

ANÁLISE DE SÉRIES TEMPORAIS PARA PREVISÃO DE DEMANDA NO INSS

Trabalho de Conclusão de Curso
apresentado como parte dos requisitos para
obtenção do grau de Especialista em Ciência de
Dados.

Aluno: Elias Fernandes de Oliveira

Orientador: Prof. Me. Maxwell Sarmiento

Brasília – DF

Setembro/2022



Escola Nacional de Administração Pública

ANÁLISE DE SÉRIES TEMPORAIS PARA PREVISÃO DE DEMANDA NO INSS

Autor: Elias Fernandes de Oliveira
Controladoria-Geral da União

Palavras-chave: Instituto Nacional do Seguro Social (INSS), séries temporais, demanda, previsão.

O trabalho realizado abordou a gestão de demandas no âmbito do INSS, tendo como objetivo a identificação de método apropriado para previsão de demandas futuras. Foram utilizados dados sobre demandas mensais do INSS e aplicadas técnicas de análise de séries temporais. Para organização e análise dos dados, utilizou-se linguagem de programação Python. Como resultado, foram identificados modelos de previsão com desempenho satisfatório nos testes realizados, havendo argumentos que justificam a priorização do modelo de Suavização Exponencial Simples (SES). Ademais, constatou-se que a série histórica avaliada foi influenciada por fatores externos relevantes, o que pode ter causado distorções nos dados e, conseqüentemente, impactado nas análises realizadas, sendo oportuna a realização de análises adicionais após o incremento de novas observações na série.

1 Introdução

Poucos órgãos públicos são tão conhecidos e estão tão presentes no cotidiano dos cidadãos brasileiros como o Instituto Nacional do Seguro Social (INSS), Autarquia responsável pela concessão de uma série benefícios e serviços relacionados à Seguridade Social e, em especial, à Previdência Social. Ao final de 2021, a folha de pagamento do instituto totalizava mais de 36 milhões de beneficiários a um custo anual de R\$ 684 bilhões¹.

Tamanha relevância vem acompanhada de constante escrutínio acerca da qualidade e da tempestividade dos serviços prestados pelo INSS, especialmente em virtude do impacto social associado às atribuições do Instituto. Nesse contexto, um tema constantemente abordado, especialmente por órgãos de controle e pela imprensa, é a capacidade da Autarquia em atender tempestivamente sua demanda, comumente tratada como a “fila do INSS”.

A “fila do INSS”, historicamente caracterizada como uma fila física de pessoas em busca de atendimento nas agências da previdência social (APS) transformou-se, nos últimos anos, em uma fila digital, haja vista a mudança de paradigma introduzida pela Transformação Digital² do INSS, que teve como objetivos tanto facilitar o acesso dos cidadãos aos serviços quanto flexibilizar a distribuição das demandas entre os servidores da Autarquia. A partir de então, grande parte da demanda passou a ser apresentada via canais digitais de atendimento, sem a necessidade de atendimentos presenciais ou agendamentos.

Tais mudanças propiciaram ao cidadão maior celeridade e conforto na apresentação de requerimentos ao INSS. Por outro lado, representaram um grande desafio de gestão ao Instituto, que precisou reorganizar sua força de trabalho e capacitar servidores para trabalhar com a análise de requerimentos recebidos via canais digitais de atendimento. Ademais, a flexibilização dos canais de solicitação permitiu a entrada de requerimentos que compunham uma demanda reprimida pela capacidade de atendimento das APS.

Assim, apesar das mudanças empreendidas pela gestão do INSS representarem avanços em relação ao modelo anterior, o Instituto segue enfrentando dificuldades na gestão de suas

¹ Números obtidos em consulta ao Painel de Pagamentos de Benefícios Previdenciários, organizado pela Controladoria-Geral da União. Disponível em: <https://www.gov.br/cgu/pt-br/centrais-de-conteudo/paineis/beneficiosprevidenciarios>

² Mais informações em: <https://www.gov.br/inss/pt-br/centrais-de-conteudo/publicacoes/apresentacoes/plano-digital-inss-apresentacao-nova-proposta-02-7-lorena-seme-final-pdf>

demandas – constata-se, por exemplo, que existiam cerca de 1,8 milhões de requerimentos de benefício aguardando uma decisão da Autarquia ao final de 2021³.

Diante do exposto, e de forma a contribuir com a discussão apresentada, realizou-se estudo sobre a demanda de requerimentos apresentados mensalmente ao INSS. Entende-se que a referida análise é relevante para subsidiar ações relacionadas à alocação de pessoas e à adoção de medidas para incremento da produtividade. Ademais, levantamentos preliminares indicaram que o INSS não possui metodologia consolidada para projeção de demandas futuras.

A partir do tema proposto para a pesquisa, e adotando como referência teorias relacionadas ao estudo de séries temporais, foram realizadas análises visando a identificação de um modelo que possa auxiliar o INSS a prever e, conseqüentemente, gerir suas demandas.

A opção pela aplicação de técnicas de análise e modelagem de séries temporais em detrimento de métodos alternativos, a exemplo da aplicação da teoria das filas, deve-se, principalmente, aos dados disponíveis quando da realização da pesquisa, que consistiam na série histórica de demandas recebidas mensalmente pelo INSS. Trata-se de informação suficiente para proposição de modelos de previsão de séries temporais; por outro lado, modelos baseados na teoria das filas pressupõe informações adicionais sobre o sistema de fila, como detalhes sobre a quantidade de atendentes e sobre os tempos de atendimento.

2 Fundamentação Teórica

2.1 Demandas do INSS

Conforme dispõe a Portaria PRES/INSS nº 1.372, de 28.10.2021, as Centrais de Análise de Benefício⁴ (CEAB) são unidades responsáveis pelo tratamento das solicitações encaminhadas ao INSS, voltadas “à análise de requerimentos de reconhecimento de direitos em todas as suas fases, compensação previdenciária, apuração de indício de irregularidade, manutenção de

³ <https://agenciabrasil.ebc.com.br/economia/noticia/2022-01/previdencia-social-inicia-2022-com-desafio-de-zerar-filas>

⁴ As CEAB são classificadas de acordo com o tipo de demanda a ser tratada: Demandas Judiciais (Ceab/DJ); Manutenção e Cadastro - Ceab/MAN; e Reconhecimento de Direitos - Ceab/RD. Há ainda uma CEAB para tratamento de demandas relacionadas ao Regime Próprio de Previdência Social (RPPS) da União, mas que possui atribuições e organização específicas, sendo gerida de forma diversa das demais CEAB, motivo pelo qual não foi incluída no escopo deste trabalho.

benefícios, atualização de cadastro e demandas judiciais”. As CEAB são, portanto, as estruturas responsáveis pelo tratamento de grande parte das demandas por serviços apresentadas pelos cidadãos ao INSS, possuindo influência direta, por exemplo, sobre a concessão de benefícios como aposentadorias e pensões.

Além das CEAB, outro importante componente no contexto da já mencionada Transformação Digital do INSS foi o Sistema Gerenciador de Tarefas (GET)⁵ – sistema utilizado para gestão das demandas no âmbito do Instituto (tratadas como tarefas), realizando registros de entrada, dos trâmites e das conclusões relacionadas aos diversos tipos de serviço prestados. A partir da implementação do referido sistema, dados sobre demandas recebidas pelo INSS tornaram-se mais acessíveis, tornando mais propícia a realização de estudos e análises para apoiar a gestão da Autarquia.

Por meio do GET é possível, por exemplo, identificar a quantidade de tarefas criadas diariamente, mensalmente ou anualmente, com detalhamentos variados sobre o tipo de tarefa, a origem da demanda, dentre outros recortes possíveis. Destaque-se, por sua vez, que apenas em meados de 2018 o sistema passou a ser utilizado de forma integral pelo INSS, o que pode fragilizar dados referentes a períodos anteriores.

2.2 Séries temporais: características e análise exploratória

O presente trabalho prestou-se à aplicação de técnicas de análise de séries temporais junto aos dados de demandas mensais do INSS. Sobre o tema, Vishwas e Patel definem:

[...] uma série temporal é uma coleção de dados armazenados de forma cronológica. Análises estatísticas e matemáticas realizadas para encontrar padrões ocultos e informações relevantes em dados como esses são chamadas análises de séries temporais. Técnicas de modelagem de séries temporais são utilizadas para compreender padrões do passado nos dados e para tentar prever horizontes futuros (VISHWAS; PATEL, 2020, tradução nossa)⁶.

Os referidos autores destacam, ainda, que séries temporais que incorporam registros de uma única característica ou variável são chamadas séries temporais univariadas. Por outro lado,

⁵ Sobre o sistema, o INSS (BRASIL, 2019) afirma tratar-se de um “sistema provedor da infraestrutura que, desde dezembro de 2016, está à disposição do INSS para estabelecer de forma plena sua nova proposta de gestão pública colaborativa. Permite que qualquer entidade parceira atue efetivamente como um posto avançado do instituto, de forma segura e eficiente. É a plataforma fundamental para a proposta de gestão pública colaborativa do INSS”.

⁶ Texto original: A time series is a collection of data points that are stored with respect to their time. Mathematical and statistical analysis performed on this kind of data to find hidden patterns and meaningful insight is called time-series analysis. Time-series modeling techniques are used to understand past patterns from the data and try to forecast future horizons.

caso os registros incorporem mais de uma característica ou variável, a série temporal é caracterizada como multivariada.

Auffarth (2021) aponta que a análise de séries temporais envolve diversas técnicas de exploração de dados, como a visualização de distribuições, a análise de tendências e de padrões cíclicos e sazonais e a análise de relacionamentos entre variáveis de interesse. Conforme entendimento consolidado nos referenciais teóricos, uma série temporal é composta pela agregação dos seguintes elementos:

- a) Tendência: representa o grau de variação da média dos dados da série com o decorrer do tempo. A tendência demonstra o comportamento de longo prazo da série, podendo apresentar padrões de crescimento ou de decréscimo;
- b) Sazonalidade: caracteriza-se pela ocorrência de padrões de flutuação em intervalos regulares de tempo, muito comum, por exemplo, em dados sobre economia e clima;
- c) Variações Cíclicas: caracterizam-se por flutuações periódicas observadas em séries temporais, alterando o comportamento da série, mas que ocorrem em frequência menor se comparadas com a flutuação sazonal – normalmente, com intervalos superiores a um ano;
- d) Ruído: são variações irregulares na série de dados, que não apresentam padrão identificável. Enquanto os demais componentes da série temporal podem ser identificados e utilizados para modelagem e previsão, o ruído é o componente não modelável, compondo grande parte dos erros existentes em modelos de previsão.

Posto isso, a literatura confere especial destaque à etapa de exploração e pré-processamento dos dados de séries temporais, tanto para identificação de características básicas do conjunto de dados – como origem dos dados, existência de falhas ou erros de coleta e forma de agregação e disponibilização das informações – quanto para descrição de aspectos específicos à análise de séries temporais, dentre os quais destacam-se:

- a) Estacionariedade: de acordo com Nielsen (2021), “*uma série temporal estacionária é aquela que tem propriedades estatísticas razoavelmente estáveis ao longo do tempo, sobretudo no que diz respeito à média e à variância*”. Muitos modelos estatísticos tradicionais de séries temporais pressupõem uma série temporal estacionária;
- b) Sazonalidade: “*é qualquer tipo de comportamento recorrente no qual a frequência é estável. Pode ocorrer em muitas frequências diferentes e ao mesmo tempo [...]*”.

Identificar e lidar com a sazonalidade faz parte do processo de modelagem” (NIELSEN, 2021);

- c) Autocorrelação: refere-se à ideia de que “*um valor em uma série temporal em um determinado ponto no tempo pode estar correlacionado com o valor em outro ponto no tempo*” (NIELSEN, 2021).

A literatura especializada descreve a existência de análises e testes para avaliar o comportamento das séries em relação aos aspectos supracitados. Quanto à estacionariedade, além da inspeção visual por meio de análises gráficas, há dois testes de hipótese comumente citados: o teste de Dickey-Fuller aumentado (ADF) e o teste de Kwiatkowski-Phillips-Schmidt-Shin (KPSS), que testam a existência de raízes unitárias nas séries temporais (apesar de possuírem hipóteses nulas opostas).

Quanto à sazonalidade e à autocorrelação dos dados, são comumente analisadas por meio da representação gráfica das funções de autocorrelação⁷ e de autocorrelação-parcial⁸ da série temporal.

2.3 Modelos para previsão em séries temporais

A partir da realização da análise exploratória dos dados, é possível avaliar a aplicabilidade de diferentes modelos para previsão de séries temporais. Modelos estatísticos e de aprendizado de máquina (inclusive métodos de aprendizado profundo, ou *deep learning*) são comumente citados pela literatura especializada⁹ como opções para realização de previsões em séries temporais. Ademais, conforme destacado por Nau (2020), há casos em que modelos simples podem ser mais adequados para a realização de previsões.

⁷ De acordo com Auffarth (2021), a função de autocorrelação é um método para determinar a existência de correlação entre valores sucessivos de uma série.

⁸ A função de autocorrelação parcial apresenta a correlação entre um valor em determinado instante e uma de suas defasagens, retirando os efeitos das outras defasagens.

⁹ A exemplo das obras mencionadas nas “Referências Bibliográficas”.

2.3.1 Suavização Exponencial

Os métodos de suavização exponencial originam-se dos estudos realizados por Robert Goodell Brown, publicados em 1956. Posteriormente, Charles Holt e Peter Winters apresentaram contribuições para o desenvolvimento de novos algoritmos baseados em suavização exponencial.

A técnica de suavização exponencial utiliza uma média ponderada dos valores passados para prever valores futuros. De acordo com Vishwas e Patel, a ideia principal dessa técnica é atribuir maior peso a valores recentes de uma série temporal, sendo muito utilizada para modelos de previsão de curto prazo e para series temporais que se movem lentamente ao longo do tempo.

Os métodos de suavização exponencial descritos pela literatura são a suavização exponencial simples (SES), utilizada quando não existem padrões claros de tendência e sazonalidade na série; a suavização exponencial dupla (ou suavização exponencial de Holt), que agrega um componente para suavização de séries que possuem tendência, e; a suavização exponencial tripla (ou suavização exponencial de Holt-Winters), que agrega componentes para suavização de séries com tendência e sazonalidade.

2.3.2 Modelo Autorregressivo (AR), Modelo de Média Móvel (MA) e Modelos Autorregressivos Integrados de Média Móvel (ARIMA).

Modelos AR, MA e ARIMA são muito utilizados para realização de previsões em séries temporais univariadas, e levam em conta o comportamento passado da série temporal para previsão de valores futuros.

Um modelo AR realiza uma regressão em valores passados para prever valores futuros, e pressupõe que série temporal seja estacionária. Sua modelagem depende da identificação da ordem do modelo AR (p), em que “ p ” define as unidades de tempo anteriores a serem incorporadas ao modelo.

Um modelo MA baseia-se “*em um processo em que o valor em cada ponto no tempo é uma função dos termos de ‘erro’ do valor do passado recente, cada qual independente dos outros*” (NIELSEN, 2021). O modelo MA também pressupõe uma série estacionária, e sua ordem MA (q) é definida por um valor “ q ”, que representa o total de defasagens a serem consideradas no modelo.

Por fim, o modelo ARIMA permite a combinação dos modelos AR e MA, reconhecendo que a mesma série temporal pode ter a dinâmica dos dois modelos citados (o que leva à identificação de um modelo ARMA). A inclusão de um termo integrado (d) na modelagem ARMA, que representa a quantidade de vezes que a série temporal deve ser diferenciada para gerar

estacionariedade, resulta em um modelo ARIMA (p, d, q). Conforme destacado por Nielsen (2021), “*Os modelos ARIMA continuam a proporcionar um desempenho próximo a tudo que há de mais moderno, sobretudo quando se trata de conjuntos pequenos de dados, em que o aprendizado de máquina mais sofisticado ou modelos de aprendizado profundo não são os melhores*”.

Para identificação dos valores de “p”, “d” e “q” de um modelo ARIMA, deve-se avaliar se a série é ou não estacionária, o que definirá o valor de “d”. Por sua vez, “p” e “q” podem ser obtidos, respectivamente, em consulta aos resultados das funções de autocorrelação parcial (FACP) e de autocorrelação (FAC) da série temporal. Existem, ainda, bibliotecas Python que automatizam a seleção de parâmetros para um modelo ARIMA, utilizando critérios específicos para otimização.

2.3.3 Floresta Aleatória

O modelo de Floresta Aleatória, baseado em aprendizado de máquina, possui origem em métodos de árvore de decisão, em que o processo de previsão é realizado em “*um passo de cada vez e de um jeito não linear*” (NIELSEN, 2021). O modelo de floresta aleatória utiliza a média dos resultados de várias árvores de decisão paralelas para efetuar tarefas de regressão, apresentando como vantagem a redução do risco de sobreajuste do modelo.

Destaque-se, ainda, que o modelo de Floresta Aleatória pressupõe que características sobre os dados estejam disponíveis para realização de previsões sobre um valor alvo, o que surge com um dificultador em séries temporais univariadas – nesses casos, faz-se necessária a geração de características relacionados aos valores da série para utilização do modelo. Uma alternativa citada pela literatura é criação de uma tabela com o valor original da série e com outras colunas contendo sequências defasadas da mesma série (conforme a quantidade de defasagens estabelecidas). Há ainda a possibilidade de utilização de estatísticas da série, como médias móveis, bem como a decomposição de informações temporais em dias, meses, anos, dentre outras possibilidades.

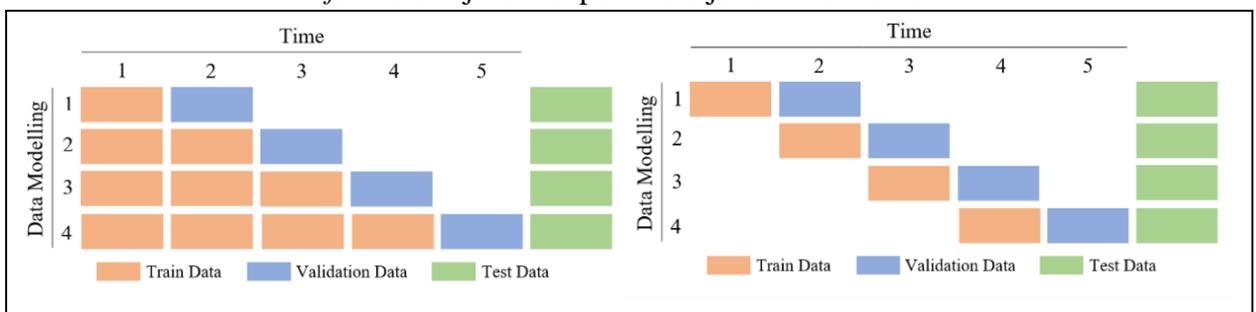
2.4 Validação e teste

Uma boa forma de testar um modelo e de comparar sua performance preditiva contra outros modelos é realizando uma validação *out-of-sample*, o que significa separar os dados em dois grupos: um grupo de treino, utilizado para ajuste do modelo, e um grupo de teste, que será

utilizado para mensurar os erros de previsão. Em séries temporais, em virtude da existência de dependência entre as ocorrências em diferentes períodos, é importante que os dados mais antigos sejam utilizados para treinamento, enquanto os dados mais recentes comporão o grupo de teste.

Além disso, para obter mais segurança de que o modelo gerado possa ser generalizado para outros períodos da série, permitindo uma visão mais realista sobre sua performance, é amplamente recomendada a realização de validações ao longo da série. Considerando as características das séries temporais, recomenda-se a utilização da validação *walk forward*, que é baseada na separação da série em grupos de validação sem alteração da ordem cronológica dos dados. Os grupos de validação podem ser gerados a partir de janelas expansivas, em que os dados de treinamento vão sendo acumulados a cada nova rodada de validação, ou a partir de janelas deslizantes, quando os grupos de dados possuem tamanhos iguais, mas diferem no ponto de início de cada período de treinamento/validação. A figura 1 exemplifica os métodos de validação descritos: à esquerda, uma ilustração da utilização de janelas expansivas; à direita, uma representação de grupos de validação gerados por janelas deslizantes.

Figura 1. Exemplo de técnicas de seleção de grupos para realização da validação *walk forward* – janela expansiva e janela deslizante.



Fonte: Extraído de <https://audhiaprilliant.medium.com/walk-forward-optimization-cross-validation-technique-for-time-seraubaies-data-61739f58f2c0>.

A partir dos resultados obtidos junto aos dados de validação e de teste, é possível definir uma medida para aferir a qualidade das previsões realizadas pelo modelo. As métricas de erro resumizam as diferenças entre os valores previstos e os valores reais (Auffarth, 2021), sendo úteis para uma análise comparativa entre modelos. Ainda de acordo com Auffarth, as métricas de erro mais populares são o Erro Quadrático Médio, ou MSE (sigla para *Mean Squared Error*), o Erro Absoluto Médio, ou MAE (sigla para *Mean Absolute Error*) e a Raiz Quadrada do Erro Quadrático Médio, ou RMSE (sigla para *Root Mean Squared Error*).

3 Metodologia

3.1 Extração e tratamento dos dados

Para atingimento dos objetivos propostos no trabalho, foram utilizados dados agregados sobre requerimentos apresentados aos INSS, presentes na base de dados que dá suporte ao GET. Foram extraídas todas as tarefas criadas entre julho/2018 e julho/2022, segregadas por mês, por tipo de requerimento e por localidade de solicitação. A partir da referida extração, foram selecionados apenas as tarefas a serem tratadas pelas CEAB, de acordo com os códigos presentes no Anexo I da Portaria PRES/INSS nº 1.286, de 05.04.2021¹⁰ (última alteração pela Portaria PRES/INSS nº 1.373, de 29 de outubro de 2021).

Ainda, a demanda mensal, inicialmente representada como tarefas criadas, foi convertida em pontos criados com base nas pontuações¹¹ estabelecidas no Anexo I da mesma Portaria.

3.2 Análise exploratória dos dados

Realizados os referidos ajustes, procedeu-se à aplicação de técnicas de estatística descritiva e de visualização de dados como um primeiro passo para identificação de características da série sob estudo.

Posteriormente, após a realização de ajustes e agrupamentos nos dados, foram aplicados métodos exploratórios específicos para séries temporais, como a aplicação de testes de hipótese para verificação da estacionariedade da série temporal (testes ADF e KPSS) e a plotagem das funções de autocorrelação e de autocorrelação parcial.

3.3 Seleção de modelos

A partir das características da série avaliada, foram elencados modelos para treinamento e avaliação quando à sua capacidade preditiva.

De forma a realizar projeções para um período de 12 meses, a série foi subdividida da seguinte forma: 77% dos dados (37 meses iniciais) foram separados para treinamento dos modelos,

¹⁰ Foram considerados, como de responsabilidade das CEAB, os códigos presentes nas tabelas 01, 02, 03, 04 e 05 do Anexo I da Portaria.

¹¹ A pontuação das tarefas é definida pelo INSS de acordo com a complexidade de cada atividade, e é utilizada para mensurar a produtividade dos servidores.

sendo os outros 23% (12 meses finais) reservados para realização de testes sobre sua capacidade de previsão. Ademais, de forma a melhor estimar a qualidade dos modelos, aplicou-se o método de validação *walk forward* junto aos dados de treinamento (foram utilizadas janelas expansivas para seleção dos grupos).

O desempenho dos modelos foi avaliado e comparado com base nas seguintes métricas:

- a) Raiz Quadrada do Erro Quadrático Médio, ou RMSE (sigla para *Root Mean Squared Error*): a RMSE é mensurada a partir da média do quadrado dos erros de previsão de cada ponto, o que impede que erros com sinais distintos se cancelem. Calcula-se, então, a raiz quadrada do valor encontrado, obtendo-se uma métrica na mesma escala dos valores das previsões, o que torna sua interpretação mais intuitiva. Quanto menor o valor da RMSE, melhor o desempenho do modelo;
- b) Erro Percentual Absoluto Médio, ou MAPE (sigla para *Mean Absolute Percentage Error*): é uma medida estatística sobre a acurácia do modelo, medida em termos percentuais, calculada a partir da razão dos erros de previsão em relação aos valores reais. Quanto mais próximo de zero, melhor o desempenho do modelo.

3.4 Ferramentas

Para organização e análise dos dados, utilizou-se linguagem de programação Python, via *Jupyter Notebook*¹². As bibliotecas Python utilizadas foram: NumPy, pandas, scikit-learn, Matplotlib, seaborn, sktime, SciPy e statsmodels.

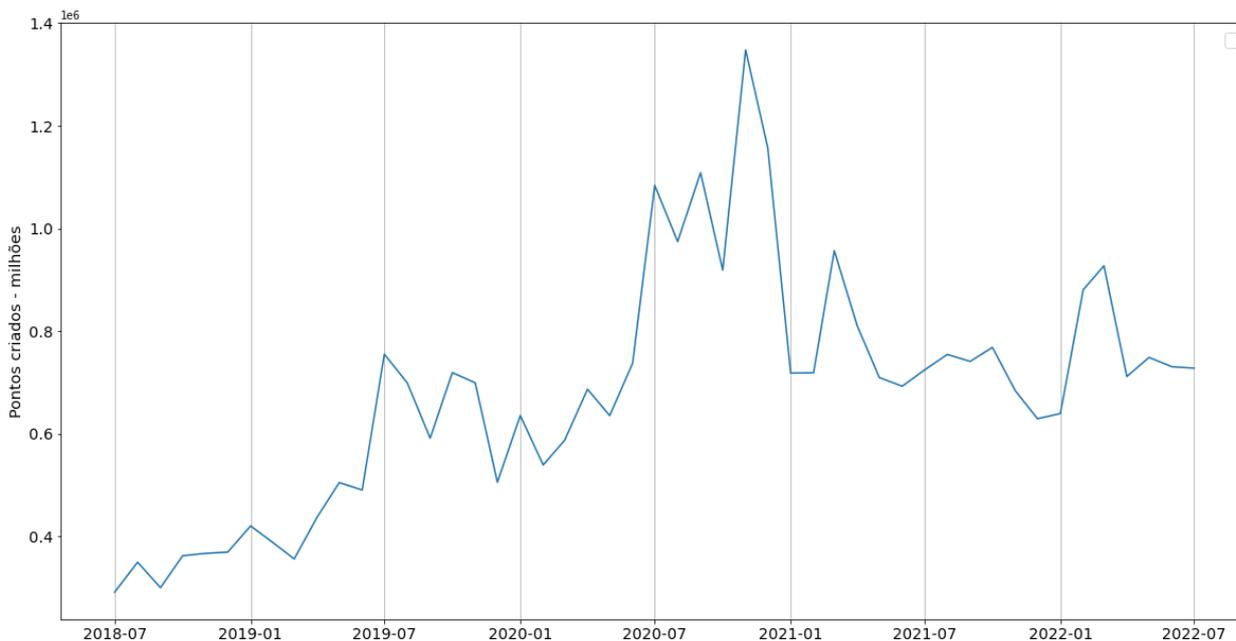
4 Resultados e Discussão

Os dados sobre demandas mensais no âmbito das CEAB¹³ foram agrupados de forma a identificar o total de tarefas criadas mensalmente no INSS. A figura 2 ilustra o comportamento da série de dados no período avaliado.

¹² Jupyter Notebook é uma aplicação web para criação e compartilhamento de documentos com códigos de programação e textos explicativos.

¹³ CEAB/RD, CEAB/DJ, CEAB/RDAI, CEAB/MAN.

Figura 2. Série histórica de dados sobre demandas recebidas pelo INSS no âmbito das CEAB



Fonte: Elaborado pelo Autor a partir de dados obtidos junto ao INSS.

A representação gráfica da série temporal apresenta informações relevantes a respeito do conjunto de dados. Inicialmente, verifica-se que a porção inicial da série possui valores mais baixos se comparadas com outros pontos do gráfico (ou seja, encontra-se em um nível mais baixo). Essa situação, além de sugerir uma possível tendência na série, reflete possíveis dificuldades inerentes ao início da implementação da nova sistemática de gestão de demandas/tarefas, inaugurada pelo projeto de Transformação Digital do INSS. Outra possibilidade é que a implementação da nova sistemática tenha sido realizada de forma gradual, o que pode explicar, por exemplo, a mudança de nível observada entre julho de 2018 e julho de 2019.

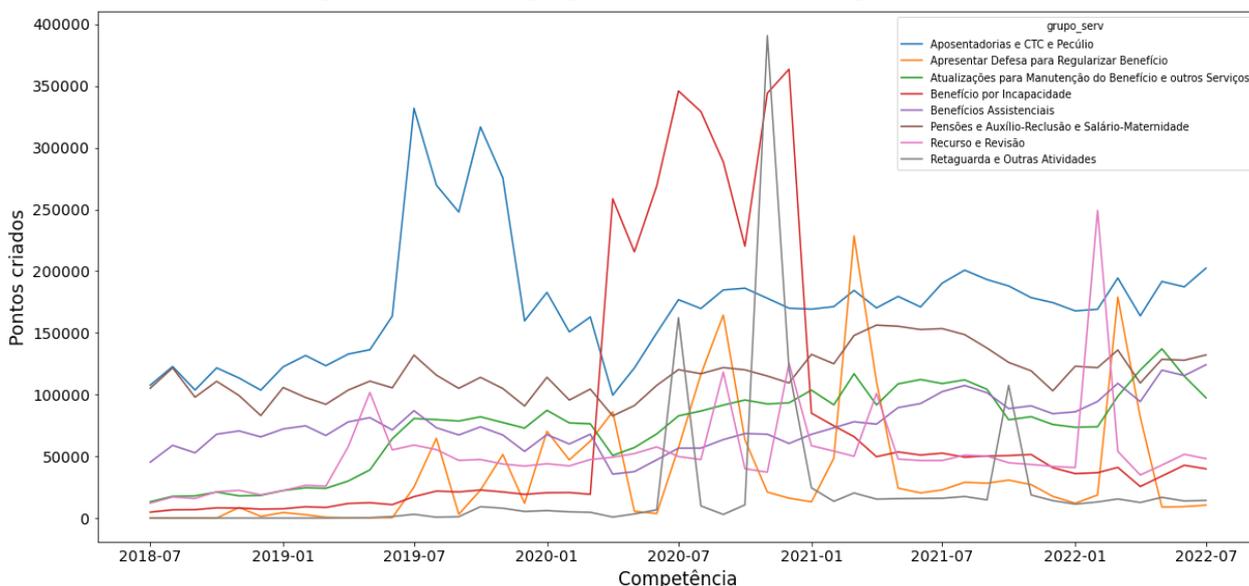
Chama atenção, ainda, a existência de um período de forte alta iniciado no ano de 2020 e que se estende até dezembro do referido exercício. Trata-se de período marcado pelo início e posterior agravamento da Emergência de Saúde Pública de Importância Internacional (ESPII) ocasionada pelo coronavírus SARS-CoV-2 (COVID-19)¹⁴, que trouxe diversas consequências de ordem sanitária, social e econômica ao Brasil. É provável, portanto, que o período de alta seja

¹⁴ Mais informações em <https://www.paho.org/pt/covid19>.

reflexo do momento de crise enfrentado pela sociedade, quando mais cidadãos solicitaram os serviços da Previdência Social.

Posto isso, e de forma a melhor compreender os motivos que levaram ao forte aumento na demanda durante o primeiro ano da emergência sanitária (COVID-19), realizou-se análise segregada da série histórica de cada um dos grupos de serviço que compõem a demanda mensal no âmbito das CEAB. A figura 3 apresenta um panorama para os 8 (oito) grupos de serviço mais representativos no período avaliado.

Figura 3. Série histórica de dados sobre demandas recebidas pelo INSS no âmbito das CEAB para os 8 (oito) grupos de serviço mais representativos.



Fonte: Elaborado pelo Autor a partir de dados obtidos junto ao INSS.

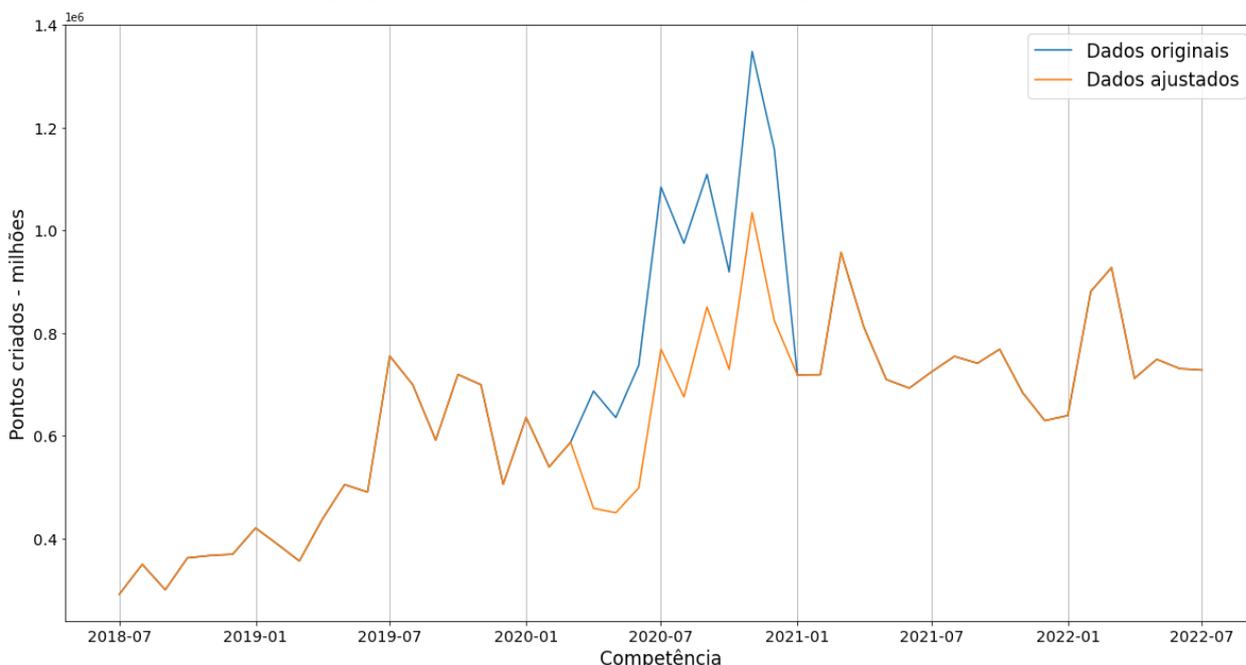
Adotando como enfoque o comportamento das séries no ano de 2020, chama a atenção o comportamento dos serviços do tipo “Benefício por Incapacidade” e “Retaguarda e Outras Atividades”, que apresentaram picos proeminentes. O grupo de serviço “Apresentar Defesa para Regularizar Benefício” também apresenta picos em 2020, mas que não são tão discrepantes em relação aos demais picos observados ao longo da série. Os demais grupos de serviço avaliados não demonstram comportamento atípico no ano de 2020, à exceção de uma queda mais acentuada em abril de 2020. Por fim, alguns grupos de serviço apresentam fortes picos em 2022 (“Recurso e Revisão” e “Apresentar Defesa para Regularizar Benefício”).

Aprofundadas as análises, apenas para grupo de serviço “Benefício por Incapacidade” identificou-se comportamento claramente atípico no ano de 2020, não havendo indicação de que

o comportamento se repetirá no futuro próximo. O fenômeno pode ser explicado pela publicação da Lei nº 13.982, de 02.04.2020, que flexibilizou os procedimentos de concessão do auxílio-doença. Assim, o auxílio-doença, que em condições normais não seria analisado por servidores do INSS, mas sim por Peritos Médicos Federais, passou a ser concedido de forma administrativa no âmbito do Instituto, sem a necessidade de realização de perícia médica. Por esse motivo, os dados referentes ao grupo de serviço “Benefício por Incapacidade” foram tratados, por meio da substituição de valores: os dados registrados entre abril de 2020 e dezembro de 2020 foram substituídos pela média dos demais valores da série (anteriores e posteriores) – com isso, a média do grupo de dados caiu de 78.594 pontos criados por mês para 30.180 pontos criados por mês.

Quanto aos demais grupos de serviço, eventuais comportamentos irregulares não puderam ser diretamente atribuídos à crise sanitária de 2020, de forma que não foram tratados. A figura 4 ilustra os resultados da alteração realizada no conjunto de dados, que culminou na redução do nível da série para o ano de 2020.

Figura 4. Comparativo da série histórica antes e após o tratamento de dados referentes ao grupo de serviço “Benefício por Incapacidade”.

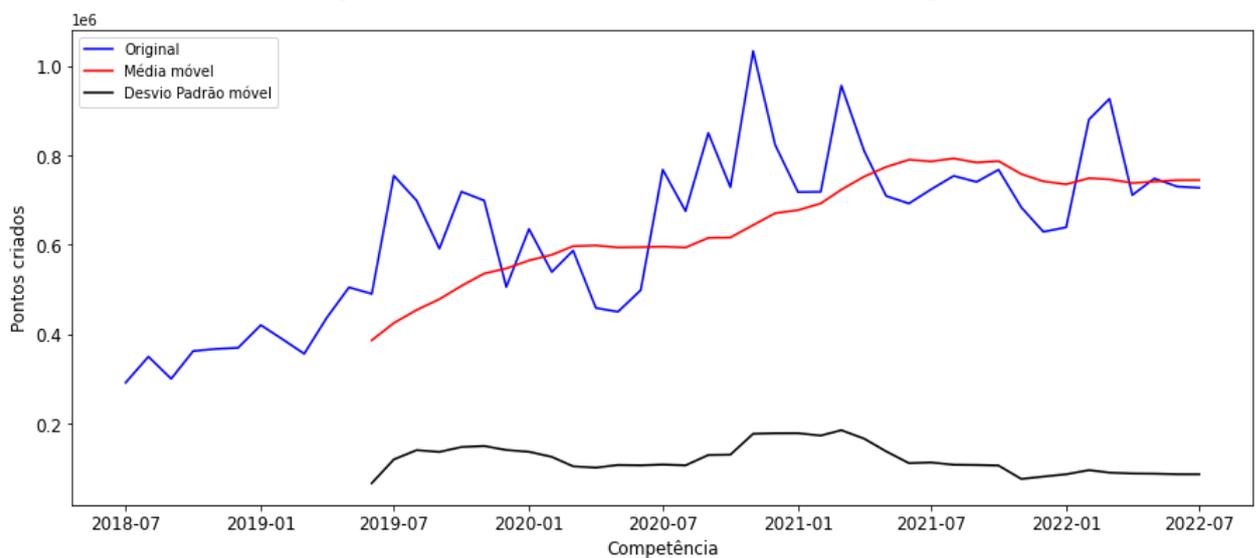


Fonte: Elaborado pelo Autor a partir de dados obtidos junto ao INSS.

A série tratada foi, então, avaliada quanto à estacionariedade, à sazonalidade e quanto à existência de autocorrelação entre os dados.

Realizou-se, inicialmente, inspeção visual sobre as médias móveis dos dados e do desvio padrão dos dados, considerando uma janela de 12 meses. Conforme figura 5, apesar do desvio padrão apresentar comportamento relativamente estável ao redor de uma média constante, é possível notar crescimento na média dos valores mensais da série, o que sugere que a série não é estacionária. Uma análise mais detida, entretanto, demonstra certa estabilidade da média móvel nos últimos meses da série temporal, o que pode indicar que as variações na média sejam reflexo tanto do período inicial de implementação do processo de Transformação Digital quanto do momento atípico ocasionado pela emergência sanitária (COVID-19).

Figura 5. Médias móveis dos dados e do desvio padrão.



Fonte: Elaborado pelo Autor a partir de dados obtidos junto ao INSS.

A estacionariedade foi avaliada, ainda, por meio dos testes de hipótese ADF e KPSS, cujos resultados estão registrados na Tabela 1.

Tabela 1. Resultados dos testes de hipótese para verificação da estacionariedade.

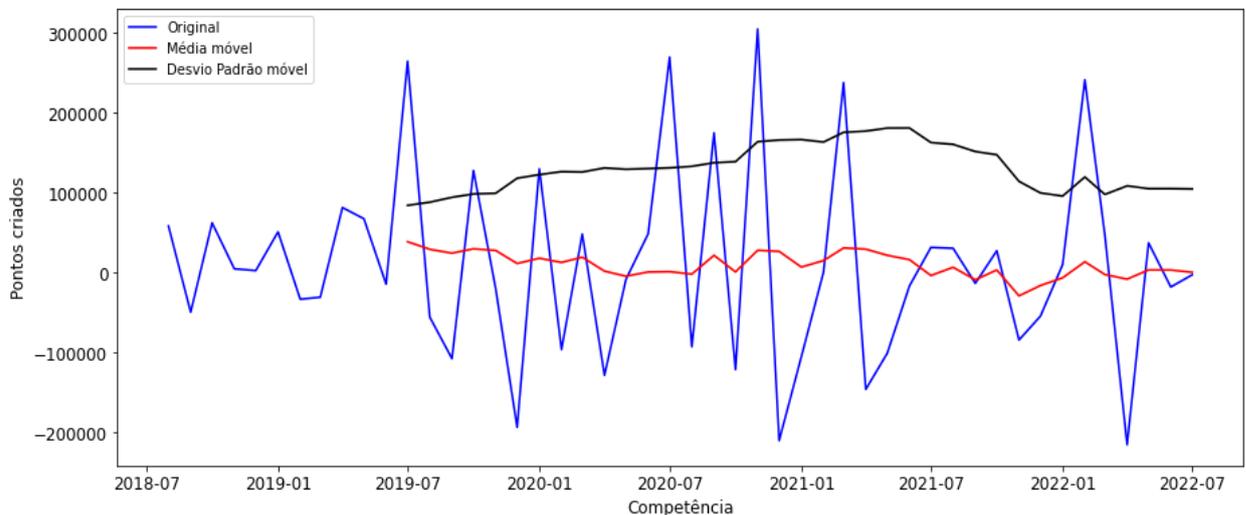
Teste	Resultados			
	Estatística do teste	p-valor	Valor crítico (5%)	Valor crítico (1%)
ADF (H ₀ : série não estacionária)	-2,805	0,057	-2,923	-3,574
KPSS (H ₀ : série estacionária)	0,798	0,010	0,463	0,739

Fonte: Elaborado pelo Autor a partir de dados obtidos junto ao INSS.

Embora o teste ADF não permita uma conclusão categórica sobre o comportamento dos dados, o teste KPSS sugere que a série não é estacionária com um nível de confiança de 99%. Assim, assumiu-se a série como não estacionária.

Tomou-se, então, a primeira diferença¹⁵ da série temporal para que os dados adquirissem padrão estacionário, o que foi corroborado pelos resultados dos testes ADF (p-valor tendendo a zero) e KPSS (p-valor = 0,1).

Figura 7. Médias móveis dos dados e do desvio padrão, após realização da primeira diferença.

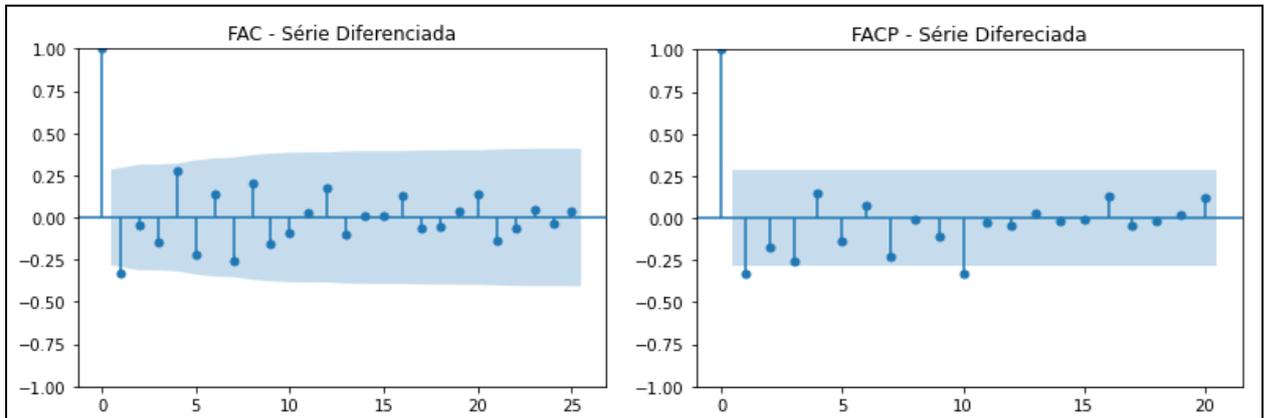


Fonte: Elaborado pelo Autor a partir de dados obtidos junto ao INSS.

Realizada a diferenciação da série, foram plotados os resultados das funções de autocorrelação (FAC) e de autocorrelação parcial (FACP) para avaliação quanto à existência de ciclos sazonais e de autocorrelação nos dados (Figura 8).

¹⁵ Consiste em tomar a diferença entre duas observações consecutivas. Uma nova série é gerada a partir do valor das diferenças.

Figura 8. Resultados da FAC e da FACP da série temporal diferenciada.



Fonte: Elaborado pelo autor a partir de dados obtidos junto ao INSS.

Quanto à sazonalidade, não foi possível identificar nenhum padrão relevante, uma vez que não há correlações periódicas relevantes após a primeira defasagem (*lag* 1). Quanto à autocorrelação, a FAC apresenta apenas uma correlação relevante no *lag* 1 (mesmo assim, em valor muito próximo do limite de significância). Já a FACP apresenta correlação significativa nos *lags* 1 e 10 (também em valores muito próximos ao limite de corte). Em nenhum dos casos há uma queda acentuada nas correlações após o *lag* 1.

Realizadas as análises exploratórias, foram identificadas as seguintes informações acerca da série de dados:

- É provável que tanto os meses iniciais da série temporal (período de implementação da nova metodologia de gestão de tarefas) quanto os meses referentes ao ano de 2020 (ano de maior impacto da emergência sanitária causada pelo COVID-19) tenham apresentado comportamento divergente daquele que seria originalmente esperado, uma vez que existiam fatores externos relevantes influenciando os dados nos referidos períodos;
- A série de dados possui quantidade limitada de registros (49 pontos no total) e grande flutuação nos dados, como mudanças de nível e picos em determinados meses, o que torna menos indicada a aplicação de modelos de previsão mais complexos;
- A série torna-se estacionária com a realização de diferenciação nos dados (primeira diferença);
- Não há evidência de sazonalidade na série;
- Há baixa correlação entre os pontos da série temporal, havendo uma correlação fraca na primeira defasagem tanto da FAC quando da FACP.

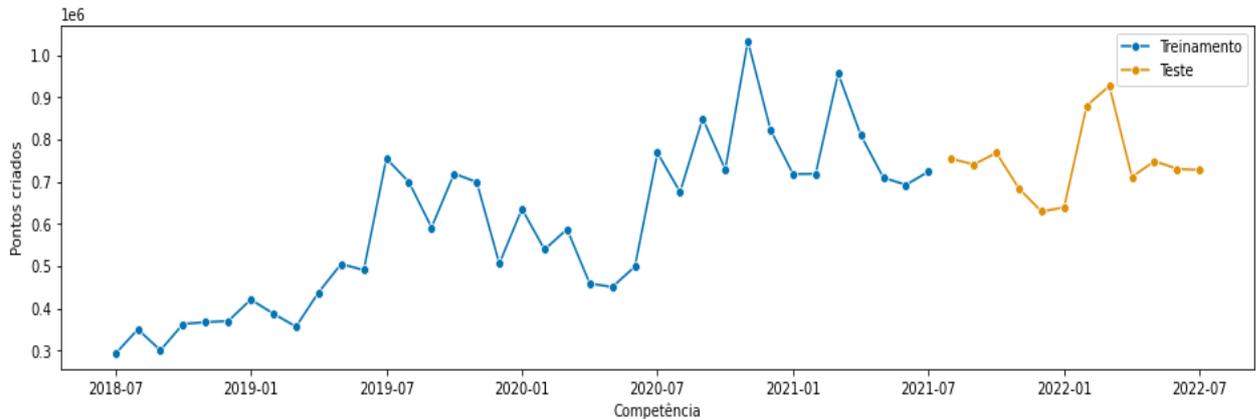
Diante do exposto, optou-se por priorizar a utilização de métodos menos complexos na etapa de seleção e avaliação de modelos para previsão da série temporal. Inicialmente, de forma a estabelecer uma linha de base para comparação, utilizou-se o valor mais recente da série como previsão para os meses seguintes (doravante denominado modelo *Naive*). Feito isso, procedeu-se ao treinamento e avaliação dos seguintes modelos:

- a) Suavização Exponencial Simples (SES): dentre os modelos de suavização exponencial, optou-se pela suavização exponencial simples devido à ausência de padrões sazonais nos dados e, ainda, pela aparente estabilização da tendência nos últimos meses da série temporal, assumindo-se, assim, que série não teria uma tendência futura bem definida. O modelo foi implementado por meio da função *ExponentialSmoothing* da biblioteca “sktime”;
- b) ARIMA: considerando a necessidade de diferenciação para que a série adquira um comportamento estacionário, realizou-se o ajuste de modelos ARIMA levando em conta os resultados da FAC e da FACP. Assim, foram avaliados os modelos ARIMA (1,1,0), ARIMA (0,1,1), ARIMA (1,1,1), além da utilização de uma função automatizada para especificação de modelos ARIMA. Foram utilizadas as funções *ARIMA* e *AutoARIMA* da biblioteca “sktime”;
- c) Floresta Aleatória: aplicou-se um modelo de floresta aleatória para avaliar se haveria ganho relevante em qualidade com a utilização de técnicas de aprendizado de máquina. Utilizou-se a função *RandomForestRegressor*, da biblioteca “scikit-learn”, em conjunto com a função *make_reduction*¹⁶ da biblioteca “sktime”;

Conforme já exposto, a série de dados foi dividida em grupos de treino e teste (Figura 9). O grupo de treinamento foi utilizado para ajuste e validação dos modelos, enquanto o grupo de teste teve por intuito avaliar o desempenho dos modelos para realização de previsões futuras.

¹⁶ Utilizada para criação de uma versão tabular da série temporal com sequências defasadas da mesma série (foram utilizadas duas defasagens) e para adoção de estratégia recursiva na previsão de dados de 12 meses a frente.

Figura 9. Identificação dos grupos de dados utilizados para treinamento e para teste.



Fonte: Elaborado pelo autor a partir de dados obtidos junto ao INSS.

Utilizou-se o método de validação *walk forward* com janela expansiva para avaliação da capacidade de generalização dos modelos: a partir do 18º mês da série, os modelos realizavam previsões para os seis (6) meses subsequentes, avançando um período no tempo a cada rodada de validação (tendo em vista a utilização de janelas expansivas, um número maior de dados era utilizado para treinamento a cada rodada de validação). Por fim, como critério de comparação entre os diferentes modelos, adotou-se a Raiz Quadrada do Erro Quadrático Médio (RMSE). A Tabela 2 resume os resultados do processo de validação dos modelos.

Tabela 2. Resultados da validação *walk forward* sobre os dados de treinamento.

Último mês de treinamento (cutoff)	Desempenho dos modelos na validação (RMSE)						
	Naive	SES ¹⁷	AutoARIMA ¹⁸	ARIMA (0,1,1)	ARIMA (1,1,1)	ARIMA (1,1,0)	Floresta Aleatória ¹⁹
2019-12	70.708,10	76.729,28	70.708,10	129.459,25	83.633,22	102.696,50	155.386,31
2020-01	137.599,75	125.031,87	137.599,75	208.365,12	170.520,20	163.390,32	152.149,13
2020-02	122.106,66	117.464,17	126.387,53	134.534,04	126.364,68	111.103,51	119.993,89
2020-03	159.695,01	161.289,95	159.695,01	134.570,17	134.143,68	135.727,46	157.459,75
2020-04	248.640,90	213.339,83	248.640,90	156.345,29	167.198,59	194.235,96	191.071,01
2020-05	349.567,20	333.898,60	349.567,20	295.680,64	305.572,61	323.433,39	278.020,94
2020-06	334.963,97	345.020,75	334.963,97	297.760,77	302.670,68	314.286,86	304.321,79
2020-07	124.516,78	182.392,94	124.516,77	120.617,80	120.457,98	117.455,52	220.965,37
2020-08	176.799,61	179.237,53	176.799,61	139.566,73	139.195,36	137.589,12	230.109,46
2020-09	125.984,58	128.404,94	125.984,58	160.946,71	147.834,22	138.688,06	253.299,27
2020-10	163.616,69	149.670,38	144.781,73	140.272,29	131.646,99	127.410,50	270.846,01
2020-11	259.645,95	172.151,19	265.314,29	324.331,64	269.517,64	265.314,29	160.987,77
2020-12	108.131,21	132.385,14	235.359,19	272.811,30	240.958,03	235.359,19	109.568,16
2021-01	104.983,78	92.281,40	96.572,67	182.765,03	148.015,18	114.563,10	160.401,16
RMSE Médio	177.640,01	172.092,71	185.492,24	192.716,20	177.694,93	177.232,41	197.470,00

Fonte: Elaborado pelo autor a partir de dados obtidos junto ao INSS.

Destaque-se, inicialmente, que as métricas de erro indicam variação pouco relevante na qualidade dos modelos. De modo geral, todos apresentaram desempenho pouco satisfatório no processo de validação. Não obstante o exposto, o modelo de SES obteve a melhor média geral nos valores da métrica de erro, seguido pelo modelo linha de base (Naive) e pelos modelos ARIMA. É interessante destacar, ainda, que: (i) todos os modelos apresentaram resultados ruins para previsões dentro do ano de 2020, período impactado pela emergência sanitária (COVID-19), e; (ii) conforme os dados se afastam do ano de 2020, os erros de previsão do modelo de SES apresentam-se, de modo geral, menores do que os erros dos demais modelos.

¹⁷ Utilizou-se a parametrização padrão da função *ExponentialSmoothing* da biblioteca “sktime”.

¹⁸ Utilizou-se a parametrização padrão da função *AutoARIMA* da biblioteca “sktime”.

¹⁹ Utilizou-se a parametrização padrão tanto para a função *RandomForestRegressor*, da biblioteca “scikit-learn”, quanto para a função *make_reduction*, da biblioteca “sktime”.

Em virtude dos resultados obtidos durante a etapa de treinamento e validação, todos²⁰ os modelos foram submetidos à etapa de teste, momento em foram comparadas previsões dos modelos com dados reais do grupo de teste (12 meses). Os resultados foram avaliados a partir de duas métricas de erro: o RMSE e o MAPE. A Tabela 3 apresenta os resultados das métricas para cada modelo testado, enquanto a Figura 10 ilustra o comportamento real da série em comparação aos valores previstos pelos modelos.

Tabela 3. Resultados das métricas de erro para as previsões realizadas pelos modelos.

Métrica	Modelos			
	Naive	SES ²¹	AutoARIMA ²²	Floresta Aleatória ²³
RMSE	85.359,04	85.650,43	86.294,24	103.946,01
MAPE	7,70%	7,73%	7,82%	8,96%

Fonte: Elaborado pelo autor a partir de dados obtidos junto ao INSS.

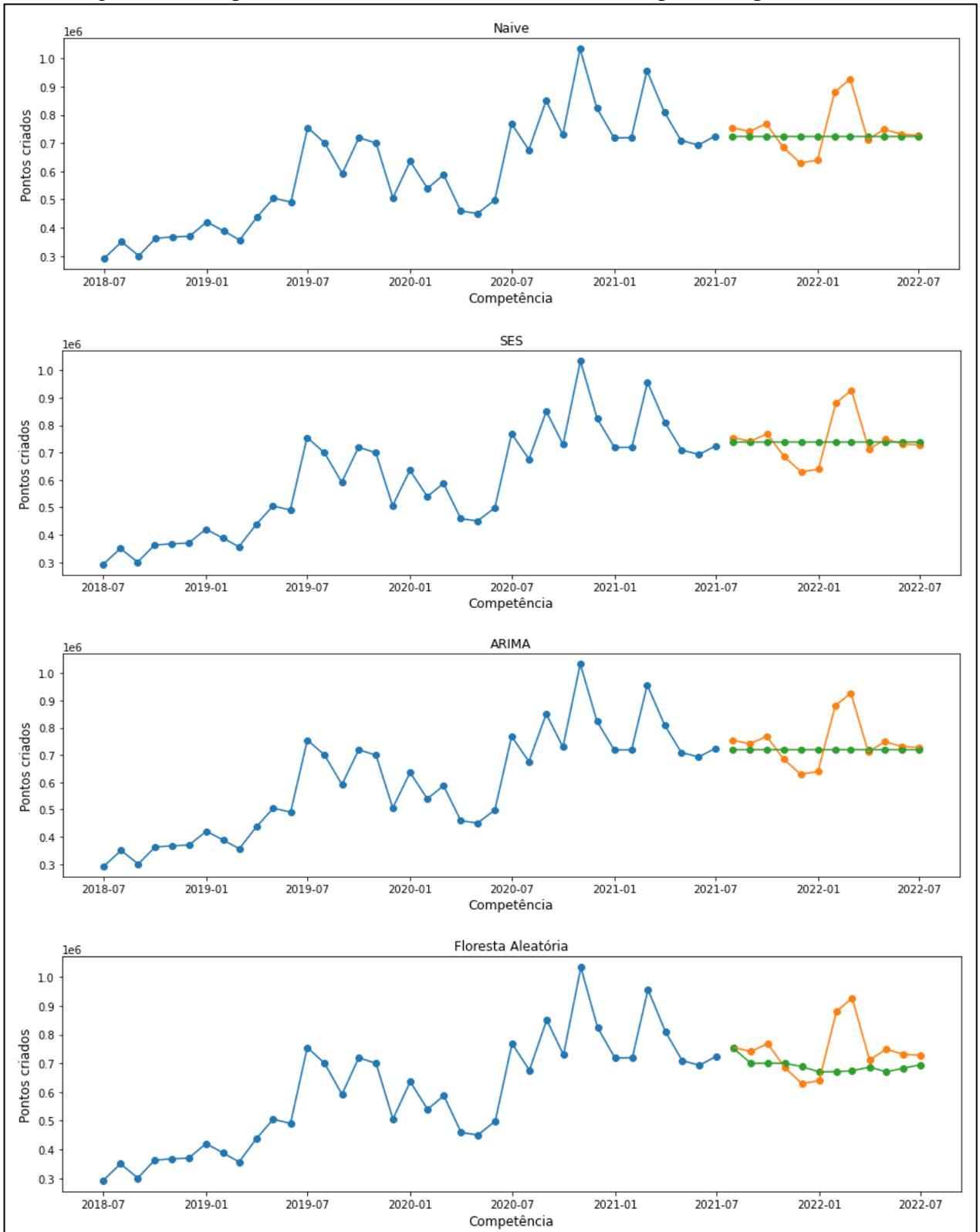
²⁰ Em relação aos modelos ARIMA, considerando que não houve diferença relevante no desempenho dos modelos treinados e validados, optou-se por utilizar apenas a função AutoARIMA da biblioteca “sktime”, que realiza busca automatizada para otimização de parâmetros de um modelo ARIMA, considerando as características dos dados avaliados. Assim, os modelos ARIMA parametrizados manualmente não foram incluídos na etapa de teste.

²¹ Utilizou-se a parametrização padrão da função *ExponentialSmoothing* da biblioteca “sktime”.

²² Utilizou-se a parametrização padrão da função *AutoARIMA* da biblioteca “sktime”.

²³ Utilizou-se a parametrização padrão da função *RandomForestRegressor*, da biblioteca “scikit-learn”, e da função *make_reduction*, da biblioteca “sktime”.

Figura 10. Comparativo entre os valores reais e os valores previstos pelos modelos.



Fonte: Elaborado pelo autor a partir de dados obtidos junto ao INSS.

Tanto os valores calculados para as métricas de erro quanto a plotagem comparativa demonstram que os modelos avaliados apresentaram desempenho muito semelhante em produção, exceção feita ao modelo de Floresta Aleatória, que apresentou erros um pouco maiores, o que sugere um desempenho inferior. Destaque-se que os valores de MAPE próximos a 8% (o que, grosso modo, representa o percentual médio de desvio dos valores previstos em relação aos valores reais), indicam que os modelos foram capazes de fazer boas previsões para o período de 12 meses.

Chama atenção, ainda, o fato de que o desempenho dos modelos foi superior na etapa de testes em comparação à etapa de validação. Essa situação é reflexo das grandes variações observadas nos dados de treino, causadas por fatores extraordinários relacionados aos dados de 2020, principalmente. Trata-se de situação que sugere cautela na interpretação dos resultados dos modelos, havendo a possibilidade de que o comportamento mais atual da série temporal esteja adquirindo características distintas daquelas observadas nos dados históricos.

Feitas tais ponderações, a decisão pela priorização e utilização de um dos modelos depende, basicamente, da existência de justificativa que demonstre vantagem em relação ao modelo linha de base (Naive), cuja utilização não exige esforço computacional ou de análise. Assim, levando em conta apenas o desempenho observado nos testes, seria intuitivo optar pelo modelo Naive. Frise-se, entretanto, que uma análise mais pormenorizada indica riscos na utilização do modelo Naive para previsões de 12 meses, uma vez que movimentos bruscos na série (para mais ou para menos) poderiam distorcer suas previsões, que são baseadas exclusivamente no último registro disponível na série temporal. Assim, entende-se que seria mais prudente a utilização do modelo de SES, pois: (i) apesar de simples, é um método que suaviza a ocorrência de outliers na série, mitigando riscos de impactos relevantes nas previsões; (ii) a SES apresentou desempenho um pouco melhor na etapa de treinamento e validação dos modelos, o que pode indicar uma melhor capacidade de generalização, e; (iii) o modelo de SES permite a customização do parâmetro de suavização (α), que define o peso a ser dado às observações mais distantes.

Assim, verifica-se que a utilização de técnicas de análise de séries temporais permitiu a identificação de modelos que podem auxiliar o INSS a prever e a gerir suas demandas futuras, havendo indicativos de que o modelo de SES seria o mais adequado dentre os modelos avaliados e que apresentaria ganhos não desprezíveis em relação a técnicas simplificadas de previsão. Registre-se, entretanto, que em virtude da quantidade limitada de observações na série temporal analisada, bem como da aparente influência de fatores externos relevantes sobre os dados do

período avaliado, seria oportuna a realização de novas análises após o acúmulo de mais dados, o que permitiria avaliar o comportamento da série em contextos mais estáveis.

5 Conclusão

O trabalho realizado abordou a gestão de demandas no âmbito do INSS, tema recorrente no debate público, especialmente devido à enorme dificuldade da Autarquia em atender de forma tempestiva os cidadãos que procuram seus serviços.

Diante do exposto, e de forma a contribuir com a discussão, realizou-se estudo sobre a demanda de requerimentos apresentados mensalmente ao INSS no âmbito das CEAB, utilizando-se de técnicas de análise de séries temporais com o objetivo de identificar um modelo que pudesse auxiliar o INSS a prever e, conseqüentemente, a gerir suas demandas.

A partir dos resultados obtidos, destaque-se, inicialmente, haver risco de que a série temporal analisada tenha apresentado comportamento divergente daquele que seria originalmente esperado, especialmente em virtude da ocorrência de fatos atípicos que impactaram a procura pelos serviços do INSS entre os anos de 2020 e 2021 (COVID-19). Não obstante o exposto, os modelos selecionados apresentaram desempenho satisfatório nas validações e testes realizados, haja vista os resultados mensurados pelas métricas de desempenho utilizadas. Ademais, em que pese os modelos avaliados terem apresentado desempenho semelhante nos testes, foram identificadas vantagens na utilização do modelo de SES, haja vista ser uma solução de baixa complexidade e que mitigaria riscos inerentes à utilização do modelo linha de base, que realiza previsões utilizando o último valor da série, exclusivamente.

Conclui-se, assim, que a utilização de técnicas de análise de séries temporais produziu resultados satisfatórios no que diz respeito à previsão de demandas futuras no âmbito do INSS, havendo argumentos para priorização da aplicação do modelo de Suavização Exponencial Simples em detrimento dos demais modelos e de técnicas simplificadas de previsão.

Por fim, considerando a quantidade limitada de períodos na série temporal analisada, bem como considerando a influência de fatores externos relevantes sobre os dados do período avaliado, seria desejável que novas análises fossem realizadas após o incremento de novas observações na série histórica, o que poderia resultar na identificação de modelos mais ajustados aos dados e, conseqüentemente, mais precisos e com maior capacidade de generalização.

6 Referências bibliográficas

AUFFARTH, Ben. **Machine Learning for Time-Series with Python**. Birmingham: Packt Publishing, 2021.

BRASIL. Ministério do Desenvolvimento Social. Instituto Nacional do Seguro Social. **INSS Digital: Uma nova forma de atender**. Brasília, 2019.

NAU, Robert. **Statistical forecasting: notes on regression and time series analysis**. Duke University, 2020. Disponível em: <https://people.duke.edu/~rnau/411home.htm>

NIELSEN, Aileen. **Análise Prática de Séries Temporais – Predição com Estatística e Aprendizado de Máquina**. Rio de Janeiro: Alta Books, 2021.

VISHWAS, B V; PATEL, Ashish. **Hands-on Time Series Analysis with Python: From Basics to Bleeding Edge Techniques**. Nova Iorque: Apress, 2020.