

Comitê de máquinas aplicado à previsão de séries temporais meteorológicas

Arthur R. dos Santos* Lucas Camilotti**,***
Eduardo de F. R. Loures** Mathues H. D. M. Ribeiro****
Roberto Z. Freire****

* *Volvo do Brasil, Curitiba, PR, Brasil*
(e-mail: arthurroseira95@gmail.com)

** *Programa de Pós-Graduação em Engenharia de Produção e Sistemas (PPGEPS), Pontifícia Universidade Católica do Paraná (PUCPR), Curitiba, PR, Brasil, (lucas.camilotti@pucpr.edu.br, eduardo.loures@pucpr.br)*

*** *Wert Solutions, Curitiba, PR, Brasil*

**** *Universidade Tecnológica Federal do Paraná (UTFPR)*
(e-mail: mribeiro@utfpr.edu.br, robertofreire@utfpr.edu.br)

Abstract: The research field of weather forecasting is dominated by numerical models. Despite presenting mathematical modeling of atmospheric dynamics very close to reality, these models have their limitations, as the modeling is based on approximations and a large volume of data is necessary to generate better results. This work proposes the minimization of the prediction error of numerical models through the application of machine learning (ML) algorithms: Random Forest (RF), Support Vector Regression (SVR), and Multilayer Perceptron (MLP) considering an ensemble forecasting methodology. The proposed strategy differs from the classical approaches by weighting the final results of the ML models to favor the best predictor. The results showed that this methodology reduces the variance of the results in the application of a cross-validation analysis and that the combination of a numerical model with ML algorithms can generate an improvement of up to 23% in the coefficient of determination and 29% in the root mean squared error of the predictions when compared with a purely numerical model.

Resumo: No âmbito da previsão meteorológica, modelos numéricos dominam esta linha de pesquisa. Apesar de apresentarem uma modelagem matemática da dinâmica atmosférica muito próxima da realidade, tais modelos possuem suas limitações, pois baseiam-se em aproximações e na necessidade de grandes volumes de dados para gerar melhores resultados. Este trabalho propõe a aplicação de algoritmos de aprendizado de máquina: Floresta aleatória (RF), Regressão por Vetores de Suporte (SVR) e Rede Neural Perceptron Multicamadas (MLP), onde buscou-se melhorar o desempenho preditivo de modelos numéricos fazendo uso da metodologia baseada em comitê de máquinas. A estratégia proposta neste estudo difere-se das estratégias clássicas pela ponderação dos resultados dos modelos de forma a favorecer o melhor preditor. Os resultados mostraram que essa metodologia reduz a variância dos resultados na aplicação de uma validação cruzada e que a combinação de um modelo numérico com algoritmos de aprendizado de máquina pode gerar uma melhora de até 23% no coeficiente de determinação e 29% na raiz do erro médio quadrático das previsões em relação aos resultados do modelo puramente numérico.

Keywords: ensemble forecasting; machine learning; time series prediction; weather forecasting

Palavras-chaves: comitê de máquinas, aprendizado de máquina, previsão de séries temporais, previsão climática.

1. INTRODUÇÃO

Em decorrência da grande influência do clima nas atividades humanas, um ramo específico da ciência voltado para o desenvolvimento de métodos relacionados a previsão climática foi criado, a meteorologia. Com altos investimentos em informatização e automação nas análises, há um imensurável volume de dados gerados por meio das observações climáticas. O processamento dessas amostras é realizado considerando supercomputadores administra-

dos por organizações específicas, com o intuito de serem utilizados no que são chamados de modelos preditivos numéricos. Tais modelos visam reproduzir a dinâmica da atmosfera utilizando conceitos da física. Apesar de hoje existir uma capacidade computacional elevada, a grande quantidade de dados e a carga de processamento requerida por estes modelos acabam limitando o uso de métodos numéricos com estas características (Ren et al., 2021; Cho et al., 2020a).

O avanço tecnológico dos últimos anos trouxe ganhos a outras áreas do conhecimento humano, como é o caso da inteligência artificial e da subárea relacionada aos algoritmos de aprendizado de máquina. A combinação desta tecnologia com o atual estado dos sistemas de análise climática resultou em sistemas inteligentes de previsão. Aplicações recentes em meteorologia envolvem a combinação dos métodos numéricos com o objetivo de aprimorar o poder de previsão destes sistemas sem realizar um aumento significativo na quantidade de dados, fazendo uso da técnica chamada de *Ensemble Forecasting* (Cho et al., 2020a; Cifuentes et al., 2020).

O objetivo deste trabalho é aplicar métodos de aprendizado de máquina para a previsão de séries temporais meteorológicas, procurando realizar um estudo comparativo entre os algoritmos tradicionalmente utilizados em diversas outras áreas de aplicação (como por exemplo finanças e epidemiologia), juntamente com a estratégia conhecida por comitê de máquinas (do inglês, *ensemble learning*). Os modelos de aprendizado de máquina Floresta Aleatória (do inglês, *random forest* RF), Regressão por Vetores de Suporte (do inglês, *support vector regression* SVR) e Rede Neural Perceptron Multicamadas (do inglês, *multilayer perceptron* MLP) foram selecionados para este estudo e serão avaliados considerando os seguintes aspectos: adaptabilidade sobre as características das séries temporais meteorológicas, flexibilidade na implementação e volume de processamento computacional necessário.

A contribuição deste estudo baseia-se na aplicação de modelos de aprendizado de máquina associada ao comitê de máquinas. Sendo um método que, além de apresentar uma melhoria do desempenho preditivo dos modelos numéricos, valida uma nova metodologia de previsão que reduz a variância nos resultados dos modelos de aprendizado de máquina quando utilizados sobre dados meteorológicos em relação à sua aplicação de forma individualizada.

2. TRABALHOS CORRELATOS

A combinação de modelos baseados em *numerical weather prediction* (NWP) com aprendizado de máquina foi abordado no trabalho de Du (2019). Neste trabalho a previsão do fluxo do vento foi realizada considerando métodos numéricos e três tipos de algoritmos de aprendizado de máquina, sendo eles: *Artificial Neural Networks* (ANN), SVR e *Gaussian Process* (GP). Em seguida, o *ensemble* foi criado combinando resultados por meio do modelo *Bayesian Model Average* (BMA). A previsão foi realizada com um horizonte de previsão de 1 passo frente. O *ensemble* de técnicas provou ser 28% mais eficiente que os outros modelos testados em termos de erro percentual.

Cho et al. (2020a) propuseram um estudo comparativo entre alguns modelos de aprendizado de máquina na previsão da temperatura máxima e mínima do ar um passo à frente. Eles utilizam o princípio de *Ensemble Forecasting* unindo modelos numéricos NWP e os modelos de aprendizado de máquina *Floresta Aleatória* (RF), MLP e SVR. Os resultados alcançados foram de 0,86 no RMSE para o modelo de *Ensemble Forecasting*.

Já no estudo apresentado por Singla et al. (2021), uma metodologia de previsão de radiação solar utilizando de-

composição *Wavelet* para séries temporais em combinação com uma rede neural BiLSTM foi apresentada. Neste estudo, foi realizada a previsão da radiação solar 24 horas à frente e o aprendizado conjunto se dá pelas redes BiLSTM autônomas treinadas e alocadas para cada subsérie. Finalmente, os valores previstos de cada subsérie das redes BiLSTM são reconstruídos para fornecer a previsão solar final. O modelo proposto reduz o RMSE médio mensal na faixa de 26,04% a 58,89%.

No trabalho realizado por Markovics and Mayer (2022) foi realizado um estudo que compara 24 modelos de aprendizado de máquina para previsão de energia fotovoltaica para o dia seguinte com base em previsões meteorológicas de modelos numéricos, testadas para conjuntos de dados de resolução de 15 minutos com duração de dois anos de 16 usinas fotovoltaicas na Hungria. Os resultados mostram que os dois modelos mais precisos são a regressão do cume de kernel e o perceptron multicamada, que resultam em uma diminuição de 13,1% da raiz do erro médio quadrático (RMSE).

3. DESCRIÇÃO DO PROBLEMA

O modelo de previsão numérica LDAPS (*Local Data Assimilation and Prediction System*) é gerenciado pela *Korea Meteorological Administration* (KMA) e utiliza dados das 25 *Automatic Weather Stations* (AWS) localizadas na cidade de Seul, conforme demonstrado na Figura 1, para gerar previsões de diversas variáveis climáticas (como por exemplo, a temperatura máxima e mínima do ar, umidade relativa do ar, nebulosidade e até a incidência de raios solares).

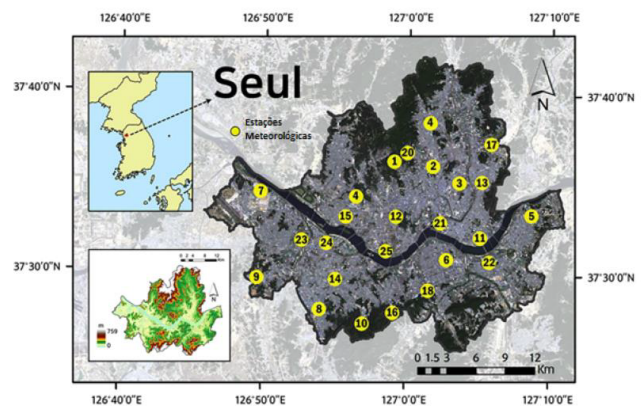


Figura 1. Localização geográfica das estações meteorológicas (Adaptado de CHO et al., 2020).

A base de dados utilizada neste estudo, foi coletada pela KMA durante o verão de cada ano entre 2013 até 2017. As variáveis foram geradas considerando 3 fontes diferentes, sendo elas: variáveis climáticas coletadas por estações do tempo automáticas (dados *in situ*), previsões do modelo numérico LDAPS e dados auxiliares não climáticos relacionadas à geografia de cada estação. O conjunto de todas as 23 variáveis está descrito na Tabela 1 e, ao todo, foram coletadas 7.753 amostras no período de 5 anos. Os dados estão disponíveis publicamente no repositório da Universidade da Califórnia Irvine (Cho et al., 2020b).

Tabela 1. Variáveis da base de dados da KMA.

	Nomenclatura	Descricao (unidade)
Variáveis do modelo LDAPS	LDAPS_RHmin	Umidade relativa do ar m nima do proximo dia (%)
	LDAPS_RHMax	Umidade relativa do ar maxima do proximo dia (%)
	LDAPS_Tmax_lapse	Temperatura do ar maxima do proximo dia ($^{\circ}C$)
	LDAPS_Tmin_lapse	Temperatura do ar m nima do proximo dia ($^{\circ}C$)
	LDAPS_WS	Velocidade media do vento no proximo dia ($m=s$)
	LDAPS_LH	Fluxo de calor latente do proximo dia ($W=m^2$)
	LDAPS_CC1	1 ^a previsao de nebulosidade (0-5 h) (%)
	LDAPS_CC2	2 ^a previsao de nebulosidade (6-11 h) (%)
	LDAPS_CC3	3 ^a previsao de nebulosidade (12-17 h) (%)
	LDAPS_CC4	4 ^a previsao de nebulosidade (18-23 h) (%)
	LDAPS_PPT1	1 ^a previsao de precipitacao (0-5 h) (%)
	LDAPS_PPT2	2 ^a previsao de precipitacao (6-11 h) (%)
	LDAPS_PPT3	3 ^a previsao de precipitacao (12-17 h) (%)
	LDAPS_PPT4	4 ^a previsao de precipitacao (18-23 h) (%)
In-situ	Solar Radiation	Radiacao solar do dia ($wh=m^2$)
	Present_Tmax	Temperatura Maxima Atual ($^{\circ}C$)
	Present_Tmin	Temperatura M nima Atual ($^{\circ}C$)
	Next_Tmax	Temperatura Maxima Atual do proximo dia ($^{\circ}C$)
	Next_Tmin	Temperatura M nima do proximo dia ($^{\circ}C$)
Auxiliares	lat	Latitude ($^{\circ}$)
	lon	Longitude ($^{\circ}$)
	DEM	Elevacao (m)
	Slope	Inclinacao ($^{\circ}$)

Este trabalho de pesquisa teve como objetivo a aplicação de algoritmos de aprendizado de máquina em conjunto com os dados gerados pelo modelo LDAPS para prever a temperatura máxima e mínima para o próximo dia na cidade de Seoul (horizonte de previsão de um-passo-a-frente). Como procedimento metodológico, foram definidas duas abordagens de comparação: na primeira, os modelos foram comparados individualmente em relação aos resultados de previsão do modelo LDAPS e na segunda etapa, foi implementada a técnica de validação cruzada sobre os algoritmos selecionados, onde 8 subconjuntos de dados (derivados da base de dados original), foram selecionados variando o número de observações utilizadas na fase de treinamento e teste. Com isso, o resultado é a média da validação da previsão sobre as observações de teste de cada um dos 8 subconjuntos.

4. MODELOS DE PREVISÃO

Nesta seção serão descritos os 3 modelos de previsão de séries temporais adotados. Também será feita uma breve introdução à metodologia de *Ensemble Forecasting* considerado a implementação do modelo denominado *Multi-Model Ensemble* (MME).

4.1 Regressão por Vetores de Suporte (SVR)

O método máquina de vetores suporte (do inglês, *Support Vector Machine* (SVM) (Boser et al., 1992), é utilizado para classificação e possui resultado comparável ou até superior às redes neurais artificiais em diversas aplicações (Bratsas et al., 2019). Nestes modelos, a ideia é encontrar uma função, também pode ser chamada de hiperplano, que melhor se adapte a um determinado conjunto de dados de treino de forma a atingir pelo menos um determinado valor de desvio dos valores objetivo.

Para possibilitar a determinação do melhor plano, os dados são mapeados em um espaço dimensional de maior ordem. Para reduzir o custo computacional, utiliza-se uma função auxiliar chamada *Kernel*, que permite que as operações sejam realizadas no espaço característico implícito dos dados sem a necessidade de computar as coordenadas dos dados neste plano. A função Kernel denominada função de base radial (RBF), é expressa em (1).

$$K(X_1; X_2) = \exp \frac{kX_1 - X_2 k^2}{2^2} \quad (1)$$

A seleção dos hiperparâmetros da SVR pode ser formalizada como um problema de otimização. Nesse caso pode-se considerar 2 parâmetros de regularização, sendo eles o número de vetores de suporte (C) e a folga (γ). Neste estudo, o algoritmo utilizado para parametrização da SVR foi a evolução diferencial (DE – *Differential Evolution*), proposto por Storn and Price (1997). Por possuir um mecanismo de busca estocástico, é eficaz também quando a função objetivo é descontínua.

4.2 Floresta Aleatória (RF)

O modelo *Floresta Aleatória* representa uma generalização do algoritmo árvores de decisão (do inglês, *Decision Trees* DT), onde o conceito de floresta vem justamente pelo fato de que o modelo é formado por um conjunto de árvores formadas de forma aleatória e o resultado da previsão vem a ser o valor médio das previsões de todas as árvores. O algoritmo de treinamento das árvores decisórias faz uso da técnica de reamostragem (do inglês, *Bootstrap Sampling*), que consiste em criar inúmeras amostras da base original com reposição. Um dos motivos para a utilização desta técnica estar ficando cada vez mais comum é o aumento da capacidade computacional, assim, mais permutações de amostras são possíveis durante o treinamento.

A aplicação da técnica *Bootstrap Sampling* em algoritmos de grande variância, como é o caso da DT, é conhecida por *Bootstrap Aggregation*, ou ainda chamado de *Bagging*. Esse método é utilizado para prevenir o sobre ajuste no resultado de previsão e foi introduzido primeiramente por Breiman (1996). O conceito por trás desta técnica está em combinar predições de diversos algoritmos de aprendizado para criar uma saída mais precisa.

4.3 Rede Neural Perceptron Multicamadas (MLP)

Nesta arquitetura de rede neural artificial, ocorre uma transformação linear da entrada pelos valores dos pesos, o viés e, na sequência, a não linearidade é introduzida

pela função de ativação. A relação entre os pesos atribuídos para cada variável de entrada e o bias de cada neurônio pode ser representado em (2), que descreve como o somatório do produto entre o sinal de entrada x_i com o respectivo peso w_i mais o valor do bias b , para todas as N entradas, resultando na saída do neurônio que será aplicado na função de ativação da rede f .

$$Y = \sum_{i=1}^N (w_i x_i) + b \quad (2)$$

Nesta aplicação de uma MLP-NN, a função de ativação utilizada foi a função ReLu (*Rectified Linear Unit*). A vantagem da utilização desta função é que os neurônios não são ativados ao mesmo tempo, isso implica que um neurônio será desativado apenas quando a saída da transformação linear for igual a zero. A função ReLu pode ser visualizada em (3). A arquitetura da rede constitui-se apenas de uma camada oculta com 8 neurônios, assim como na camada de entrada. em decorrência do horizonte de previsão (um passo à frente), a camada de saída possui apenas um neurônio.

$$f(x) = \begin{cases} 0; & \text{if } x < 0, \\ x; & \text{if } x \geq 0. \end{cases} \quad (3)$$

4.4 Multi-Model Ensemble (MME)

De forma geral, este modelo origina-se do conceito de aprendizagem conjunta, onde o produto final nada mais é do que uma combinação dos resultados de um conjunto de algoritmos. Os algoritmos que formam este conjunto podem ser de mesma origem, como um conjunto de redes neurais (*ensemble* homogêneo) ou de origem diferente (*ensemble* heterogêneo) como é o caso do conjunto deste trabalho, onde os algoritmos baseiam-se em diferentes conceitos em sua forma de aprendizagem. A aplicação deste método não visa somente melhores resultados de previsão como também a redução da variância nas previsões em relação aos algoritmos aplicados de forma individual.

Em outras palavras, para produzir uma previsão, deve-se utilizar algum método para unificar os resultados de cada modelo para cada amostra da base de dados. A combinação mais simples e comum dos membros do grupo é a média aritmética do, que é calculada como uma média simples (também conhecido como *ensemble* conservador). Porém no caso desta aplicação foi utilizado um método de média ponderada que favorece o melhor preditor, ação que é realizada pelo termo multiplicador P conforme demonstrado em (4):

$$EM = \frac{\sum_{i=1}^P (X_i P_i)}{\sum_{i=1}^P P_i} \quad (4)$$

A seleção do melhor preditor nada mais é do que a ordenação de cada modelo de aprendizado de máquina treinado de acordo com o seus resultados de previsão em relação às métricas utilizadas para validação, ou seja, será atribuído maior dentro do conjunto ao algoritmo que melhor se adaptar ao conjunto de dados e apresentar os melhores resultados. Com isso a formação do modelo

MME é realizado ao final da implementação e validação dos demais. Os valores dos pesos são definidos de forma empírica, através da análise do efeito de cada peso na composição do resultado final.

5. CONFIGURAÇÃO DOS MÉTODOS DE PREVISÃO

Com base na descrição dos conceitos e procedimentos de implementação e treinamento dos algoritmos de aprendizado de máquina. Definiu-se os seguintes Hiperparâmetros para cada um dos 3 modelos aplicados neste estudo (SVR, RF, MLP). Os valores estão dispostos na Tabela 2.

Tabela 2. Configuração dos Algoritmos de aprendizado de máquina

	Hiperparâmetro	Valor
SVR	Vetores de Suporte (C)	49.32 (Temp. Max) e 1.09 (Temp. Min)
	Folga (γ)	0.42 (Temp. Max) e 0.49 (Temp. Min)
RF	Numero de Arvores	500 (Temp. Max e Min)
MLP	Camadas Ocultas	2 (Temp. Max e Min)
	Neurônios Camadas Ocultas	80 (Temp. Max e Min)
	Dropout	25% (Temp. Max e Min)
	Epcas	200 (Temp. Max e Min)
	Funcao de Optimizacao	Adam (Temp. Max e Min)

5.1 Multi-Model Ensemble

O modelo MME, foi definido como sendo a média ponderada dos resultados de previsão de cada modelo, o que caracteriza o conjunto preditivo como heterogêneo. Vale ressaltar que não foram treinados novos modelos com parâmetros diferentes para formar este conjunto, a configuração dos algoritmos utilizados nesta etapa foi mantida. Utilizando a Eq. (4), os pesos definidos para cada modelo estão dispostos na Tabela 3.

Tabela 3. Parametrização Modelo MME.

Modelo	Peso Temp. maxima	Peso Temp. minima
SVR	1,00	1,00
RF	2,70	2,60
MLP	1,30	1,40

Os pesos do modelo MME, descritos na Tabela 3 foram definidos da mesma forma que número de neurônios das camadas do modelo multilayer perceptron. Neste caso, foi feito uma grade combinatória envolvendo os três pesos, iniciando em 0 até 3 com variação de 0,1.

5.2 Critérios de avaliação de desempenho

Como métodos de avaliação dos algoritmos treinados para esta aplicação, foram usados a raiz do erro médio quadrático (*RMSE*) e o coeficiente de correlação linear (R^2). Estas duas métricas foram representadas em (5) e (6), respectivamente.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}; \quad (5)$$

$$R^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sum_{i=1}^n (y_i - \bar{y})^2}; \quad (6)$$

onde \hat{y}_i representam os valores preditos, y_i os valores atuais, e \bar{y} a media dos valores reais.

6. RESULTADOS

Nesta seção são apresentados os resultados obtidos quanto ao treinamento dos algoritmos em duas etapas distintas conforme descrito na seção 3. Na primeira abordagem os modelos de aprendizado de máquina são comparados com o modelo LDAPS, na sequência será demonstrado os resultados obtidos através da aplicação do método de validação cruzada. A avaliação dos modelos ocorreu da mesma maneira, considerando os valores de $RMSE$ e R^2 das previsões obtidas. Na Tabela 4 estão descritos os resultados obtidos na primeira etapa, destacando, em negrito, os melhores resultados obtidos para cada métrica.

Tabela 4. Resultados da primeira etapa de aplicação dos modelos.

	Temperatura máxima		Temperatura m nima	
	R^2	$RMSE$	R^2	$RMSE$
LDAPS	0,72	1,87	0,74	1,20
SVR	0,72	1,76	0,84	1,14
RF	0,70	1,85	0,91	0,85
MLP	0,71	1,65	0,88	0,96
MME	0,68	1,76	0,87	0,89

Percebe-se que não houve uma melhora perceptível nos resultados de previsão para a temperatura máxima dentre os modelos aplicados em nenhuma das métricas de avaliação selecionadas. Porém no caso da temperatura mínima houve um aumento considerável em relação ao coeficiente de determinação em cerca de 23% (0,91 com o modelo RF), este aumento de R^2 acarretou uma melhora de 0,35 no $RMSE$. Analisando os resultados apresentados nas Figuras 2 e 3, nota-se as diferenças entre as duas séries temporais (temperaturas máximas e mínimas). Diferente do que acontece com a temperatura máxima, não há diferenças tão abruptas de temperatura mínima em um intervalo curto de tempo, ou seja, é uma série histórica mais constante, o que favorece a atenuação do erro de previsão.

Na segunda abordagem para avaliação dos modelos foi aplicado o método de validação cruzada, onde houve uma subdivisão múltipla do conjunto de dados original de forma aleatória. Dessa forma os subconjuntos diferenciam-se no volume de dados tanto para treinamento quando para a validação. Por se tratar de um estudo voltado para séries temporais destaca-se a importância de se manter a ordem cronológica dos dados, independentemente do segmento de tempo utilizado para aplicação no modelo.

Tabela 5. Resultados da aplicação dos modelos por meio da validação cruzada.

	Temperatura máxima		Temperatura m nima	
	R^2	$RMSE$	R^2	$RMSE$
SVR	0:61±0:06	1:95±0:09	0:73±0:06	1:20±0:11
RF	0:69±0:11	1:74±0:15	0:82±0:13	0:95±0:17
MLP	0:67±0:04	1:67±0:06	0:75±0:07	1:48±0:12
MME	0:68±0:05	1:73±0:06	0:79±0:08	1:13±0:11

Conforme pode ser analisado nos resultados da Tabela 5, houve alteração nos resultados de todos os modelos quando confrontados com os resultados apresentados na Tabela 1.

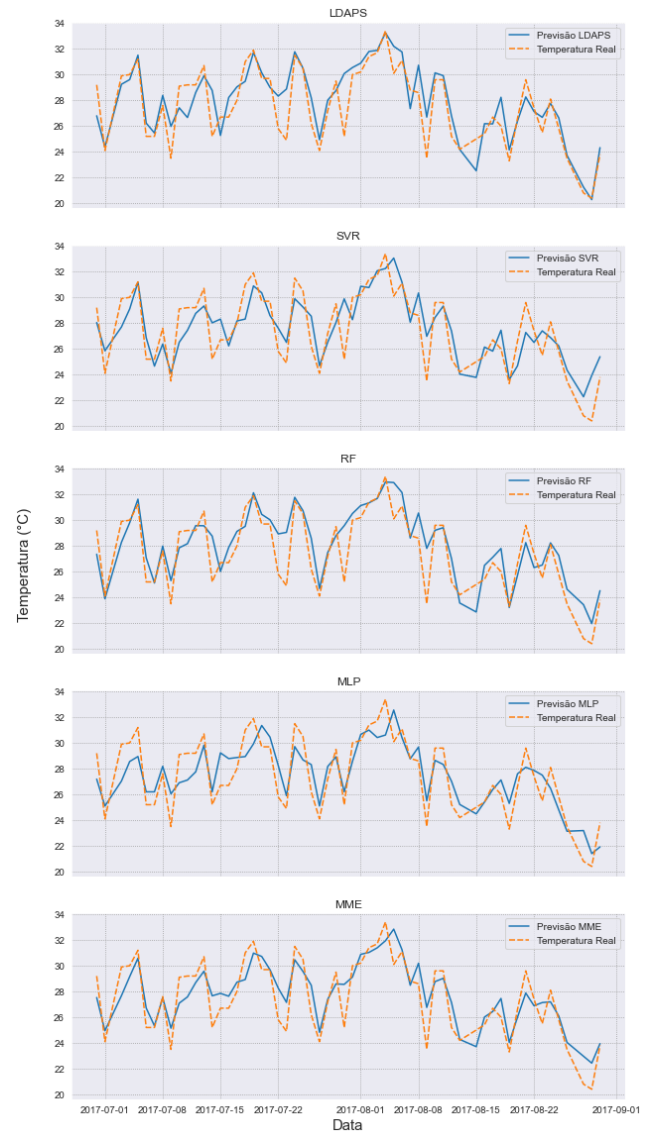


Figura 2. Resultados de previsão de temperatura máxima em comparação com os valores reais.

A mudança mais significativa foi no algoritmo SVR que obteve o maior decréscimo de desempenho, onde os piores resultados de validação foram obtidos nos subconjuntos de dados com menos volume de amostras. O algoritmo RF e MLP demonstraram desempenho similar que às demais validações, sendo que o algoritmo RF apresentou uma leve melhora no resultado, mas também uma alta variabilidade com valores de R^2 entre 0:64 e 0:90 para a validação dos subconjuntos da temperatura mínima. De maneira geral, houve uma mudança na capacidade preditiva dos modelos conforme os subconjuntos de dados foram alterados, principalmente pelo fato da variação do volume de amostras, onde alguns modelos, como por exemplo o SVR, se apresentaram mais sensíveis a quantidades menores de dados.

7. CONCLUSÕES

Este trabalho teve como premissa o estudo das características de séries temporais meteorológicas multivariadas, com o objetivo final de gerar um estudo comparativo entre

