ConBRepro
XVI CONGRESSO BRASILEIRO DE ENGENHARIA DE PRODUÇÃO
EVENTO ON-LINE
02 a 04 de dezembro de 2026
Inovação e Transformação Digital como Estratégia Competitiva nas Engenharias

Predição do Erro de Previsão em Séries Temporais: Uma Comparação de Representações por Atributos

Marcelo Ruy
FAGEN – UFU

Resumo: A previsão de séries temporais subsidia decisões de planejamento em diferentes contextos organizacionais. Como nenhum método apresenta desempenho superior em todas as situações, abordagens de meta-aprendizado utilizam atributos extraídos das séries para caracterizá-las e antecipar o desempenho dos métodos de previsão. Nesse contexto, este trabalho teve como objetivo comparar a capacidade dos atributos dos conjuntos tsfeatures e tsfresh, isoladamente e em conjunto, de predizer o erro de previsão do método ES-RNN nas séries anuais, trimestrais e mensais da Competição M4. O sMAPE do ES-RNN foi calculado com base nos valores futuros e nas previsões para todo o horizonte de cada série, enquanto os atributos foram extraídos exclusivamente dos dados históricos. Após o pré-processamento, as bases foram divididas em conjuntos de treino, validação e teste, nas proporções de 60%, 20% e 20%. Para cada periodicidade e representação, ajustou-se um modelo de random forest, cujos hiperparâmetros foram selecionados por busca aleatória. Os resultados mostraram que todas as representações por atributos superaram o modelo de referência baseado na média do sMAPE de teste. O tsfresh apresentou desempenho superior ao tsfeatures nas três periodicidades. A combinação dos conjuntos produziu ganhos marginais nas séries trimestrais e mensais e desempenho ligeiramente inferior nas anuais, indicando complementaridade limitada entre as representações.

Palavras-chave: Séries temporais, Meta-aprendizado, Extração de atributos, Previsão de séries temporais, Erro de previsão.

Forecast Error Prediction in Time Series: A Comparison of Feature-Based Representations

Abstract: Time series forecasting supports planning decisions across different organizational contexts. Since no single method performs best in all situations, meta-learning approaches use features extracted from time series to characterize them and anticipate the performance of forecasting methods. In this context, this study aimed to compare the ability of tsfeatures and tsfresh features, both separately and jointly, to predict the forecast error of ES-RNN method for the yearly, quarterly, and monthly series of the M4 Competition. The ES-RNN sMAPE was calculated from the future values and forecasts over the entire forecasting horizon of each series, whereas the features were extracted exclusively from historical data. After preprocessing, the datasets were split into training, validation, and test sets using proportions of 60%, 20%, and 20%, respectively. For each frequency and feature representation, a random forest model was fitted, with its hyperparameters selected through random search. The results showed that all feature-based representations outperformed the benchmark model based on the mean test sMAPE. tsfresh outperformed tsfeatures across all three frequencies. Combining both feature sets yielded marginal improvements for the quarterly and monthly series and slightly poorer performance for the yearly series, indicating limited complementarity between the representations.

Keywords: Time series, Meta-learning, Feature extraction, Time series forecasting, Forecast error.

1. Introdução

O recente avanço das tecnologias de informação tem feito com que muitas organizações gerem e coletem diariamente grandes quantidades de dados ordenados no tempo (KANG *et al.*, 2020), formando a chamada série temporal. Essas séries podem ser contínuas, quando o registro é feito de maneira ininterrupta, ou discretas, quando as observações são realizadas em instantes determinados, geralmente separados por intervalos regulares (MORETTIN; TOLOI, 2018).

Uma série temporal é composta de um ou mais componentes básicos: tendência, sazonalidade, variações cíclicas e variações irregulares. A tendência é uma trajetória de aumento ou decréscimo persistente e de longo prazo nos dados. A sazonalidade é um padrão que se repete com uma periodicidade conhecida e fixa e que ocorre quando a série é afetada por fatores como a época do ano ou o dia da semana. O ciclo também apresenta recorrência, mas possui periodicidade não fixa, normalmente longa e às vezes até desconhecida. Tradicionalmente o ciclo é agrupado à tendência e ambos são analisados conjuntamente. As variações irregulares são o elemento que confere à série seu comportamento estocástico e são modeladas como uma sequência de variáveis aleatórias com alguma distribuição de probabilidade (HYNDMAN; ATHANASOPOULOS, 2021).

Adicionalmente, as séries temporais distinguem-se pela dependência existente entre observações sucessivas. Em geral, valores temporalmente próximos tendem a apresentar maior associação do que aqueles separados por intervalos mais longos. Essa dependência pode ser expressa pela autocorrelação serial, que quantifica a associação linear entre os valores da série e suas versões defasadas no tempo (MORETTIN; TOLOI, 2018).

Os principais objetivos de análise de séries temporais são: descrever sua dinâmica; investigar seu mecanismo gerador; identificar padrões periódicos nos dados; e executar previsões de seus valores futuros (MORETTIN; TOLOI, 2018). Na prática, este último tópico é o mais utilizado, principalmente em se tratando de séries econômicas, financeiras e organizacionais. Previsões são requeridas em muitas situações, como planejar a demanda de determinado produto ou o nível de estoque de certo item. Consequentemente, a precisão das previsões é um fator crítico para reduzir custos e fornecer um melhor serviço aos clientes (MAKRIDAKIS; HIBON, 2000).

Uma importante vertente de pesquisas em previsão de séries temporais dedica-se ao desenvolvimento e à comparação de métodos preditivos. A diversidade de abordagens disponíveis pode ser observada, por exemplo, nas Competições M4 e M5, nas quais numerosos métodos foram avaliados em grandes conjuntos de séries (MAKRIDAKIS *et al.*, 2018, 2022). Entretanto, nenhum método apresenta desempenho superior em todas as situações. Surge, assim, uma questão recorrente tanto na pesquisa quanto na prática: qual é o método mais adequado às características dos meus dados? (PETROPOULOS *et al.*, 2014). Esta questão corresponde ao problema de seleção de algoritmos, que pode ser abordado por meio do meta-aprendizado (PETELIN *et al.*, 2023).

O meta-aprendizado procura identificar relações entre as características das tarefas e o desempenho dos algoritmos, de modo a selecionar ou combinar as estratégias mais apropriadas para cada tipo de dado (WANG *et al.*, 2009). Em se tratando de previsões, essa abordagem geralmente se apoia em atributos extraídos das próprias séries, os quais sintetizam propriedades como tendência, sazonalidade, dependência temporal, dentre outras. Além de subsidiarem a seleção ou a ponderação de métodos preditivos (MONTERO-MANSO *et al.*, 2020; TALAGALA *et al.*, 2023; TALAGALA *et al.*, 2022), esses atributos têm sido empregados na simulação de séries com características específicas (KANG *et al.*, 2020; KANG *et al.*, 2017; SMITH-MILES; BOWLY, 2015) e na identificação

de séries atípicas ou anômalas (HYNDMAN *et al.*, 2015). Nesse contexto, Petelin *et al.* (2023) investigaram a capacidade de atributos extraídos pelos conjuntos tsfresh e catch22 de predizer o desempenho dos métodos de previsão participantes da Competição M4. Para isso, os autores utilizaram o sMAPE como variável de interesse e avaliaram diferentes algoritmos de aprendizado de máquina como metamodelos, incluindo random forest (RF), XGBoost, KNN e perceptron multicamadas. Além da capacidade preditiva superior do tsfresh, o estudo constatou que a importância dos atributos pode variar conforme o metamodelo e o método de previsão considerados.

Embora abrangente, o estudo não incluiu o conjunto de atributos tsfeatures. Diante dessa lacuna, o presente artigo tem como objetivo comparar a capacidade dos atributos extraídos pelo tsfeatures e pelo tsfresh, isoladamente e em conjunto, de predizer o sMAPE do método ES-RNN de Smyl (2020) nas séries anuais, trimestrais e mensais da Competição M4.

Para isolar o efeito dos conjuntos de atributos, utilizou-se apenas o RF, que apresentou o melhor desempenho preditivo entre os quatro metamodelos avaliados por Petelin *et al.* (2023). Um modelo de RF foi ajustado separadamente para cada periodicidade, evitando combinar séries com diferentes estruturas temporais. A escolha dessas periodicidades decorre de sua predominância em aplicações práticas na área de Gestão.

A análise concentra-se no ES-RNN, algoritmo vencedor da Competição M4. Essa delimitação é respaldada pela alta correlação observada pelos autores entre os desempenhos dos competidores, isto é, séries relativamente fáceis ou difíceis de prever tendem a apresentar comportamento semelhante entre diferentes métodos.

O interesse pelo tsfeatures se justifica por seu uso em métodos de seleção e combinação de modelos de previsão, como os *frameworks* FFORMS (TALAGALA *et al.*, 2023), FFORMA (MONTERO-MANSO *et al.*, 2020) e FFORMPP (TALAGALA *et al.*, 2022), bem como no método GRATIS de geração de séries temporais com características controladas (KANG *et al.*, 2020). Em particular, o FFORMA alcançou a segunda colocação na Competição M4. Esse resultado reforça a relevância prática desses atributos e motiva sua comparação com o conjunto mais amplo fornecido pelo tsfresh.

Para alcançar tais objetivos, o trabalho estrutura-se em cinco seções, incluindo esta introdução. A segunda apresenta o referencial teórico; a terceira descreve os procedimentos metodológicos; a quarta apresenta e discute os resultados; e a quinta reúne as considerações finais.

2. Referencial Teórico

2.1 A Competição M4 e o Método ES-RNN

A Competição M4 foi realizada em 2018 com o propósito de ampliar e atualizar as evidências acerca da precisão dos métodos de previsão. Foram utilizadas 100 mil séries temporais reais em seis domínios: microeconômico (25,1%), setorial (18,8%), macroeconômico (19,4%), financeiro (24,5%), demográfico (8,7%) e outros (3,4%). Além dos métodos estatísticos tradicionais, a competição incorporou abordagens baseadas em *machine learning* (ML) e avaliou tanto previsões pontuais quanto intervalos de previsão. Quanto à periodicidade, o conjunto compreendeu 23 mil séries anuais, 24 mil trimestrais, 48 mil mensais, 359 semanais, 4.227 diárias e 414 horárias. Os horizontes de previsão foram definidos de acordo com cada frequência: 6 anos, 8 trimestres, 18 meses, 13 semanas, 14 dias e 48 horas (MAKRIDAKIS *et al.*, 2018).

Os resultados reforçaram a eficácia da combinação de métodos, pois 12 dos 17 procedimentos mais precisos empregaram alguma forma de combinação, predominantemente entre abordagens estatísticas. Os dois primeiros colocados utilizaram

estratégias híbridas que integravam métodos estatísticos e de ML. Em contrapartida, os métodos baseados exclusivamente em ML obtiveram resultados inferiores, pois nenhum superou o benchmark de combinação e apenas um apresentou maior precisão que o método ingênuo dessazonalizado. Esses achados indicaram que os maiores avanços não decorreram da substituição dos métodos estatísticos, mas de sua integração com técnicas de ML (MAKRIDAKIS *et al.*, 2018).

O método ES-RNN de Smyl (2020) obteve a primeira colocação na M4, ao combinar modelos de suavização exponencial com redes neurais recorrentes do tipo LSTM. Ele é composto por três elementos principais: dessazonalização e normalização adaptativa das séries, geração das previsões e combinação dos resultados. No primeiro deles, fórmulas de suavização exponencial em espaço de estados são utilizadas para estimar os níveis e os componentes sazonais de cada série, permitindo que os dados sejam dessazonalizados e normalizados antes de serem processados pela rede neural. Esse procedimento mostrou-se especialmente relevante porque as séries da competição foram disponibilizadas sem informações de calendário e apresentavam padrões sazonais provenientes de diferentes domínios. Após o pré-processamento, uma rede LSTM global, treinada conjuntamente sobre numerosas séries, produz as previsões para todo o horizonte, que são posteriormente transformadas de volta a sua escala original e ressazonalizadas. Por fim, o método combina previsões provenientes de diferentes execuções, modelos e épocas de treinamento, com o propósito de reduzir as incertezas associadas aos parâmetros e aumentar a robustez dos resultados. Desse modo, o ES-RNN apresenta caráter simultaneamente híbrido e hierárquico: híbrido por integrar a modelagem estatística da suavização exponencial ao aprendizado de máquina, e hierárquico por conciliar parâmetros locais, específicos de cada série, com parâmetros globais compartilhados entre diferentes séries, possibilitando o aprendizado cruzado sem desconsiderar suas particularidades.

2.2 Medidas de Erro de Previsão

Um erro de previsão é a diferença entre o valor observado e o previsto. Segundo Hyndman e Athanasopoulos (2021), a precisão das previsões pode ser avaliada de três maneiras: por meio dos erros dependentes de escala, dos erros porcentuais e dos erros escalonados. Os erros dependentes de escala estão na mesma unidade de medidas da série original. Muito embora sejam mais simples e intuitivos, a desvantagem é que não se pode comparar erros advindos de séries que estejam em unidades diferentes. Os erros mais comuns deste tipo são o erro absoluto médio (*mean absolute error* – MAE) e a raiz quadrada do erro quadrático médio (*root mean squared error* – RMSE).

Os erros percentuais não possuem o inconveniente anterior, pois são livres de unidade. Entretanto, são indefinidos quando $y_t = 0$ e só são aplicáveis em séries medidas em uma escala de razão. Os dois principais são o erro percentual absoluto médio (*mean absolute percentage error* – MAPE) e o erro percentual absoluto médio simétrico (*symmetric mean absolute percentage error* – sMAPE), cuja fórmula é dada pela equação (1):

$$sMAPE = \frac{2}{h} \cdot \sum_{t=n+1}^{n+h} \frac{\left|Y_t - \hat{Y}_t\right|}{|Y_t| + \left|\hat{Y}_t\right|} \qquad (1)$$

Uma alternativa, sem os problemas anteriores, são os erros escalonados, por exemplo, o erro escalonado absoluto médio (*mean absolute scaled error* – MASE). A despeito disto, este artigo utiliza o sMAPE para manter a comparabilidade com o estudo de Petelin *et al.* (2023).

2.3. Representações de Séries Temporais por Atributos

Atributos de séries temporais são medidas escalares obtidas por meio de operações matemáticas que sintetizam suas propriedades. Por estarem frequentemente associados a

fundamentos teóricos, como a análise de Fourier e a teoria da informação, eles permitem interpretar padrões das séries e compreender aspectos dos processos que as geram. Além disso, possibilitam comparar sistematicamente diferentes métodos e identificar, de maneira empírica, aqueles mais adequados a determinado problema, reduzindo a subjetividade e o esforço envolvidos na seleção manual (HENDERSON; FULCHER, 2025).

Segundo Henderson e Fulcher (2025), o tsfeatures reúne 62 atributos comumente utilizados em econometria e previsão, incorporados progressivamente ao longo dos trabalhos de seus autores. Neste estudo, foram considerados os 42 atributos empregados no FFORMA, com exclusão de dois que permanecem constantes em análises realizadas separadamente por periodicidade ("*nperiods*" e "*seasonal_period*"). Assim, o conjunto compreende 40 características para séries trimestrais e mensais e 32 para séries anuais, uma vez que os atributos sazonais não se aplicam a séries com frequência igual a 1.

O tsfresh é composto por 783 atributos que descrevem propriedades como autocorrelação, entropia, quantis, transformadas rápidas de Fourier e características distribucionais. O conjunto também incorpora o procedimento FRESH (*FeatuRe Extraction based on Scalable Hypothesis tests*), que utiliza testes de hipóteses para filtrar os atributos extraídos e controlar a proporção de características irrelevantes (HENDERSON; FULCHER, 2025).

Além dos conjuntos anteriores, há outras bibliotecas para a extração de atributos de séries temporais. O catch22 reúne 22 atributos representativos, selecionados entre mais de 7.700 características do conjunto hctsa para reduzir a redundância sem comprometer o desempenho em tarefas de classificação de séries. O TSFEL, originalmente desenvolvido para dados inerciais, disponibiliza 156 atributos dos domínios estatístico, temporal, espectral e de *wavelets*. Já o Kats, desenvolvido pelo Facebook Research, oferece recursos para previsão, detecção de anomalias e extração de 40 atributos, incluindo medidas de tendência, autocorrelação e decomposição STL (HENDERSON; FULCHER, 2025).

3. Metodologia

Esta seção descreve os procedimentos de obtenção, pré-processamento e análise dos dados. Todas as etapas foram realizadas no software R, versão 4.5.3 (R CORE TEAM, 2026), com o auxílio de pacotes computacionais específicos.

As séries temporais foram obtidas por meio do pacote M4comp2018 (MONTERO-MANSO *et al.*, 2019), que reúne todas as séries da Competição M4, seus valores futuros e as previsões submetidas pelos 25 participantes mais bem classificados. Desse conjunto, foram selecionados os dados históricos, os valores futuros e as previsões pontuais do ES-RNN para as séries anuais, trimestrais e mensais.

Com base nos valores futuros e nas respectivas previsões, calculou-se o sMAPE do ES-RNN sobre todo o horizonte de previsão de cada série (seção 2.1), conforme a Equação (1). Os atributos dos conjuntos tsfeatures e tsfresh, por sua vez, foram extraídos exclusivamente dos dados históricos. Esse procedimento preserva a natureza prospectiva da tarefa e evita o vazamento de informações, pois os valores do período de teste não estariam disponíveis no momento em que as previsões fossem realizadas. Os atributos do tsfeatures foram calculados com o pacote homônimo (HYNDMAN *et al.*, 2019), enquanto os atributos do tsfresh foram obtidos por meio do pacote theft (HENDERSON, 2022).

Na sequência, realizou-se o pré-processamento dos dados, que compreendeu o tratamento de valores ausentes e infinitos, bem como a exclusão de atributos com variância zero. Como o algoritmo random forest não requer a normalização dos atributos (GÉRON, 2023), esse procedimento não foi realizado. As séries anuais receberam tratamento distinto das

trimestrais e mensais, pois não permitem o cálculo dos atributos sazonais do tsfeatures, resultando em um conjunto menor de características (seção 2.3).

Para as séries trimestrais e mensais, foram obtidos valores válidos para os 40 atributos do tsfeatures, com exceção de uma série mensal que gerou erro durante a extração e foi excluída. O tsfresh, por sua vez, produziu uma quantidade elevada de valores ausentes e infinitos, concentrados em determinados atributos, o que inviabilizou a simples exclusão de todas as séries com dados inválidos. Com base na análise dos quantis amostrais, foram inicialmente excluídos os atributos com mais de 1% de valores ausentes ou infinitos. Esse limiar foi definido de maneira empírica, buscando equilibrar a retenção de atributos e a preservação do número de séries. Após a consolidação dessa etapa, eliminaram-se, sucessivamente, as séries com valores ausentes, as séries com valores infinitos e os atributos com variância zero. Esse processo resultou na exclusão de 412 atributos e 342 séries trimestrais, além de 271 atributos e 862 séries mensais. Ao final, permaneceram 23.658 séries trimestrais (98,6% do total) e 47.138 séries mensais (98,2%), representadas por 40 atributos do tsfeatures e, respectivamente, 372 e 513 atributos do tsfresh.

Para as séries anuais, foram calculados os 32 atributos aplicáveis do tsfeatures, alguns dos quais produziram valores ausentes. Como esse conjunto é relativamente reduzido, adotou-se um limiar menos restritivo, excluindo-se inicialmente os atributos com mais de 3% de valores inválidos, de modo a preservar todos os atributos do tsfeatures. Em seguida, foram eliminadas, em etapas sucessivas, as séries com valores ausentes, as séries com valores infinitos e os atributos com variância zero. Ao todo, foram excluídas 3.049 séries e 415 atributos, restando 19.951 séries anuais (86,7% do total), representadas por 32 atributos do tsfeatures e 369 do tsfresh.

Concluído o pré-processamento, os dados foram organizados para a análise de regressão por random forest. A etapa anterior resultou em três bases, correspondentes às séries anuais, trimestrais e mensais, cada uma contendo o sMAPE do ES-RNN e os respectivos atributos dos conjuntos tsfeatures e tsfresh. Cada base foi dividida, por amostragem aleatória simples, em conjuntos de treino, validação e teste, nas proporções de 60%, 20% e 20%, respectivamente. Para assegurar a reprodutibilidade do particionamento, foram utilizadas sementes previamente definidas, uma para cada periodicidade. A partir de cada base, foram consideradas três representações: tsfeatures, tsfresh e a combinação de ambos. As mesmas partições foram mantidas nas três representações.

A modelagem por random forest foi realizada com o pacote mlr3 (LANG *et al.*, 2019), utilizando a implementação do pacote ranger. Os hiperparâmetros foram ajustados por busca aleatória, mediante a avaliação de 100 configurações (RHYS, 2020). Cada configuração foi ajustada aos dados de treino e avaliada na partição de validação, adotando-se a minimização do RMSE como critério de seleção. A otimização contemplou os seguintes hiperparâmetros: (*i*) *mtry* – número de atributos selecionados aleatoriamente como candidatos em cada divisão das árvores, variando de 1 até o total de atributos da respectiva representação; (*ii*) *min.node.size* – número mínimo de observações permitido em um nó terminal, variando de 1 a 30; (*iii*) *sample.fraction* – proporção das observações utilizada na construção de cada árvore, variando de 0,50 a 1,00; (*iv*) *max.depth* – profundidade máxima das árvores, variando de 5 a 30. O hiperparâmetro *num.trees*, correspondente ao número de árvores, não foi submetido à otimização e permaneceu fixado em 500, como solução de compromisso entre estabilidade preditiva e custo computacional (BERK, 2020). Após a seleção da configuração com o menor RMSE de validação, o modelo correspondente foi avaliado na partição de teste. O desempenho foi mensurado pelo RMSE, MAE e R^2 e comparado ao de um modelo de referência baseado na média do sMAPE do conjunto de teste. Os resultados dessa avaliação são apresentados na seção seguinte.

4. Resultados e Discussão

Aplicando-se o procedimento descrito na seção anterior, foram avaliados 12 modelos para estimar o sMAPE do ES-RNN, considerando as três periodicidades e quatro configurações de atributos, conforme apresentado na Tabela 1. Para cada periodicidade, foram ajustados três modelos de regressão por random forest (RF), utilizando os atributos do tsfeatures, do tsfresh e de ambos os conjuntos. A configuração “Nenhum” corresponde ao modelo de referência sem variáveis explicativas, cuja previsão é dada pela média do sMAPE observada no conjunto de teste.

TABELA 1 – Desempenho dos atributos e do modelo de referência, segundo a periodicidade

Atributos	Séries Anuais			Séries Trimestrais			Séries Mensais		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
Nenhum	0,1472	0,0949	—	0,1257	0,0786	—	0,1378	0,1002	—
tsfeatures	0,1197	0,0729	0,3390	0,0948	0,0546	0,4322	0,0914	0,0532	0,5602
tsfresh	0,1079	0,0648	0,4622	0,0813	0,0451	0,5823	0,0814	0,0463	0,6509
Ambos	0,1084	0,0649	0,4575	0,0805	0,0446	0,5905	0,0808	0,0460	0,6560

Fonte: Dados da pesquisa

O primeiro resultado observado na Tabela 1 é que, nas três periodicidades, todas as representações por atributos apresentaram RMSE e MAE inferiores aos do modelo de referência. Além disso, os modelos com atributos produziram valores positivos e moderados de R^2, indicando desempenho superior à previsão constante. Esses resultados mostram que as propriedades extraídas dos dados históricos contêm informações capazes de distinguir séries nas quais o ES-RNN tende a apresentar menor ou maior sMAPE. Como os atributos foram extraídos exclusivamente do histórico e o sMAPE foi calculado sobre os horizontes futuros, sua capacidade preditiva foi avaliada fora da amostra, e não apenas de forma descritiva.

Também se observa na Tabela 1 que a capacidade preditiva, medida pelo R^2 dos metamodelos, aumenta com a frequência das séries. O desempenho melhora das séries anuais para as trimestrais e mensais nos três conjuntos de atributos. Possíveis explicações estão relacionadas (*i*) ao tamanho das séries temporais: em média, séries de maior frequência possuem mais observações, o que pode favorecer a estimação das características da dinâmica dos dados; e (*ii*) à quantidade de observações por ciclo anual: séries mensais possuem 12 observações, trimestrais 4 e anuais apenas 1, proporcionando maior riqueza para a identificação de padrões sazonais intra-anuais, que sequer podem ser capturados em séries com frequência anual.

Petelin *et al.* (2023) reportaram RMSE de aproximadamente 0,18 para o modelo de referência, valor superior aos observados na Tabela 1. Essa diferença pode estar relacionada ao fato de os autores terem considerado conjuntamente todos os métodos da Competição M4, enquanto o presente estudo utilizou apenas o ES-RNN. Além disso, os autores combinaram todas as periodicidades em um único conjunto, ao passo que, neste trabalho, o modelo de referência foi calculado separadamente para as séries anuais, trimestrais e mensais. Como essas periodicidades apresentam níveis de erro de previsão distintos, uma única previsão pela média tende a ser menos adequada ao conjunto agregado, contribuindo para o maior RMSE observado.

Esses autores também relataram RMSE próximo de 0,11 para o RF baseado no tsfresh, valor superior aos obtidos para as séries trimestrais e mensais deste estudo e próximo ao observado para as séries anuais (ver Tabela 1). Embora os resultados não sejam diretamente comparáveis, pelas diferenças metodológicas entre os trabalhos, uma possível explicação seria a modelagem adotada. Enquanto Petelin *et al.* (2023) agregaram

diferentes periodicidades em um único metamodelo, neste artigo foi ajustado um RF separadamente para cada frequência. Essa abordagem pode favorecer a aprendizagem ao reduzir a heterogeneidade associada a diferentes estruturas temporais e distribuições de erro, desde que cada grupo disponha de séries suficientes para o ajuste de um modelo específico, como é o caso das séries anuais, trimestrais e mensais, que perfazem 95% dos dados da Competição M4.

Outro resultado evidenciado pela Tabela 1 é a superioridade consistente do tsfresh sobre o tsfeatures nas três periodicidades e métricas avaliadas. Esse resultado é convergente com o observado por Petelin *et al.* (2023), que também encontraram desempenho superior do tsfresh em relação ao catch22 na predição do erro dos métodos de previsão da Competição M4. Assim, nos dois estudos, o tsfresh apresentou maior capacidade preditiva que os conjuntos de atributos comparados diretamente, embora em contextos experimentais distintos. Esse desempenho sugere que o conjunto mais amplo de atributos do tsfresh captura informações relacionadas ao erro de previsão que não são integralmente representadas por conjuntos mais compactos, como tsfeatures e catch22. Contudo, o tsfeatures também apresentou capacidade preditiva relevante, mesmo utilizando um número substancialmente menor de atributos.

Por fim, a combinação dos dois conjuntos proporcionou ganhos diminutos em relação ao uso isolado do tsfresh nas séries trimestrais e mensais, enquanto, nas anuais, resultou em desempenho ligeiramente inferior. Esses resultados indicam que os atributos do tsfeatures acrescentam pouca informação preditiva quando os atributos do tsfresh já estão presentes, sugerindo complementaridade limitada entre as duas representações. Assim, parte substancial da informação capturada pelo tsfeatures parece estar representada, direta ou indiretamente, no tsfresh.

5. Conclusões

Este trabalho comparou a capacidade de duas representações de séries temporais, tsfeatures e tsfresh, de forma isolada e conjuntamente, de predizer o sMAPE do método ES-RNN nas séries anuais, trimestrais e mensais da Competição M4. Para isso, foram utilizados modelos de regressão por random forest específicos para cada periodicidade e representação, a partir dos atributos extraídos dos dados históricos. Cada configuração foi ajustada aos dados de treino, otimizada na partição de validação e avaliada no conjunto de teste por meio do RMSE, MAE e R^2, tendo como referência um modelo sem atributos, cuja previsão corresponde à média do sMAPE observada nos dados de teste.

Os resultados indicam que os atributos de séries temporais contêm informação relevante para antecipar o erro do ES-RNN, uma vez que todas as representações superaram o modelo de referência nas três periodicidades. O tsfresh apresentou desempenho superior ao tsfeatures em todas as periodicidades, enquanto a combinação de ambos produziu apenas ganhos marginais nas séries trimestrais e mensais e desempenho ligeiramente inferior nas anuais. Assim, para a tarefa analisada, o tsfresh isoladamente mostrou-se a alternativa mais adequada, embora o tsfeatures também tenha apresentado capacidade preditiva relevante utilizando um número substancialmente menor de atributos. Observou-se ainda maior capacidade de predição do erro nas séries mensais, seguidas pelas trimestrais e anuais.

Como principal contribuição, o estudo amplia a investigação sobre o uso de atributos de séries temporais na predição do desempenho de métodos de previsão ao incorporar o tsfeatures à comparação com o tsfresh e avaliar também sua combinação. Os resultados fornecem evidências de que a escolha da representação influencia a capacidade preditiva dos metamodelos e sugerem que a ampliação do conjunto de atributos nem sempre produz

ganhos proporcionais. Nesse sentido, o estudo oferece subsídios para a escolha entre representações mais abrangentes, potencialmente mais informativas, e representações mais compactas e parcimoniosas, especialmente em abordagens de meta-aprendizado voltadas à previsão.

O estudo apresenta, contudo, limitações que restringem a generalização dos resultados. A análise concentrou-se no ES-RNN, no random forest e nas séries anuais, trimestrais e mensais da Competição M4, não permitindo extrapolar diretamente as conclusões para outros métodos de previsão, metamodelos, periodicidades ou conjuntos de dados. Além disso, foi utilizada uma única divisão em treino, validação e teste, não sendo possível avaliar a variabilidade dos resultados decorrente de diferentes particionamentos.

Como trabalhos futuros, recomenda-se ampliar a análise para outros métodos de previsão e algoritmos de meta-aprendizado, bem como empregar procedimentos de reamostragem que permitam avaliar a estabilidade dos resultados. Também seria relevante investigar a capacidade preditiva e o custo computacional das representações, considerando o tempo de extração dos atributos, o consumo de memória e o tempo de ajuste dos metamodelos. Essa análise permitiria verificar se o menor número de atributos do tsfeatures se traduz efetivamente em vantagens computacionais e estabelecer em quais condições uma representação mais compacta pode ser preferível à maior capacidade preditiva proporcionada pelo tsfresh.

Referências

BERK, R. A. **Statistical Learning from a Regression Perspective**. 3. ed. Cham: Springer International Publishing, 2020.

GÉRON, A. **Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow:** concepts, tools, and techniques to build intelligent systems. 3. ed. Sebastopol: O'Reilly Media, 2023.

HENDERSON, T. **theft:** Tools for Handling Extraction of Features from Time Series, 2022. Disponível em: https://CRAN.R-project.org/package=theft. Acesso em: 29 ago. 2026.

HENDERSON, T.; FULCHER, B. D. Feature-Based Time-Series Analysis in R using the Theft Ecosystem. **The R Journal**, v. 17, n. 3, p. 43–68, 21 out. 2025.

HYNDMAN, R. *et al.* **tsfeatures:** Time Series Feature Extraction, 2019. Disponível em: https://CRAN.R-project.org/package=tsfeatures. Acesso em: 29 ago. 2026.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting:** principles and practice. 3. ed. Melbourne, Australia: OTexts, 2021.

HYNDMAN, R. J.; WANG, E.; LAPTEV, N. Large-Scale Unusual Time Series Detection. *In*: 2015 IEEE INTERNATIONAL CONFERENCE ON DATA MINING WORKSHOP (ICDMW). **2015 IEEE International Conference on Data Mining Workshop (ICDMW)**. Atlantic City, NJ, USA: IEEE, nov. 2015. Disponível em: http://ieeexplore.ieee.org/document/7395871/. Acesso em: 26 ago. 2026.

KANG, Y.; HYNDMAN, R. J.; LI, F. GRATIS: GeneRAting TIme Series with diverse and controllable characteristics. **Statistical Analysis and Data Mining: The ASA Data Science Journal**, v. 13, n. 4, p. 354–376, ago. 2020.

KANG, Y.; HYNDMAN, R. J.; SMITH-MILES, K. Visualising forecasting algorithm performance using time series instance spaces. **International Journal of Forecasting**, v. 33, n. 2, p. 345–358, abr. 2017.

LANG, M. *et al.* mlr3: A modern object-oriented machine learning framework in R. **Journal of Open Source Software**, v. 4, n. 44, p. 1903, 11 dez. 2019.

MAKRIDAKIS, S.; HIBON, M. The M3-Competition: results, conclusions and implications. **International Journal of Forecasting**, v. 16, n. 4, p. 451–476, out. 2000.

MAKRIDAKIS, S.; SPILIOTIS, E.; ASSIMAKOPOULOS, V. The M4 Competition: Results, findings, conclusion and way forward. **International Journal of Forecasting**, v. 34, n. 4, p. 802–808, out. 2018.

MAKRIDAKIS, S.; SPILIOTIS, E.; ASSIMAKOPOULOS, V. M5 accuracy competition: Results, findings, and conclusions. **International Journal of Forecasting**, v. 38, n. 4, p. 1346–1364, out. 2022.

MONTERO-MANSO, P. *et al.* FFORMA: Feature-based forecast model averaging. **International Journal of Forecasting**, v. 36, n. 1, p. 86–92, jan. 2020.

MONTERO-MANSO, P.; NETTO, C.; TALAGALA, T. **M4comp2018:** Data from the M4-Competition, 2019. Disponível em: https://github.com/carlanetto/M4comp2018.

MORETTIN, P. A.; TOLOI, C. M. **Análise de Séries Temporais:** modelos lineares univariados. 3. ed. São Paulo: Editora Edgard Blücher, 2018. v. 1.

PETELIN, G.; CENIKJ, G.; EFTIMOV, T. Towards understanding the importance of time-series features in automated algorithm performance prediction. **Expert Systems with Applications**, v. 213, p. 119023, mar. 2023.

PETROPOULOS, F. *et al.* “Horses for Courses” in demand forecasting. **European Journal of Operational Research**, v. 237, n. 1, p. 152–163, ago. 2014.

R CORE TEAM. **R:** A Language and Environment for Statistical Computing. Vienna, Austria. R Foundation for Statistical Computing, 2026. Disponível em: https://www.R-project.org/.

RHYS, H. **Machine Learning with R, the Tidyverse, and Mlr**. New York: Manning Publications, 2020.

SMITH-MILES, K.; BOWLY, S. Generating new test instances by evolving in instance space. **Computers & Operations Research**, v. 63, p. 102–113, nov. 2015.

SMYL, S. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. **International Journal of Forecasting**, v. 36, n. 1, p. 75–85, jan. 2020.

TALAGALA, T. S.; HYNDMAN, R. J.; ATHANASOPOULOS, G. Meta-learning how to forecast time series. **Journal of Forecasting**, v. 42, n. 6, p. 1476–1501, set. 2023.

TALAGALA, T. S.; LI, F.; KANG, Y. FFORMPP: Feature-based forecast model performance prediction. **International Journal of Forecasting**, v. 38, n. 3, p. 920–943, jul. 2022.

WANG, X.; SMITH-MILES, K.; HYNDMAN, R. J. Rule induction for forecasting method selection: Meta-learning the characteristics of univariate time series. **Neurocomputing**, v. 72, n. 10–12, p. 2581–2594, jun. 2009.