructured & amorphous material Inc.), conforme recebidos e tendo a água deionizada como fluido de base durante todo o procedimento experimental (AMJAD et al., 2018).

A contribuição do aquecimento sensível e do calor latente de vaporização dos nanofluidos foi revelado no contexto de seu desempenho fototérmico. Os resultados mostram que todos os nanofluidos possuem maior absorção de energia solar do que o fluido base e com destaque o nanofluido de prata revelou-se o melhor entre todos, devido à sua forte natureza de ressonância plasmônica (AMJAD et al., 2018).

De acordo com a pesquisa de Bandarra Filho et al. (2014), os resultados mostraram melhor capacidade entre a demais de conversão fototérmica das partículas de prata, mesmo que estas estejam em concentrações baixas. Na concentração de partículas de prata de 1,62, 3,25 e 6,5 ppm pode ser obtido na temperatura mais elevada um aumento de até 52%, 93% e 144%, respectivamente, de energia térmica armazenada.

Usando o nanofluido de prata em água para um estudo experimental que buscava determinar a eficiência térmica de um coletor solar de tubo evacuado por termosifão, a prata aumentou a eficiência desse coletor entre 20% e 40% (OZSOY; CORUMLU, 2018).

Um sistema de absorção direta baseado em nanopartículas oferece uma alternativa promissora aos coletores solares convencionais. Isso pode se dar investigando experimentalmente as características de conversão fototérmica de uma das nanopartículas plasmônicas, ou seja, a prata, em condições realistas. Os nanofluidos de prata estáveis são formulados através de um homogeneizador de alta pressão e os experimentos são conduzidos sob a luz do sol em um telhado com testes em execução contínua por aproximadamente 10h (BANDARRA FILHO et al., 2014).

Os resultados de Bandarra Filho et al. (2014) mostraram que as partículas de prata têm excelente capacidade de conversão fototérmica, mesmo em concentrações muito baixas. O aumento de até 144% na energia térmica armazenada pode ser obtido na temperatura de pico para uma concentração de partículas de 6,5 ppm. O desempenho de conversão fototérmica mostra um comportamento transiente e é melhor alcançado no período de radiação inicial, devido à baixa perda de calor e ao forte efeito de ressonância plasmônica de superfície dos nanofluidos de prata. A taxa de absorção específica inicial (SAR) quase constante é obtida para

concentrações de nanopartículas de até 6,5 ppm, mas diminui significativamente em concentrações mais altas, o que está associado ao aumento das interações partícula-partícula.

2.3 Sistemas térmicos utilizando nanofluidos

A condutividade térmica é uma propriedade importante nos estudos sobre nanofluidos. Os nanofluidos possuem condutividades térmicas maiores que as do fluido base. Ao inserir nanopartículas no fluido base, se alteram as propriedades térmicas e assim a transferência de calor também é alterada (BUONOMO et al., 2015; SURESH et al., 2011).

Suresh et al. (2011) evidenciam que a condutividade térmica dos nanofluidos varia de acordo com tamanho, forma e tipo de material das nanopartículas. Como exemplo há os nanofluidos com nanopartículas metálicas, que têm maior condutividade térmica que as nanopartículas não metálicas.

Análises empíricas revelaram que a condutividade térmica depende da condutividade térmica do fluido de base e das nanopartículas. Além da temperatura, são relevantes a fração de volume, a forma e a área de superfície das nanopartículas (BUONOMO et al., 2015).

2.4 Modelagem

Variados métodos, a exemplo da simulação numérica e do método experimental, são usados como alternativas na compreensão da complexa relação entre a eficiência do coletor solar e os parâmetros físicos térmicos. O baixo custo é considerado uma das vantagens da simulação numérica. No entanto, os resultados numéricos não são perfeitos devido à simplificação e aproximação dos problemas de interesse. Por outro lado, o método experimental tem a possibilidade de oferecer um resultado mais satisfatório. Porém, o mesmo pode apresentar custo elevado e limitado número de dados experimentais, quanto ao resultado (LIU et al., 2012).

A literatura existente quanto à previsão e a modelagem de sistemas de aproveitamento de energia solar térmica e renovável considera em linhas gerais que estão disponíveis para modelar sistemas térmicos solares, dois métodos: o primeiro, através da análise dos fenômenos termodinâmicos dentro do sistema; e o segundo, por meio de técnicas de inteligência computacional (AHMAD; REYNOLDS; REZGUI, 2018).

2.5 Aprendizado de máquina

Considerado um subcampo da ciência da computação, o aprendizado de máquina é classificado como um método de inteligência artificial, possuindo variadas finalidades e tendo como vantagem ser um método em que um modelo pode resolver problemas impossíveis de representação por algoritmos explícitos (VOYANT et al., 2017).

O desempenho dos algoritmos de aprendizado de máquina, para alcançar um resultado excelente, precisam ser aprimorados em algum conjunto de parâmetros (ALADE; ABD RAHMAN; SALEH, 2019).

A análise de regressão é um método de investigação estatística que visa apresentar o impacto de variáveis independentes selecionadas em uma variável dependente. A regressão linear é utilizada para avaliar a dependência linear das variáveis. Particularmente, se mostra uma ferramenta útil para prever respostas quantitativas (FANG; LAHDELMA, 2016; JAMES et al., 2013).

Os modelos de aprendizado de máquina encontram relações entre entradas e saídas, mesmo que a representação explícita seja impossível; tal característica possibilita a aplicação de modelos de aprendizado de máquina em muitos casos, por exemplo, em reconhecimento de padrões, problemas de classificação, filtragem de spam e em mineração de dados e problemas de previsão. Particularmente, a classificação e a mineração de dados são interessantes neste domínio, visto que é necessário trabalhar com grandes conjuntos de dados e a tarefa de pré-processamento e preparação de dados pode ser realizada pelos modelos de aprendizado de máquina (VOYANT et al., 2017).

Diferentes abordagens de aprendizagem de máquina são usadas na previsão ou classificação e preparação de dados, quais sejam: aprendizagem supervisionada, aprendizagem não supervisionada, aprendizagem semissupervisionada e aprendizagem por *ensemble*, descritos a seguir (DE VRIES; THIERENS, 2021; LAN et al., 2021; VOYANT et al., 2017).

Nos modelos de previsão/classificação e preparação, os algoritmos de aprendizado de máquina aprendem com os dados. Assim, é fundamental escolher as informações certas e prepará-las devidamente para permitir que o problema seja resolvido eficientemente. A preparação dos elementos específicos essencial para um conjunto de dados necessita das especificidades dos mesmos, como os tipos de variáveis, bem como os algoritmos que serão usados para modelá-los que podem impor expectativas ou requisitos aos dados (VOYANT et al., 2017).

Com os modelos de aprendizagem supervisionada os recursos e rótulos de todas as amostras são aplicados, o que seria capaz de gerar uma representação apropriada ao aprender a relação entre os recursos e rótulos. Na aprendizagem não supervisionada, são construídos apenas aprendendo as regularidades dos recursos de dados entre as amostras não rotuladas. Já nos modelos da aprendizagem semisupervisionada, diferentemente da aprendizagem supervisionada e não supervisionada, se faz uso de amostras rotuladas e não rotuladas simultaneamente. Assim, as informações contidas nas amostras não rotuladas fornecem dicas úteis sobre como os dados são distribuídos no espaço do recurso (LAN et al., 2021).

A aprendizagem por *ensemble* ou sistema de classificação múltipla, é uma coleção de aprendizes. Esses aprendizes são treinados usando dados de entrada e quando novos exemplos são classificados, suas previsões são combinadas para chegar a uma previsão simultânea. Um bom *ensemble* gera previsões com uma precisão superior à de qualquer um dos classificadores independentemente (DE VRIES; THIERENS, 2021).

3 MATERIAIS E MÉTODOS

3.1 Descrição do aparato de testes

O arranjo experimental da parede solar contém um suporte feito a partir de perfis de alumínio e duas células de teste (Figura 1). Sua estrutura foi revestida com tinta na cor branca com o objetivo de evitar aquecimento e transferência de calor para as células por meio do alumínio. Internamente, foram inseridas canaletas para a fixação das placas de vidro. Com a finalidade de impedir possíveis vibrações, esteios foram utilizados na estrutura para uma melhor fixação no solo.

Figura 1 – Foto do aparato de teste - "parede solar". Nesta situação específica, a célula superior estava preenchida com nanofluido de TiO_2 e a inferior com água.



Fonte: Santos (2019).

As duas células de teste escolhidas e fixadas no suporte possuem formato retangular constituído de duas placas de vidro branco temperado e transparente. Cada placa de vidro possui 1000 mm de comprimento, 500 mm de altura e 8 mm de espessura, separadas por um espaço de 5 mm onde foram inseridos os nanofluidos e a água deionizada, de acordo com cada teste. Cada célula de teste foi preenchida com um volume de 2,5 litros. O espaço nas canaletas entre a placa de vidro e os perfis de alumínio foram preenchidos com silicone para evitar vazamentos e firmar o vidro na estrutura.

Para aferição das temperaturas dos fluidos foram utilizados 6 termopares do tipo K com precisão de $0,1^{\circ}\text{C}$, frequência de 1 Hz e calibração em banho termostático contra um termômetro de vidro, onde três foram inseridos na célula inferior e mais três na superior. Outro termopar também foi instalado na parte externa da estrutura para medir dados da temperatura ambiente. Para aquisição dos dados foi usado um Datalogger modelo RX12 OMEGA. Usando os equipamentos piranômetro e anemômetro, respectivamente, houve o monitoramento dos dados da radiação solar global e velocidade do vento. As nanopartículas foram dispersas no fluido base através de agitação ultrassônica (AMORIM NETO, 2020; SANTOS, 2019).

Os experimentos foram realizados no Laboratório de Energia Solar e Gás Natural (LESGN) da Universidade Federal do Ceará na cidade de Fortaleza (Latitude: $03^{\circ} 43' 02''$ Sul, Longitude: $38^{\circ} 32' 35''$ Oeste) e investigaram o comportamento dos perfis de temperatura de quatro concentrações de nanopartículas de dióxido de titânio (12,5 ppm; 25 ppm; 50 ppm; 75 ppm) no período entre outubro e dezembro de 2018, das 06:00 às 18:00 horas, e de nanopartículas de prata (0,8125 ppm; 1,625 ppm; 3,25 ppm; 6,5 ppm) entre junho e outubro de 2019, das 5:30 às 21:30 horas, dispersas em água deionizada em um sistema de absorção direta de radiação solar. Para a análise foram selecionados dados de duas condições de céu distintas, dias de céu claro e nublado.

3.2 Modelo físico aplicado à parede solar

Para a execução do balanço de energia do sistema de testes foram utilizadas várias considerações e equacionamentos adaptados da obra de Duffie e Beckman (2013). O balanço de energia mostra a distribuição da energia solar incidente em ganho útil, perdas térmicas e perdas ópticas para a parede solar.

O calor armazenado no decorrer do período de testes é dado pela variação de entalpia do sistema, que neste caso acontece a volume constante, sendo a própria variação de energia interna do sistema (Equação 1).

$$Q = mc_p [T_{(i)} - T_{(t)}] \quad (1)$$

Nesta, Q é igual à diferença entre a radiação solar absorvida e a perda térmica, multiplicados pela área útil das placas de vidro, definida pela Equação 2:

$$Q = A[G - U(T_f - T_{ar})] \quad (2)$$

Onde G representa a radiação solar absorvida pela célula de teste por unidade de área, que é igual à diferença entre a radiação solar incidente e as perdas ópticas. A energia térmica perdida pelas células de teste para o ambiente por condução, convecção e radiação pode ser demonstrada como o produto do coeficiente de transferência de calor U vezes a diferença entre a temperatura média do fluido T_f e a temperatura ambiente T_{ar} , multiplicados pela área útil da célula de teste.

Simultaneamente, os processos de transferência de calor atuantes na parede solar foram analisados com o método computacional meta-heurístico de Busca Harmônica, utilizado para encontrar as propriedades ópticas de absorvidade do fluido e albedo do solo, envolvidas no processo de variação de energia térmica, a partir dos dados medidos nos experimentos, conforme explicado adiante.

A Busca Harmônica, do inglês Harmony Search, é um algoritmo heurístico fundamentado na similaridade do fenômeno artificial de um grupo musical em busca da melhor harmonia. Tal busca ocorre com a combinação de dados existentes e a geração de novos dados que são combinados para formar possíveis soluções (GEEM; KIM; LOGANATHAN, 2001).

A Heurística compreende um conjunto de métodos para solucionar um problema onde as variáveis em questão utilizam a experiência obtida no decurso das iterações. A meta-heurística combina diferentes conceitos inteligentemente para explorar o espaço de busca, de modo que estratégias de aprendizagem são utilizadas para estruturar informações e encontrar soluções eficientes e praticamente ótimas (OSMAN; LAPORTE, 1996).

Na fase de calibração do modelo físico deste trabalho, foi utilizado o método proposto por Askarzadeh e Zebarjadi (2014) para seleção dos parâmetros da Busca Harmônica. Essa calibração se constituiu em determinar parâmetros das variáveis das absorvidades dos fluidos e albedo do solo e propriedades físicas relacionadas ao balanço de energia que, ou não se encontraram valores consistentes na literatura, ou esses valores simplesmente não resultaram em desempenho consistente do modelo, quando comparados aos resultados experimentais usados como referência.

Após os cálculos de balanço, um índice de correlação de Pearson (R) de 0,73 foi encontrado entre os valores medidos para a variação de energia interna e os simulados pela meta-heurística de Busca Harmônica, a fim de verificar a eficácia da aplicação do método. O coeficiente de correlação de Pearson (R), também chamado de coeficiente de correlação

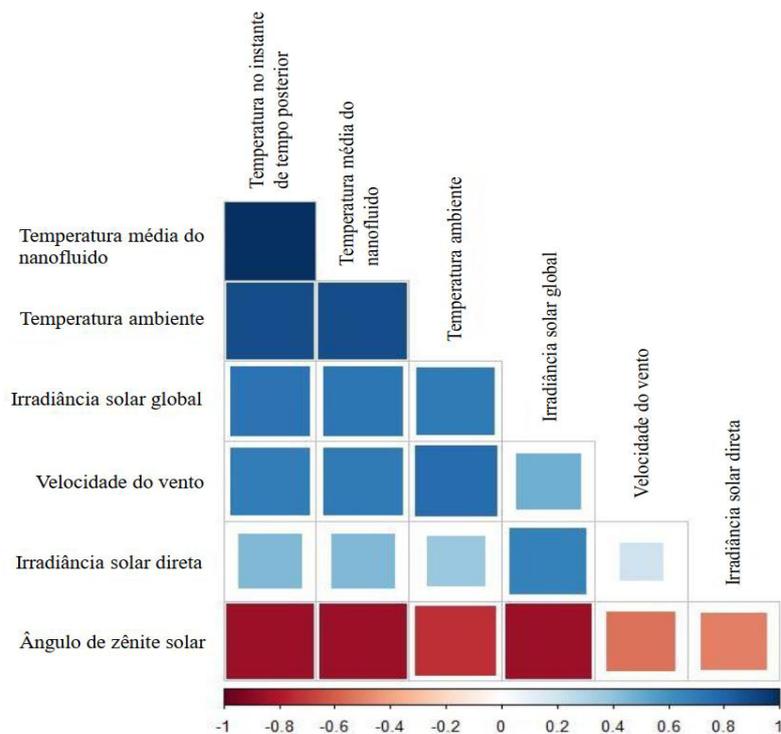
produto-momento ou o R de Pearson, mede o grau da correlação linear entre duas variáveis quantitativas. O coeficiente de correlação Pearson varia de -1 a 1, onde valores do seu módulo localizados entre 0,70 até 1 são considerados como fortes.

Para maiores detalhamentos sobre o equacionamento e a solução obtida do balanço de energia para o modelo, recomenda-se o trabalho de Santos (2019).

3.3 Correlação entre os preditores

A correlação entre os preditores se baseia nas variáveis do modelo físico que se mostram mais relevantes. Os dados de entrada da modelagem numérica são: temperatura no instante de tempo posterior, temperatura ambiente, temperatura média do nanofluido, irradiância solar direta, irradiância solar global, velocidade do vento e ângulo de zênite solar, onde todos eles foram usados para estimar a temperatura do nanofluido. Como mostrado na Figura 2, suas interrelações são vistas na matriz de correlação entre os preditores da modelagem, variando de -1 a 1, onde 1 e -1 mostram a melhor correlação, positiva e negativa, respectivamente.

Figura 2 – Matriz de correlação entre os preditores da modelagem.



Fonte: Elaborado pelo autor.

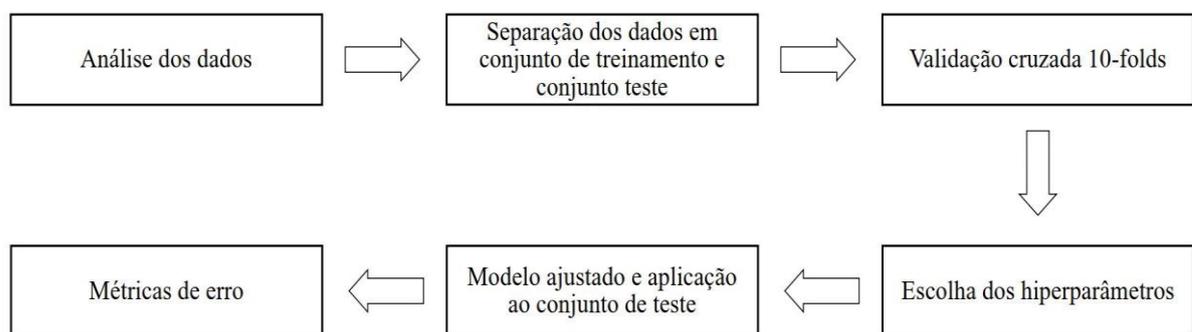
3.4 Modelos Numéricos

Nessa direção, abordar-se-á a linguagem de programação R (versão 4.0.2) e RStudio (versão 1.3.1073) como interface, que foram usados para a análise estatística das métricas de erro, sendo o RMSE utilizado para a escolha do modelo com melhor desempenho.

Os dados foram separados em conjunto de treinamento e conjunto teste. Os dados do conjunto de treinamento foram utilizados para treinar os métodos calculando os hiperparâmetros ideais destes (AHMAD; REYNOLDS; REZGUI, 2018).

Foi utilizada uma validação cruzada de 10-fold (10 subconjuntos) para o conjunto de treinamento, onde esta fornece uma indicação melhor do quão bem o modelo se sairá com novos dados, e por meio das várias divisões, esta acaba testando o modelo no conjunto inteiro. O conjunto de teste foi utilizado para validar a modelagem numérica após a escolha dos hiperparâmetros ótimos (Figura 3).

Figura 3 – Organização do conjunto de dados.



Fonte: Elaborado pelo autor.

Foram utilizados neste trabalho sete modelos numéricos para estimativa de temperatura e variação de temperatura: Regressão de Ridge, LASSO, *ElasticNet*, kNN, XGBoost, MARS e Modelo agregado *Ensemble*. Dentre eles, dois modelos foram utilizados como métodos de *ensemble* (combinados): Ridge e MARS, descritos a seguir.

3.4.1 Regressão de Ridge

Preliminarmente a regressão de Ridge precisa de uma transformação de normalização das variáveis preditoras. Isso é realizado subtraindo de cada valor a média estimada do preditor e dividindo cada valor pelo desvio padrão estimado do preditor. Deste

modo, todos os valores ficam na mesma escala e têm um desvio padrão de aproximadamente 1. A regressão de Ridge estabelece uma penalidade no tamanho total dos coeficientes, de forma que modelos com coeficientes menores tendem a ser escolhidos. Assim, preditores mais importantes terão coeficientes mais altos e a penalidade manterá outros coeficientes baixos. Modelos de Ridge reduzem o RSS (soma dos resíduos quadrados), sendo um parâmetro de regulação que equilibra a importância de um erro baixo com a importância de coeficientes pequenos. Um modelo diferente é produzido com cada escolha de λ , que é seu hiperparâmetro de calibração (AL-OBEIDAT; SPENCER; ALFANDI, 2020).

Na Equação 3, a variável y representa a resposta, x são os preditores e β representam os coeficientes de regressão. Nesta, $(\lambda \sum_{j=1}^p \beta_j^2)$ é a norma L_2 da segunda parcela da equação (JAMES et al., 2013).

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p \beta_j^2 = \text{RSS} + \lambda \sum_{j=1}^p \beta_j^2 \quad (3)$$

A regressão de Ridge reduz os coeficientes dos preditores de forma que as variáveis com menor colaboração para o resultado tenham seus coeficientes próximos de zero. A redução dos coeficientes é alcançada penalizando o modelo de regressão com um termo ligado à norma L_2 , que é a soma dos coeficientes quadrados (JAMES et al., 2013).

O valor da penalidade pode ser ajustado com precisão usando uma constante chamada λ . Quando $\lambda = 0$, o termo de penalidade não tem efeito e a regressão de Ridge produzirá os coeficientes de mínimos quadrados clássicos. Entretanto, à medida em que λ aumenta para infinito, o impacto da penalidade de contração cresce e os coeficientes de regressão da Ridge ficarão próximos de zero (AL-OBEIDAT; SPENCER; ALFANDI, 2020; JAMES et al., 2013).

A vantagem da regressão de Ridge é o bom desempenho em comparação com o método dos mínimos quadrados ordinários (regressão linear), em uma situação de dados multivariados extensos com o número de preditores maior do que o número de observações. Como desvantagem a regressão de Ridge incluirá todos os preditores no modelo final, por menor que seja a influência destes (JAMES et al., 2013).

A regressão de Ridge reduz os coeficientes aproximados de zero, porém não definirá nenhum deles precisamente para zero. A regressão LASSO possibilita reduzir os coeficientes de regressão para zero, sendo uma alternativa para solucionar essa dificuldade

(JAMES et al., 2013).

3.4.2 LASSO

Do inglês Menor Redução Absoluta e Operador de Seleção, o LASSO reduz os coeficientes de regressão até zero, penalizando o modelo de regressão com o termo da norma L_1 , que é a soma dos coeficientes absolutos, conforme a Equação 4 (JAMES et al., 2013).

$$\text{RSS} + \lambda \sum_{j=1}^p |\beta_j| \quad (4)$$

A penalidade tem o objetivo de forçar algumas das estimativas dos coeficientes que contribuem menos ao modelo serem exatamente iguais a zero. Assim, o LASSO também pode ser considerado alternativa aos métodos de seleção de subconjuntos de variáveis a fim de reduzir a complexidade do modelo. Assim como na regressão de Ridge, isso se dá selecionando um bom valor de λ (AL-OBEIDAT; SPENCER; ALFANDI, 2020; JAMES et al., 2013).

A regressão LASSO tem como vantagem sobre a regressão Ridge a produção de modelos mais simples e mais compreensíveis, por integrar um conjunto menor de preditores. Contudo, nem a regressão Ridge nem o LASSO irão dominar um ao outro. Normalmente o LASSO pode ter um desempenho melhor em situações onde alguns dos preditores contém coeficientes grandes e os demais coeficientes pequenos (JAMES et al., 2013).

3.4.3 ElasticNet

O *ElasticNet* proporciona um modelo de regressão que é penalizado tanto com a norma L_1 como com a norma L_2 . Consequentemente, reduz efetivamente os coeficientes (como na regressão de Ridge) e define alguns coeficientes em zero (como em LASSO), expresso na Equação 5 (KIM; LEE, 2021).

$$\text{RSS} + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \quad (5)$$

Como o Ridge tem baixo viés e alta variância para o erro de predição, pode levar ao *overfitting*. Em contraste com a Ridge, o LASSO tem alta polarização e baixa variância para o erro de predição, o que pode levar a um *underfitting*. Normalmente, o problema de

underfitting pode ser resolvido aumentando a complexidade do modelo. Mas quando a complexidade do modelo aumenta consideravelmente, surge o problema do *overfitting*. O *ElasticNet* foi desenvolvido para equilibrar o viés e a variância para o erro de predição e melhorar algumas limitações do LASSO. Assim, o *ElasticNet* tende a atingir a complexidade ideal do modelo, reduzindo o erro total de previsão (KIM; LEE, 2021).

3.4.4 *kNN*

K-Vizinhos mais próximos (do inglês K-Nearest Neighbors, kNN) é um algoritmo que compreende dois parâmetros: o número de vizinhos (K) e a medida de dissimilaridade ou similaridade usada para descobrir os vizinhos mais próximos. A distância Euclidiana (Equação 6) é a medida mais utilizada para determinar os vizinhos, sendo a abordada neste trabalho.

$$d(x, y) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad (6)$$

Em relação ao parâmetro K (número de vizinhos), variadas alternativas podem ser empregadas para determinar o valor mais adequado por tarefa. Uma opção é avaliar o algoritmo kNN no conjunto de validação, aderindo diferentes valores para K. O valor de K que chegar à melhor precisão será selecionado para as instâncias de teste (JAMES et al., 2013).

3.4.5 *XGBoost*

Aumento de gradiente extremo (XGBoost) é um algoritmo avançado de aprendizado de máquina baseado em árvore de decisão. É uma implementação da máquina de *gradient boosting* de Friedman (2002) criada por Chen e Guestrin (2016), que tem sido amplamente utilizada em problemas de previsão. A ideia central do XGBoost é a criação de novos modelos para prever os resíduos de modelos anteriores e, seguidamente, juntar todos os modelos para fazer as previsões finais objetivando minimizar a função de perda. (ZHONG et al., 2020).

O XGBoost usa uma estrutura de aumento de gradiente e também é um método de conjunto baseado em árvore de decisão. Com a estrutura da árvore, a previsão final é calculada somando as pontuações em todas as folhas, como expresso na Equação 7. Nesta, \hat{y}_i é a predição, B é número de árvores e f_k é o conjunto das florestas possíveis (HUANG et al., 2020).

$$\hat{y}_i = \sum_{b=1}^B f_b(x_i), f_b \in F \quad (7)$$

O XGBoost faz melhorias na função objetivo de otimização, ou seja, aprimora a função de perda e punição de complexidade, conforme Equação 8 (HUANG et al., 2020).

$$F(X) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{b=1}^B \Omega(f_b) \quad (8)$$

Na qual X são os parâmetros de otimização, l é função de perda, Ω é função de regulamentação e n o número de amostras.

3.4.6 MARS

Spline de Regressão Adaptativa Multivariada (MARS) é uma técnica de regressão que pode ser vista como uma extensão de modelos lineares que automaticamente modelam a não linearidade e as interações entre variáveis.

O modelo MARS possui melhores características de interpretação. Sua expressão matemática tem menos coeficientes e é conveniente para aplicação. Tal característica pode ser importante para acelerar os cálculos e economizar recursos computacionais sem deteriorar a precisão da estimativa. Além disso, há a possibilidade de o algoritmo selecionar automaticamente variáveis importantes e abandonar variáveis redundantes, evitando o *overfitting* e *underfitting*. O MARS pode ser representado em uma forma que identifica separadamente as contribuições aditivas e aquelas associadas a diferentes interações multivariáveis, conforme Equação 9 (FRIEDMAN, 1991; LI et al., 2019).

$$f(X) = \beta_0 + \sum_{m=1}^M \beta_m B_m(X) \quad (9)$$

Onde X é um vetor de amostra de variáveis preditoras de entrada. Essa função consiste em um parâmetro de interseção β_0 e a soma ponderada dos pesos β_m de uma ou mais funções básicas. O modelo também pode ser utilizado como uma soma ponderada de funções de base B do conjunto que abrange todos os valores de cada preditor, ou seja, esse conjunto consiste em uma função de base B e parâmetro t (localizações dos nós), para cada valor distinto

em cada variável preditora (FRIEDMAN, 1991; LI et al., 2019).

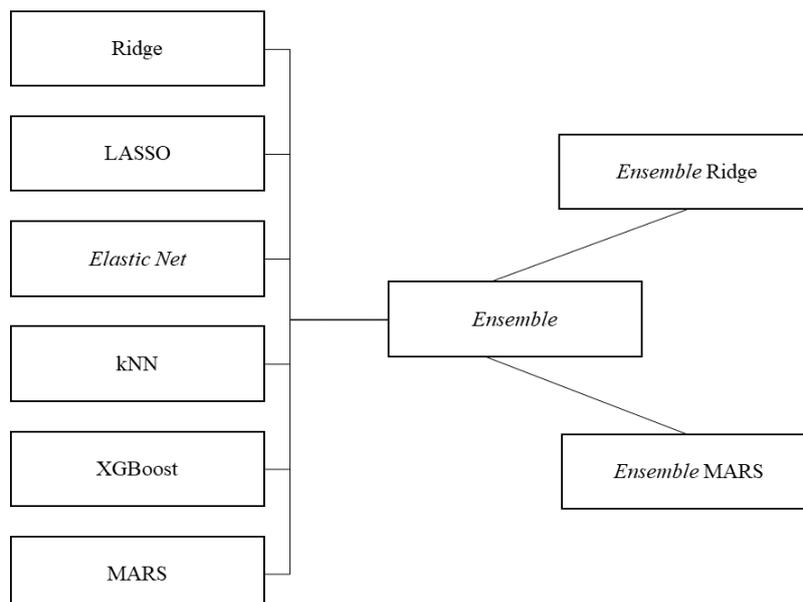
3.4.7 Modelo agregado Ensemble

O *ensemble* é uma abordagem de aprendizado de máquina que integra vários modelos básicos para gerar a saída final. O mesmo se popularizou devido ao seu excelente desempenho de generalização. Os modelos de *ensemble* geralmente resultam em um desempenho bem superior se comparados aos indivíduos que os compõem (AHMAD; REYNOLDS; REZGUI, 2018; FAN; XIAO; WANG, 2014).

Geralmente, a aprendizagem em *ensemble* pode ser sucinta como uma abordagem de duas etapas. A primeira etapa pretende desenvolver uma série de modelos básicos de forma paralela ou sequencial, onde os modelos básicos desenvolvidos de forma paralela têm poucas conexões entre si. Isso ocorre diferentemente dos modelos de base desenvolvidos de forma sequencial, em que um modelo de base tem influência na geração de seu modelo de base subsequente. Já a segunda etapa utiliza modelos básicos para gerar o resultado final por meio de certos esquemas de combinação: a votação por maioria e os métodos de média ponderada compõem alguns exemplos (AHMAD; REYNOLDS; REZGUI, 2018; FAN; XIAO; WANG, 2014).

Neste trabalho foram usadas duas abordagens de *ensemble* paralelo (Figura 4), a primeira baseada na regressão de Ridge e a segunda baseada na regressão MARS.

Figura 4 – Modelo agregado *Ensemble*.



Fonte: Elaborado pelo autor.

3.4.8 Métricas de erros

Quanto à análise de desempenho dos modelos para verificar se os mesmos se apresentavam satisfatoriamente no conjunto de teste, foram utilizadas as métricas: raiz do erro quadrático médio (RMSE), coeficiente de determinação (R^2), erro absoluto médio (MAE) e erro médio por viés (MBE), conforme apresentado nas Equações 10, 11, 12 e 13, respectivamente.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (11)$$

$$\text{MAE} = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N} \quad (12)$$

$$\text{MBE} = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)}{N} \quad (13)$$

Onde N é o número de amostras, y_i o valor medido e \hat{y}_i e o valor previsto da amostra.

4 RESULTADOS E DISCUSSÕES

Nesta seção são apresentados os resultados dos modelos de aprendizagem de máquina avaliando o comportamento dos perfis de temperatura em estimativa de temperatura final e variação de temperatura de quatro concentrações de nanopartículas de dióxido de titânio (12,5 ppm; 25 ppm; 50 ppm; 75 ppm) e de nanopartículas de prata (0,8125 ppm; 1,625 ppm; 3,25 ppm; 6,5 ppm), dispersas em água deionizada em um sistema de absorção direta de radiação solar.

Nas Tabelas 1, 2, 3, 4, 5, 6, 7 e 8, são destacados os resultados do modelo físico de Santos (2019) e o modelo de aprendizado de máquina com melhor resultado de RMSE, tanto para estimativa de temperatura final como para variação de temperatura nas quatro concentrações de nanopartículas.

4.1 TiO₂ – 12,5 ppm

Na concentração de 12,5 ppm em estimativa de temperatura final, cinco modelos computacionais apresentaram desempenho superior ao modelo físico, representando 63% dos testados. O modelo *Ensemble* MARS apresentou o melhor RMSE, como mostrado na Tabela 1, com valor de 0,164 °C seguido por XGBoost com 0,165 °C com uma diferença de 0,36% entre os dois melhores modelos. Assim, o modelo *Ensemble* MARS apresentou um desempenho 57,04% superior ao modelo físico.

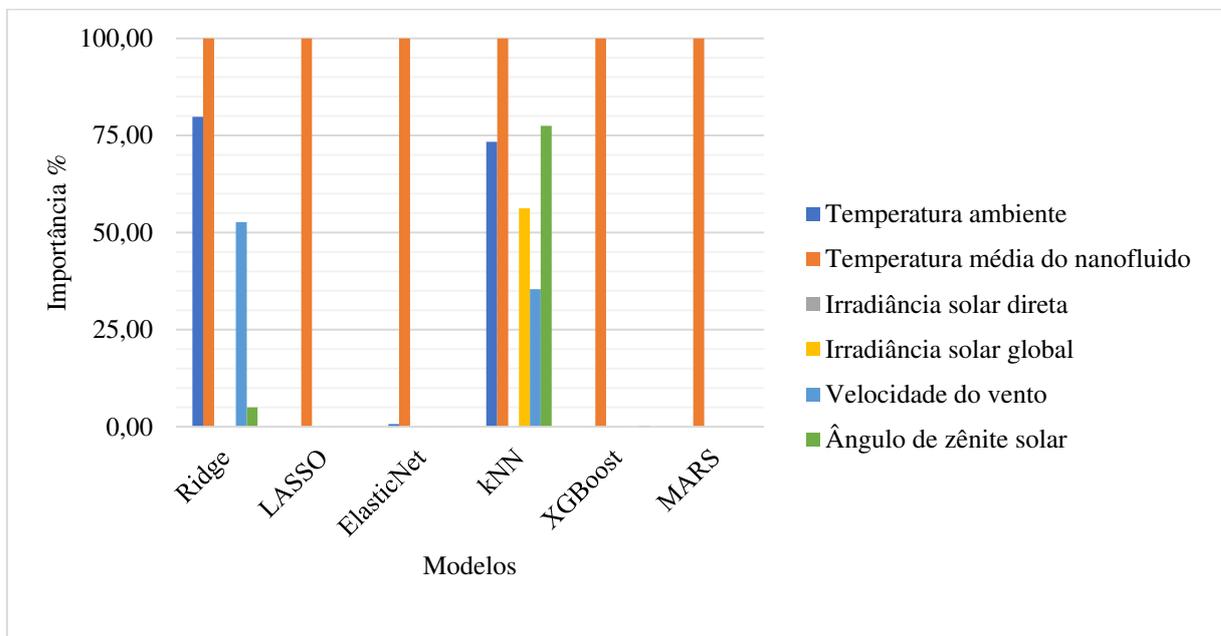
A temperatura média do nanofluido foi o preditor usado em todos os modelos computacionais, com sua importância em 100%, como mostrado no Gráfico 1. O modelo com melhor desempenho fez o uso exclusivo do preditor média do nanofluido, diferentemente dos modelos Ridge e kNN, os quais fizeram uso de todos os preditores com suas diferentes porcentagens de importância.

Tabela 1 – TiO₂ - 12,5 ppm (Temperatura final).

Modelos	RMSE (°C)	R ²	MAE	MBE
Modelo Físico	0,382	0,993	0,297	-0,043
Ridge	1,197	0,968	0,962	-0,340
LASSO	0,208	0,999	0,180	-0,052
ElasticNet	0,258	0,999	0,227	-0,074
kNN	1,122	0,970	0,867	-0,086
XGBoost	0,166	0,999	0,125	-0,011
MARS	0,170	0,999	0,126	-0,022
Ensemble Ridge	0,409	0,996	0,322	-0,091
Ensemble MARS	0,164	0,999	0,124	-0,010

Fonte: Elaborado pelo autor.

Gráfico 1 – Relação de importância dos preditores em 12,5 ppm (Temperatura final).



Fonte: Elaborado pelo autor.

Na concentração de nanopartículas em 12,5 ppm para estimativa de variação de temperatura, todos os modelos computacionais apresentaram desempenho superior ao modelo físico, como mostrado na Tabela 2. Entre o modelo com pior e melhor desempenho é observada uma variação do RMSE de 0,018 °C com 4,83%, entre eles. O modelo MARS apresentou o melhor RMSE, de 0,142 °C, sendo 62,78% superior ao modelo físico.

O modelo MARS fez o uso do preditor irradiância global, mostrado no Gráfico 2, com sua importância em 100%, ao contrário dos outros modelos testados. Com exceção dos preditores irradiância direta e velocidade do vento, todos foram usados em diferentes porcentagens de importância. Os modelos de *ensemble* utilizaram a média do nanofluido como

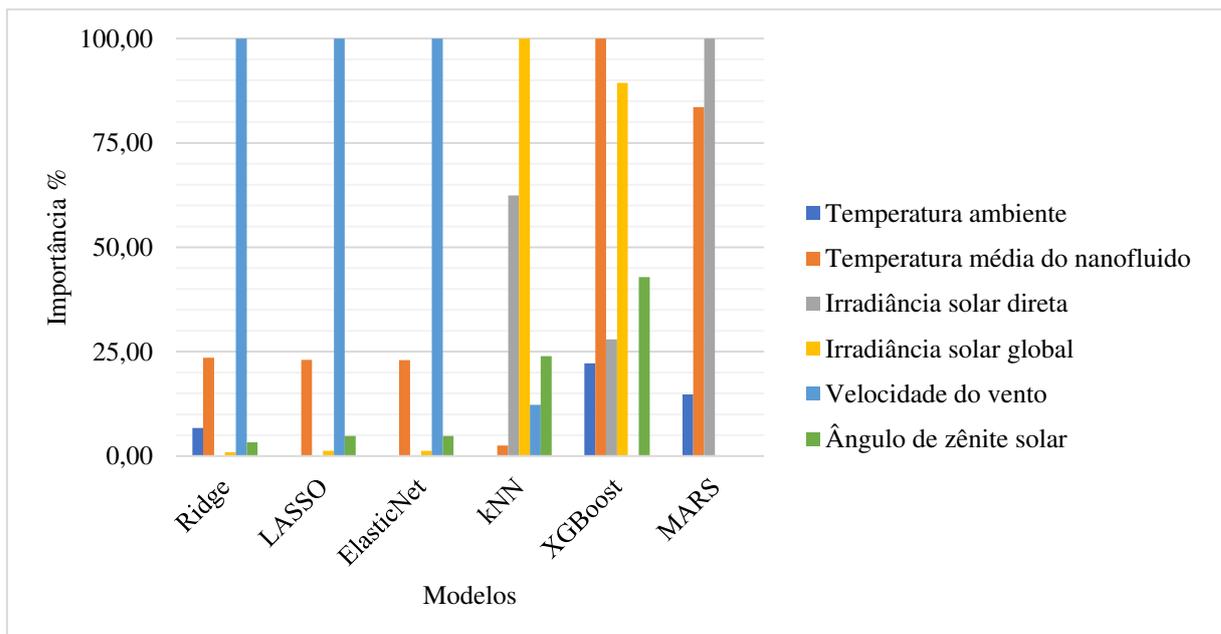
o preditor de maior importância em 100%.

Tabela 2 – TiO₂ - 12,5 ppm (Variação de temperatura).

Modelos	RMSE (°C)	R ²	MAE	MBE
Modelo Físico	0,382	0,993	0,297	-0,043
Ridge	0,161	0,108	0,122	-0,001
LASSO	0,161	0,113	0,123	0,002
ElasticNet	0,161	0,113	0,123	0,002
kNN	0,152	0,224	0,110	-0,020
XGBoost	0,149	0,254	0,110	-0,001
MARS	0,142	0,293	0,105	0,000
Ensemble Ridge	0,149	0,255	0,110	-0,002
Ensemble MARS	0,157	0,223	0,117	0,003

Fonte: Elaborado pelo autor.

Gráfico 2 – Relação de importância dos preditores em 12,5 ppm (Variação de temperatura).



Fonte: Elaborado pelo autor.

4.2 TiO₂ – 25 ppm

O melhor RMSE na concentração de nanopartículas em 25 ppm para estimativa de temperatura final foi obtido no modelo *ensemble* MARS com 0,136 °C, sendo 65,10% superior ao modelo físico, representado na Tabela 3. Cinco dos modelos testados foram superiores ao modelo físico, onde tiveram uma variação de 15,49%. Três modelos testados apresentaram resultados inferiores ao modelo físico.

Como observado no Gráfico 3, o preditor média do nanofluido foi usado com sua

importância em 100% em todos modelos testados. O modelo com melhor resultado *ensemble* MARS fez uso exclusivo da média do nanofluido. Os modelos com desempenho inferior fizeram uso de mais preditores em diferentes porcentagens de importância, como no modelo kNN.

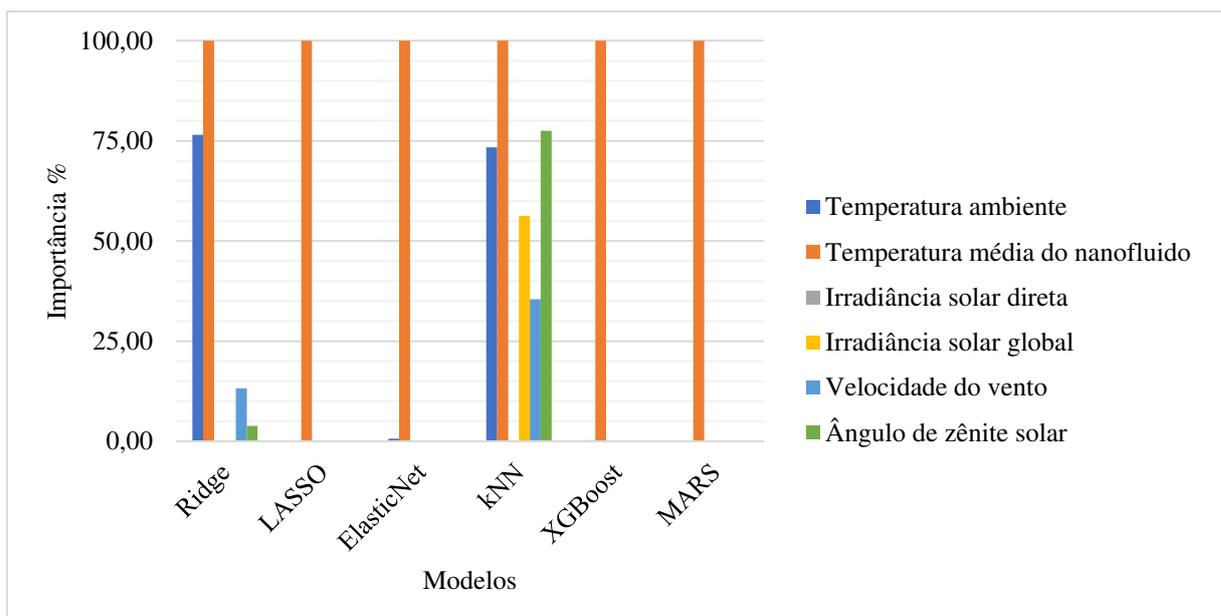
Os dois modelos agregados *ensemble* Ridge e *ensemble* MARS apresentaram diferentes resultados. O modelo com melhor desempenho fez uso exclusivo do modelo XGBoost, enquanto o *ensemble* Ridge fez uso de todos os modelos com importância em 100% para o modelo *ElasticNet*.

Tabela 3 – TiO₂ - 25 ppm (Temperatura final).

Modelos	RMSE (°C)	R ²	MAE	MBE
Modelo Físico	0,395	0,994	0,296	0,011
Ridge	0,670	0,992	0,554	0,130
LASSO	0,198	0,999	0,166	0,022
<i>ElasticNet</i>	0,258	0,999	0,227	-0,074
kNN	1,122	0,967	0,867	-0,086
XGBoost	0,137	1,000	0,099	0,024
MARS	0,170	0,999	0,126	-0,022
<i>Ensemble</i> Ridge	1,197	0,968	0,962	-0,340
<i>Ensemble</i> MARS	0,137	1,000	0,099	0,024

Fonte: Elaborado pelo autor.

Gráfico 3 – Relação de importância dos preditores em 25 ppm (Temperatura final).



Fonte: Elaborado pelo autor.

O modelo XGBoost apresentou o resultado RMSE de 0,125 °C na concentração de nanopartículas em 25 ppm para estimativa de variação de temperatura, tendo desempenho superior de 68,10% em relação ao modelo físico, representado na Tabela 4. Todos os modelos testados apresentaram eficiência superior ao modelo físico, com uma variação de RMSE em 0,018 °C, equivalente a 4,64% entre Ridge e XGBoost.

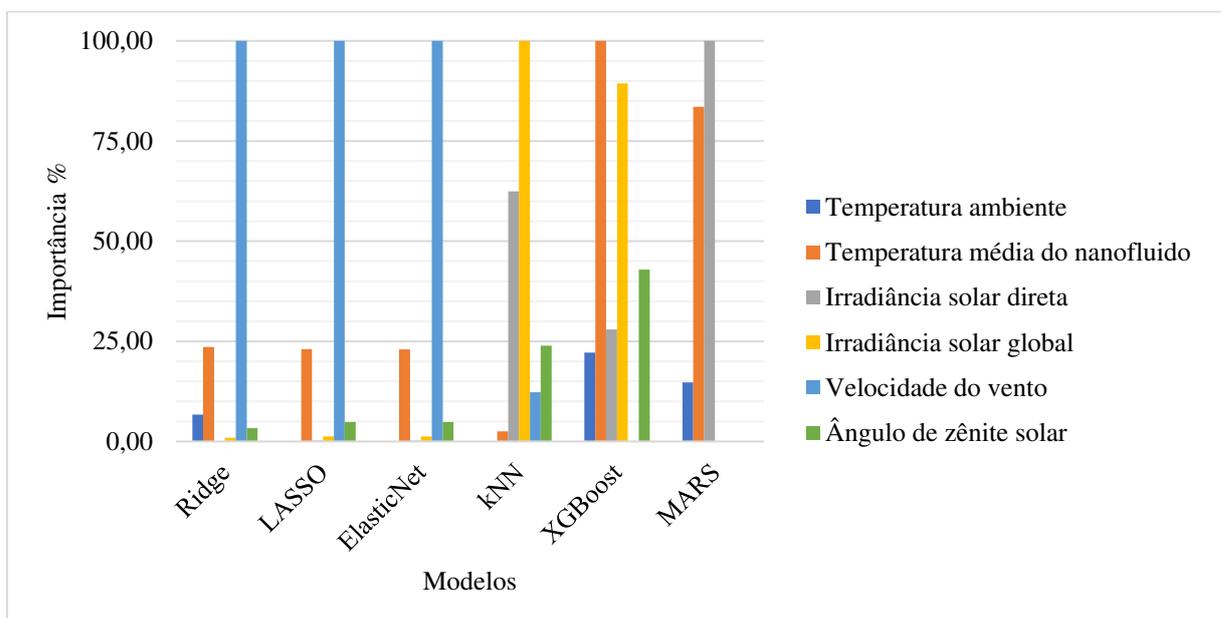
Como apresentado no Gráfico 4, o modelo XGBoost não fez exclusão de preditores, entretanto a média do nanofluido teve importância de 100% como o principal preditor usado nesse modelo. Os modelos lineares fizeram uso de todos os preditores, sem exclusão, com uma importância em 100% para velocidade do vento, ocasionando poucas variações entre os valores de RMSE.

Tabela 4 – TiO₂ - 25 ppm (Variação de temperatura).

Modelos	RMSE (°C)	R ²	MAE	MBE
Modelo Físico	0,395	0,994	0,296	0,011
Ridge	0,144	0,195	0,102	-0,002
LASSO	0,144	0,195	0,102	-0,003
ElasticNet	0,144	0,195	0,102	-0,003
kNN	0,134	0,326	0,092	-0,002
XGBoost	0,126	0,427	0,088	0,008
MARS	0,131	0,370	0,099	0,019
Ensemble Ridge	0,128	0,421	0,089	0,008
Ensemble MARS	0,136	0,395	0,092	0,006

Fonte: Elaborado pelo autor.

Gráfico 4 – Relação de importância dos preditores em 25 ppm (Variação de temperatura).



Fonte: Elaborado pelo autor.

4.3 TiO₂ – 50 ppm

O modelo *ensemble* MARS apresentou um RMSE de 0,196 °C na concentração de nanopartículas em 50 ppm para estimativa de temperatura final, sendo 54,06% superior ao modelo físico, Tabela 5. Seis dos modelos testados obtiveram resultados melhores se comparados ao modelo físico; Rigde e kNN foram os modelos com os piores valores de RMSE.

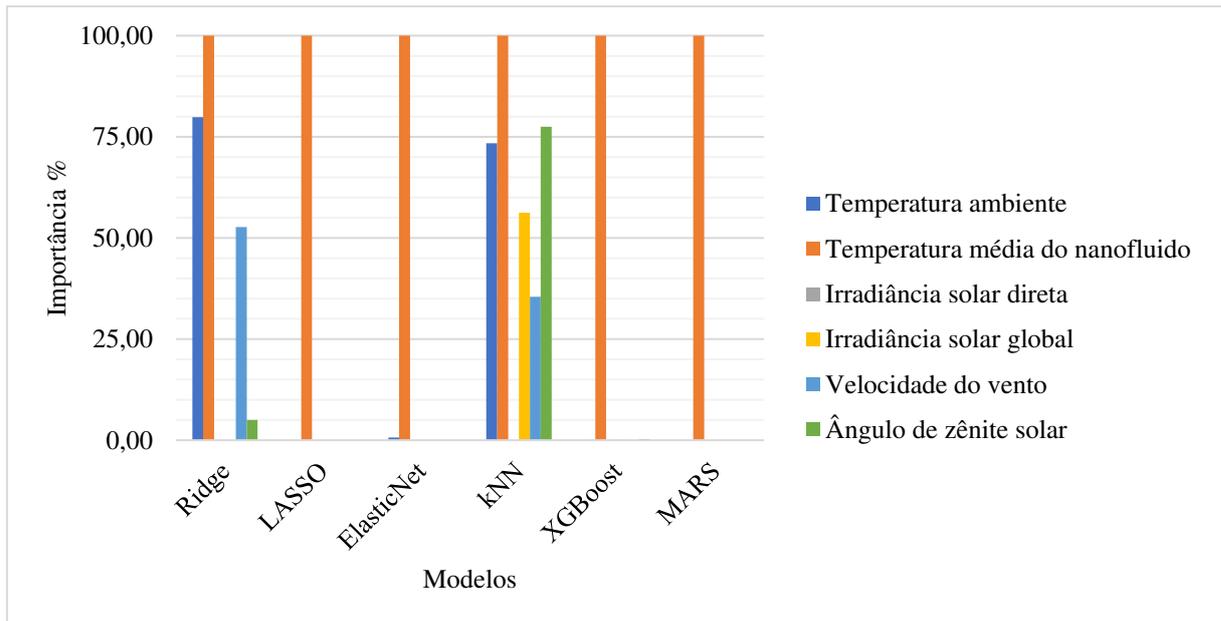
O preditor com maior importância foi a média do nanofluido para todos os modelos testados, como observa-se no Gráfico 5. Os modelos Ridge e kNN apresentaram os piores resultados, fizeram um uso maior de preditores com apenas uma exclusão cada, diferentemente dos modelos com os melhores desempenhos, onde foi feito uso exclusivo de um preditor com sua importância em 100%. Modelos como XGBoost e *ElasticNet* fizeram uso de mais preditores, onde a média do nanofluido foi o preditor com maiores porcentagens de importância em 100%, e os demais preditores variando de 0,01% a 0,68% de importância.

Tabela 5 – TiO₂ - 50 ppm (Temperatura final).

Modelos	RMSE (°C)	R ²	MAE	MBE
Modelo Físico	0,458	0,991	0,390	0,085
Ridge	0,928	0,984	0,740	0,255
LASSO	0,246	0,999	0,211	0,006
<i>ElasticNet</i>	0,304	0,999	0,271	0,011
kNN	1,260	0,979	0,919	0,601
XGBoost	0,197	0,999	0,140	0,021
MARS	0,222	0,999	0,162	0,001
<i>Ensemble</i> Ridge	0,311	0,998	0,245	0,145
<i>Ensemble</i> MARS	0,197	0,999	0,140	0,021

Fonte: Elaborado pelo autor.

Gráfico 5 – Relação de importância dos preditores em 50 ppm (Temperatura final).



Fonte: Elaborado pelo autor.

Na estimativa de variação de temperatura em 50 ppm, como mostra a Tabela 6, todos os modelos apresentaram RMSE menor que o modelo físico. O modelo XGBoost com 0,191 °C equivalente a 58,19%, obteve o melhor desempenho. A variação do RMSE ficou em 0,026 °C, igual a 5,82%, entre os modelos.

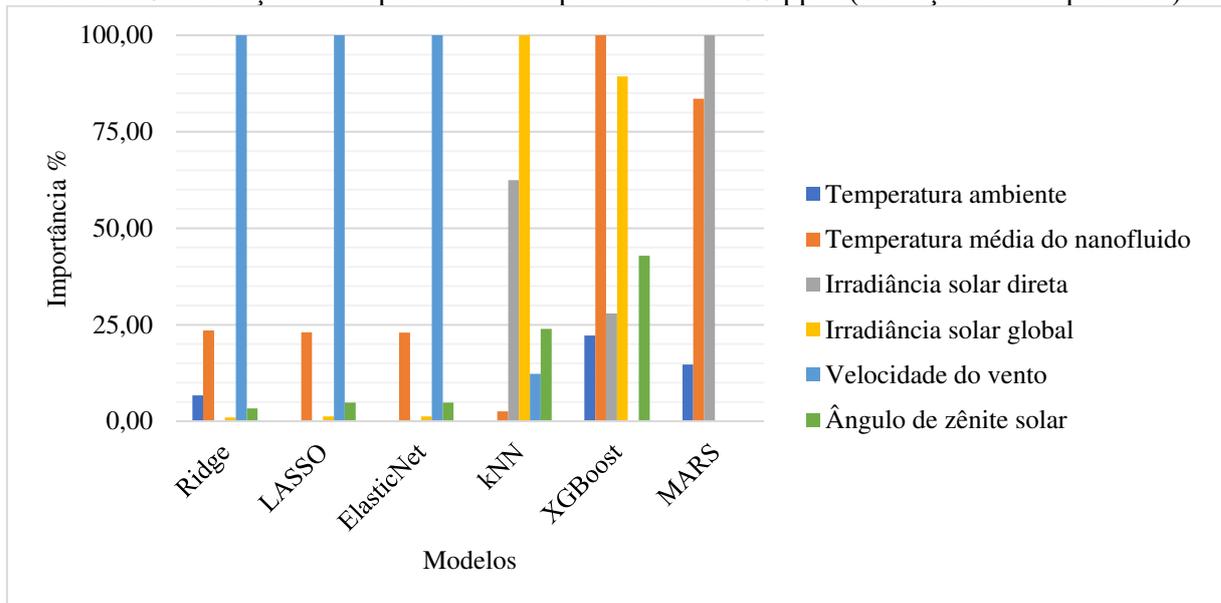
O modelo XGBoost, sem exclusão fez o uso de todos os preditores com uma maior porcentagem de importância para o preditor média do nanofluido, apresentado no Gráfico 6 com sua utilização em 100%, mas com pouca utilização do preditor velocidade do vento. Diferentemente aconteceu com os modelos lineares, onde o preditor velocidade do vento teve sua porcentagem de importância em 100%, utilizando todos os preditores sem exclusão, como o modelo XGBoost.

Tabela 6 – TiO₂ - 50 ppm (Variação de temperatura).

Métodos	RMSE (°C)	R ²	MAE	MBE
Modelo Físico	0,458	0,991	0,390	0,085
Ridge	0,217	0,054	0,157	-0,008
LASSO	0,218	0,051	0,159	-0,011
<i>ElasticNet</i>	0,218	0,050	0,159	-0,011
kNN	0,197	0,222	0,139	0,004
XGBoost	0,192	0,299	0,133	0,020
MARS	0,192	0,264	0,139	0,014
<i>Ensemble Ridge</i>	0,194	0,295	0,134	0,022
<i>Ensemble MARS</i>	0,203	0,282	0,139	0,027

Fonte: Elaborado pelo autor.

Gráfico 6 – Relação de importância dos preditores em 50 ppm (Variação de temperatura).



Fonte: Elaborado pelo autor.

4.4 TiO₂ – 75 ppm

Na concentração de 75 ppm, dois modelos, Ridge e kNN, não apresentaram desempenho melhor que o modelo físico para estimativa de temperatura final, observado na Tabela 7. O modelo com melhor desempenho ficou para *ensemble* MARS, com um RMSE de 0,184 °C, equivalente a 60,43% superior ao modelo físico. Entre os seis modelos com melhores resultados o RMSE variou 0,099 °C, igual a 21,32 %.

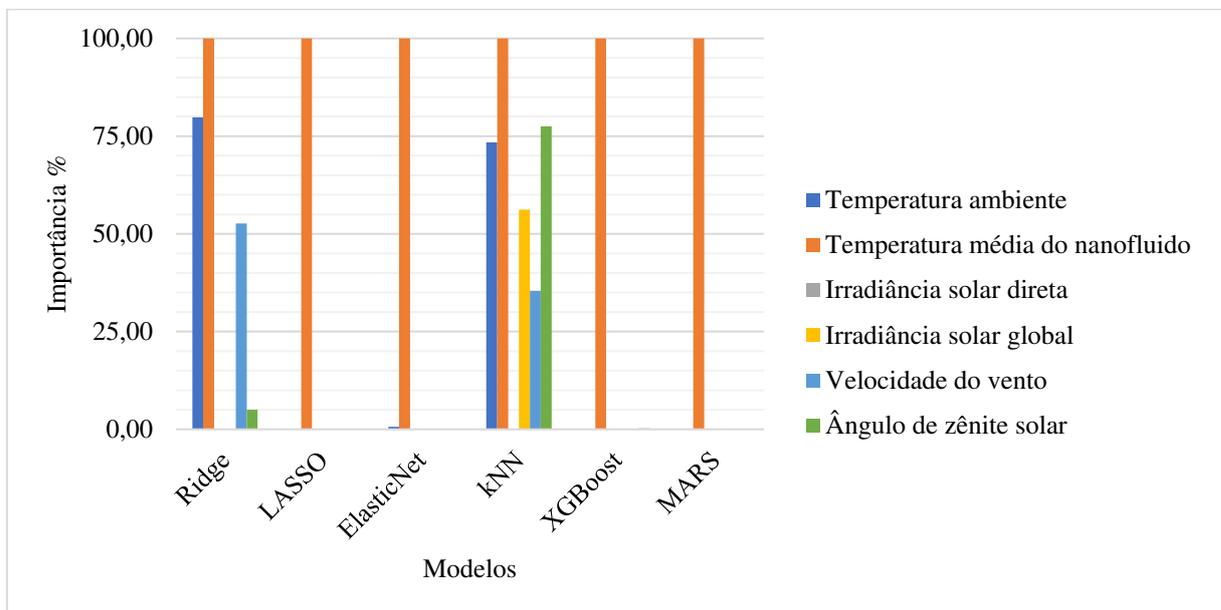
Como apresentado no Gráfico 7, a média do nanofluido foi o preditor usado com sua importância em 100% por todos os modelos testados. Os modelos com pior resultado fizeram o uso de vários preditores em diferentes porcentagens de importância, onde os modelos Ridge e kNN apresentaram um uso acima dos 70% de importância para o preditor temperatura ambiente, ao contrário dos modelos com melhor desempenho.

Tabela 7 – TiO₂ - 75 ppm (Temperatura final).

Modelos	RMSE (°C)	R ²	MAE	MBE
Modelo Físico	0,465	0,991	0,381	-0,041
Ridge	0,903	0,985	0,770	0,054
LASSO	0,220	0,999	0,184	0,001
<i>ElasticNet</i>	0,283	0,999	0,248	0,001
kNN	0,882	0,984	0,646	0,025
XGBoost	0,186	0,999	0,125	0,000
MARS	0,217	0,999	0,147	-0,013
<i>Ensemble Ridge</i>	0,274	0,999	0,216	0,015
<i>Ensemble MARS</i>	0,184	0,999	0,124	0,002

Fonte: Elaborado pelo autor.

Gráfico 7 – Relação de importância dos preditores em 75 ppm (Temperatura final).



Fonte: Elaborado pelo autor.

Como mostrado na Tabela 8, na concentração de 75 ppm para estimativa de variação temperatura todos os modelos obtiveram desempenho superior a 60% em relação ao modelo físico, onde o modelo kNN ficou com melhor RMSE de 0,146 °C, equivalente a 68,56% superior ao modelo físico.

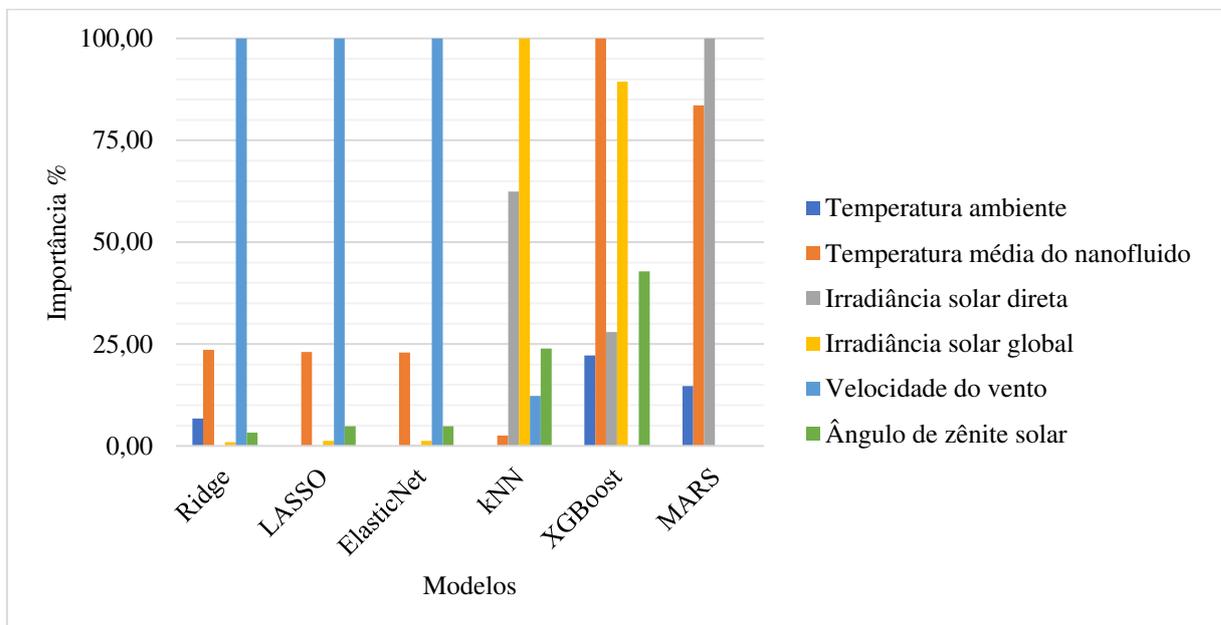
O Gráfico 8 traz a importância em porcentagens dos preditores, onde a velocidade do vento foi utilizada com sua importância em 100% por três modelos sem exclusão de preditores. Outros três modelos foram usados na média do nanofluido com sua importância em 100%, ao contrário do modelo com melhor resultado, onde se utilizou a irradiância direta com sua importância em 100% e sem exclusão de preditores.

Tabela 8 – TiO₂ - 75 ppm (Variação de temperatura).

Modelos	RMSE (°C)	R ²	MAE	MBE
Modelo Físico	0,465	0,991	0,381	-0,041
Ridge	0,182	0,120	0,127	0,004
LASSO	0,182	0,121	0,127	0,004
ElasticNet	0,182	0,121	0,127	0,004
kNN	0,146	0,429	0,100	0,000
XGBoost	0,151	0,401	0,104	0,004
MARS	0,152	0,386	0,106	0,006
Ensemble Ridge	0,152	0,394	0,105	0,004
Ensemble MARS	0,165	0,328	0,113	0,004

Fonte: Elaborado pelo autor.

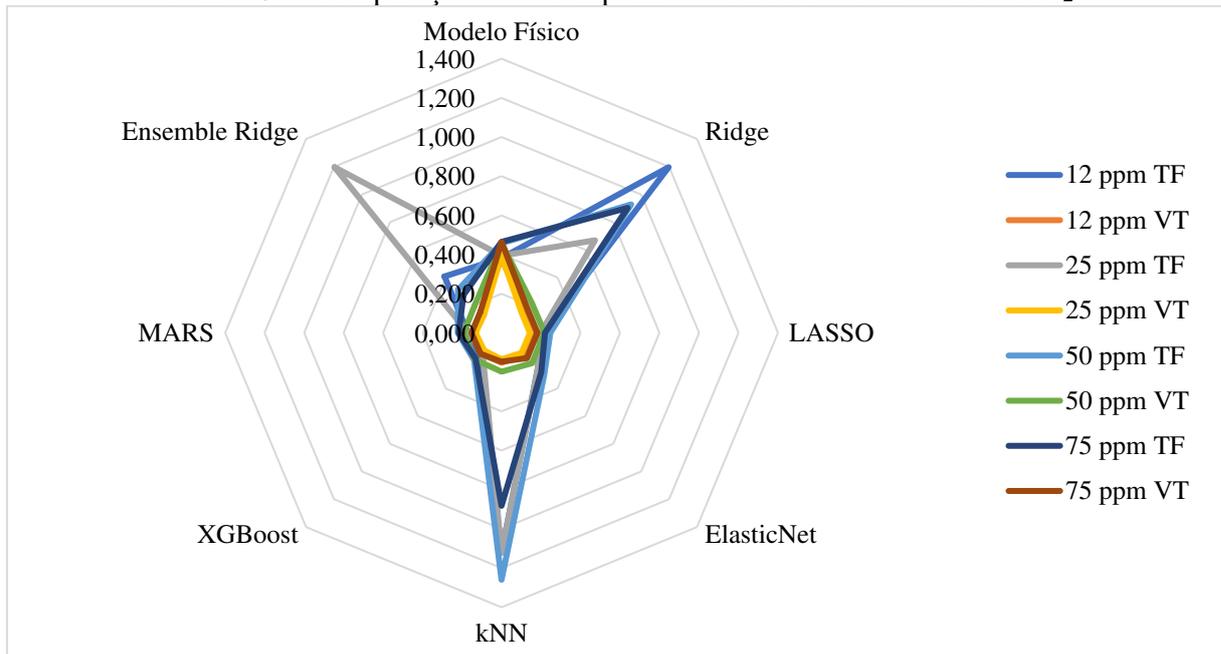
Gráfico 8 – Relação de importância dos preditores em 75 ppm (Variação de temperatura).



Fonte: Elaborado pelo autor.

Os resultados mostram que o modelo computacional teve maior eficiência em estimativa de variação de temperatura, onde os modelos testados foram superiores ao modelo físico nas quatro concentrações de nanofluido dióxido de titânio. Uma das dificuldades do modelo físico é o baixo índice de absorção de radiação e alta difusividade térmica do vidro apresentados por Santos (2019).

O modelo agregado *ensemble* Ridge mostrou resultados inferiores se comparado ao modelo físico em duas concentrações, 12,5 ppm e 25 ppm. No entanto, em maiores concentrações como 50 ppm e 75 ppm o modelo apresentou seus melhores resultados para estimativa de temperatura final.

Gráfico 9 – Comparação de desempenho do RMSE dos modelos - TiO₂

Fonte: Elaborado pelo autor.

Os modelos Ridge e kNN para estimativa de temperatura final e variação de temperatura nas diferentes concentrações de dióxido de titânio, apresentaram desempenhos inferiores comparados a outros modelos nas quatro concentrações (Gráfico 9). Estes modelos fizeram uso de um maior número de preditores e com poucas exclusões, em diferentes porcentagens de importância, ao contrário dos modelos com os melhores desempenhos.

4.5 Prata – 0,8125 ppm

Na concentração de 0,8125 ppm em estimativa de temperatura final, três modelos mostraram maior eficiência e com resultados próximos. O modelo *Ensemble* MARS apresentou o menor RMSE, como mostrado na Tabela 9, com valor de 0,230 °C seguido por XGBoost e MARS. Em trabalho semelhante usando Floresta Aleatórias, desenvolvido por Amorim Neto (2020) foi encontrado um valor RMSE de 0,562 °C e Ensemble MARS apresentando um desempenho superior de 59,02%.

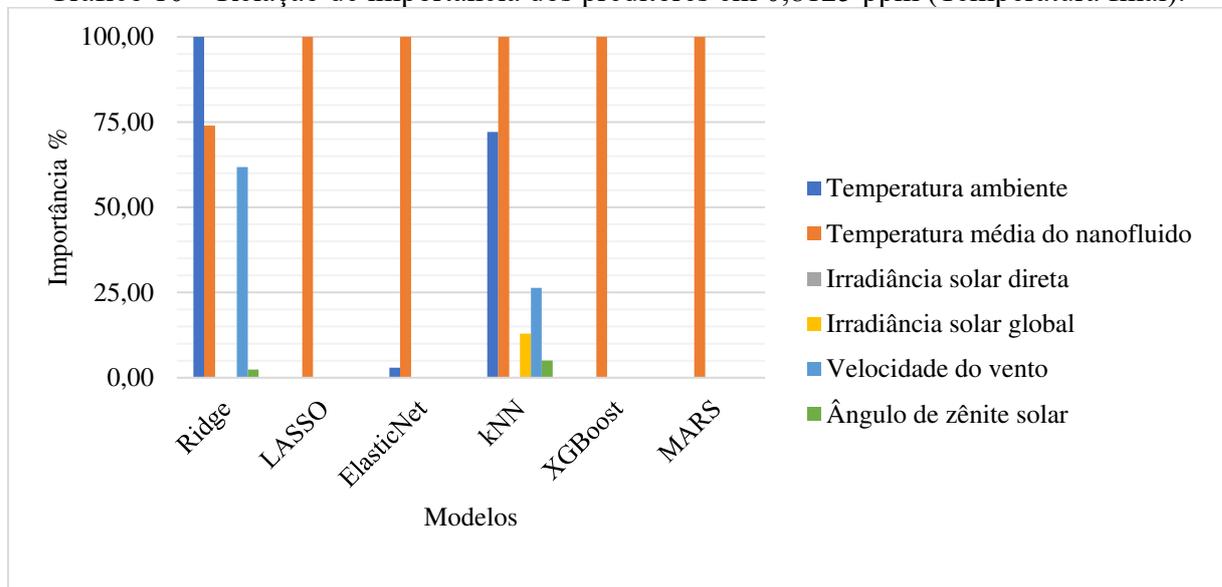
Como mostrado no Gráfico 10, modelos como Ridge e kNN que apresentaram desempenho inferiores de RMSE, fizeram um maior uso de preditores em diferentes porcentagens de importância, ao contrário dos modelos com os melhores RMSE, onde a temperatura média do nanofluido foi o preditor usado com maior frequência e com sua porcentagem de importância em 100%.

Tabela 9 – Prata - 0,8125 ppm (Temperatura final).

Modelos	RMSE (°C)	R ²	MAE	MBE
Ridge	2,056	0,969	1,746	0,122
LASSO	0,398	1,000	0,365	-0,207
<i>ElasticNet</i>	0,458	1,000	0,412	-0,222
kNN	2,152	0,966	1,631	-0,312
XGBoost	0,235	1,000	0,165	-0,061
MARS	0,245	1,000	0,185	-0,048
<i>Ensemble Ridge</i>	0,712	0,997	0,603	-0,143
<i>Ensemble MARS</i>	0,230	1,000	0,163	-0,060

Fonte: Elaborado pelo autor.

Gráfico 10 – Relação de importância dos preditores em 0,8125 ppm (Temperatura final).



Fonte: Elaborado pelo autor.

Como apresentado na Tabela 10, os quatro primeiros resultados de RMSE mostraram resultados semelhantes para estimativa de variação de temperatura, onde os modelos LASSO e *ElasticNet* apresentaram os maiores RMSE. Os quatro últimos resultados mostram os modelos com melhor desempenho, onde *Ensemble MARS* aparece com RMSE de 0,157 °C, se comparando a *ElasticNet* mostra uma eficiência de 45,49%.

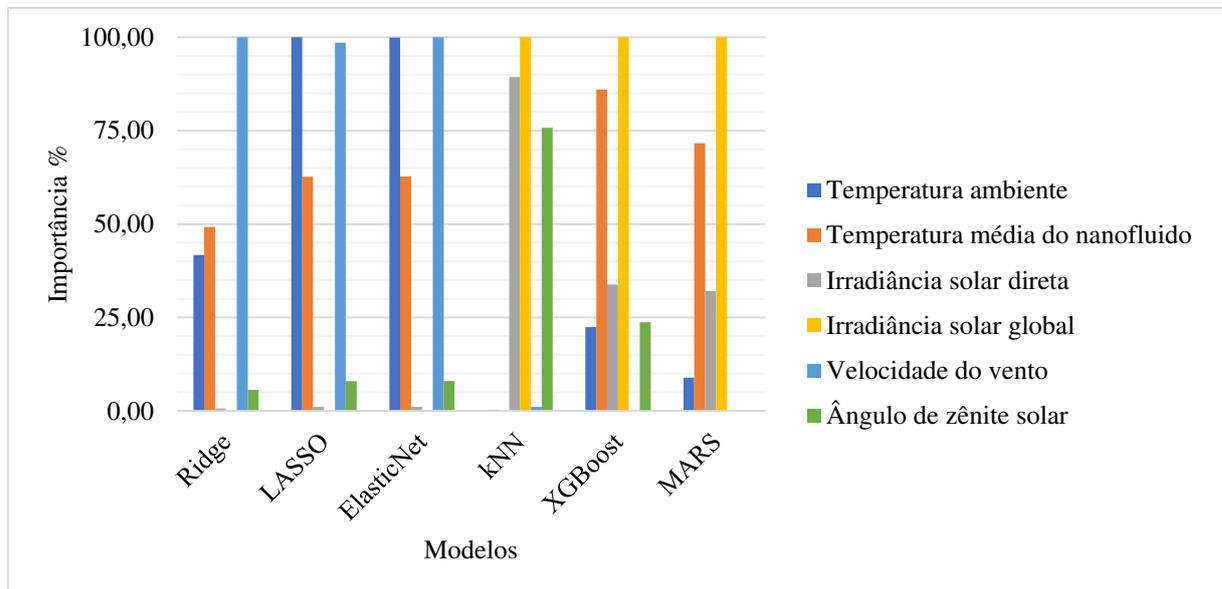
O preditor irradiância solar global foi usado por cinco modelos com sua porcentagem de importância em 100% (Gráfico 11). Os modelos Ridge e *ElasticNet* fizeram uso de cinco preditores, utilizando como preditor de maior importância a velocidade do vento e não fizeram o uso da irradiância solar global, ao contrário dos modelos com os menores RMSE.

Tabela 10 – Prata - 0,8125 ppm (Variação de temperatura).

Modelos	RMSE (°C)	R ²	MAE	MBE
Ridge	0,270	0,027	0,204	-0,003
LASSO	0,287	0,019	0,217	0,001
<i>ElasticNet</i>	0,287	0,019	0,217	0,001
kNN	0,233	0,146	0,158	-0,045
XGBoost	0,159	0,581	0,121	-0,017
MARS	0,182	0,471	0,131	-0,033
<i>Ensemble Ridge</i>	0,184	0,435	0,134	-0,029
<i>Ensemble MARS</i>	0,157	0,578	0,120	-0,008

Fonte: Elaborado pelo autor.

Gráfico 11 – Relação de importância dos preditores em 0,8125 ppm (Variação de temperatura).



Fonte: Elaborado pelo autor.

4.6 Prata – 1,625 ppm

Os modelos *Ensemble MARS* e XGBoost apresentaram os menores RMSE com 0,173 °C (Tabela 11), equivalente a 28,09% superior, como encontrado em trabalho semelhante de Abdul Hamid et al. (2016). O modelo Ridge apresentou resultado distante do melhor resultado encontrado na estimativa de temperatura final de 1,625 ppm.

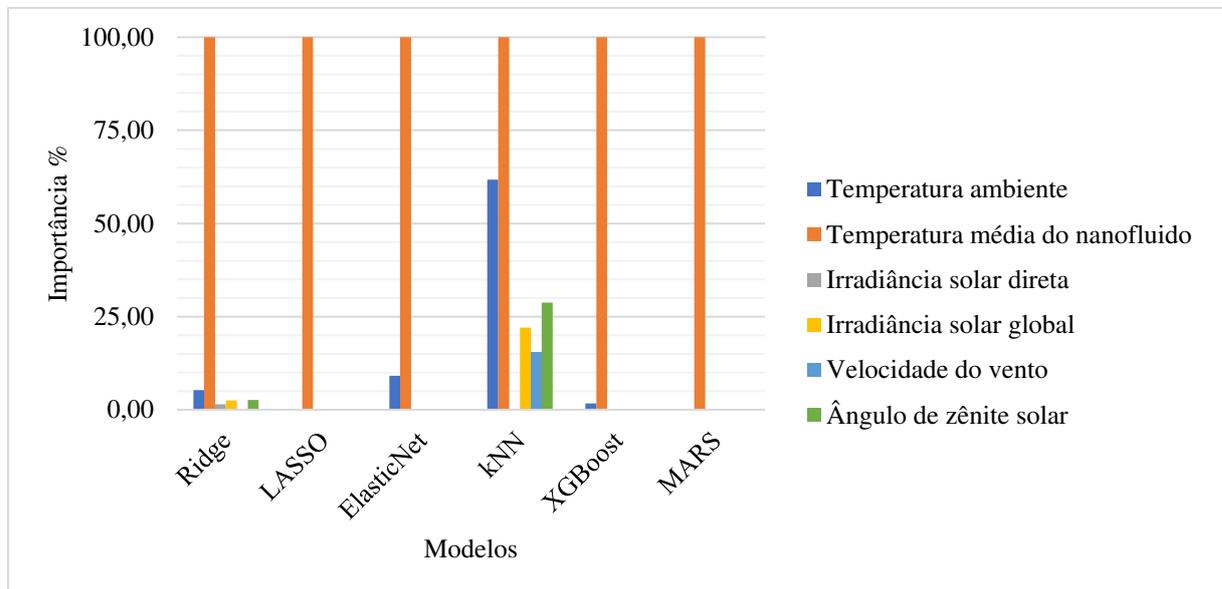
Os modelos Ridge e kNN fizeram uso da maior quantidade de preditores com apenas uma exclusão cada – velocidade do vento e irradiância solar direta, respectivamente. A temperatura média do nanofluido foi o preditor usado com sua importância em 100% por todos os modelos nessa concentração, como mostrado no Gráfico 12.

Tabela 11 – Prata - 1,625 ppm (Temperatura final).

Modelos	RMSE (°C)	R ²	MAE	MBE
Ridge	10,727	0,681	7,788	-7,564
LASSO	0,350	1,000	0,317	-0,049
<i>ElasticNet</i>	0,410	1,000	0,359	-0,071
kNN	6,880	0,662	5,200	2,727
XGBoost	0,173	1,000	0,127	-0,002
MARS	0,228	1,000	0,170	-0,001
<i>Ensemble Ridge</i>	1,430	0,994	1,076	-0,859
<i>Ensemble MARS</i>	0,173	1,000	0,127	-0,002

Fonte: Elaborado pelo autor.

Gráfico 12 – Relação de importância dos preditores em 1,625 ppm (Temperatura final).



Fonte: Elaborado pelo autor.

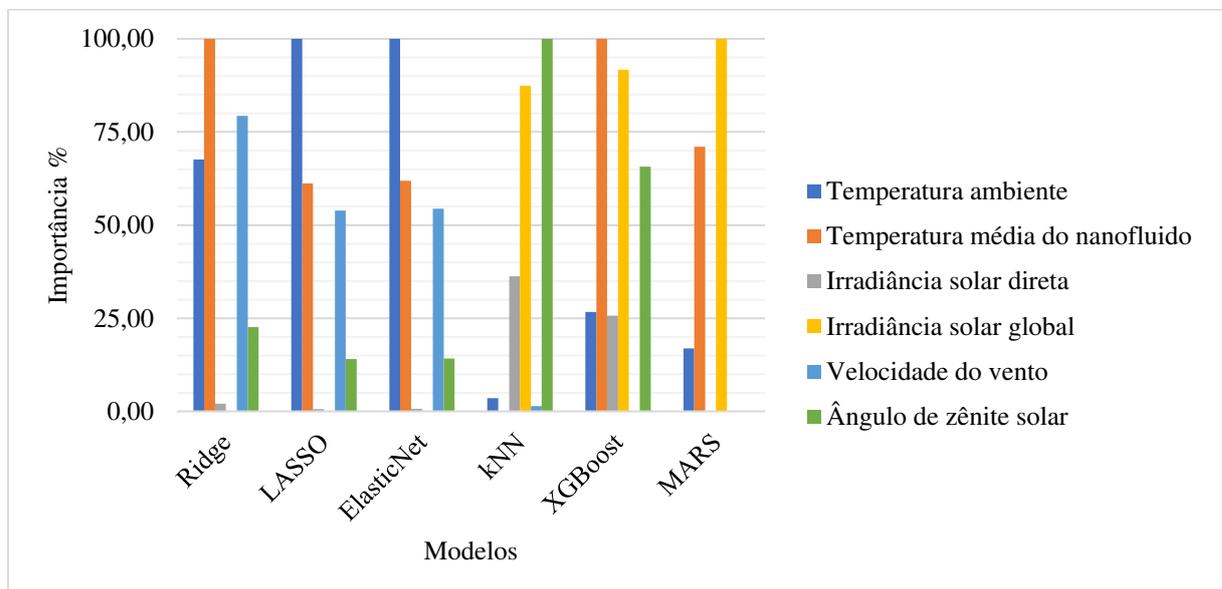
Na concentração de 1,625 ppm para estimativa de variação de temperatura (Tabela 12), os quatro últimos resultados apresentam os menores RMSE, respectivamente, bem como resultados semelhantes, variando apenas 3,13%, desempenhos estes próximos dos encontrados na concentração de 0,8125 ppm, onde *Ensemble MARS* mostrou o melhor resultado. O modelo *Ensemble MARS* usou como base o modelo XGBoost com a temperatura média do nanofluido como preditor de maior porcentagem de importância (Gráfico 13).

Tabela 12 – Prata - 1,625 ppm (Variação de temperatura).

Modelos	RMSE (°C)	R ²	MAE	MBE
Ridge	1,785	0,003	1,338	-1,246
LASSO	2,629	0,002	1,943	-1,844
<i>ElasticNet</i>	2,610	0,002	1,929	-1,830
kNN	0,360	0,086	0,271	-0,203
XGBoost	0,131	0,674	0,100	-0,011
MARS	0,141	0,624	0,106	-0,002
<i>Ensemble Ridge</i>	0,166	0,552	0,126	-0,062
<i>Ensemble MARS</i>	0,136	0,648	0,105	-0,009

Fonte: Elaborado pelo autor.

Gráfico 13 – Relação de importância dos preditores em 1,625 ppm (Variação de temperatura).



Fonte: Elaborado pelo autor.

4.7 Prata – 3,25 ppm

Os modelos *Ensemble MARS*, XGBoost, MARS para estimativa de temperatura final obtiveram os melhores resultados de RMSE, variando cerca de 0,024 °C (Tabela 13). Comparando somente os modelos de *Ensemble* na mesma concentração, a variação foi de 0,508 °C. O modelo *Ensemble MARS* mostrou uma eficiência de 67,43% em relação ao modelo *Ensemble Ridge*.

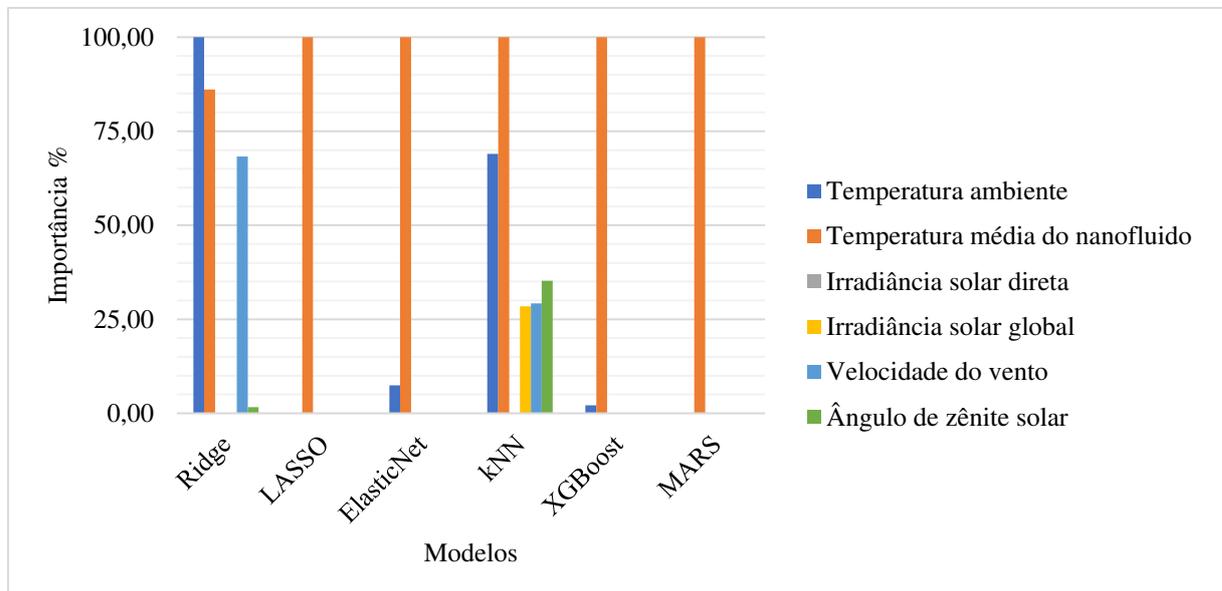
Como apresentado no Gráfico 14, apenas o modelo Ridge não utilizou a temperatura média do nanofluido com sua porcentagem de importância variável em 100%. No comparativo entre os modelos de *Ensemble*, Ridge fez a exclusão do modelo kNN e utilizou com maior porcentagem de importância *ElasticNet*, ao contrário do MARS que fez uso do XGBoost como modelo base.

Tabela 13 – Prata - 3,25 ppm (Temperatura final).

Modelos	RMSE (°C)	R ²	MAE	MBE
Ridge	1,787	0,984	1,501	-0,373
LASSO	0,389	1,000	0,353	-0,059
<i>ElasticNet</i>	0,485	0,999	0,434	-0,097
kNN	2,768	0,957	1,906	-0,916
XGBoost	0,250	1,000	0,154	-0,044
MARS	0,270	1,000	0,210	-0,004
<i>Ensemble Ridge</i>	0,754	0,998	0,627	-0,244
<i>Ensemble MARS</i>	0,246	1,000	0,152	-0,043

Fonte: Elaborado pelo autor.

Gráfico 14 – Relação de importância dos preditores em 3,25 ppm (Temperatura final).



Fonte: Elaborado pelo autor.

Como indica a Tabela 14, em estimativa variação de temperatura na concentração de 3,25 ppm, os modelos Ridge, LASSO e *ElasticNet* apresentaram os resultados semelhantes de RMSE, como pode ser analisado no Gráfico 15, em que a velocidade do vento foi o preditor utilizado com sua porcentagem de importância variável em 100%.

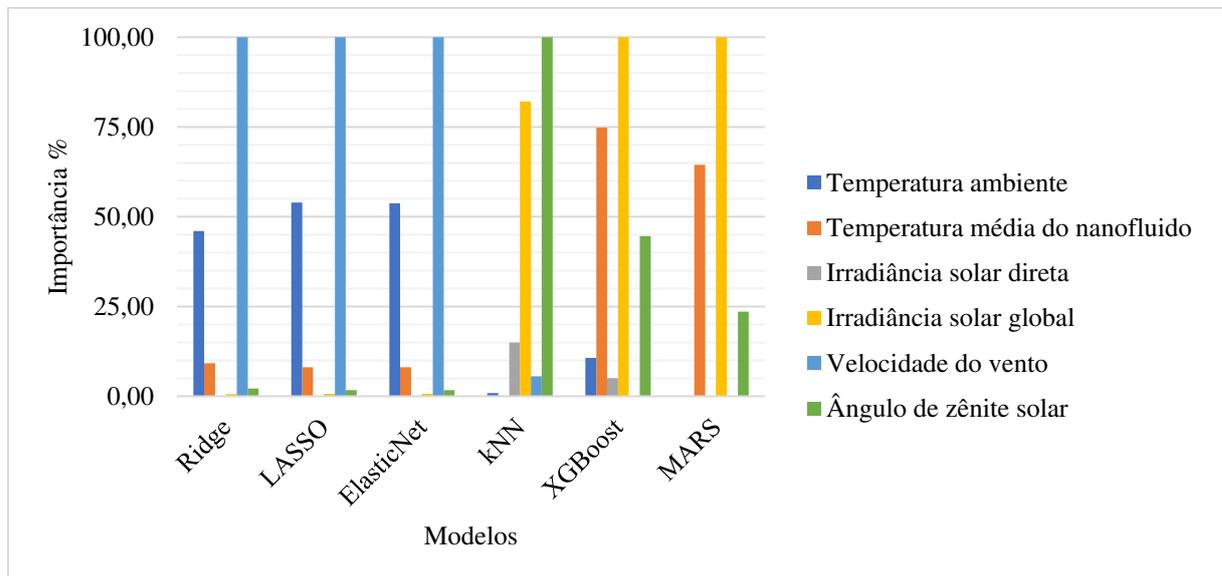
Os quatros últimos modelos apresentados na Tabela 14, mostram resultados próximos variando cerca de 0,01 °C, onde fizeram o uso do preditor irradiância solar global com sua porcentagem de importância em 100%. O modelo MARS com RMSE de 0,119 °C apresentou o melhor desempenho.

Tabela 14 – Prata - 3,25 ppm (Variação de temperatura).

Modelos	RMSE (°C)	R ²	MAE	MBE
Ridge	0,249	0,184	0,197	0,030
LASSO	0,249	0,195	0,198	0,034
<i>ElasticNet</i>	0,249	0,195	0,199	0,034
kNN	0,155	0,692	0,112	-0,031
XGBoost	0,125	0,796	0,085	-0,019
MARS	0,119	0,834	0,086	-0,027
<i>Ensemble Ridge</i>	0,124	0,804	0,085	-0,022
<i>Ensemble MARS</i>	0,129	0,777	0,088	-0,017

Fonte: Elaborado pelo autor.

Gráfico 15 – Relação de importância dos preditores em 3,25 ppm (Variação de temperatura).



Fonte: Elaborado pelo autor.

4.8 Prata – 6,5 ppm

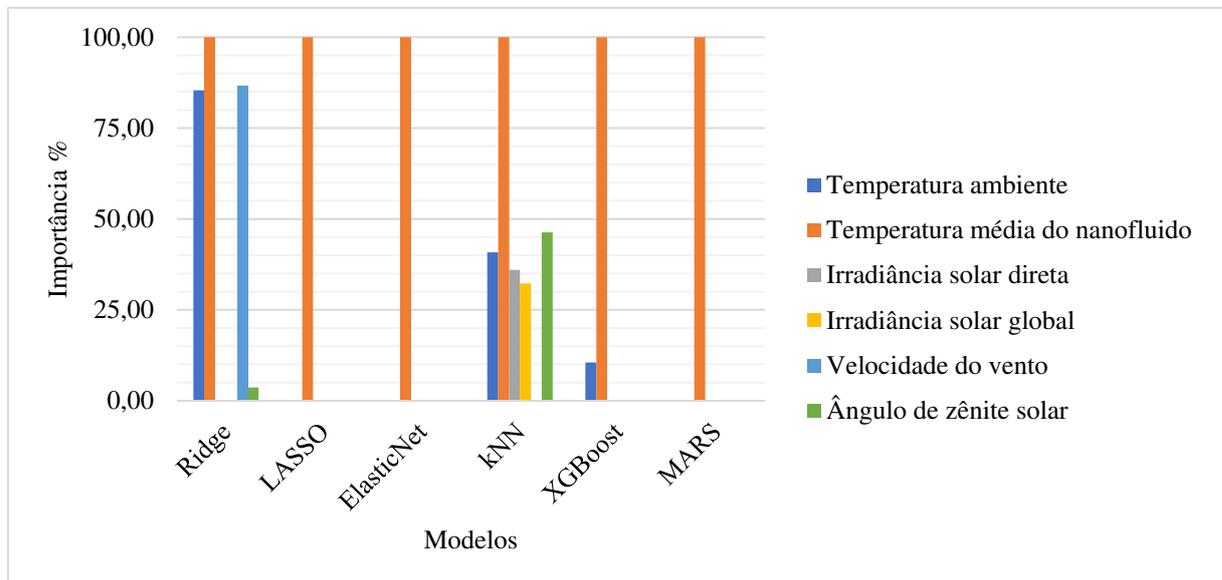
A Tabela 15 mostra que na estimativa de temperatura final o modelo MARS obteve o melhor desempenho RMSE de 0,239 °C. Todos os modelos apresentados fizeram uso da temperatura média do nanofluido com sua porcentagem de importância variável em 100% (Gráfico 16). Os modelos Ridge e kNN fizeram exclusão de apenas um preditor – irradiância solar global e velocidade do vento, respectivamente; ao contrário dos modelos com os melhores desempenhos nessa concentração, que fizeram uso exclusivo da temperatura média do nanofluido.

Tabela 15 – Prata - 6,5 ppm (Temperatura final).

Modelos	RMSE (°C)	R ²	MAE	MBE
Ridge	2,460	0,978	2,030	0,128
LASSO	0,462	1,000	0,403	0,124
<i>ElasticNet</i>	0,623	1,000	0,539	0,095
kNN	3,515	0,934	2,535	0,219
XGBoost	0,718	0,997	0,596	0,099
MARS	0,239	1,000	0,183	0,009
<i>Ensemble Ridge</i>	1,052	0,995	0,838	0,126
<i>Ensemble MARS</i>	0,718	0,997	0,596	0,596

Fonte: Elaborado pelo autor.

Gráfico 16 – Relação de importância dos preditores em 6,5 ppm (Temperatura final).



Fonte: Elaborado pelo autor.

Os desempenhos dos modelos na concentração 6,5 ppm pra estimativa de variação de temperatura mostraram resultados próximos dos encontrados na concentração 3,25 ppm, valores semelhantes apresentados no trabalho de Abdul Hamid et al. (2016), onde o modelo XGBoost mostrou uma melhoria de 22,89% (Tabela 16). Entre os modelos de *Ensemble* ocorreu uma variação de 0,015 °C, correspondendo a 7,44 %.

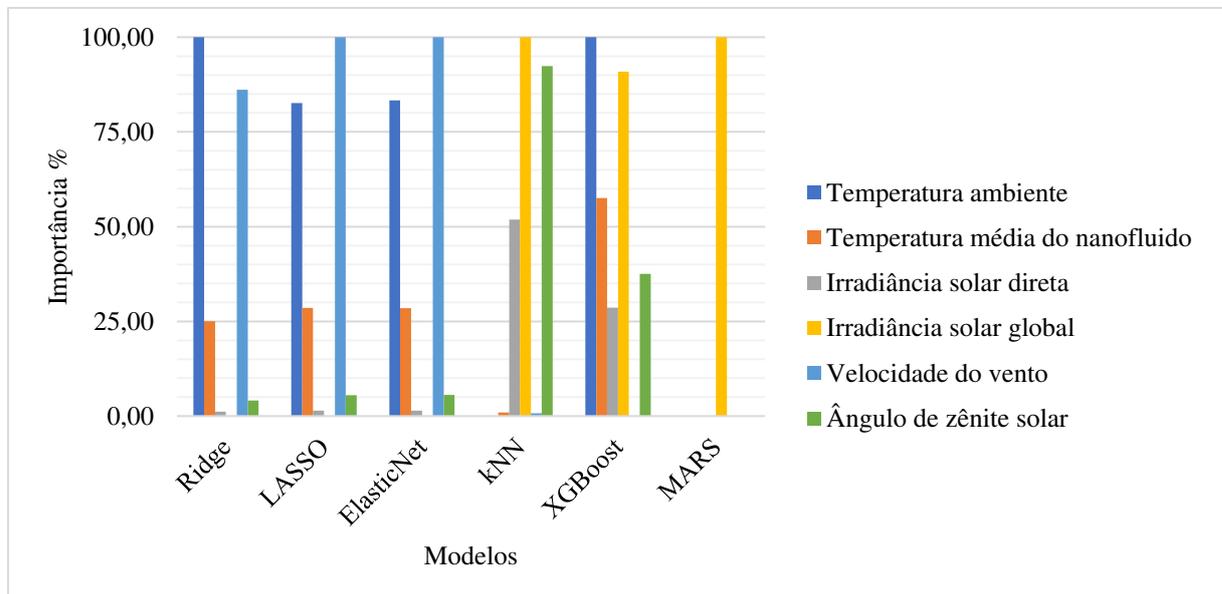
O modelo XGBoost fez o uso preditor temperatura ambiente em sua porcentagem de importância variável em 100%, com apenas a exclusão do preditor velocidade do vento como mostra o Gráfico 17. Os modelos *Ensemble* usaram com maior importância o modelo XGBoost.

Tabela 16 – Prata - 6,5 ppm (Variação de temperatura).

Modelos	RMSE (°C)	R ²	MAE	MBE
Ridge	0,296	0,123	0,236	-0,038
LASSO	0,321	0,115	0,254	-0,041
<i>ElasticNet</i>	0,320	0,115	0,253	-0,040
kNN	0,260	0,439	0,190	-0,017
XGBoost	0,185	0,695	0,132	0,050
MARS	0,314	0,461	0,202	0,123
<i>Ensemble Ridge</i>	0,202	0,612	0,147	0,043
<i>Ensemble MARS</i>	0,187	0,696	0,133	0,051

Fonte: Elaborado pelo autor.

Gráfico 17 – Relação de importância dos preditores em 6,5 ppm (Variação de temperatura).

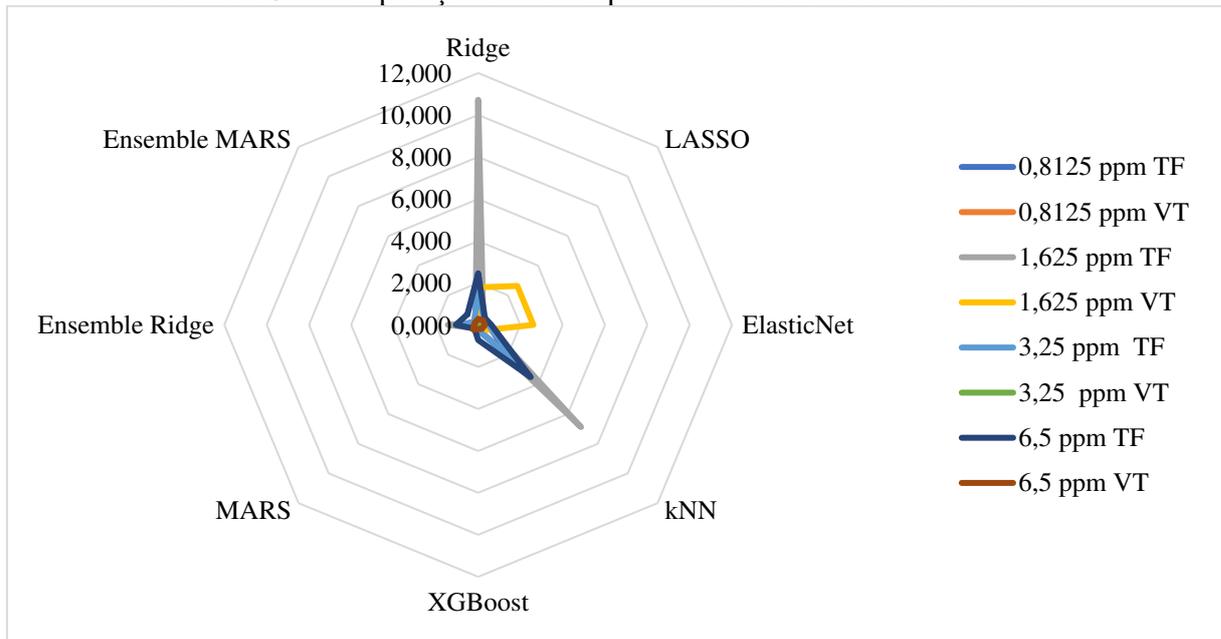


Fonte: Elaborado pelo autor.

Os resultados mostram que o modelo computacional para nanopartículas de prata teve maior eficiência em estimativa de variação de temperatura. O modelo agregado *Ensemble MARS* apresentou os menores RMSE em 62,5% dos casos testados, onde o modelo trabalha com exclusão dos resultados dos modelos menos eficientes.

Os resultados encontrados são condizentes com o estudo de Hadavimoghaddam et al. (2022), que fizeram uma pesquisa para comparar o desempenho e acurácia dos modelos mais famosos da literatura para predição.

Gráfico 18 – Comparação de desempenho do RMSE dos modelos - Prata



Fonte: Elaborado pelo autor.

Os modelos Ridge e kNN apresentaram desempenhos inferiores na concentração de 1,625 ppm, para estimativa de temperatura final. Foram obtidos os maiores valores de RMSE em nanopartículas de prata, tais modelos trabalham com maior uso de preditores e com poucas exclusões, mas de maneira geral os modelos demonstraram desempenhos próximos de RMSE, independente da concentração (Gráfico 18).

5 CONCLUSÃO

No trabalho foi realizada a análise computacional do desempenho de modelos de aprendizado de máquina para a estimativa de temperatura em coletores solares de quatro concentrações de nanopartículas de dióxido de titânio (12,5 ppm; 25 ppm; 50 ppm; 75 ppm) e nanopartículas de prata (0,8125 ppm; 1,625 ppm; 3,25 ppm; 6,5 ppm), dispersas em água deionizada em um sistema de absorção direta de radiação solar. No estudo com o dióxido de titânio o modelo computacional mostrou uma eficiência superior ao modelo físico em todos os testes para estimativa de variação de temperatura. Já para a análise das nanopartículas de prata mostrou resultados semelhantes ao identificado na literatura e condizentes aos encontrados em nanopartículas de dióxido de titânio.

Os resultados apontam que, dentre os modelos analisados, o modelo kNN mostrou o melhor desempenho RMSE para estimativa de variação de temperatura. O melhor RMSE foi encontrado na concentração de 75 ppm de dióxido de titânio, com valor de 0,146 °C. Assim, o modelo computacional apresentou uma eficiência 68,56% superior em relação ao modelo físico. O melhor desempenho RMSE nos estudos de nanopartículas de prata foi encontrado na concentração de 1,625 ppm em estimativa de variação de temperatura com o modelo *Ensemble* MARS, com cerca de 0,136 °C.

O modelo agregado *ensemble* MARS apresentou os melhores resultados nas quatro concentrações de nanofluido de dióxido de titânio para estimativa de temperatura final, no qual fez o uso do preditor média do nanofluido com sua porcentagem de importância em 100%. O modelo XGBoost apresentou os melhores desempenhos em duas concentrações de nanofluido dióxido de titânio para 25 ppm e 50 ppm em estimativa de variação de temperatura, com a média do nanofluido como principal preditor em porcentagem de importância.

O modelo agregado *Ensemble* MARS, que trabalha com exclusão dos resultados dos modelos menos eficientes, apresentou os menores RMSE em 62,5%, seguido por MARS 25,0% e XGboost 12,5% nos casos testados em nanopartículas de prata.

O modelo computacional se mostrou eficiente em comparação ao modelo físico, alcançando o objetivo desta pesquisa para análises de estimativa de variação de temperatura nas concentrações de nanopartículas de dióxido de titânio. Nas concentrações de nanopartículas de prata os resultados são condizentes com a literatura, se mostrando assim adequado para o objetivo desse estudo.

REFERÊNCIAS

- ABDUL HAMID, K.; AZMI, W. H.; MAMAT, Rizalman; USRI, N. A. Thermal conductivity enhancement of TiO₂ nanofluid in water and ethylene glycol (EG) mixture. **Indian Journal of Pure & Applied Physics (IJPAP)**, [S. l.], v. 54, n. 10, p. 651–655, 2016. DOI: 10.56042/IJPAP.V54I10.11802. Disponível em: <http://op.niscpr.res.in/index.php/IJPAP/article/view/11802/955>. Acesso em: 1 jul. 2022.
- AĞBULUT, Ümit; GÜREL, Ali Etem; ERGÜN, Alper; CEYLAN, İlhan. Performance assessment of a V-trough photovoltaic system and prediction of power output with different machine learning algorithms. **Journal of Cleaner Production**, [S. l.], v. 268, p. 122269, 2020. DOI: 10.1016/J.JCLEPRO.2020.122269.
- AHMAD, Muhammad Waseem; REYNOLDS, Jonathan; REZGUI, Yacine. Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. **Journal of Cleaner Production**, [S. l.], v. 203, p. 810–821, 2018. DOI: 10.1016/j.jclepro.2018.08.207.
- AL-OBEIDAT, Feras; SPENCER, Bruce; ALFANDI, Omar. Consistently accurate forecasts of temperature within buildings from sensor data using ridge and lasso regression. **Future Generation Computer Systems**, [S. l.], v. 110, p. 382–392, 2020. DOI: 10.1016/j.future.2018.02.035. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0167739X17327541>. Acesso em: 23 mar. 2021.
- ALADE, Ibrahim Olanrewaju; ABD RAHMAN, Mohd Amiruddin; SALEH, Tawfik A. Predicting the specific heat capacity of alumina/ethylene glycol nanofluids using support vector regression model optimized with Bayesian algorithm. **Solar Energy**, [S. l.], v. 183, p. 74–82, 2019. DOI: 10.1016/j.solener.2019.02.060.
- ALI, Naser; TEIXEIRA, Joao A.; ADDALI, Abdulmajid. **A Review on Nanofluids: Fabrication, Stability, and Thermophysical Properties**. *Journal of Nanomaterials* Hindawi Limited, , 2018. DOI: 10.1155/2018/6978130.
- AMJAD, Muhammad; JIN, Haichuan; DU, Xiaoze; WEN, Dongsheng. Experimental photothermal performance of nanofluids under concentrated solar flux. **Solar Energy Materials and Solar Cells**, [S. l.], v. 182, p. 255–262, 2018. DOI: 10.1016/j.solmat.2018.03.044.
- AMORIM NETO, Juarez Pompeu De. **Modelagem numérica e experimental de uma “parede solar” utilizando nanofluidos de prata e dióxido de titânio**. 2020. Universidade Federal do Ceará, Fortaleza, 2020. Disponível em: <http://www.repositorio.ufc.br/handle/riufc/52003>. Acesso em: 8 mar. 2021.
- ASKARZADEH, Alireza; ZEBARJADI, Masoud. Wind power modeling using harmony search with a novel parameter setting approach. **Journal of Wind Engineering and Industrial Aerodynamics**, [S. l.], v. 135, p. 70–75, 2014. DOI: 10.1016/J.JWEIA.2014.10.012.
- BANDARRA FILHO, Enio Pedone; MENDOZA, Oscar Saúl Hernandez; BEICKER, Carolina Lau Lins; MENEZES, Adonis; WEN, Dongsheng. Experimental investigation of a silver nanoparticle-based direct absorption solar thermal system. **Energy Conversion and**

Management, [*S. l.*], v. 84, p. 261–267, 2014. DOI: 10.1016/j.enconman.2014.04.009.

BAZDIDI-TEHRANI, Farzad; KHABAZIPUR, Arash; VASEFI, Seyed Iman. Flow and heat transfer analysis of TiO₂/water nanofluid in a ribbed flat-plate solar collector. **Renewable Energy**, [*S. l.*], v. 122, p. 406–418, 2018. DOI: 10.1016/j.renene.2018.01.056.

BUONOMO, Bernardo; MANCA, Oronzio; MARINELLI, Lorenzo; NARDINI, Sergio. Effect of temperature and sonication time on nanofluid thermal conductivity measurements by nano-flash method. **Applied Thermal Engineering**, [*S. l.*], v. 91, p. 181–190, 2015. DOI: 10.1016/j.applthermaleng.2015.07.077.

CHEN, Tianqi; GUESTRIN, Carlos. XGBoost: A Scalable Tree Boosting System. **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**, New York, NY, USA, p. 785–794, 2016. DOI: 10.1145/2939672.

DE VRIES, Sjoerd; THIERENS, Dirk. A reliable ensemble based approach to semi-supervised learning. **Knowledge-Based Systems**, [*S. l.*], v. 215, p. 106738, 2021. DOI: 10.1016/J.KNOSYS.2021.106738.

DUFFIE, John A.; BECKMAN, William A. **Solar Engineering of Thermal Processes**. 4. ed. Hoboken: John Wiley & Sons, 2013.

FAN, Cheng; XIAO, Fu; WANG, Shengwei. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. **Applied Energy**, [*S. l.*], v. 127, p. 1–10, 2014. DOI: 10.1016/J.APENERGY.2014.04.016.

FANG, Tingting; LAHDELMA, Risto. Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. **Applied Energy**, [*S. l.*], v. 179, p. 544–552, 2016. DOI: 10.1016/j.apenergy.2016.06.133.

FRIEDMAN, Jerome H. Multivariate Adaptive Regression Splines. **The Annals of Statistics**, [*S. l.*], v. 19, n. 1, p. 1–67, 1991. DOI: 10.1214/AOS/1176347963.

FRIEDMAN, Jerome H. Stochastic gradient boosting. **Computational Statistics & Data Analysis**, [*S. l.*], v. 38, n. 4, p. 367–378, 2002. DOI: 10.1016/S0167-9473(01)00065-2.

GEEM, Zong Woo; KIM, Joong Hoon; LOGANATHAN, G. V. A New Heuristic Optimization Algorithm: Harmony Search: <http://dx.doi.org/10.1177/003754970107600201>, [*S. l.*], v. 76, n. 2, p. 60–68, 2001. DOI: 10.1177/003754970107600201. Disponível em: <https://journals.sagepub.com/doi/10.1177/003754970107600201>. Acesso em: 7 jul. 2021.

GUPTA, Hemant Kumar; AGRAWAL, Ghanshyam Das; MATHUR, Jyotirmay. An experimental investigation of a low temperature Al₂O₃-H₂O nanofluid based direct absorption solar collector. **Solar Energy**, [*S. l.*], v. 118, p. 390–396, 2015. DOI: 10.1016/j.solener.2015.04.041.

HADAVIMOGHADDAM, Fahimeh; ATASHROUZ, Saeid; REZAEI, Farzaneh; MUNIR, Muhammad Tajammal; HEMMATI-SARAPARDEH, Abdolhossein; MOHADDESPOUR, Ahmad. Modeling thermal conductivity of nanofluids using advanced correlative approaches: Group method of data handling and gene expression programming. **International Communications in Heat and Mass Transfer**, [*S. l.*], v. 131, p. 105818, 2022. DOI:

10.1016/J.ICHEATMASSTRANSFER.2021.105818.

HUANG, Jiun Chi; TSAI, Yi Chun; WU, Pei Yu; LIEN, Yu Hui; CHIEN, Chih Yi; KUO, Chih Feng; HUNG, Jeng Fung; CHEN, Szu Chia; KUO, Chao Hung. Predictive modeling of blood pressure during hemodialysis: a comparison of linear model, random forest, support vector regression, XGBoost, LASSO regression and ensemble method. **Computer Methods and Programs in Biomedicine**, [S. l.], v. 195, p. 105536, 2020. DOI: 10.1016/J.CMPB.2020.105536.

JAMES, Gareth; WITTEN, Daniela; HASTIE, Trevor; TIBSHIRANI, Robert. **An Introduction to Statistical Learning**. New York, NY: Springer, 2013. v. 103 DOI: 10.1007/978-1-4614-7138-7.

KHANLARI, Ataollah; SÖZEN, Adnan; ŞIRIN, Ceylin; TUNCER, Azim Doğu; GUNGOR, Afsin. Performance enhancement of a greenhouse dryer: Analysis of a cost-effective alternative solar air heater. **Journal of Cleaner Production**, [S. l.], v. 251, p. 119672, 2020. DOI: 10.1016/J.JCLEPRO.2019.119672.

KIM, Hansu; LEE, Tae Hee. A robust elastic net via bootstrap method under sampling uncertainty for significance analysis of high-dimensional design problems. **Knowledge-Based Systems**, [S. l.], v. 225, p. 107117, 2021. DOI: 10.1016/j.knosys.2021.107117.

LAN, Xixi; ZOU, Changchun; KANG, Zhihong; WU, Xingwei. Log facies identification in carbonate reservoirs using multiclass semi-supervised learning strategy. **Fuel**, [S. l.], v. 302, p. 121145, 2021. DOI: 10.1016/J.FUEL.2021.121145.

LI, Danny H. W.; CHEN, Wenqiang; LI, Shuyang; LOU, Siwei. Estimation of hourly global solar radiation using Multivariate Adaptive Regression Spline (MARS) – A case study of Hong Kong. **Energy**, [S. l.], v. 186, p. 115857, 2019. DOI: 10.1016/j.energy.2019.115857.

LIU, Qibin; YANG, Minlin; LEI, Jing; JIN, Hongguang; GAO, Zhichao; WANG, Yalong. Modeling and optimizing parabolic trough solar collector systems using the least squares support vector machine method. **Solar Energy**, [S. l.], v. 86, n. 7, p. 1973–1980, 2012. DOI: 10.1016/j.solener.2012.01.026.

OSMAN, Ibrahim H.; LAPORTE, Gilbert. Metaheuristics: A bibliography. **Annals of Operations Research** 1996 63:5, [S. l.], v. 63, n. 5, p. 511–623, 1996. DOI: 10.1007/BF02125421. Disponível em: <https://link.springer.com/article/10.1007/BF02125421>. Acesso em: 7 jul. 2021.

OZSOY, Ahmet; CORUMLU, Vahit. Thermal performance of a thermosyphon heat pipe evacuated tube solar collector using silver-water nanofluid for commercial applications. **Renewable Energy**, [S. l.], v. 122, p. 26–34, 2018. DOI: 10.1016/j.renene.2018.01.031.

SANTOS, Rogério Feliciano de Moura. **Balanco de energia e avaliação heurística de parâmetros de absorção de um sistema solar fototérmico baseado no nanofluido dióxido de titânio-água deionizada**. 2019. Universidade Federal do Ceará, Fortaleza, 2019. Disponível em: <http://www.repositorio.ufc.br/handle/riufc/43062>. Acesso em: 20 fev. 2020.

SURESH, S.; VENKITARAJ, K. P.; SELVAKUMAR, P.; CHANDRASEKAR, M. Synthesis of Al₂O₃-Cu/water hybrid nanofluids using two step method and its thermo physical properties. **Colloids and Surfaces A: Physicochemical and Engineering Aspects**, [S. l.], v.

388, n. 1–3, p. 41–48, 2011. DOI: 10.1016/j.colsurfa.2011.08.005.

VOYANT, Cyril; NOTTON, Gilles; KALOGIROU, Soteris; NIVET, Marie Laure; PAOLI, Christophe; MOTTE, Fabrice; FOUILLOY, Alexis. **Machine learning methods for solar radiation forecasting: A review** *Renewable Energy* Elsevier Ltd, , 2017. DOI: 10.1016/j.renene.2016.12.095.

WANG, Han; YANG, Weimin; CHENG, Lisheng; GUAN, Changfeng; YAN, Hua. Chinese ink: High performance nanofluids for solar energy. *Solar Energy Materials and Solar Cells*, [S. l.], v. 176, p. 374–380, 2018. DOI: 10.1016/j.solmat.2017.10.023.

YU, Wei; XIE, Huaqing. A Review on Nanofluids: Preparation, Stability Mechanisms, and Applications. *Journal of Nanomaterials*, [S. l.], v. 2012, p. 17, 2012. DOI: 10.1155/2012/435873.

ZHONG, Wei; HUANG, Wei; LIN, Xiaojie; LI, Zhongbo; ZHOU, Yi. Research on data-driven identification and prediction of heat response time of urban centralized heating system. *Energy*, [S. l.], v. 212, p. 118742, 2020. DOI: 10.1016/J.ENERGY.2020.118742.