

UNIVERSIDADE FEDERAL DE VIÇOSA

NEULLER ALVES PEREIRA

**MODELOS DE MACHINE LEARNING NA PREVISÃO DA
EVAPOTRANSPIRAÇÃO DE REFERÊNCIA EM CURTO PRAZO**

**VIÇOSA – MINAS GERAIS
2022**

NEULLER ALVES PEREIRA

**MODELOS DE MACHINE LEARNING NA PREVISÃO DA
EVAPOTRANSPIRAÇÃO DE REFERÊNCIA EM CURTO PRAZO**

Relatório final, apresentado à Universidade Federal de Viçosa, como parte das exigências para a obtenção do título de Engenheiro Agrícola e Ambiental.

Orientador: Rubens Alves de Oliveira

Coorientador: Andre Luiz de Freitas Coelho

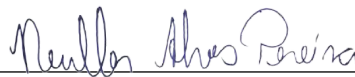
NEULLER ALVES PEREIRA

**MODELOS DE MACHINE LEARNING NA PREVISÃO DA
EVAPOTRANSPIRAÇÃO DE REFERÊNCIA EM CURTO PRAZO**

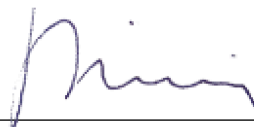
Relatório final, apresentado a Universidade Federal de Viçosa, como parte das exigências, para obtenção do título de Engenheiro Agrícola e Ambiental.

APROVADO: 02 de agosto de 2022.

Assentimento:



Neuller Alves Pereira
Autor



Rubens Alves de Oliveira
Orientador

“Confia ao Senhor as tuas obras, e teus pensamentos serão estabelecidos”

(Provérbios 16:3)

RESUMO

A obtenção da evapotranspiração de referência (ET_o) é útil em diversos processos relacionados com planejamento e gestão dos recursos hídricos, zoneamento agroclimático, projeto e manejo de irrigação, dentre outros. A possibilidade de se estimar a ET_o antes de sua ocorrência, abre grandes margens para a otimização dos diversos processos que dependem desta variável. Pensando nisso, objetivou-se com o presente trabalho analisar o desempenho do modelo Facebook Prophet e do framework Auto-sklearn na previsão de ET_o de curto prazo. Para isso, valores históricos de ET_o foram calculados com dados obtidos da estação meteorológica do Instituto Nacional de Meteorologia localizada no município de Viçosa (MG), e os modelos foram treinados com base em dados climáticos registrados no dia anterior à ocorrência da ET_o. Além disso, para verificar o potencial da utilização de variáveis que podem ser obtidas a partir de modelos de previsão do tempo, fez-se a adição de variáveis climáticas reais do dia de ocorrência da evapotranspiração ao conjunto de treinamento. Os melhores modelos de regressão obtidos com a execução do Auto-sklearn foram também treinados individualmente para cada uma das estações e meses do ano. Para o cenário de treinamento de um modelo anual, a adição das variáveis reais do dia de ocorrência da ET_o propiciou a diminuição da variação da raiz quadrada do erro médio (RMSE) de 0,73 para 0,47 (35,6%), e aumento do coeficiente de determinação (R²) de 0,61 para 0,84 (37,7%). Quando os modelos foram treinados mensalmente, a adição dessas variáveis levou à melhoria de desempenho em alguns meses e piora em outros. Para o treinamento baseado nas estações do ano, todos os indicadores apresentaram melhoria com a adição das variáveis do dia de previsão. Em todos os cenários de treinamentos anuais, os modelos de regressão encontrados com o Auto-Sklearn apresentaram melhores valores RMSE e R² que o modelo de *time series* Facebook Prophet. A utilização de um único modelo para todos os dias do ano apresentou resultados mais satisfatórios que os modelos treinados de forma mensal e modelos com base nas quatro estações do ano.

Palavras-chave: Auto-sklearn. Facebook Prophet. Dados climáticos, Evaporação, Transpiração.

ABSTRACT

The knowledge of the reference evapotranspiration (ET_o) is useful in many processes involving water resources planning and management, climate zoning, irrigation operation, and sizing, among others. The possibility to estimate the ET_o before its occurrence opens opportunities for optimizing the many processes that depend on this variable. Thinking about that, this work has the objective to analyze the performance of the Facebook Prophet model and the framework Auto-sklearn at short-term ET_o prediction. This way, historical values of ET_o were calculated, using the meteorological data obtained from the Brazilian National Institute of Meteorology station located in the municipality of Viçosa, Minas Gerais, Brazil, and the models have been trained with the weather data from the day before ET_o occurrence. Besides that, to look into the applicability of variables that can be obtained from weather prediction models, climate data from the evapotranspiration day of occurrence was added to the training set. The best regression models obtained with the Auto-sklearn application were trained as well individually for each of the stations and months of the year. For the training of the annual model, the addition of the evapotranspiration day variables has contributed to the decrease of root-mean-square deviation (RMSE) from 0,73 to 0,46 (35,6%) and the increase of coefficient of determination (R²) from 0,61 to 0,84 (37,7%). When the models were trained monthly, the addition of these variables led to improved performance on some months and worse on others. At training based on seasons, all the indicators showed improvement with the addition of prediction day variables. Considering all annual model training scenarios, the regression models encountered by Auto-sklearn presented better values of RMSE and R² than the Facebook Prophet time series model. Using only one model for all the days of the year presented more satisfying results than the models trained monthly and based on the four seasons models.

Keywords: Auto-sklearn. Facebook Prophet. Climatic data, Evaporation, Transpiration.

SUMÁRIO

1. INTRODUÇÃO	7
2. MATERIAL E MÉTODOS	10
2.1. Cálculo da evapotranspiração de referência	10
2.2. Algoritmo para coleta dos dados e cálculo da ETo	10
2.3. Limpeza e processamento dos dados	11
2.4. Cenários de análise	12
3. RESULTADOS E DISCUSSÃO	14
4. CONCLUSÕES	24
REFERÊNCIAS	25

1. INTRODUÇÃO

Evapotranspiração é o termo utilizado para representar o processo pelo qual a água é transportada da superfície terrestre para a atmosfera. Em resumo, este fenômeno pode ser dividido em duas partes: a evaporação direta e a transpiração das plantas. Quando a evapotranspiração é obtida em uma superfície padrão hipotética coberta com uma cultura (grama) que possui altura de 0,12 metros, albedo de 0,23 e está em situação de ausência de estresse hídrico, tem-se a chamada evapotranspiração de referência (ET_o) (CARVALHO et al., 2011).

A obtenção da evapotranspiração das culturas é muito importante para o acompanhamento da perda de água da superfície terrestre de uma determinada região, sendo útil em processos relacionados com planejamento e gestão dos recursos hídricos, zoneamento agroclimático, projeto e manejo de irrigação, dentre outros. Na literatura podem ser encontrados coeficientes para culturas que, quando aplicados sobre os valores de evapotranspiração de referência, permitem estimar o valor da evapotranspiração real dessas culturas (SOUZA et al., 2010).

Existem diversos métodos que possibilitam a obtenção da evapotranspiração de referência de forma empírica ou determinística. Dentre eles, o método de Penman-Monteith-FAO 56 é tido como padrão pela Organização das Nações Unidas para Alimentação e Agricultura (FAO) (ALLEN et al., 2006). Estudos recentes, buscam resultados de evapotranspiração com métodos inovadores, com diferentes climas em várias partes do mundo (ISLAN et al., 2021; CUNHA et al., 2021; HUERTA et al., 2022; NDULUE & RAJAN, 2021) tornando possível determinar os valores de ET_o com precisão em diferentes condições climáticas.

Do ponto de vista de sistemas de irrigação, tecnologias eficientes que possibilitem a previsão das lâminas de água a serem aplicadas podem auxiliar usuários e órgãos gestores no planejamento da alocação dos recursos hídricos. As previsões de evapotranspiração em curto prazo, que geralmente são mais precisas, possibilitam o processo de tomada de decisão antecipada, abrindo margens para otimização do uso dos recursos nos projetos e no manejo da irrigação (TRAORE et al., 2016).

No decorrer do tempo, ferramentas de inteligência artificial vêm sendo aplicadas na previsão da evapotranspiração de referência em diferentes regiões do planeta. Alguns dos principais modelos utilizados são as redes neurais artificiais, a modelagem *fuzzy* e a computação evolucionária (SANIKHANI et al., 2019).

Dentre os principais desafios existentes no processo de previsão da evapotranspiração de referência, pode-se citar a exigência de grande poder computacional e a necessidade de

obtenção de dados climáticos em quantidade e qualidade suficientes para desenvolvimento dos modelos. Quando são utilizadas variáveis advindas de previsão do tempo, os erros associados a esses dados devem ser analisados com cautela para que a ETo obtida seja confiável (TRAORE et al., 2016).

Segundo Taylor e Letham (2018), o *Facebook Prophet* é um modelo de regressão aditiva que, com um esforço computacional relativamente baixo, possui grande capacidade para detectar padrões de mudança em variáveis que se alteram de forma sazonal. Conforme apresentado na documentação do modelo, ele permite a adição de diversos tipos de sazonalidade (diária, semanal, mensal e personalizada), além de possibilitar a incorporação do efeito de eventos que ocorrem em épocas específicas e de variáveis que ajudem na previsão do fenômeno em estudo (regressores). O modelo apresenta ainda recursos como validação cruzada e otimização de hiperparâmetros, que auxiliam nos processos de treinamento e validação.

De acordo com Feurer et al. (2019), o *framework Auto-sklearn* é uma ferramenta que possibilita a seleção de modelos clássicos de *machine learning* implementados no *framework sklearn*, além do ajuste de hiperparâmetros. Segundo os autores, o *Auto-sklearn* otimiza o tempo de treinamento dos modelos com a utilização de conhecimentos extraídos de *datasets* similares. Como resultado de sua execução, o *framework* retorna os modelos que tiveram melhor desempenho para a métrica de avaliação escolhida, bem como o conjunto de hiperparâmetros de cada um.

No Brasil, o Instituto Nacional de Meteorologia (INMET) disponibiliza uma base aberta com dados climáticos históricos coletados por estações meteorológicas em diferentes regiões do país (INMET, 2022). Tais dados, quando coletados e processados de forma correta, formam uma rica fonte de dados que podem ser aplicados no ajuste de modelos de previsão de evapotranspiração de referência.

Embora não estejam presentes em todas as cidades do país, as estações meteorológicas do INMET estão espalhadas pelas cinco regiões do Brasil, podendo elas serem automáticas, convencionais ou pluviômetros automáticos. As estações automáticas são aquelas que possuem a capacidade de registrar os dados com frequência horária, disponibilizando-os na base de dados da instituição sem a necessidade de intervenção humana. Nessas estações são coletadas informações referentes a temperatura, umidade, temperatura de ponto de orvalho, pressão atmosférica, velocidade, direção e rajada do vento, radiação solar e precipitação. O INMET armazena os registros históricos das coletas em sua base de dados e os disponibiliza de forma aberta e interativa através de gráficos e tabelas mantidas em seu portal. Além disso, o instituto também disponibiliza os dados através de uma *Application Programming Interface* (API), que

permite o acesso aos dados históricos com a utilização de requisições HTTP, possibilitando sua integração com diversas ferramentas de desenvolvimento *web*.

Com este trabalho, objetivou-se analisar o desempenho do modelo *Facebook Prophet* e do *framework Auto-sklearn* na previsão de ETo de curto prazo com base em dados coletados por estações meteorológicas do INMET. A hipótese adotada foi que variáveis climáticas observadas no dia anterior poderiam ser úteis na previsão da evapotranspiração diária. Além disso, esperava-se que a utilização de mais de um modelo ao longo do ano diminuiria o erro quando comparado à utilização de um único modelo, e que variáveis meteorológicas que podem ser obtidas nos sistemas de previsão do tempo do INMET teriam potencial para melhorar o desempenho dos modelos de regressão.

2. MATERIAL E MÉTODOS

2.1. Cálculo da evapotranspiração de referência

Para treinamento dos modelos, a evapotranspiração de referência diária foi calculada com base no método de *Penman-Monteith-FAO* (ALLEN et al., 2006). Para tanto, foram utilizadas as equações apresentadas na circular técnica da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) (CONCEIÇÃO, 2006), intitulada “Roteiro de cálculo da evapotranspiração de referência pelo método de Penman-Monteith-FAO”. A estação escolhida para obtenção dos dados está localizada no município de Viçosa, em Minas Gerais. Os dados foram coletados para o período compreendido entre os meses de novembro de 2005 e abril de 2022.

2.2. Algoritmo para coleta dos dados e cálculo da ETo

Com o objetivo de realizar a coleta dos dados das estações do INMET, um programa de computador foi desenvolvido utilizando a linguagem de programação *python*. Este programa utilizou o módulo *Requests*, que permite a execução de requisições HTTP com essa linguagem de programação, para acessar a API mantida pelo instituto e obter os dados de uma determinada estação. A implementação do algoritmo se deu através de uma classe intitulada “*ETo_Calculator*”, que tem as equações para cálculo da ETo (Equações 1 a 6) implementadas em seus métodos, além de possuir funções dedicadas para fazer requisições na API do INMET, e para processamento dos dados brutos coletados. Dessa forma, para proceder a coleta de dados para a estação desejada, a classe é instanciada com a passagem do código da estação e o período de cálculo em seus parâmetros. Ao ser executado, o programa retorna os dados coletados pela estação meteorológica, bem como os valores de ETo calculados em todos os dias contidos no período escolhido.

$$E_{to} = \frac{0,408\Delta(Rn - G) + \frac{\gamma 900U_2(e_s - e_a)}{T + 273}}{\Delta + \gamma(1 + 0,34U_2)} \quad (1)$$

em que:

ETo = evapotranspiração de referência calculada pelo método de Penman-Monteith-FAO.

Δ = declividade da curva de pressão de vapor em relação à temperatura ($kPa^{\circ}C^{-1}$).

Rn = saldo de radiação na superfície ($MJ m^{-2} dia^{-1}$).

G = fluxo de calor no solo ($MJm^{-2}dia^{-1}$).

γ = constante psicrométrica ($kPa^{\circ}C^{-1}$).

U_2 = velocidade do vento a 2 metros de altura (ms^{-1}).

e_s = pressão de saturação de vapor (kPa).

e_a = pressão atual de vapor (kPa).

T = temperatura do ar média ao longo do dia ($^{\circ}C$).

$$\Delta = \frac{4098 \left[0,6108 \exp\left(\frac{17,27 T}{T + 237,3}\right) \right]}{(T + 237,3)^2} \quad (2)$$

$$e_s = 0,6108 \exp\left(\frac{17,27 T}{T + 237,3}\right) \quad (3)$$

$$e_a = \frac{e_s UR}{100} \quad (4)$$

$$\gamma = 0,665 \cdot 10^{-3} Patm \quad (5)$$

$$Patm = 101,3 * \left(\frac{293 - 0,0065 z}{293}\right)^{5,26} \quad (6)$$

em que:

z = altitude do local (m).

2.3. Limpeza e processamento dos dados

De uma forma geral, verificou-se que as estações do INMET apresentaram falhas de medições e pausas para manutenção ao longo do seu tempo de funcionamento. Nesses períodos, houve a ausência parcial ou total de informações na base de dados do Instituto. Dessa forma, os dados coletados foram submetidos a análises e processados para remoção de falhas. Além disso, os valores negativos de radiação global medidos pela estação foram descartados. Variáveis calculadas com base em dados meteorológicos e que são usadas no cálculo da ETo, como o saldo de radiação e a pressão de saturação de vapor, foram calculadas e adicionadas ao conjunto de *features*. Os valores de ETo calculados foram então deslocados em um dia, de

forma que cada valor de ETo fosse agrupado com as condições climáticas registradas no dia anterior. Por fim, a evapotranspiração de referência calculada no dia anterior também foi adicionada ao conjunto de *features*.

Em seguida, as variáveis que seriam utilizadas no treinamento dos modelos (variáveis meteorológicas do dia anterior) foram submetidas a análise de correlação para verificar seu comportamento em função da ETo. Ao fim desse processo, *features* cujo módulo da correlação era próximo de zero foram removidas da base de treinamento dos modelos, restando um total de 15 variáveis do dia anterior associadas à evapotranspiração do dia atual, sendo elas: temperaturas máxima, média e mínima, saldo de radiação, radiação global corrigida, declividade da curva de pressão de vapor, pressão de saturação de vapor, pressão vapor média, sensação térmica, temperatura de ponto de orvalho máxima, umidade mínima, umidade instantânea média, velocidade do vento medida, velocidade do vento corrigida e evapotranspiração de referência do dia anterior.

Após a análise de correlação, as *features* restantes foram transformadas com a utilização da função *MinMaxScaler* do *framework Sklearn* (PEDREGOSA et al., 2011), que, para cada uma das variáveis, executa o ajuste das dimensões para uma escala de zero a um, respeitando as proporções das dimensões originais. Tal transformação se deu com o objetivo de atenuar a variabilidade de magnitude das diferentes variáveis meteorológicas utilizadas no treinamento. Em seguida, fez-se o ranqueamento das *features* com a utilização da função de eliminação recursiva de variáveis (RFE) do *framework Sklearn*, tendo o modelo *Random Forest Regressor* como estimador de importância. Os dados foram então divididos em grupos de treinamento e validação. Para treinamento foram utilizados os dados anteriores a janeiro de 2019, e para validação de janeiro de 2019 a abril de 2022.

2.4. Cenários de análise

Uma vez que as *features* foram ranqueadas em ordem de importância, procedeu-se o treinamento do modelo de *times series Facebook Prophet*, apresentado por Taylor & Letham (2018), e do *framework de automated machine learning Auto-sklearn*, apresentado por Feurer et al. (2019). Esses modelos foram treinados recursivamente com a adição de uma *feature* a cada treinamento, de forma que no primeiro apenas a variável mais importante foi utilizada, e no último todas as quinze estavam presentes. Em seguida, fez-se uma análise da diminuição da raiz quadrada do erro médio (RMSE) e do aumento do coeficiente de determinação (R^2) em função da adição das variáveis no treinamento. Dado o fato de que as condições climáticas

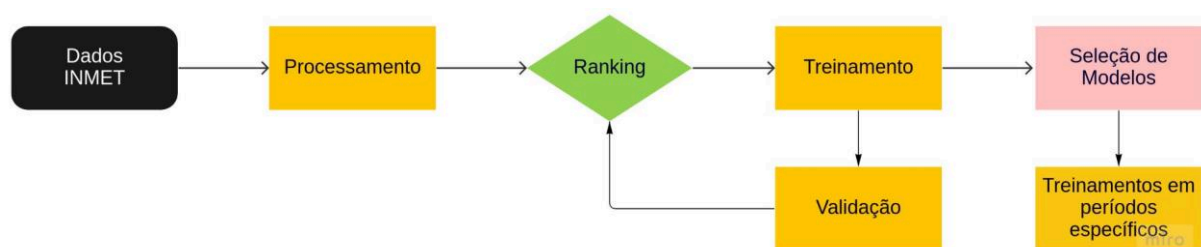
sofrem variações ao longo do ano, no modelo *Prophet* a sazonalidade semanal foi desabilitada, e a anual mantida. Procedeu-se então a otimização de hiperparâmetros.

No processo de treinamento do *Auto-Sklearn*, a minimização do erro quadrático médio foi escolhido como métrica de ranqueamento dos modelos. Dentre todos os retornados, um modelo de regressão foi escolhido para dar prosseguimento com os experimentos. Esse modelo foi então utilizado nos seguintes cenários: um único treinamento para todos os dias do ano, um treinamento individual para cada estação do ano, e um para cada mês do ano.

Com o objetivo de verificar o potencial da utilização de variáveis que podem ser obtidas a partir de modelos de previsão do tempo, foram adicionadas entre as *features* de treinamento as seguintes variáveis do dia de ocorrência da ETO: temperatura máxima, temperatura mínima, umidade relativa máxima e umidade relativa mínima. A previsão de tais variáveis é realizada diariamente pelo INMET, e os valores disponibilizados através de uma API pública. Como não foi encontrada nenhuma base de dados que armazenasse os valores históricos preditos, foram utilizados os dados reais medidos no dia em que se desejava fazer a previsão de ETO.

Por fim, o processo de pré-seleção com base em correlação foi repetido, sendo que a umidade relativa máxima foi descartada durante o processo. As demais variáveis foram então adicionadas ao conjunto das *features* do dia anterior, o ranqueamento com a função de RFE executado, e os modelos treinados nos mesmos cenários apresentados anteriormente. Na Figura 1 estão esquematizadas de forma simplificada as etapas de treinamento dos modelos.

Figura 1 - Fluxo simplificado de treinamento dos modelos



Fonte: Próprio Autor.

3. RESULTADOS E DISCUSSÃO

Na Tabela 1, as variáveis meteorológicas utilizadas no treinamento dos modelos foram ordenadas por importância elencada na execução da função de RFE.

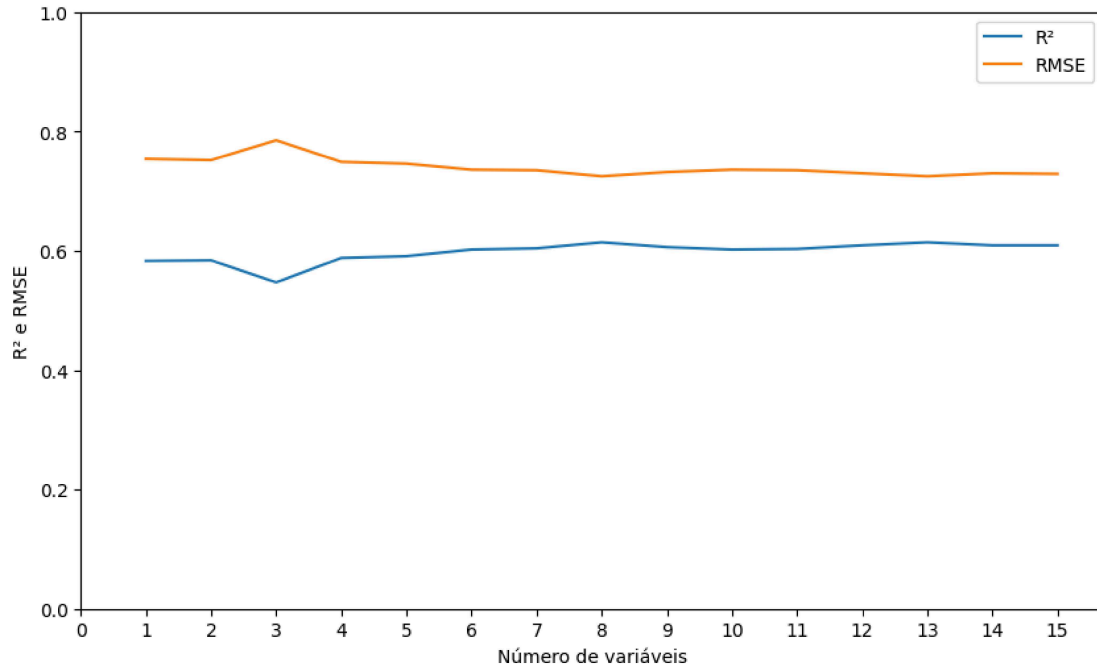
Tabela 1 - Ranking de variáveis de acordo com o peso obtido na execução da função de eliminação recursiva de variáveis (RFE)

Variável	Ranking
Evapotranspiração de referência do dia anterior	1
Saldo de Radiação	2
Radiação global corrigida	3
Declividade da curva de pressão de vapor	4
Pressão média de saturação de vapor	5
Temperatura média	6
Sensação térmica	7
Temperatura mínima	8
Pressão atual de vapor	9
Velocidade do vento corrigida	10
Temperatura máxima	11
Temperatura do ponto de orvalho máxima	12
Umidade instantânea média	13
Velocidade do vento	14
Umidade relativa mínima	15

Fonte: Próprio Autor.

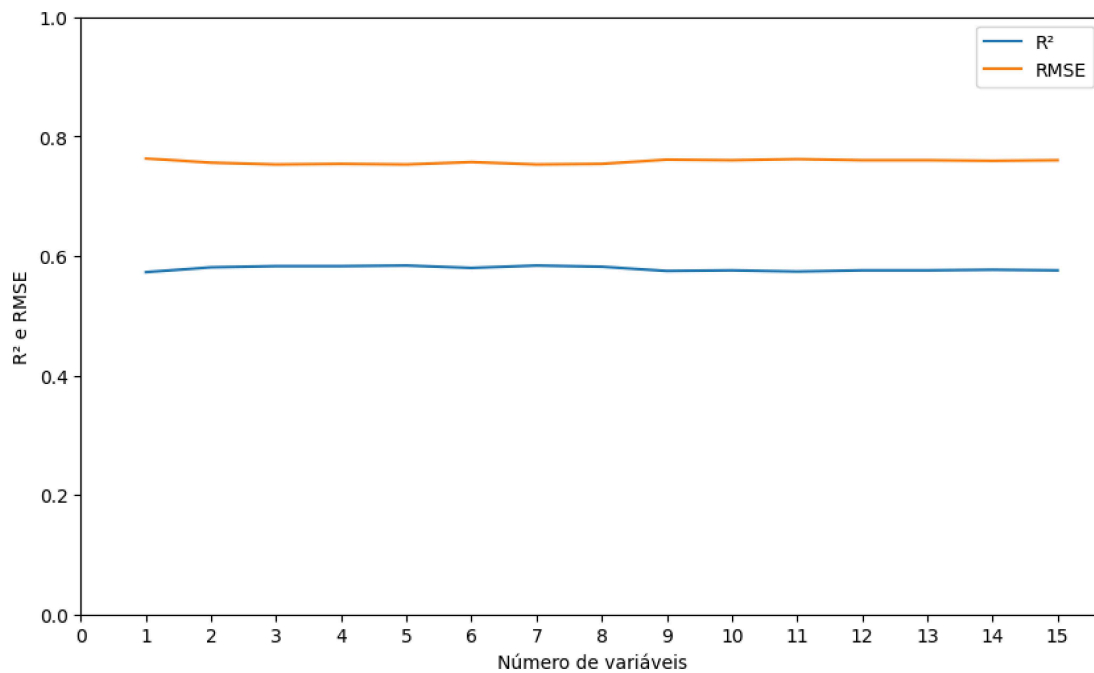
Nas Figuras 2 e 3, estão apresentadas as variações do RMSE e do R^2 com a adição de cada uma das variáveis da Tabela 1 ao conjunto de treinamento dos modelos.

Figura 2 - Variação do coeficiente de determinação (R^2) e da raiz quadrada do erro médio (RMSE) em função da adição de variáveis no treinamento com o *Auto-Sklean*



Fonte: Próprio Autor.

Figura 3 - Variação do coeficiente de determinação (R^2) e da raiz quadrada do erro médio (RMSE) em função da adição de variáveis no treinamento com o *Prophet*

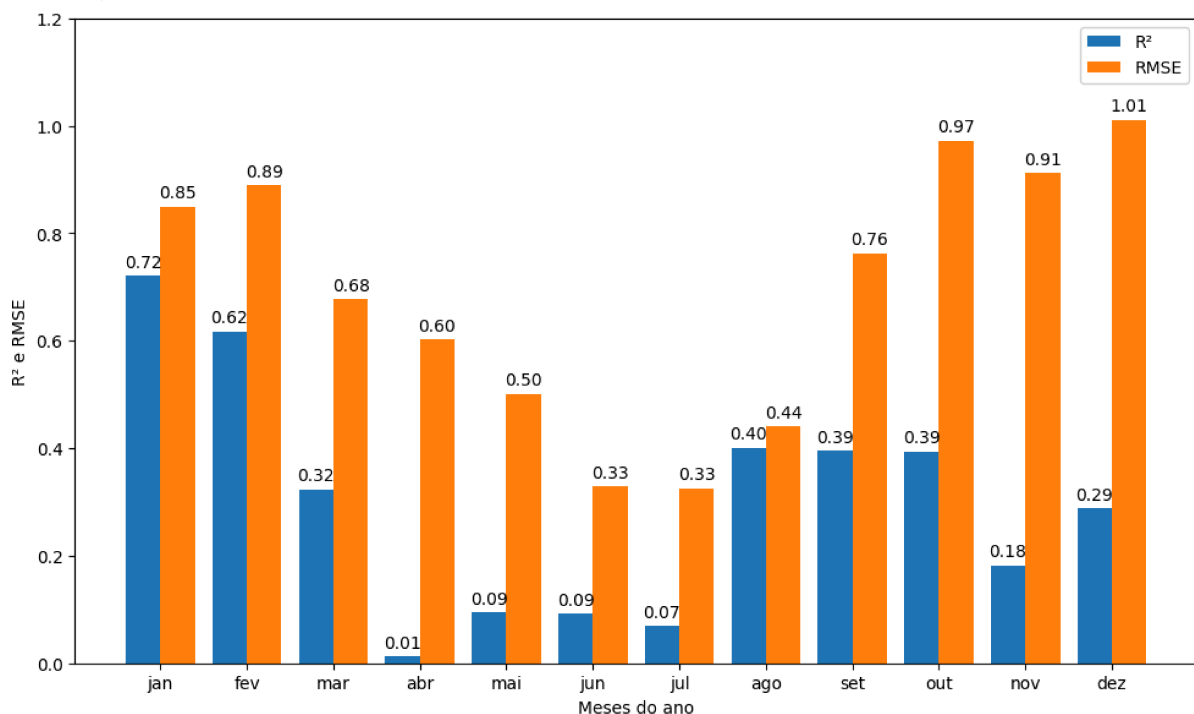


Fonte: Próprio Autor.

Na Figura 2, para o *Auto-Sklearn*, pode-se observar um ligeiro aumento no coeficiente de determinação e diminuição do erro médio durante a adição das variáveis 1 a 8, seguido por estabilização. Na adição da terceira variável (radiação global corrigida) houve uma piora nos dois indicadores analisados, mostrando que, apesar de essa variável estar entre as mais importantes quando se tem o *Random Forest Regressor* como estimador, isso não ocorreu no treinamento dos modelos com o *Auto-Sklearn*. Dentre todos os modelos treinados o *ARDRegression* apresentou melhores valores de RMSE e R^2 ao longo da adição das variáveis meteorológicas, e, portanto, foi escolhido para dar prosseguimento com as análises. Para o *Facebook Prophet*, a variação dos valores de R^2 e RMSE foi praticamente insignificante ao longo da adição das variáveis ao treinamento (Figura 3).

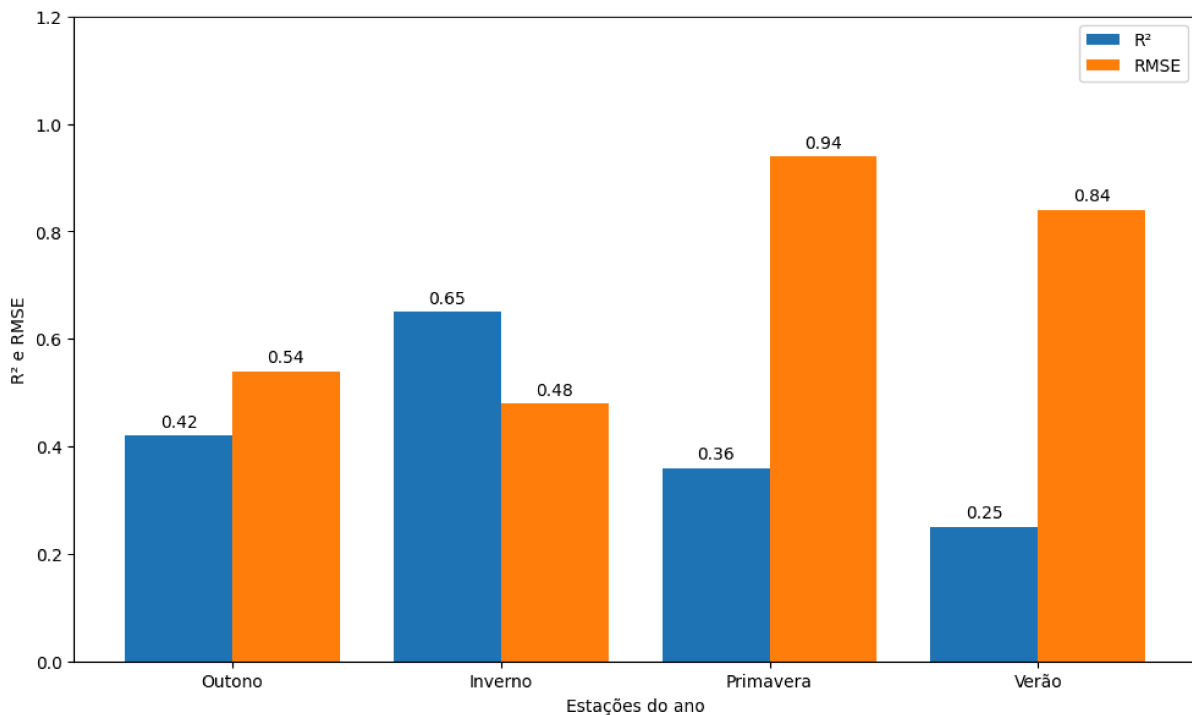
Na Figura 4, estão apresentados os valores de RMSE e R^2 encontrados para cada mês do ano no cenário de treinamento do *ARDRegression* com dados mensais isolados. Na Figura 5, esses resultados são apresentados para o treinamento de um modelo para cada estação do ano.

Figura 4 - Valores do coeficiente de determinação (R^2) e da raiz quadrada do erro médio (RMSE) encontrados no treinamento com dados mensais isolados utilizando o modelo *ARDRegression* e as oito variáveis selecionadas



Fonte: Próprio Autor.

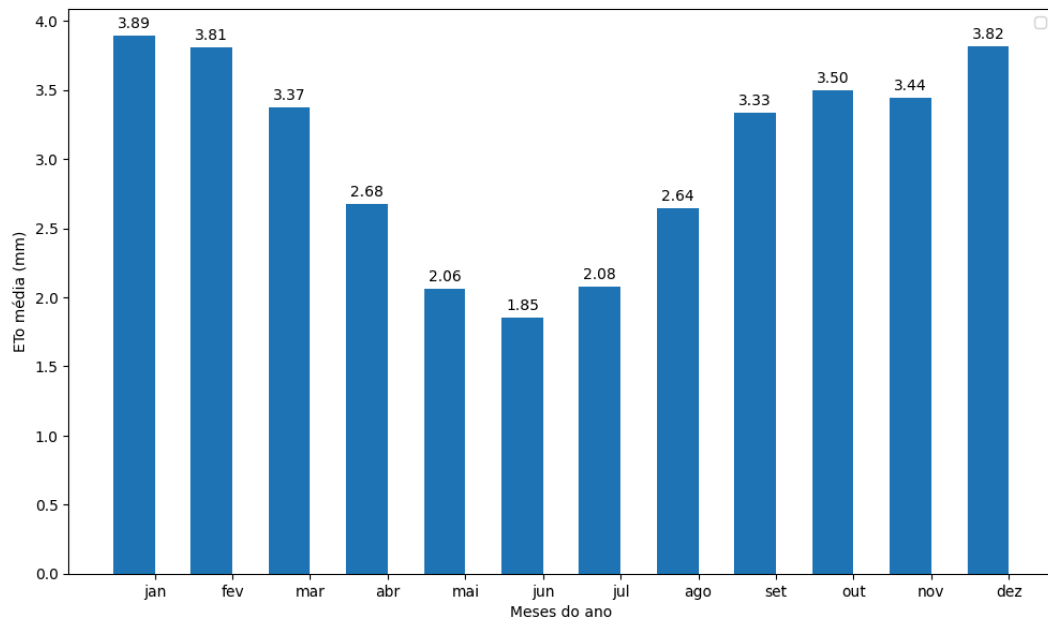
Figura 5 - Valores do coeficiente de determinação (R^2) e da raiz quadrada do erro médio (RMSE) encontrados no treinamento baseado na estação do ano utilizando o modelo *ARDRegression* e as oito variáveis selecionadas



Fonte: Próprio Autor.

Na Figura 4 pode-se observar que, no treinamento mensal, houve grande variabilidade de valores de RMSE e R^2 . O maior valor de R^2 foi verificado no mês de janeiro, e o menor RMSE nos meses de junho e julho. Excetuando-se janeiro e fevereiro, os valores de R^2 encontrados foram menores que os obtidos no treinamento de um modelo único para todos os dias do ano (Figura 2). A Figura 4 mostra que os valores de R^2 foram praticamente nulos nos meses de abril, maio, junho e julho, apontando que os modelos desses meses são pouco explicativos para os dados previstos. Acredita-se que os baixos valores de R^2 encontrados nesses meses estejam associados com as variáveis utilizadas no treinamento do modelo, que, conforme apresentado na Tabela 1, estão em sua maioria relacionadas com radiação solar e temperatura, variáveis que tendem a sofrer menores variações nestes meses do ano, diminuindo a capacidade do modelo de detectar mudanças. Na Figura 6 é apresentada a evapotranspiração média diária observada em cada mês do ano. Como o RMSE é uma métrica que possui a mesma escala das variáveis previstas, acredita-se que a diminuição do erro nesses meses não está associada a uma melhora no desempenho dos modelos, mas sim ao fato de a evapotranspiração ser menor.

Figura 6 - Evapotranspiração de referência média diária em cada mês do ano



Fonte: Próprio Autor.

Na Tabela 2, está apresentado o resultado do ranqueamento de variáveis após a adição de temperatura e umidade máximas e mínimas do dia de ocorrência da evapotranspiração de referência. Dentre essas variáveis, apenas a umidade relativa máxima foi descartada na análise prévia de correlação e por isso não aparece na Tabela 2. Conforme esperado, as variáveis reais do dia da evapotranspiração estiveram entre as mais importantes de acordo com o resultado da função de RFE, sendo que a temperatura máxima ficou em primeiro lugar, a umidade relativa mínima em terceiro e a temperatura mínima em décimo. Pode-se notar, também, que a adição dessas variáveis influenciou na ordem de importância das demais. A evapotranspiração de referência do dia anterior, por exemplo, que na Tabela 1 aparece como a variável mais importante, na Tabela 2, passou a apresentar importância inferior ao saldo de radiação, à radiação global corrigida e à temperatura máxima.

Tabela 2 – Ranking de variáveis de acordo com o peso obtido na execução da função de eliminação recursiva de variáveis (RFE), após adição de variáveis do dia de previsão

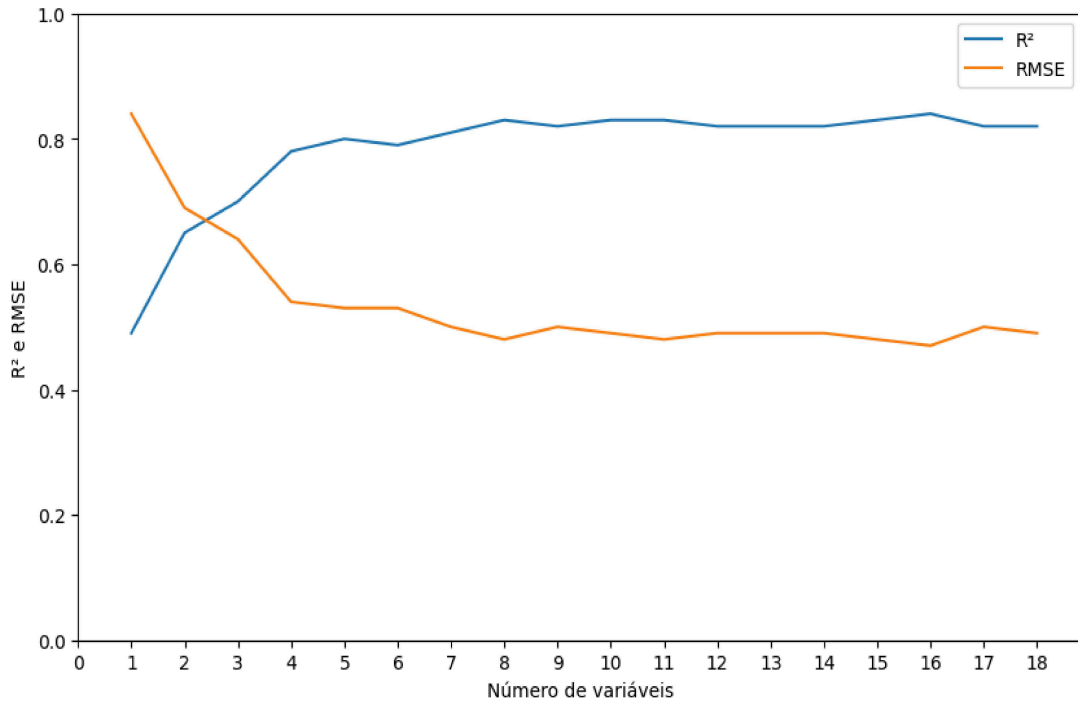
Variável	Ranking
Temperatura máxima dia atual	1
Saldo de Radiação	2
Umidade relativa mínima do dia atual	3
Radiação global corrigida	4
Temperatura máxima	5
Evapotranspiração de referência do dia anterior	6
Temperatura mínima	7
Umidade instantânea média	8
Velocidade do vento corrigida	9
Temperatura mínima dia atual	10
Umidade relativa mínima	11
Pressão média de saturação de vapor	12
Pressão atual de vapor	13
Temperatura do ponto de orvalho máxima	14
Sensação térmica	15
Velocidade do vento	16
Temperatura média	17
Declividade da curva de pressão de vapor	18

Fonte: Próprio Autor.

Nas Figuras 7 e 8 podem ser observadas as variações dos valores de RMSE e R^2 na execução recursiva do *Auto-Sklearn* e do *Prophet* com a adição das variáveis apresentadas na Tabela 2. Analisando as Figuras 2, 3, 7 e 8 pode-se observar que, para ambos os modelos, a variação passou a ser mais significativa após a adição dessas variáveis. No treinamento do *Auto-Sklearn*, o modelo com melhor desempenho encontrado foi o *Histogram-based Gradient Boosting Regression Tree* (HGBRT). Para o *Prophet*, os melhores resultados foram obtidos com a utilização das 6 variáveis mais importantes. O *Auto-Sklearn* apresentou melhoras

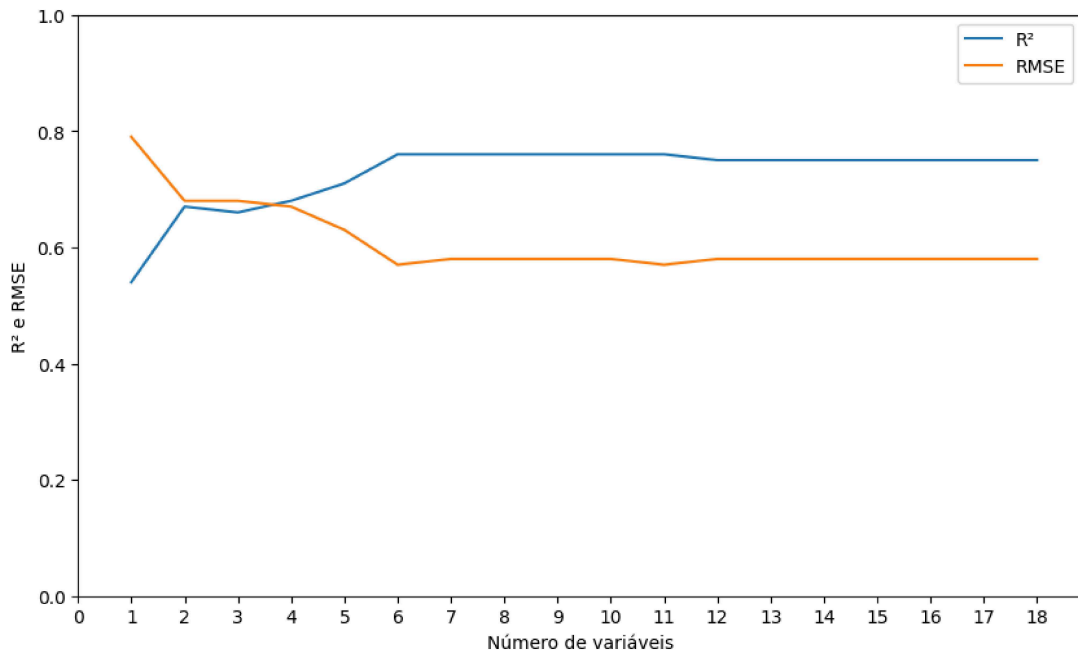
significativas até a adição da 16ª variável e, portanto, elas foram utilizadas na análise dos demais cenários.

Figura 7 - Variação do coeficiente de determinação (R^2) e da raiz quadrada do erro médio (RMSE) para o *Auto-Sklearn* após a adição das variáveis do dia de previsão



Fonte: Próprio Autor.

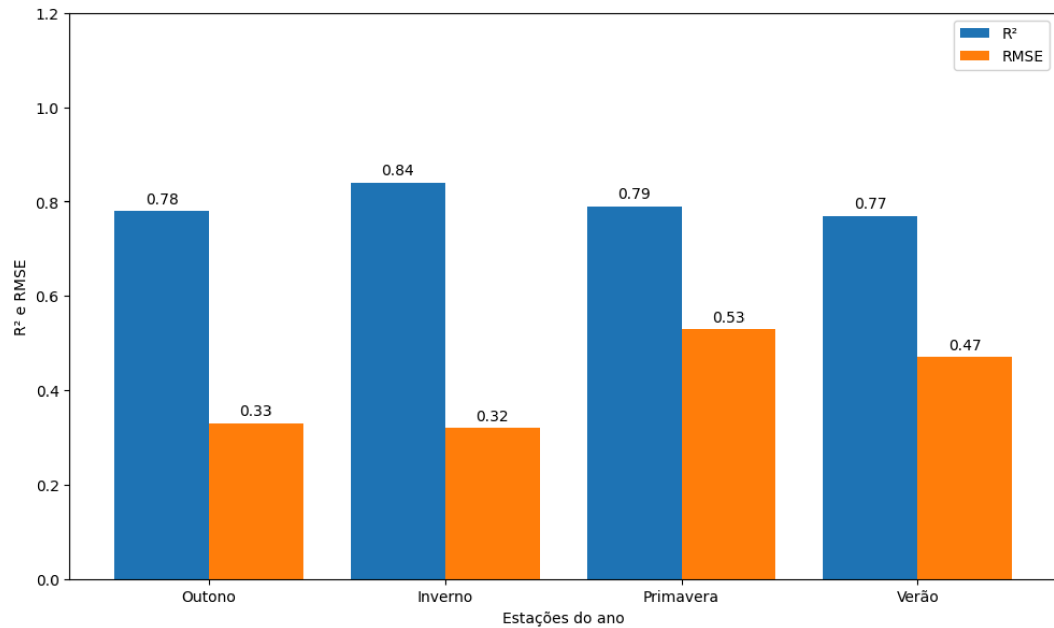
Figura 8 - Variação do coeficiente de determinação (R^2) e da raiz quadrada do erro médio (RMSE) para o *Facebook Prophet* após a adição das variáveis do dia de previsão



Fonte: Próprio Autor.

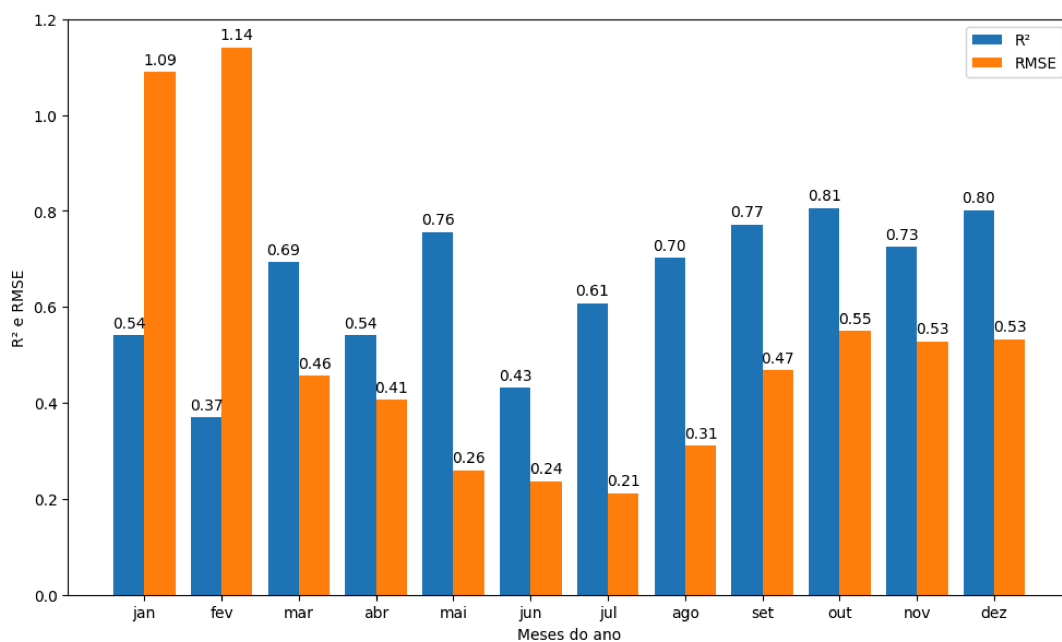
Nas Figuras 9 e 10, são apresentados resultados do treinamento do HGBRT baseado em estações do ano e mensal após a adição das novas variáveis.

Figura 9 - Valores do coeficiente de determinação (R^2) e da raiz quadrada do erro médio (RMSE) no treinamento com base na estação do ano utilizando o modelo HGBRT após adição das variáveis do dia de previsão



Fonte: Próprio Autor.

Figura 10 - Valores do coeficiente de determinação (R^2) e da raiz quadrada do erro médio (RMSE) no treinamento com dados mensais isolados utilizando o modelo HGBRT após a adição das variáveis do dia de previsão



Fonte: Próprio Autor.

Quando comparado com a utilização de um único modelo anual de previsão (Figura 7), o treinamento para o período de inverno (Figura 9) apresentou o mesmo valor de R^2 , e nas demais estações e meses do ano (Figura 10) os valores foram inferiores. Acredita-se que isso ocorreu em decorrência da menor quantidade de dados disponíveis para treinamento com base nas estações do ano e mensal. Para o RMSE, houve uma diminuição para os modelos de outono e inverno, além dos meses abril, maio, junho, julho e agosto. Para os demais meses e estações, observou-se um aumento do erro. Conforme mencionado anteriormente, acredita-se que a diminuição do RMSE nos meses citados tenha ocorrido devido à menor ETo a eles associada.

Para o cenário de treinamento de um modelo anual único (Figuras 2 e 7), a adição das variáveis do dia da previsão propiciou a diminuição do RMSE de 0,73 para 0,47 (35,6%), e aumento do R^2 de 0,61 para 0,84 (37,7%). Quando os modelos foram treinados mensalmente (Figuras 4 e 9), a adição dessas variáveis levou à diminuição do R^2 e aumento do RMSE nos meses de janeiro e fevereiro, nos demais meses ambos os indicadores melhoraram. Para o treinamento com base nas estações do ano (Figuras 5 e 9), todos os indicadores apresentaram melhoria com a adição dos dados do dia de previsão. Tais resultados evidenciam o grande potencial dessas variáveis, que podem ser obtidas com modelos de previsão do tempo, para a melhoria da previsão da ETo. Como foram usados apenas dados reais neste trabalho, a influência dos erros presentes nesses modelos de previsão do tempo pode ter um impacto significativo e deve ser estudada com cautela.

Em todas as comparações, a utilização de um único modelo para todos os dias do ano apresentou resultados mais satisfatórios, em termos de R^2 , que os modelos treinados de forma mensal e modelos com base nas estações do ano. Acredita-se que a quantidade de dados disponíveis para treinamento tenha influenciado no resultado, uma vez que a base de dados utilizada é aproximadamente doze vezes menor, para os modelos mensais, e quatro vezes menor, para os modelos baseados em estações do ano.

Embora a variabilidade climática tenha grande influência na magnitude da ETo, limitando a comparação com resultados obtidos para regiões diferentes, verifica-se que os valores de RMSE aqui encontrados se assemelham aos observados em trabalhos que implementaram metodologias diferentes para previsão de Eto em curto prazo. De Oliveira (2019) aplicou redes neurais convolucionais na previsão de séries temporais de evapotranspiração de referência, utilizando dados coletados da estação meteorológica do INMET situada no município de Sete Lagoas (MG). De uma forma geral, as redes neurais treinadas pelo autor apresentaram valores de RMSE variando entre 0,72 e 0,80 para o horizonte de previsão de um dia. Alves Sobrinho et al. (2011) implementaram uma Rede Neural Artificial

para estimar a ETo em função da temperatura do ar diária. Para isso, dados foram coletados de uma estação meteorológica automática localizada no município de Dourados (MS), e os valores de RMSE obtidos estiveram entre 0,41 e 0,70. Sendo assim, pode-se observar que a utilização de variáveis meteorológicas do dia anterior e de modelos de previsão do tempo, se apresenta como promissora para previsão da ETo de curto prazo.

4. CONCLUSÕES

Em todos os cenários de treinamento com base anual, o desempenho do modelo de regressão encontrado com o *Auto-Sklearn* foi superior, em termos de RMSE e R^2 , ao modelo de *time series Facebook Prophet*. O emprego de variáveis do dia anterior para previsão de evapotranspiração de referência levou à ocorrência RMSEs semelhantes aos encontrados na literatura para modelos que utilizam estratégias diferentes, mostrando que o *Auto-Sklearn* e o *Facebook Prophet* apresentaram resultados satisfatórios na previsão da ETo de curto prazo. Além disso, estes resultados evidenciam o grande potencial que as variáveis do dia anterior têm para serem utilizadas em modelos de previsão da ETo.

Em termos globais, a aplicação de um único modelo para previsão de ETo em todos os dias do ano, resultou em um desempenho superior à utilização de modelos treinados de forma mensal e modelos com base nas quatro estações do ano. Acredita-se que a quantidade de dados disponíveis para treinamento tenha tido influência significativa sobre isso.

A utilização das variáveis temperatura máxima, temperatura mínima e umidade relativa mínima, que podem ser facilmente obtidas com a utilização de modelos públicos de previsão do tempo, apresenta grande potencial para melhorar o desempenho de algoritmos de *Machine Learning* na previsão de evapotranspiração de referência de curto prazo.

Para trabalhos futuros, sugere-se a avaliação dos erros resultantes da aplicação de dados obtidos diretamente da API de previsão do tempo do INMET sobre a estimativa da ETo, sendo necessária, para isso, a formação de uma base com os dados previstos diariamente pelo Instituto. Pode-se também desenvolver um trabalho de disponibilização dos modelos treinados para serem utilizados em aplicações práticas, utilizando *pipelines* com integração com as APIs do INMET. Além disso, para verificar se o incremento na quantidade de dados melhora o desempenho dos modelos treinados, pode-se conduzir experimentos com a combinação de dados de estações situadas em cidades próximas.

REFERÊNCIAS

- ALLEN, R. G. et al. A recommendation on standardized surface resistance for hourly calculation of reference ETo by the FAO56 Penman-Monteith method. *Agricultural water management*, 81(1-2), 1-22, 2006. <https://doi.org/10.1016/j.agwat.2005.03.007>
- ALVES SOBRINHO, T. et al. Estimativa da evapotranspiração de referência através de redes neurais artificiais. *Revista Brasileira de Meteorologia*, v. 26, p. 197-203, 2011.
- CARVALHO, L. G. DE et al. Evapotranspiração De Referência: Uma Abordagem Atual De Diferentes Métodos De Estimativa. *Pesquisa Agropecuária Tropical*, v. 41, n. 3, p. 456–465, 2011.
- CONCEIÇÃO, M. AF. Roteiro de cálculo da evapotranspiração de referência pelo método de Penman-Monteith-FAO. *Embrapa Uva e Vinho. Circular Técnica*, 2006.
- CUNHA, A. C. et al. Influence of the estimated global solar radiation on the reference evapotranspiration obtained through the Penman-Monteith FAO 56 method. *Agricultural Water Management*, 243, 106491, 2021. <https://doi.org/10.1016/j.agwat.2020.106491>
- DE OLIVEIRA, P. Previsão de séries temporais de evapotranspiração de referência com redes neurais convolucionais. Dissertação de mestrado. Universidade Federal de Minas Gerais – UFMG. Programa de Pós-Graduação em Engenharia Elétrica – PPGEE. 2019.
- FEURER, M. et al. Auto-sklearn: Efficient and Robust Automated Machine Learning. Automated Machine Learning. **The Springer Series on Challenges in Machine Learning**. Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-05318-5_6
- HUERTA, et al. PISCOeo_pm, a reference evapotranspiration gridded database based on FAO Penman-Monteith in Peru. *Scientific data*, 9(1), 1-18, 2022. <https://doi.org/10.1038/s41597-022-01373-8>
- INMET. Instituto Nacional de Meteorologia. [s.l.: s.n.], 2022. Disponível em: <<https://portal.inmet.gov.br/manual>>.
- ISLAM, S. et al. Performance evaluation of FAO Penman-Monteith and best alternative models for estimating reference evapotranspiration in Bangladesh. *Heliyon*, 7(7), e07487, 2021. <https://doi.org/10.1016/j.heliyon.2021.e07487>
- NDULUE, E. et al. Performance of the FAO Penman-Monteith equation under limiting conditions and fourteen reference evapotranspiration models in southern Manitoba. *Theoretical and Applied Climatology*, 143(3), 1285-1298, (2021). <https://doi.org/10.1007/s00704-020-03505-9>
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. **The Journal of machine Learning research**, v. 12, p. 2825-2830, 2011.
- SANIKHANI, H. et al. Temperature-based modeling of reference evapotranspiration using several artificial intelligence models: application of different modeling scenarios. **Theoretical and applied climatology**, v. 135, n. 1, p. 449-462, 2019.

TAYLOR, S. J.; LETHAM, B. Forecasting at scale. **The American Statistician**, v. 72, n. 1, p. 37-45, 2018.

TRAORE, S. et al. Deployment of artificial neural network for short-term forecasting of evapotranspiration using public weather forecast restricted messages. **Agricultural Water Management**, v. 163, p. 363-379, 2016.