



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS QUIXADÁ
CURSO DE GRADUAÇÃO EM ENGENHARIA DE SOFTWARE

JOSÉ IAGO DA SILVA LIMA

**FERRAMENTA PARA AUXÍLIO À ANÁLISE DE EVASÃO ATRAVÉS DE
INTERVALOS DE PREDIÇÃO DE SÉRIES TEMPORAIS**

QUIXADÁ

2026

JOSÉ IAGO DA SILVA LIMA

FERRAMENTA PARA AUXÍLIO À ANÁLISE DE EVASÃO ATRAVÉS DE INTERVALOS
DE PREDIÇÃO DE SÉRIES TEMPORAIS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia De Software do CAMPUS QUIXADÁ da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia De Software.

Orientador: Prof. Dr. Criston Pereira de Souza.

QUIXADÁ

2026

JOSÉ IAGO DA SILVA LIMA

FERRAMENTA PARA AUXÍLIO À ANÁLISE DE EVASÃO ATRAVÉS DE INTERVALOS
DE PREDIÇÃO DE SÉRIES TEMPORAIS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia De Software do CAMPUS QUIXADÁ da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia De Software.

Aprovada em: 28/01/2026

BANCA EXAMINADORA

Prof. Dr. Criston Pereira de Souza (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Elvis Miguel Galeas Stancanelli
Universidade Federal do Ceará (UFC)

Prof. Dr. Paulo de Tarso Guerra Oliveira
Universidade Federal do Ceará (UFC)

AGRADECIMENTOS

Primeiramente, agradeço a Deus, por me conceder força, sabedoria e resiliência para superar todos os desafios encontrados ao longo dessa jornada e por me permitir chegar até este momento tão importante da minha vida.

Agradeço, também, aos meus pais, **José Wagner Alves de Lima** e **Antonia Francisca Ferreira da Silva**, por todo o amor, apoio e incentivo ao longo da minha vida e, especialmente, durante minha jornada acadêmica. Foram eles que sempre acreditaram em mim, me motivaram nos momentos difíceis e me deram forças para continuar mesmo diante dos desafios.

Agradeço ao meu irmão e melhor amigo, **José Ian da Silva Lima**, que, muitas vezes sem saber das dificuldades que eu enfrentava, conseguia me alegrar e tornar os momentos mais leves.

Durante essa trajetória enfrentei diversos obstáculos, como o período da pandemia, greve universitária e a perda do meu querido tio **Antonio** que infelizmente não conseguiu me ver chegando ao fim da graduação. Ainda assim, com o apoio da minha família, consegui seguir em frente e chegar até aqui.

Concluir uma graduação não é uma tarefa fácil. Trata-se de uma caminhada longa, desafiadora e, muitas vezes, cansativa, especialmente quando precisamos sair da casa dos pais para enfrentar uma nova rotina, uma nova cidade e responsabilidades maiores. No entanto, tudo se torna mais possível quando temos pessoas que nos apoiam e acreditam em nossos sonhos. Embora longe, meus pais sempre se fizeram presentes na minha jornada, sempre me incentivando e apoiando e comemorando cada pequena vitória durante a graduação.

E, por fim, não posso deixar de agradecer a todas as outras pessoas que me ajudaram durante toda a minha graduação, aos meus vizinhos que me ajudaram quando cheguei em Quixadá, aos meus amigos que fiz durante a graduação, e à minha namorada **Yana Duarte**, que sempre me incentivou e me ajudou durante o período de escrita do TCC.

Por isso, deixo aqui minha profunda gratidão à minha família, que esteve ao meu lado em todos os momentos e foi fundamental para que eu pudesse concluir esta importante etapa da minha vida.

“Quem tem um 'porquê' para viver suporta
quase qualquer 'como'.” (NIETZSCHE, 1889)

RESUMO

A evasão universitária configura-se como um desafio recorrente no contexto das instituições de ensino superior, demandando soluções tecnológicas que auxiliem no acompanhamento e na organização de informações institucionais. Diante desse cenário, este trabalho tem como objetivo o desenvolvimento da ferramenta web denominada *Oráculo*, voltada ao apoio à exploração, visualização e organização de dados relacionados à evasão acadêmica, com foco em facilitar o acesso às informações e subsidiar processos de tomada de decisão. A solução foi concebida a partir de uma arquitetura moderna, escalável e desacoplada, fundamentada no paradigma cliente-servidor. Nessa abordagem, o backend atua como provedor de serviços por meio de uma API REST, enquanto o frontend é responsável pela interação com o usuário final, permitindo evolução independente entre as camadas e maior facilidade de manutenção do sistema. O backend foi desenvolvido utilizando o micro-framework Flask, integrado ao ecossistema Python, enquanto o frontend foi implementado com Next.js e React, garantindo uma interface reativa, responsiva e adequada para navegadores modernos. A persistência dos dados é realizada por meio de um banco de dados relacional leve, com abstração do acesso por meio de um mapeador objeto-relacional. A ferramenta foi projetada para execução em ambientes de computação em nuvem e adota práticas modernas de engenharia de software, como versionamento de código, integração contínua e entrega contínua. Como resultado, o *Oráculo* apresenta-se como uma solução modular e extensível, alinhada aos objetivos do trabalho, contribuindo para o apoio institucional no acompanhamento da evasão universitária por meio de uma abordagem tecnológica integrada.

Palavras-chave: ferramenta web; arquitetura desacoplada; evasão universitária; API REST; visualização de dados.

ABSTRACT

University dropout represents a recurring challenge in the context of higher education institutions, requiring technological solutions that support the monitoring and organization of institutional information. In this context, this work aims to develop the web-based tool called *Oráculo*, designed to support the exploration, visualization, and organization of data related to academic dropout, with the goal of facilitating access to information and supporting decision-making processes. The solution was conceived based on a modern, scalable, and decoupled architecture, grounded in the client–server paradigm. In this approach, the backend acts as a service provider through a RESTful API, while the frontend is responsible for user interaction, enabling independent evolution of system layers and easier system maintenance. The backend was developed using the Flask microframework within the Python ecosystem, while the frontend was implemented with Next.js and React, ensuring a responsive and reactive interface suitable for modern web browsers. Data persistence is handled through a lightweight relational database, with access abstraction provided by an object–relational mapping layer. The tool was designed to operate in cloud computing environments and adopts modern software engineering practices, such as version control, continuous integration, and continuous delivery. As a result, *Oráculo* stands out as a modular and extensible solution aligned with the objectives of this work, contributing to institutional support for monitoring university dropout through an integrated technological approach.

Keywords: web-based tool; decoupled architecture; university dropout; RESTful API; data visualization.

LISTA DE FIGURAS

Figura 1 – Fluxograma dos processos metodológicos.	28
Figura 2 – Painel de indicadores.	29
Figura 3 – Decomposição da série temporal do CH (Centro de Humanidades); a linha vertical indica o começo da pandemia.	32
Figura 4 – Interface de upload e ingestão de arquivos CSV.	42
Figura 5 – Tela de ajuste dos dados das séries temporais.	42
Figura 6 – Tela final de resultados da ferramenta Oráculo.	43
Figura 7 – Gráfico de decomposição da série temporal do campus da UFC de Quixadá.	51
Figura 8 – Gráfico de autocorrelação (ACF) da UFC de Quixadá.	51
Figura 9 – Gráfico de autocorrelação parcial (PACF) da UFC de Quixadá.	52
Figura 10 – Gráfico de predição da UFC de Quixadá.	52
Figura 11 – Gráfico de decomposição da série temporal da unidade CT (Centro de Tecnologia)	53
Figura 12 – Gráficos de autocorrelação (ACF) da unidade CT	54
Figura 13 – Gráficos de autocorrelação parcial (PACF) da unidade CT	54
Figura 14 – Gráficos da predição da unidade CT	54
Figura 15 – Gráfico de decomposição da série temporal da unidade CH (Centro de Humanidades) da UFC	55
Figura 16 – Gráficos de autocorrelação (ACF) da unidade CH	55
Figura 17 – Gráficos de autocorrelação parcial (PACF) da unidade CH	56
Figura 18 – Gráficos da predição da unidade CH	56
Figura 19 – Gráfico de decomposição da série temporal do campus FEAAC (Faculdade de Economia, Administração, Atuária e Contabilidade) da UFC	57
Figura 20 – Gráficos de autocorrelação (ACF) da unidade FEAAC	57
Figura 21 – Gráficos de autocorrelação parcial (PACF) da unidade FEAAC	58
Figura 22 – Gráficos da predição da unidade FEAAC	58
Figura 23 – Gráfico de decomposição da série temporal de toda a UFC	59
Figura 24 – Gráficos de autocorrelação (ACF) de toda a UFC	59
Figura 25 – Gráficos de autocorrelação parcial (PACF) de toda a UFC	60
Figura 26 – Gráficos da predição de toda a UFC	60

LISTA DE TABELAS

Tabela 1 – Relatório comparativo de desempenho do modelo ARIMA por unidade acadêmica	61
--	----

LISTA DE QUADROS

Quadro 1 – Comparação entre os Trabalhos Relacionados e a Proposta	27
Quadro 2 – Tecnologias e bibliotecas utilizadas no projeto.	41
Quadro 3 – Ecossistema de Tecnologias do Backend.	48
Quadro 4 – Ecossistema de Tecnologias do Frontend.	48

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Objetivos	15
<i>1.1.1</i>	<i>Objetivo Geral</i>	<i>15</i>
<i>1.1.2</i>	<i>Objetivos Específicos</i>	<i>15</i>
<i>1.1.3</i>	<i>Organização</i>	<i>15</i>
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	Evasão Universitária	17
2.2	Séries Temporais	18
2.3	Decomposição de Séries Temporais: Método STL	18
2.4	Coleta e Preparação de Dados para Séries Temporais	19
<i>2.4.1</i>	<i>Indicadores Acadêmicos e Coleta</i>	<i>19</i>
<i>2.4.2</i>	<i>Limpeza e Tratamento de Dados</i>	<i>19</i>
<i>2.4.3</i>	<i>Transformação e Suavização de Séries</i>	<i>20</i>
2.5	Modelagem de Séries Temporais: ARIMA	20
2.6	Métricas de Avaliação de Modelos Preditivos	20
<i>2.6.1</i>	<i>Erro Absoluto Médio (MAE)</i>	<i>20</i>
<i>2.6.2</i>	<i>Erro Quadrático Médio (RMSE)</i>	<i>21</i>
<i>2.6.3</i>	<i>Erro Percentual Absoluto Médio (MAPE)</i>	<i>21</i>
2.7	Identificação de Modelos Estocásticos: ACF e PACF	21
<i>2.7.1</i>	<i>Função de Autocorrelação (ACF)</i>	<i>21</i>
<i>2.7.2</i>	<i>Função de Autocorrelação Parcial (PACF)</i>	<i>22</i>
<i>2.7.3</i>	<i>O Algoritmo Auto-ARIMA e Critérios de Informação</i>	<i>22</i>
<i>2.7.4</i>	<i>Definições Matemáticas das Funções de Autocorrelação</i>	<i>23</i>
<i>2.7.4.1</i>	<i>Cálculo da ACF</i>	<i>23</i>
<i>2.7.4.2</i>	<i>Cálculo da PACF</i>	<i>23</i>
2.8	Arquitetura de Aplicações Web Desacopladas	24
3	TRABALHOS RELACIONADOS	25
3.1	Uma Análise de Modelos de Séries Temporais para Predição de Evasão Discente: Estudo de Caso do IFCE	25

3.2	Improving Dropout Forecasting during the COVID-19 Pandemic through Feature Selection and Multilayer Perceptron Neural Network	25
3.3	Ferramenta para Predição do Desempenho Acadêmico no Ensino Superior	26
3.3.1	<i>Quadro Comparativo dos Trabalhos</i>	27
4	METODOLOGIA	28
4.1	Coleta e preparação dos dados	28
4.1.1	<i>Coleta</i>	28
4.1.2	<i>Limpeza de Dados</i>	29
4.1.3	<i>Estruturação dos dados</i>	30
4.1.4	<i>Divisão dos dados</i>	30
4.2	Extração e Análise de Tendência e Sazonalidade	31
4.2.1	<i>Decomposição da Série Temporal</i>	31
4.2.2	<i>Análise Comparativa Pré e Pandemia</i>	32
4.3	Análise de Autocorrelações	33
4.3.1	<i>Autocorrelação e Função de Autocorrelação Parcial (ACF e PACF)</i>	33
4.3.2	<i>Intervalos de Confiança das Autocorrelações</i>	33
4.4	Previsão da Taxa de Evasão Durante a Pandemia	34
4.4.1	<i>Divisão dos Dados</i>	34
4.4.2	<i>Modelagem Preditiva</i>	35
4.5	Modelagem Preditiva por Unidade Acadêmica	35
4.6	Avaliação do Modelo	36
4.7	Interpretação dos Resultados	37
4.7.1	<i>Decomposição da Série Temporal</i>	37
4.7.2	<i>Função de Autocorrelação (ACF)</i>	38
4.7.3	<i>Função de Autocorrelação Parcial (PACF)</i>	38
4.7.4	<i>Previsão da Taxa de Evasão Durante a Pandemia com o Modelo ARIMA</i>	38
4.7.4.1	<i>Interpretação dos Resultados</i>	39
4.7.4.2	<i>Intervalos de Predição</i>	39
4.7.4.3	<i>Validação do Modelo</i>	39
4.8	Arquitetura e Desenvolvimento da Ferramenta	40
4.8.1	<i>Visão Geral da Tecnologia</i>	40
4.8.2	<i>Prototipação e Design de Experiência (UI/UX)</i>	41

4.8.3	Arquitetura Desacoplada	44
4.8.4	Implementação do Backend	44
4.8.4.1	<i>Processamento de Dados e Ingestão</i>	44
4.8.4.2	<i>Modelo de Dados</i>	45
4.8.5	Serviço de Análise Preditiva	45
4.8.5.1	<i>Decomposição e Modelagem ARIMA</i>	45
4.8.5.2	<i>Estratégia de Validação</i>	45
4.8.6	Implementação do Frontend	46
4.8.6.1	<i>Fluxo Lógico e Interfaces</i>	46
4.8.6.2	<i>Gerenciamento de Estado e Persistência Local</i>	47
4.8.6.3	<i>Análise Avançada e Visualização Temporal</i>	47
4.8.7	Tecnologias e Bibliotecas Utilizadas	47
4.8.7.1	<i>Tecnologias de Backend</i>	47
4.8.7.2	<i>Tecnologias de Frontend</i>	48
4.8.8	Automação e Entrega Contínua (CI/CD)	48
4.8.9	Custo-Eficiência e Acessibilidade	49
5	RESULTADOS	50
5.1	Sazonalidade	50
5.2	Modelagem Preditiva por Unidade e de toda UFC	50
5.2.1	<i>Quixadá</i>	50
5.2.2	<i>CT - Centro de Tecnologia</i>	53
5.2.3	<i>CH - Centro de Humanidades</i>	55
5.2.4	<i>FEEAC - Faculdade de Economia, Administração, Atuária e Contabilidade</i>	56
5.2.5	<i>UFC</i>	58
5.2.6	<i>Comportamento das Previsões Constantes no Modelo ARIMA</i>	60
5.3	Avaliação Quantitativa do Desempenho Preditivo	61
5.3.1	<i>Análise por Unidade Acadêmica</i>	61
5.3.2	<i>Análise Geral (Agregada)</i>	62
5.4	Óráculo	62
6	CONCLUSÃO E TRABALHOS FUTUROS	64
	REFERÊNCIAS	66

1 INTRODUÇÃO

A evasão universitária constitui um desafio persistente para as instituições de ensino superior, com impactos significativos tanto para os estudantes quanto para as próprias instituições. De acordo com o Ministério da Educação (2020), a taxa de evasão no Brasil varia entre 20% e 50%, a depender da região e do tipo de curso (BRASIL. Ministério da Educação, 2020). Embora as IES acumulem grandes volumes de dados sobre o histórico discente, a transformação desses dados em informações estratégicas para o suporte à decisão, especialmente em contextos de crise como a pandemia de COVID-19, ainda esbarra em processos manuais. Esses processos são fragmentados e tecnicamente complexos para gestores que não possuem formação ou familiaridade com a área de ciência de dados.

A pandemia, com suas restrições e mudanças abruptas no ambiente educacional, adicionou novas camadas de complexidade à questão da evasão. Um estudo de Lima *et al.* (2021) aponta que os alunos relataram um aumento nos níveis de ansiedade e estresse devido à transição para o ensino remoto, o que contribuiu para o crescimento da desistência em diversas instituições. O contexto de incertezas e a dificuldade de acesso a recursos tecnológicos e apoio psicológico foram identificados como fatores críticos que afetaram a permanência dos estudantes na educação superior durante esse período.

Nesse contexto, as arquiteturas modernas de software oferecem a oportunidade de desacoplar o processamento estatístico intensivo da interface de usuário, permitindo que modelos complexos de séries temporais, como o ARIMA, sejam disponibilizados de forma acessível via aplicações web. A relevância deste trabalho reside no desenvolvimento de uma solução tecnológica capaz de abstrair a complexidade estatística e automatizar o fluxo de análise de indicadores acadêmicos.

Além dos desafios individuais, as instituições de ensino foram compelidas a adaptar-se rapidamente a novos modelos pedagógicos. Essa transição, muitas vezes, ocorreu sem a devida preparação de docentes e gestores, o que, segundo Sousa e Almeida (2022), pode ter comprometido a qualidade da educação oferecida e, conseqüentemente, a motivação discente. Nesse cenário, compreender as nuances dessa conjuntura torna-se essencial para o desenvolvimento de intervenções eficazes que promovam a permanência e o sucesso acadêmico.

Nesse sentido, criamos uma ferramenta para integrar tecnologia e dados institucionais em uma solução acessível e orientada ao apoio à decisão, denominada *Oráculo*. A proposta central do trabalho consiste em desenvolver uma plataforma web capaz de organizar, processar e

apresentar informações relacionadas à evasão universitária de forma automatizada e compreensível, reduzindo a dependência de processos manuais e a necessidade de conhecimento técnico especializado por parte dos gestores. A ferramenta foi concebida com foco na usabilidade, modularidade e escalabilidade, permitindo que análises complexas sejam disponibilizadas por meio de uma interface intuitiva, contribuindo para o monitoramento da evasão acadêmica e para o suporte à formulação de estratégias institucionais de permanência discente.

1.1 Objetivos

Esta seção é referente ao objetivo geral e específico do presente trabalho.

1.1.1 Objetivo Geral

Desenvolver uma ferramenta de software baseada em arquitetura web para automação do processamento e análise de séries temporais da taxa de evasão da UFC, permitindo a identificação de padrões anômalos durante o período da pandemia da COVID-19.

1.1.2 Objetivos Específicos

- Extrair e analisar a tendência e a sazonalidade da série temporal, a fim de comparar os períodos pré e pandemia.
- Calcular as autocorrelações para identificar possíveis padrões repetitivos nas taxas de evasão semestrais, construindo intervalos de confiança para determinar autocorrelações estatisticamente nulas.
- Prever a taxa de evasão no período da pandemia, utilizando apenas dados anteriores a esse período, e construir os respectivos intervalos de predição.
- Aplicar aos dados provenientes do campus de Quixadá e das três unidades com o maior número de alunos, bem como para toda a UFC.
- Desenvolver uma ferramenta para automatizar e facilitar a execução dessas análises.

1.1.3 Organização

Este trabalho está organizado da seguinte forma: o Capítulo 2 aborda a fundamentação teórica; o Capítulo 3 apresenta os trabalhos relacionados; o Capítulo 4 detalha os procedimentos metodológicos; o Capítulo 5 apresenta os resultados; e o Capítulo 6 contém as

conclusões e os trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os conceitos teóricos que sustentam as etapas metodológicas deste trabalho. São abordados temas como evasão universitária, séries temporais, preparação dos dados, o método de decomposição STL, a modelagem com o modelo ARIMA e os critérios de avaliação de modelos preditivos.

2.1 Evasão Universitária

A evasão no ensino superior é um fenômeno complexo, com múltiplas causas relacionadas a fatores acadêmicos, pessoais e institucionais. Segundo o BRASIL. Ministério da Educação (2020), a taxa de evasão no Brasil varia entre 20% e 50%, sendo impactada por condições socioeconômicas, dificuldades pedagógicas, falta de apoio e desmotivação.

Estudos como o de Tanaka *et al.* (2016) evidenciam que a desilusão com o curso, a sobrecarga emocional e o desalinhamento de expectativas podem resultar em frustração e desistência. A pandemia da COVID-19 agravou esses fatores. De acordo com Lima *et al.* (2021), a transição abrupta para o ensino remoto gerou insegurança, estresse e dificuldades de adaptação, contribuindo para o aumento da evasão.

Na Universidade Federal do Ceará (UFC), a evasão é considerada quando o aluno deixa definitivamente ou temporariamente seu curso. Pode ocorrer por cancelamento, trancamento, mudança de curso ou abandono. A mudança de curso dentro da própria instituição não é computada como evasão institucional, mas representa evasão do curso original.

Para o cálculo da evasão, aplica-se a seguinte fórmula aos dados:

$$E_t = 1 - \frac{M_{t+1} + I_{t+1}}{M_t - C_t} \quad (2.1)$$

Onde:

- E_t : Taxa de evasão no semestre t ;
- M_t : Matrículas no semestre t ;
- I_t : Ingressos no semestre t ;
- C_t : Conclusões no semestre t .

2.2 Séries Temporais

Séries temporais consistem em sequências de dados ordenadas cronologicamente, sendo úteis para identificar padrões, tendências e ciclos ao longo do tempo (Box *et al.*, 2016). Neste trabalho, são utilizadas séries semestrais da taxa de evasão da UFC entre os anos de 2013 e 2022.

A decomposição de séries temporais permite separar os seguintes componentes:

- **Tendência:** variação de longo prazo (crescimento ou declínio);
- **Sazonalidade:** padrões que se repetem em intervalos regulares;
- **Ruído:** variações aleatórias e imprevisíveis.

2.3 Decomposição de Séries Temporais: Método STL

A decomposição de séries temporais é o processo de segmentar uma série em componentes individuais para facilitar a análise de padrões subjacentes. Um dos métodos mais robustos para essa tarefa é o *Seasonal-Trend decomposition using Loess* (STL).

O método STL baseia-se no uso de suavizadores *Loess* (regressão local) para decompor a série original em componentes de tendência, sazonalidade e resíduo (Cleveland *et al.*, 1990). A suavização *Loess* é uma técnica de regressão local não paramétrica que ajusta modelos polinomiais a subconjuntos de dados definidos por uma "janela" de vizinhança. Ao atribuir pesos maiores aos pontos mais próximos do valor a ser estimado, o *Loess* permite que a curva se adapte de forma flexível às mudanças de comportamento da série sem ser distorcida por valores distantes. Diferente de métodos clássicos, o STL oferece a vantagem de permitir que o padrão sazonal mude ao longo do tempo e é menos sensível a *valores discrepantes*, o que o torna ideal para analisar dados de evasão universitária impactados por eventos externos atípicos, como a pandemia de COVID-19 (Hyndman; Athanasopoulos, 2018).

O funcionamento do STL ocorre através de um processo iterativo composto por dois loops principais: o loop interno e o loop externo.

1. Loop Interno (Estimativa de Componentes): Em cada iteração deste loop, o algoritmo executa os seguintes passos lógicos para refinar os componentes:

1. **Remoção da Tendência:** Subtrai-se a estimativa de tendência atual da série original.
2. **Suavização das Sub-séries Sazonais:** Aplica-se o suavizador *Loess* em cada sub-série de meses/semestres correspondentes ao longo dos anos.

3. **Filtragem de Passa-Baixa:** Aplica-se uma média móvel e um novo suavizador para remover ruídos da sazonalidade estimada.
4. **Cálculo da Sazonalidade Final:** Subtrai-se o resultado do filtro da etapa anterior para obter o componente sazonal puro.
5. **Cálculo da Tendência:** Subtrai-se a sazonalidade da série original e aplica-se o *Loess* para obter uma nova estimativa de tendência.

2. Loop Externo (Robustez a valores discrepantes): Após o loop interno, o algoritmo calcula o resíduo (Série Original - Tendência - Sazonalidade). No loop externo, identifica-se pontos com resíduos muito elevados (*valores discrepantes*), como os picos de evasão causados pela pandemia. O algoritmo então atribui pesos menores a esses pontos e reinicia o loop interno. Este processo garante que a tendência e a sazonalidade não sejam "puxadas" artificialmente por eventos atípicos, preservando a fidelidade da decomposição (Cleveland *et al.*, 1990).

2.4 Coleta e Preparação de Dados para Séries Temporais

2.4.1 Indicadores Acadêmicos e Coleta

A transparência na gestão acadêmica, viabilizada por portais de indicadores institucionais, permite o acompanhamento de métricas fundamentais como evasão, matrículas e retenção (UFC. Pró-Reitoria de Graduação, 2025). A organização desses indicadores de forma longitudinal e segmentada é essencial para identificar variações estruturais no comportamento discente ao longo do tempo. ¹

2.4.2 Limpeza e Tratamento de Dados

A etapa de limpeza de dados visa remover inconsistências, como registros duplicados e valores nulos. Esse processo é considerado essencial na literatura estatística para garantir a qualidade dos resultados e a validade das inferências subsequentes, evitando que dados espúrios corrompam a análise (Rubin, 1976).

¹ Dados extraídos do Painel de Indicadores da Graduação da UFC <https://tinyurl.com/painel-prograd-ufc>. Acesso em: 15 jan. 2026.

2.4.3 Transformação e Suavização de Séries

A estruturação de dados em séries temporais exige, frequentemente, a normalização de variáveis para a padronização de escalas. Adicionalmente, técnicas de suavização, como as médias móveis, são empregadas para mitigar flutuações sazonais de curto prazo e destacar a tendência de longo prazo.

2.5 Modelagem de Séries Temporais: ARIMA

O modelo *ARIMA* (*Autoregressive Integrated Moving Average*) é uma técnica clássica de análise estatística composta pelos componentes **AR** (autorregressivo), **I** (integração) e **MA** (média móvel). Sua formulação é representada pela tríade (p, d, q) , onde p indica a dependência de valores passados, d o grau de diferenciação para tornar a série estacionária e q o ajuste baseado em erros de predição anteriores (Hyndman; Athanasopoulos, 2018). Embora a aplicação do *ARIMA* tradicional exija uma análise subjetiva e manual dos gráficos de autocorrelação (ACF) e autocorrelação parcial (PACF) (Box *et al.*, 2016), tal abordagem seria inviável em uma ferramenta automatizada de larga escala. Nesse contexto, o backend da ferramenta utiliza o algoritmo *Auto-ARIMA*. Este procedimento realiza uma busca exaustiva e sistemática no espaço de estados dos parâmetros, selecionando a combinação ideal que minimiza o *Akaike Information Criterion* (AIC) (Hyndman; Khandakar, 2008). Essa automação garante que o sistema forneça predições otimizadas de forma dinâmica para diferentes conjuntos de dados, sem a necessidade de intervenção técnica manual para cada unidade acadêmica analisado.

2.6 Métricas de Avaliação de Modelos Preditivos

Três métricas principais foram utilizadas para avaliar o desempenho dos modelos: MAE, RMSE e MAPE.

2.6.1 Erro Absoluto Médio (MAE)

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (2.2)$$

O MAE expressa a média dos desvios absolutos na mesma unidade da variável analisada (Hyndman; Athanasopoulos, 2018).

2.6.2 Erro Quadrático Médio (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (2.3)$$

O RMSE penaliza mais fortemente grandes erros, sendo apropriado quando desvios elevados são críticos para a tomada de decisão (Shmueli *et al.*, 2017).

2.6.3 Erro Percentual Absoluto Médio (MAPE)

$$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (2.4)$$

O MAPE expressa o erro de previsão em porcentagem, sendo útil para a comparação do desempenho entre séries de diferentes magnitudes (Armstrong; Collopy, 1992).

2.7 Identificação de Modelos Estocásticos: ACF e PACF

A identificação da ordem dos processos autorregressivos (p) e de média móvel (q) no modelo ARIMA fundamenta-se na análise das funções de autocorrelação. Segundo a metodologia de Box *et al.* (2016), essas funções permitem extrair a estrutura de dependência temporal presente na série histórica, sendo fundamentais para a correta especificação do modelo.

2.7.1 Função de Autocorrelação (ACF)

A Função de Autocorrelação (ACF) mede a correlação linear entre a observação atual y_t e seus valores defasados no tempo y_{t-k} . No contexto do modelo ARIMA, a ACF é a ferramenta primária para identificar a ordem do componente de Média Móvel (MA). Um processo MA(q) teórico apresenta uma ACF que “corta” (torna-se estatisticamente nula) após a defasagem q (Hyndman; Athanasopoulos, 2018). Em contrapartida, para processos autorregressivos, a ACF tende a decair exponencialmente ou seguir um padrão de senoide amortecida, não apresentando um corte abrupto.

2.7.2 Função de Autocorrelação Parcial (PACF)

Diferente da ACF, a Função de Autocorrelação Parcial (PACF) remove a influência das observações intermediárias ao calcular a correlação entre y_t e y_{t-k} . Matematicamente, ela isola o impacto direto de um evento passado no presente. Esta função é essencial para determinar a ordem do componente Autorregressivo (AR). Para um modelo $AR(p)$, a PACF apresenta valores significativos até a defasagem p e cessa abruptamente após esse ponto (Morettin; Tolo, 2006).

2.7.3 O Algoritmo Auto-ARIMA e Critérios de Informação

Embora a identificação visual via ACF e PACF seja a base clássica, este trabalho utiliza o algoritmo Hyndman e Khandakar (2008) *Auto-ARIMA*. Este procedimento, implementado amplamente em pacotes estatísticos, automatiza a seleção de parâmetros através de uma busca iterativa, utilizando o Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC) como métricas de otimização principal (Hyndman; Khandakar, 2008; Hyndman; Athanasopoulos, 2018).

O AIC e o BIC são definidos pelas Equações 2.5 e 2.6, respectivamente:

$$AIC = 2k - 2\ln(\hat{L}) \quad (2.5)$$

$$BIC = k\ln(n) - 2\ln(\hat{L}) \quad (2.6)$$

Onde k representa o número de parâmetros, \hat{L} a verossimilhança máxima do modelo e n o número de observações (tamanho da amostra). A verossimilhança máxima indica o quão bem os parâmetros do modelo ajustam-se aos dados observados, maximizando a probabilidade de que o modelo descreva a série histórica real.

O algoritmo busca minimizar esses critérios, garantindo o princípio da parcimônia: a seleção do modelo com maior poder preditivo utilizando o menor número de parâmetros possível. Enquanto o AIC foca na qualidade da previsão, o BIC aplica uma penalidade mais severa à complexidade do modelo conforme o tamanho da amostra cresce, auxiliando na prevenção do *overfitting* e garantindo a capacidade de generalização do modelo para períodos futuros (Shmueli *et al.*, 2017).

2.7.4 Definições Matemáticas das Funções de Autocorrelação

Para compreender a dependência temporal, define-se primeiramente a autocovariância no lag k , denotada por γ_k . Para uma série temporal estacionária y_t com média μ , a autocovariância é dada por:

$$\gamma_k = E[(y_t - \mu)(y_{t-k} - \mu)] \quad (2.7)$$

2.7.4.1 Cálculo da ACF

A Função de Autocorrelação (ACF) no lag k , representada por ρ_k , é a autocovariância normalizada pela variância da série (γ_0). Seu valor varia no intervalo $[-1, 1]$ e é calculada conforme a Equação 2.8:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2} \quad (2.8)$$

Onde T é o tamanho da série e \bar{y} é a média amostral.

2.7.4.2 Cálculo da PACF

A Função de Autocorrelação Parcial (PACF), denotada por ϕ_{kk} , mede a correlação entre y_t e y_{t-k} após remover a dependência linear das observações de lags intermediários ($y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$). Formalmente, para um lag k , ela pode ser obtida através das equações de Yule-Walker ou via regressão:

$$y_t = \phi_{k1}y_{t-1} + \phi_{k2}y_{t-2} + \dots + \phi_{kk}y_{t-k} + \varepsilon_t \quad (2.9)$$

Onde ϕ_{kk} representa o coeficiente de autocorrelação parcial para a defasagem k . Para um processo $AR(p)$, tem-se que $\phi_{kk} = 0$ para todo $k > p$ e o ε_t denota o erro aleatório (ruído branco). O termo ε_t é assumido como uma variável com distribuição normal, média próxima de zero e sem autocorrelação significativa, representando as flutuações da taxa de evasão.

2.8 Arquitetura de Aplicações Web Desacopladas

A evolução das interfaces web permitiu a transição de sistemas monolíticos para arquiteturas desacopladas, nas quais o frontend e o backend operam de forma independente. Nessa estrutura, o backend funciona como um provedor de recursos através de uma API baseada no padrão REST (Representational State Transfer)², que utiliza o formato JSON para o tráfego de dados (Fielding, 2000).

O uso do micro-framework Flask³ no lado do servidor possibilita o processamento intensivo de modelos de séries temporais em Python⁴, enquanto o Next.js⁵ no lado do cliente garante uma interface reativa e otimizada. Essa separação facilita a manutenção e escalabilidade da ferramenta "Oráculo" em serviços de nuvem modernos.

A gestão do ciclo de vida do software baseia-se no sistema de versionamento Git⁶ (Chacon; Straub, 2014). Esta escolha permitiu a orquestração de pipelines de integração e entrega contínua (CI/CD)⁷, garantindo que as atualizações fossem validadas e implantadas de forma automática (Fowler, 2006) nos ambientes de hospedagem Vercel⁸ e Render⁹.

² Especificação do padrão REST disponível em: https://ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm.

³ Documentação oficial do Flask disponível em: <https://flask.palletsprojects.com/en/stable/>.

⁴ Documentação oficial do Python disponível em: <https://docs.python.org/3/>.

⁵ Documentação oficial do Next.js disponível em: <https://nextjs.org/docs>.

⁶ Documentação oficial do Git disponível em: <https://git-scm.com/doc>.

⁷ Visão geral sobre Integração e Entrega Contínua disponível em: <https://martinfowler.com/articles/continuousIntegration.html>.

⁸ Documentação oficial da plataforma Vercel disponível em: <https://vercel.com/docs>.

⁹ Documentação oficial da plataforma Render disponível em: <https://render.com/docs>.

3 TRABALHOS RELACIONADOS

A seguir, descrevemos três trabalhos relacionados ao tema deste estudo.

3.1 Uma Análise de Modelos de Séries Temporais para Predição de Evasão Discente: Estudo de Caso do IFCE

O artigo Filho *et al.* (2021) utiliza técnicas de séries temporais, como ARIMA, regressão linear e redes neurais autorregressivas (NNAR), para prever taxas de evasão em nove campi do Instituto Federal do Ceará (IFCE) no período de 2009 a 2017, empregando dados semestrais. A análise avalia a eficiência dessas abordagens na previsão de taxas de evasão, visando apoiar o planejamento estratégico. Os resultados são promissores, demonstrando precisão e adequação dos modelos aos dados históricos. Assim como este trabalho, o artigo aplica séries temporais para identificar padrões históricos e prever comportamentos futuros. No entanto, este estudo se diferencia ao realizar uma análise comparativa de períodos antes e durante a pandemia na UFC, enquanto o artigo do IFCE não coleta dados da pandemia, pois o corte do estudo se limita somente até 2017. Além da abordagem do estudo do IFCE, incorpora-se a análise de autocorrelações para identificar padrões sazonais, aspecto não explorado explicitamente por aquele estudo. Essa análise adicional contribui para uma compreensão mais detalhada dos padrões presentes nas taxas de evasão.

3.2 Improving Dropout Forecasting during the COVID-19 Pandemic through Feature Selection and Multilayer Perceptron Neural Network

Os alunos abandonam os estudos por diversos motivos, como dificuldades financeiras, notas baixas, mudanças de curso ou transferência para outras instituições. Esse fenômeno afeta negativamente tanto a vida acadêmica dos estudantes quanto a sustentabilidade econômica e educacional das instituições. A evasão pode comprometer a qualidade educacional, pressionar o orçamento e indicar falhas no sistema e currículo. Além disso, exige adaptações em métodos de ensino e gestão para atender às necessidades impostas pelo distanciamento social. No contexto atual, muitos dos desafios enfrentados por estudantes universitários que adotaram o ensino online durante a pandemia de COVID-19 permanecem evidentes. Um dos principais problemas foi a evasão escolar, que afetou diretamente as universidades. O ensino remoto impactou negativamente a aprendizagem dos alunos devido à falta de estímulos proporcionados pelas

interações presenciais. As taxas de evasão no ensino à distância continuam superiores às da educação presencial (Nuanmeesri *et al.*, 2022). Já este trabalho, que adota uma abordagem estatística, as análises concentram-se na detecção de anomalias e na previsão de taxas de evasão durante a pandemia. Este trabalho foca em identificar se os padrões observados estão fora do esperado para os períodos de análise. Por outro lado, o estudo de Nuanmeesri *et al.* (2022) explora um panorama mais abrangente, abordando fatores socioeconômicos e psicológicos dos alunos, sem uma ênfase exclusiva em modelagens estatísticas. A principal diferença é o foco: enquanto o artigo Nuanmeesri *et al.* (2022) analisa a evasão como um problema multifacetado, envolvendo questões sociais e institucionais, este trabalho investiga diretamente os dados de séries temporais e suas tendências, oferecendo um suporte preditivo para auxiliar na tomada de decisões baseadas em dados concretos. Ambos os estudos são complementares, pois, neste estudo, damos maior ênfase a técnicas quantitativas.

3.3 Ferramenta para Predição do Desempenho Acadêmico no Ensino Superior

O trabalho de Garcia *et al.* (ano) apresenta o desenvolvimento de uma ferramenta de software voltada para a predição do desempenho acadêmico, especificamente para identificar riscos de reprovação em disciplinas. Os autores utilizam como base o "percurso curricular" dos alunos, transformando o histórico de aprovações e reprovações em vetores de características preditivas. O sistema foi implementado em Java com integração à biblioteca Weka, disponibilizando os resultados através de uma interface web intuitiva para gestores e docentes.

A ferramenta de Garcia *et al.* (ano) obteve uma acurácia média de 86%, demonstrando que dados puramente acadêmicos são suficientes para predições robustas, sem a necessidade de atributos socioeconômicos. Uma característica marcante do software é a automação: o usuário insere apenas a lista de matriculados (arquivo CSV) e o sistema processa automaticamente o histórico para gerar a predição.

Embora ambos os trabalhos foquem no desenvolvimento de ferramentas para suporte à decisão acadêmica, este estudo se diferencia de Garcia *et al.* (ano) em dois pontos centrais: Natureza do Modelo: Enquanto o trabalho de Garcia utiliza algoritmos de classificação (Random Forest, J48) para prever o sucesso individual em disciplinas, o Oráculo utiliza séries temporais (*auto-ARIMA*) para analisar tendências institucionais agregadas. Contexto de Análise: Este trabalho foca especificamente na detecção de anomalias estatísticas geradas pela pandemia de COVID-19, um fator externo não abordado pela ferramenta citada

3.3.1 Quadro Comparativo dos Trabalhos

Nesta seção, apresentamos uma tabela comparando os três trabalhos relacionados com a presente pesquisa. Para essa comparação, adotam-se os seguintes critérios: Efeito Pandemia, que verifica se o estudo considera o impacto da COVID-19; Método, que identifica a abordagem predominante entre Estatística ou Aprendizado de Máquina; Modalidade de Ensino, para determinar se os dados referem-se ao ensino presencial ou online; Horizonte de Dados, que analisa o período histórico utilizado em anos; e, por fim, Desenvolveu Ferramenta, para indicar se o trabalho resultou em um artefato de software para suporte à decisão. O quadro a seguir resume essa comparação.

Quadro 1 – Comparação entre os Trabalhos Relacionados e a Proposta

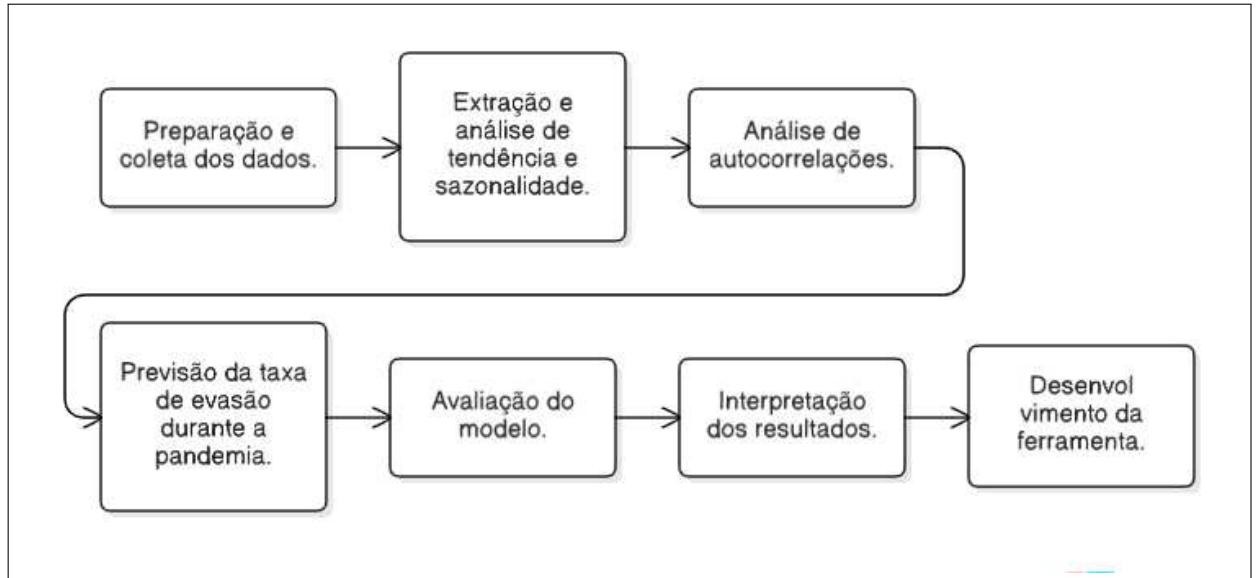
Artigo	Efeito Pandemia	Método Preditivo	Modalidade	Horizonte (Anos)	Desenvolveu Ferramenta
Filho <i>et al.</i> (2021)	Não	Aprendizado de Máquina	Presencial	8	Não
Nuanmeesri <i>et al.</i> (2022)	Sim	Aprendizado de Máquina	Presencial	30	Não
Garcia <i>et al.</i> (ano)	Não	Aprendizado de Máquina	Presencial	9	Sim
Este trabalho	Sim	Método Estatístico	Presencial	10	Sim

Fonte: Elaborado pelo autor.

4 METODOLOGIA

Este capítulo apresenta os processos metodológicos utilizados para alcançar o objetivo final do projeto. A análise será replicada para o Campus de Quixadá, para as três unidades com mais alunos e para toda a UFC, conforme o fluxograma na Figura 1.

Figura 1 – Fluxograma dos processos metodológicos.



Fonte: elaborado pelo autor.

4.1 Coleta e preparação dos dados

4.1.1 Coleta

Os dados foram coletados no painel público da Pró-Reitoria de Graduação (PROGRAD)¹ da UFC, conforme ilustrado na Figura 2, que disponibiliza informações agregadas sobre as unidades acadêmicas, como taxas de evasão, matrículas, concluintes e retenção. A seleção dos dados foi orientada pela necessidade de obter informações relevantes para a análise da evasão no período de 2013 a 2022, priorizando dados que permitissem a construção de séries temporais consistentes.

¹ Dados extraídos do Painel de Indicadores da Graduação da UFC <https://tinyurl.com/painel-prograd-ufc>. Acesso em: 15 jan. 2026.

Figura 2 – Painel de indicadores.



Fonte: Painel de indicadores da UFC.

Ao acessar o painel público da Pró-Reitoria de Graduação (PROGRAD) da Universidade Federal do Ceará, deve-se navegar até a seção referente às taxas de evasão e retenção. A seleção dos dados foi orientada pela necessidade de obter informações relevantes para a análise da evasão universitária, com foco no período de 2013 a 2022. Foram priorizados dados que possibilitassem a construção de séries temporais consistentes, especialmente as taxas de evasão por semestre.

Para a coleta dos dados, é necessário selecionar manualmente as linhas desejadas na tabela, pressionando a tecla *Ctrl* enquanto se clica em cada linha. Em seguida, deve-se clicar com o botão direito do mouse, escolher a opção “Copiar” e, posteriormente, “Copiar seleção”. Após esse procedimento, os dados devem ser colados em um editor de planilhas de preferência do pesquisador e exportados em formato .csv para posterior tratamento e análise.

4.1.2 Limpeza de Dados

Após a coleta e importação da base de dados, foi realizado um processo sistemático de limpeza com o objetivo de garantir a consistência e a qualidade das informações utilizadas neste estudo. Essa etapa é essencial em análises quantitativas, pois a presença de dados incompletos, inconsistentes ou irrelevantes pode comprometer a confiabilidade dos resultados e, consequentemente, a validade das conclusões.

Inicialmente, procedeu-se à remoção de informações que não apresentavam relevância direta para a investigação. Foram excluídas as unidades acadêmicas que não possuíam registros anteriores ao semestre de 2020.1, uma vez que tais unidades não contribuiriam para a construção de uma série temporal completa abrangendo todo o período analisado (2013 a 2022). Também foram eliminadas as unidades classificadas como temporárias, por não representarem estruturas acadêmicas permanentes da instituição.

Em seguida, foi realizada uma verificação detalhada para identificar semestres em que a taxa de evasão não estivesse registrada. Esses casos foram removidos da base de dados a fim de evitar lacunas na série temporal, que poderiam afetar o desempenho e a precisão dos modelos de previsão. Um exemplo notável dessa situação ocorreu no semestre de 2023.1, no qual nenhuma das unidades apresentava o registro da taxa de evasão, motivo pelo qual todos os dados desse período foram descartados.

Essas ações visaram garantir que apenas informações completas, consistentes e representativas permanecessem na base de dados, de modo a proporcionar uma análise mais fidedigna e a reduzir possíveis vieses decorrentes de falhas ou ausências de registro.

4.1.3 Estruturação dos dados

Após a etapa de limpeza, os dados coletados foram transformados em uma série temporal, com o índice temporal definido pelo semestre de referência. Cada ponto da série representa a taxa de evasão correspondente a um semestre, criando uma sequência cronológica contínua do período de 2013 a 2022.

Além disso, a transformação incluiu a normalização das variáveis, como a taxa de evasão, para padronizar as escalas e facilitar a aplicação de modelos estatísticos. Em casos em que a variabilidade dos dados foi muito alta, técnicas de suavização, como médias móveis, foram aplicadas para reduzir a oscilação e destacar as tendências de longo prazo. Essa abordagem é útil para identificar padrões sazonais e flutuações relacionadas a eventos específicos, como o impacto da pandemia de COVID-19.

4.1.4 Divisão dos dados

Após essa estruturação, vamos dividir os dados para nossa análise. Para isso, iremos filtrar primeiramente os dados que tenham na coluna **UNIDADE** o valor **QUIXADÁ**, criando um subconjunto de dados somente com indicadores de evasão desse campus. Logo após esse

processo, vamos filtrar as três unidades com mais alunos matriculados na soma agregada de todos os semestres analisados, que são, respectivamente, CT (Centro de Tecnologia), CH (Centro de Humanidades), FEAAC (Faculdade de Economia, Administração, Atuária e Contabilidade) e, por fim, iremos pegar todos os dados e agregar de modo que gere um conjunto de dados únicos que represente toda a UFC. Como forma de fazer a agregação de uma maneira mais justa, iremos dar pesos às unidades com base no número de matrículas, usando a moda como parâmetro para definir os pesos das unidades.

4.2 Extração e Análise de Tendência e Sazonalidade

4.2.1 Decomposição da Série Temporal

A decomposição da série temporal foi realizada com o objetivo de separar a variação total das taxas de evasão em três componentes principais: tendência, sazonalidade e componente aleatória (resíduo). Para isso, foi utilizado o método *Seasonal and Trend decomposition using Loess* (STL)², implementado na linguagem Python na versão 3.12³.

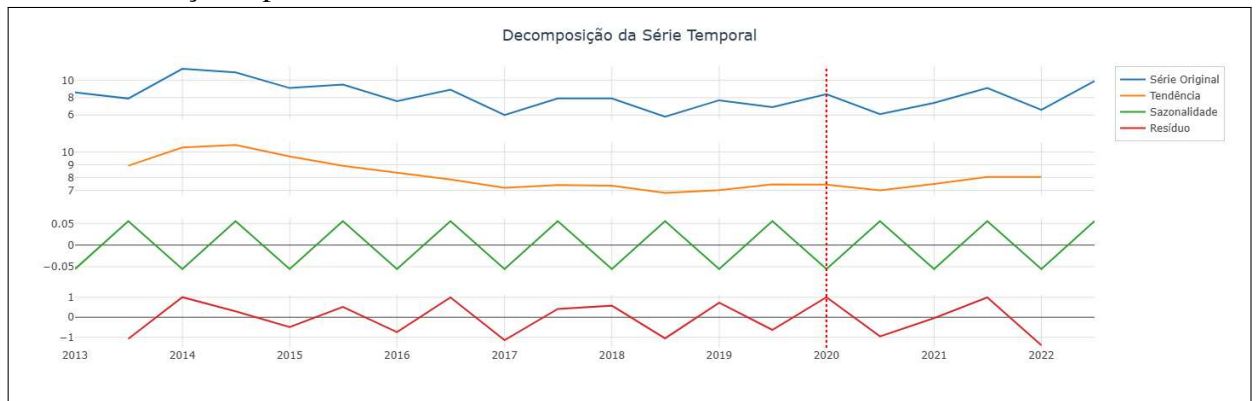
O procedimento iniciou-se com a aplicação da função `stl()`, que ajusta a tendência e a sazonalidade da série por meio de suavização local. A tendência foi obtida a partir de um ajuste suave de longo prazo, refletindo a direção geral de crescimento ou queda das taxas de evasão ao longo dos semestres. A sazonalidade foi identificada como um padrão recorrente com período fixo, correspondente ao ciclo acadêmico semestral. O componente aleatório foi obtido como o resíduo resultante da remoção da tendência e da sazonalidade dos dados originais.

A escolha do método STL deve-se à sua robustez frente a valores discrepantes e à flexibilidade para lidar com séries temporais não estacionárias e com padrões sazonais não necessariamente constantes ao longo do tempo. Além disso, a decomposição permitiu analisar separadamente o comportamento de cada componente, possibilitando comparações mais precisas entre períodos antes e durante a pandemia.

² Documentação do método STL disponível em: <https://www.statsmodels.org/stable/generated/statsmodels.tsa.seasonal.STL.html>

³ Documentação oficial do Python disponível em: <https://docs.python.org/3.12/>

Figura 3 – Decomposição da série temporal do CH (Centro de Humanidades); a linha vertical indica o começo da pandemia.



Fonte: elaborado pelo autor.

A interpretação dos componentes extraídos permitiu compreender diferentes aspectos do comportamento da evasão. A tendência indicou a direção predominante ao longo dos anos, evidenciando possíveis mudanças estruturais nas taxas de evasão. A sazonalidade mostrou oscilações recorrentes associadas ao calendário acadêmico, como aumentos ou reduções específicos de determinados semestres. Já o componente aleatório evidenciou variações não explicadas pelos padrões de longo prazo ou pela sazonalidade, podendo estar relacionadas a eventos pontuais ou fatores externos, como alterações nas políticas institucionais ou impactos socioeconômicos.

4.2.2 Análise Comparativa Pré e Pandemia

Para verificar se a pandemia da COVID-19 alterou significativamente o comportamento das taxas de evasão, foi realizada uma análise comparativa entre os períodos pré-pandemia e pandemia. A série temporal foi segmentada em dois intervalos: o primeiro, abrangendo os semestres anteriores ao início da pandemia (2013.1 até 2019.2), e o segundo, incluindo os semestres durante e após o início da pandemia (2020.1 em diante).

A comparação foi conduzida observando-se as componentes de tendência e sazonalidade obtidas na decomposição da série, de forma a identificar mudanças no comportamento de longo prazo e nos padrões sazonais. Adicionalmente, avaliou-se a magnitude e a direção das variações na tendência, buscando indícios de aceleração, desaceleração ou reversão no padrão de evasão. A análise permitiu isolar efeitos estruturais de longo prazo de possíveis impactos específicos relacionados ao contexto pandêmico.

4.3 Análise de Autocorrelações

4.3.1 Autocorrelação e Função de Autocorrelação Parcial (ACF e PACF)

A análise de autocorrelação foi realizada com o objetivo de identificar a relação entre valores da série temporal das taxas de evasão em diferentes defasagens (lags). Para isso, foram utilizados os gráficos da Função de Autocorrelação (ACF) e da Função de Autocorrelação Parcial (PACF), gerados na linguagem Python utilizando a biblioteca *statsmodels*.

O gráfico da ACF permite avaliar a presença de padrões de dependência ao longo de múltiplas defasagens, indicando se valores passados influenciam os valores futuros da série. O gráfico da PACF mede a correlação de cada defasagem com o valor atual, removendo a influência das defasagens intermediárias. Para ambos os gráficos, foram calculados intervalos de confiança de 95%, permitindo verificar a significância estatística das autocorrelações.

A interpretação dos gráficos indicou como a evasão universitária se relaciona com semestres anteriores. Picos significativos na ACF sugeriram a presença de padrões sazonais ou de longo prazo, enquanto a PACF indicou a ordem provável do componente autorregressivo (AR) necessário para capturar a dependência nos dados. No contexto da evasão, correlações significativas em determinadas defasagens podem estar associadas a fatores institucionais ou ciclos acadêmicos recorrentes, como o ingresso e a conclusão de cursos em semestres específicos, tornando a análise mais robusta e fundamentada para a modelagem preditiva.

4.3.2 Intervalos de Confiança das Autocorrelações

Para avaliar a significância estatística das autocorrelações, foram calculados intervalos de confiança de 95% para cada defasagem da série temporal. Esses intervalos indicam se uma autocorrelação observada difere significativamente de zero, considerando o acaso. Valores de autocorrelação que ultrapassam os limites do intervalo de confiança são considerados estatisticamente significativos, sugerindo dependência temporal relevante na série.

Na prática, a geração desses intervalos pela biblioteca *statsmodels*⁴ baseia-se na distribuição assintótica das autocorrelações. Sob a hipótese nula de que a série é um ruído branco, os coeficientes de autocorrelação seguem uma distribuição normal com média zero e variância $1/n$, onde n é o número de observações da série. Assim, os limites de confiança de 95% são calculados como aproximadamente $\pm 1,96/\sqrt{n}$.

⁴ Documentação oficial do statsmodels disponível em: <https://www.statsmodels.org/stable/>

Ao plotar os gráficos de ACF e PACF no Python, essas faixas de confiança são representadas visualmente (geralmente como uma área sombreada). Autocorrelações dentro dessa área indicam que a dependência observada é estatisticamente insignificante, assemelhando-se ao ruído aleatório. Por outro lado, picos que excedem esses limites sinalizam padrões consistentes, como tendências ou sazonalidades, que devem ser incorporados na modelagem. Essas informações subsidiaram a escolha dos parâmetros autorregressivos (AR) e de médias móveis (MA) no modelo selecionado pela função *auto_arima()*, garantindo que a configuração final do modelo ARIMA estivesse alinhada com a estrutura temporal observada nos dados de evasão.

No contexto das taxas de evasão, as autocorrelações significativas podem indicar que a evasão em um semestre está influenciada por semestres anteriores, refletindo padrões estruturais ou sazonais, como efeitos do calendário acadêmico ou políticas institucionais. A utilização dos intervalos de confiança permitiu diferenciar entre correlações estatisticamente relevantes e variações aleatórias, tornando a análise mais robusta e fundamentada para a modelagem preditiva.

4.4 Previsão da Taxa de Evasão Durante a Pandemia

4.4.1 Divisão dos Dados

Para possibilitar a previsão da taxa de evasão durante o período da pandemia, a série temporal utilizada neste estudo foi dividida em dois subconjuntos: um de treino e outro de teste. O conjunto de treino abrange o período de 2013 a 2019, que corresponde a um cenário estável e sem os efeitos da pandemia de COVID-19. Já o conjunto de teste engloba os anos de 2020 a 2022, período marcado diretamente pelos impactos da pandemia e, portanto, considerado essencial para a validação do modelo.

Essa divisão tem como objetivo central treinar o modelo preditivo em um contexto anterior à pandemia, em que o comportamento da evasão se desenvolveu de forma mais regular, e, posteriormente, avaliar a qualidade das previsões quando exposto a um cenário atípico. Tal abordagem possibilita a construção de comparações diretas entre os valores reais de evasão durante a pandemia e aqueles estimados pelo modelo com base apenas em informações anteriores a esse evento.

4.4.2 Modelagem Preditiva

Para a etapa de modelagem preditiva, foi utilizado o modelo *ARIMA* juntamente com o algoritmo *auto-ARIMA*, implementado por meio da biblioteca `sktime`⁵. Esse método tem a vantagem de automatizar a escolha dos parâmetros do modelo *ARIMA*, responsáveis por capturar os componentes de autocorrelação, diferenciação e médias móveis da série temporal. A seleção automática dos parâmetros ocorre com base em critérios estatísticos, como o AIC (*Akaike Information Criterion*) e o BIC (*Bayesian Information Criterion*), que avaliam a qualidade do ajuste do modelo e evitam tanto o sobreajuste quanto a subestimação da complexidade da série. Esses critérios funcionam penalizando a inclusão de parâmetros adicionais que não promovam um ganho significativo na explicação da variância dos dados. Dessa forma, evita-se o sobreajuste (*overfitting*), situação em que o modelo se torna excessivamente complexo e passa a descrever ruídos aleatórios da série histórica em vez do seu padrão estrutural, o que prejudicaria a capacidade de generalização para dados futuros. Simultaneamente, o algoritmo mitiga a subestimação da complexidade (*underfitting*) ao testar ordens superiores de parâmetros até que os resíduos da predição se comportem como um ruído branco, garantindo que toda a informação temporal relevante tenha sido extraída. O uso do *auto-ARIMA* se mostra particularmente adequado neste contexto, pois a série de evasão apresenta tanto padrões de tendência de longo prazo quanto oscilações semestrais (sazonalidade), que são difíceis de serem determinados manualmente. Assim, o algoritmo garante uma modelagem consistente e eficiente, fornecendo previsões confiáveis mesmo diante de um cenário inesperado, como a pandemia.

A escolha do algoritmo *auto-ARIMA* justifica-se pela necessidade de automação do motor de inferência da ferramenta Oráculo. Dado que o sistema deve ser capaz de realizar previsões para múltiplas unidades acadêmicas com comportamentos distintos, a seleção automática de parâmetros via minimização do AIC assegura a escalabilidade da solução e a reprodutibilidade das análises sem a necessidade de intervenção manual especializada.

4.5 Modelagem Preditiva por Unidade Acadêmica

Após a transformação dos dados em séries temporais, o próximo passo consiste na aplicação da modelagem preditiva, cujo objetivo é estimar as taxas de evasão ao longo do tempo em diferentes unidades acadêmicas. A análise foi realizada considerando o Campus de Quixadá,

⁵ Documentação oficial do `pmdarima` disponível em: <https://www.sktime.net/en/latest/users.html>

os três maiores campi da UFC e, de forma consolidada, toda a Universidade Federal do Ceará (UFC). Essa divisão permite compreender tanto particularidades locais quanto tendências gerais da instituição, ampliando a relevância e aplicabilidade dos resultados.

Para a construção do modelo, optou-se pelo *ARIMA* (AutoRegressive Integrated Moving Average), amplamente reconhecido por sua capacidade de capturar padrões de dependência temporal, oscilações sazonais e variações irregulares em séries históricas. A estimação dos parâmetros (p , d , q) foi realizada por meio do algoritmo *auto-ARIMA*, que automatiza esse processo com base na minimização do Critério de Informação de Akaike (AIC) e do Critério Bayesiano de Schwarz (BIC) vide as equações 2.5 e 2.6. Essa abordagem confere maior rigor estatístico, reduz a subjetividade na definição dos parâmetros e assegura melhor aderência às características intrínsecas da série de evasão analisada.

Além disso, durante a etapa de modelagem, foram avaliados os resíduos do modelo, buscando verificar se estes se comportam como ruído branco, sem autocorrelações significativas. Essa verificação foi feita com base no teste de Ljung-Box e na análise dos intervalos de confiança das funções de autocorrelação (ACF). A ausência de padrões nos resíduos indica que o modelo consegue capturar adequadamente a estrutura da série temporal, garantindo robustez nas previsões.

4.6 Avaliação do Modelo

A avaliação do desempenho do modelo ajustado foi realizada com base em métricas tradicionais e amplamente utilizadas na literatura de séries temporais, tais como o Erro Quadrático Médio (RMSE), o Erro Absoluto Médio (MAE) e o Erro Percentual Absoluto Médio (MAPE). O RMSE, ao considerar o quadrado dos erros antes de calcular a raiz, penaliza mais fortemente grandes desvios, sendo útil para identificar modelos que cometem erros expressivos em pontos específicos da série. Já o MAE, por representar a média dos erros absolutos, é menos sensível a valores extremos e fornece uma medida direta da magnitude do erro médio. O MAPE, por sua vez, expressa o erro médio em termos percentuais, permitindo comparar o desempenho do modelo entre séries de diferentes escalas e facilitando a interpretação dos resultados por gestores acadêmicos e formuladores de políticas institucionais.

Com o intuito de aumentar a confiabilidade do processo de avaliação e reduzir riscos de sobreajuste, a base de dados foi dividida em dois subconjuntos: um de treino, abrangendo os semestres de 2013 a 2019, e outro de teste, referente ao período de 2020 a 2022. Essa divisão se

mostrou particularmente relevante, uma vez que possibilitou avaliar a capacidade do modelo de generalizar em períodos futuros e, principalmente, em condições atípicas impostas pela pandemia de COVID-19. Dessa forma, tornou-se possível analisar se as taxas de evasão durante o período pandêmico se mantiveram dentro do esperado segundo os padrões históricos, ou se apresentaram desvios significativos.

4.7 Interpretação dos Resultados

Após o ajuste e avaliação do modelo *ARIMA*, foi realizada a análise detalhada dos resultados obtidos. Essa interpretação concentrou-se em quatro pontos principais: a decomposição da série temporal, a análise da função de autocorrelação (ACF), a função de autocorrelação parcial (PACF) e, por fim, a previsão da taxa de evasão durante o período da pandemia. Esses elementos, em conjunto, permitiram avaliar se o comportamento da evasão universitária segue padrões consistentes e previsíveis ou se foi fortemente impactado por fatores externos.

4.7.1 *Decomposição da Série Temporal*

A decomposição da série temporal foi um passo importante para compreender de forma prática a dinâmica da evasão. Essa técnica divide a série em componentes fundamentais, permitindo observar de maneira clara o que está acontecendo ao longo do tempo.

Primeiro, a análise da **tendência** mostrou se a evasão vinha crescendo, diminuindo ou permanecendo estável nos anos estudados. Por exemplo, um aumento contínuo ao longo dos semestres indicaria que o problema da evasão vem se agravando com o tempo, enquanto uma queda sugeriria avanços institucionais em retenção estudantil. Já a estabilidade mostraria que, mesmo com pequenas variações, não houve mudanças estruturais significativas no comportamento da evasão.

Além disso, a análise da **sazonalidade** buscou identificar padrões que se repetem em determinados períodos. Se fosse observado que as taxas de evasão aumentavam frequentemente em semestres iniciais do ano letivo, por exemplo, isso indicaria um padrão cíclico, possivelmente associado à adaptação de novos alunos. A ausência de sazonalidade, por outro lado, apontaria que as variações nas taxas não seguem um ciclo regular, mas sim podem estar ligadas a eventos pontuais ou a mudanças externas, como políticas públicas ou alterações socioeconômicas.

4.7.2 Função de Autocorrelação (ACF)

A função de autocorrelação (ACF) foi utilizada para verificar até que ponto os valores passados da série influenciam os valores futuros. Em termos práticos, ela mostra se existe dependência entre os semestres. Picos significativos em determinados *lags* (defasagens) indicaram que a evasão em um semestre estava relacionada com a de semestres anteriores.

Por exemplo, se a evasão de um semestre apresentasse forte correlação com a evasão do semestre imediatamente anterior, isso mostraria que choques ou variações não se dissipam rapidamente, mas se propagam ao longo do tempo. Já padrões de oscilação ou de decaimento lento sugeriram a presença de sazonalidade ou de não estacionariedade, ou seja, que a série não segue um comportamento estável ao longo do tempo. Em contrapartida, quando os valores da ACF ficaram próximos de zero, isso indicou ausência de dependência, sugerindo que a evasão naquele ponto do tempo não era explicada pelas observações passadas.

4.7.3 Função de Autocorrelação Parcial (PACF)

A análise da função de autocorrelação parcial (PACF) complementou a ACF ao identificar a influência direta de determinados *lags* na série, eliminando os efeitos indiretos de outras defasagens. Na prática, isso significa que ela mostra quais semestres passados têm um impacto realmente relevante na previsão, sem confundir o efeito acumulado de vários períodos.

Quando foram encontrados picos significativos em determinados *lags*, ficou evidente que esses pontos eram cruciais para o modelo prever corretamente a evasão. Já o rápido decaimento após um certo ponto indicou que poucos termos autorregressivos eram suficientes para explicar a série, o que confirmou a adequação do modelo ajustado automaticamente pelo método *auto-ARIMA*. Esse diagnóstico foi essencial para validar que o modelo não estava nem subajustado (perdendo informações importantes) nem superajustado (considerando ruídos como se fossem padrões).

4.7.4 Previsão da Taxa de Evasão Durante a Pandemia com o Modelo ARIMA

A previsão realizada pelo modelo *ARIMA* teve como objetivo estimar o comportamento esperado da evasão durante o período da pandemia (2020–2022), utilizando como base os dados coletados e indexados temporalmente observados entre 2013 e 2019.

4.7.4.1 *Interpretação dos Resultados*

Após o ajuste do modelo, os valores previstos para a evasão foram comparados com os valores efetivamente observados durante a pandemia. Essa comparação permitiu identificar possíveis impactos do contexto pandêmico na evasão universitária:

- **Valores observados acima do previsto:** indicam que a pandemia intensificou a evasão, possivelmente em decorrência de fatores como dificuldades acadêmicas relacionadas ao ensino remoto, problemas econômicos enfrentados pelas famílias e questões sociais, como a necessidade de conciliar estudos com responsabilidades domésticas.
- **Valores observados abaixo do previsto:** sugerem que, mesmo em meio à pandemia, a evasão apresentou redução, o que pode estar associado a políticas institucionais de apoio e retenção, como flexibilização de trancamentos, programas de assistência estudantil ou adaptação de metodologias pedagógicas.

Essa análise prática torna possível entender se os efeitos da pandemia representaram um desvio em relação ao que seria esperado apenas pela tendência histórica da evasão.

4.7.4.2 *Intervalos de Predição*

Além das previsões pontuais, foram construídos intervalos de predição com nível de confiança de 95%. Esses intervalos fornecem uma faixa dentro da qual os valores futuros teriam maior probabilidade de se situar, caso a série temporal seguisse o comportamento histórico. Assim:

- **Valores observados fora dos intervalos:** indicam que houve desvios significativos, sugerindo que fatores externos (como a pandemia) impactaram diretamente a evasão.
- **Valores observados dentro dos intervalos:** sugerem que, apesar do contexto atípico, a evasão manteve-se dentro do padrão esperado, o que pode indicar resiliência da instituição ou eficácia de medidas mitigadoras adotadas no período.

4.7.4.3 *Validação do Modelo*

A validação do modelo *ARIMA* foi realizada a partir da análise dos resíduos. Esperava-se que os resíduos se comportassem como ruído branco, isto é:

- fossem aleatórios e sem padrão aparente;
- não apresentassem autocorrelação significativa;

- exibissem distribuição aproximadamente normal, com média próxima de zero.

Quando essas condições foram atendidas, concluiu-se que o modelo conseguiu capturar adequadamente os componentes estruturais da série (tendência e possíveis padrões de sazonalidade), restando apenas flutuações aleatórias. Dessa forma, o *ARIMA* mostrou-se apropriado para a tarefa de previsão da evasão durante a pandemia.

4.8 Arquitetura e Desenvolvimento da Ferramenta

O sistema foi desenvolvido seguindo uma arquitetura cliente-servidor, estruturada de forma modular para garantir a separação de responsabilidades onde o

- **Frontend:** Representa a camada de apresentação ou interface, sendo responsável pela interação direta com o usuário. No escopo deste trabalho, foi desenvolvido utilizando a biblioteca *React*⁶ e o framework *Next.js*⁷ para assegurar uma interface dinâmica, responsiva e de alta performance. Suas principais funções incluem a captura do arquivo CSV enviado pelo usuário e a renderização intuitiva dos gráficos de previsão.
- **Backend:** Constitui a camada de processamento de dados e lógica de negócio que opera no servidor. Utiliza o micro-framework *Flask*⁸ (Python 3.12)⁹, selecionado pela sua robustez no ecossistema de ciência de dados. É encarregado de receber as requisições, realizar o tratamento e limpeza dos dados e executar o modelo preditivo *Auto-ARIMA*.
- **API RESTful:** Define o conjunto de regras e restrições que viabiliza a comunicação padronizada entre o *frontend* e o *backend*. Baseada no estilo arquitetural *REST (Representational State Transfer)*, permite que o *frontend* encaminhe dados de evasão via requisições HTTP, enquanto o servidor responde com os resultados analíticos estruturados em formato JSON.

4.8.1 Visão Geral da Tecnologia

O quadro 2 resume o *stack* tecnológico empregado em ambas as frentes de desenvolvimento.

⁶ Documentação oficial do React disponível em: <https://react.dev/reference/react>

⁷ Documentação oficial do Next.js disponível em: <https://nextjs.org/docs>

⁸ Documentação oficial do Flask disponível em: <https://flask.palletsprojects.com/en/stable/>

⁹ Documentação oficial do Python 3.12 disponível em: <https://docs.python.org/3.12/>

Quadro 2 – Tecnologias e bibliotecas utilizadas no projeto.

Componente	Tecnologia	Finalidade
Backend	Flask (Python 3.12)	Servidor e API REST
Frontend	Next.js 15 (React 19)	Interface do Usuário
Dados	SQLAlchemy / SQLite ¹	Persistência e ORM
Análise	sktime ² / statsmodels ³	Modelagem Estatística
Visualização	Plotly.js ⁴ / Recharts ⁵	Gráficos Interativos

Fonte: Elaborado pelo autor.

4.8.2 Prototipação e Design de Experiência (UI/UX)

O desenvolvimento iniciou-se com a fase de *design* de interface utilizando a ferramenta *Figma*¹⁰. Nesta etapa, elaboraram-se protótipos de alta fidelidade para mapear a jornada do usuário, garantindo uma interface intuitiva para gestores acadêmicos. O foco residiu na criação de um fluxo de trabalho simplificado: ingestão de dados brutos, parametrização da análise e visualização de indicadores. As Figuras 4, 5 e 6 demonstram a evolução do protótipo para a versão final implementada.

¹ Documentação oficial do SQLite disponível em: <https://www.sqlite.org/docs.html>

² Documentação oficial do sktime disponível em: <https://www.sktime.net/en/stable/>

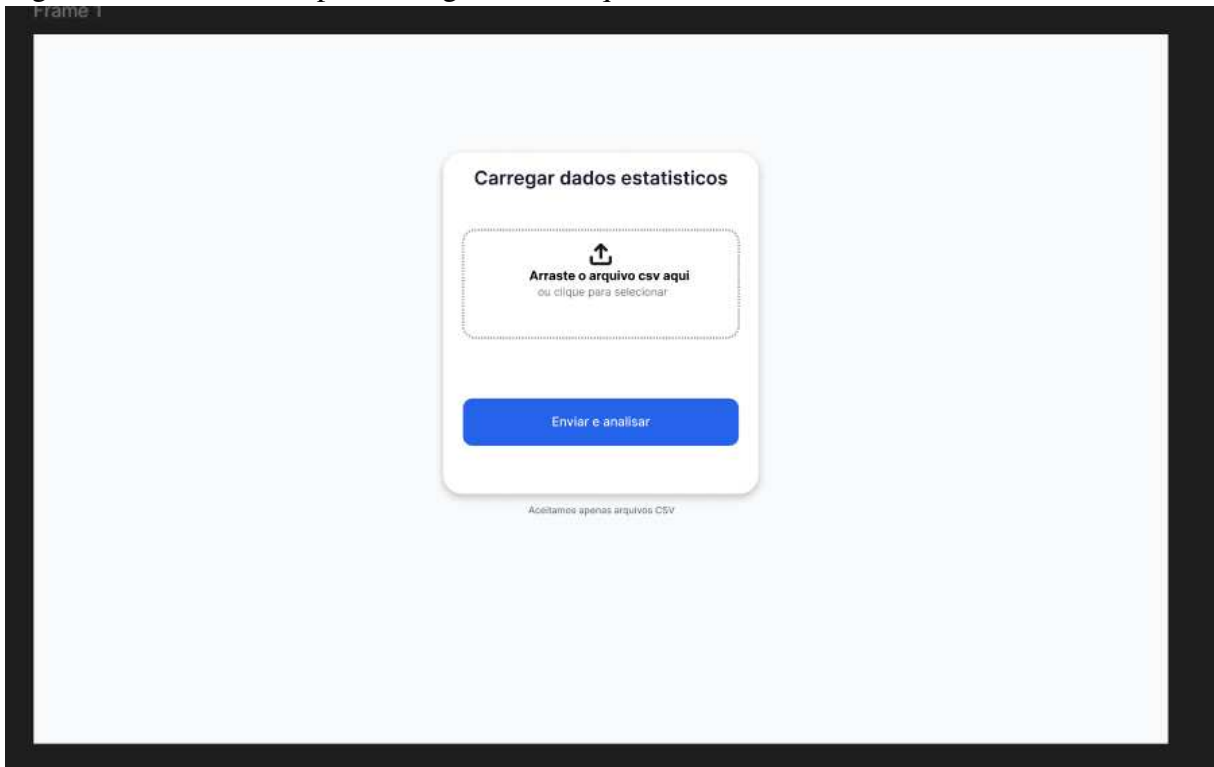
³ Documentação oficial do statsmodels disponível em: <https://www.statsmodels.org/stable/>

⁴ Documentação oficial do Plotly disponível em: <https://plotly.com/javascript/>

⁵ Documentação oficial do Recharts disponível em: <https://recharts.org/en-US/>

¹⁰ Documentação oficial do Figma disponível em: <https://help.figma.com/>

Figura 4 – Interface de upload e ingestão de arquivos CSV.



Fonte: Elaborada pelo autor.

Figura 5 – Tela de ajuste dos dados da séries temporais.

Ajustes da Análise
ID do Arquivo: d0a1a16d-8801-4d59-8910-6e824637d404

1. Selecione as Unidades

- CC
- CCA
- CH
- CRATEÚS
- CT
- FACED
- FADIR

2. Defina o Período

Semestre de Corte

Selecione um semestre... ▾

(?) Dados até este semestre serão usados para treino.

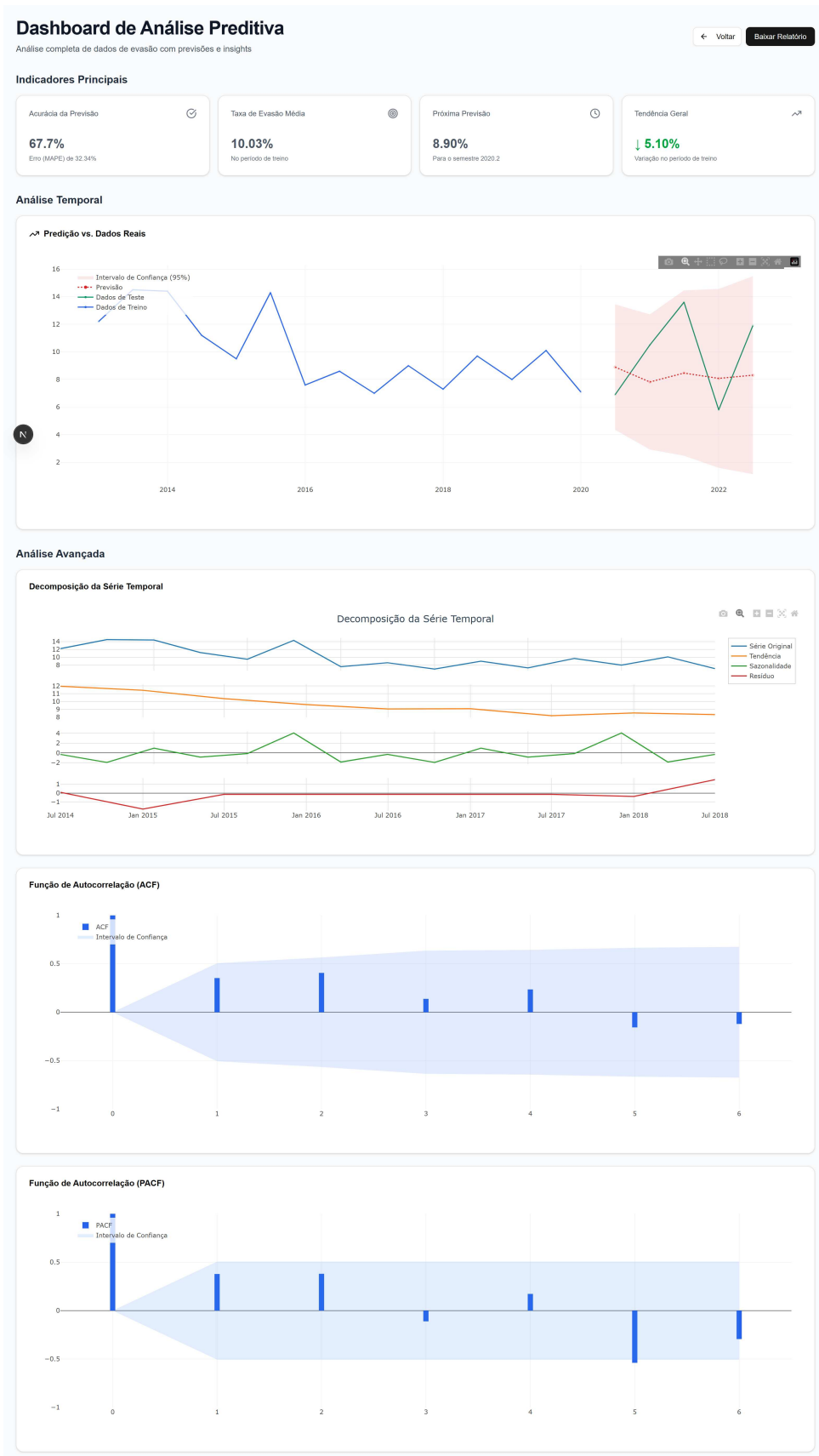
Gerar Análise →

Pré-visualização e Edição dos Dados

semestre	unidade_academica	taxa_evasao	Tx. Retenção (Prazo Padrão)	Tx. Retenção II (Prazo Máximo)	Matriculas	Ações
2013-01-01	CC	9.5	12.5%	4.9%	2633	🗑️
2013-01-01	CCA	7.8	13.5%	2.8%	1767	🗑️
2013-01-01	CH	8.6	9.9%	2.5%	2638	🗑️
2013-01-01	CT	7.7	9.5%	2.8%	3190	🗑️
2013-01-01	FACED	7.5	9.6%	1.6%	748	🗑️
2013-01-01	FADIR	2.8	2.0%	0.3%	962	🗑️

Fonte: Elaborada pelo autor.

Figura 6 – Tela final de resultados da ferramenta Oráculo.



Fonte: Elaborada pelo autor.

4.8.3 Arquitetura Desacoplada

Adotou-se uma separação rigorosa entre o *backend* e o *frontend*. O servidor de aplicação atua exclusivamente como um provedor de recursos via API RESTful, enquanto o cliente consome esses serviços de forma assíncrona. Essa abordagem promove a manutenibilidade, a escalabilidade e a independência de implantação entre as partes da aplicação.

4.8.4 Implementação do Backend

A organização do código-fonte segue o padrão de arquitetura em camadas (Routes → Services → Models), conforme ilustrado na estrutura de diretórios abaixo:

```

backend/
├── app/
│   ├── __init__.py ..... Factory function da aplicação Flask
│   ├── models/ ..... Modelos de banco de dados (SQLAlchemy)
│   ├── routes/ ..... Blueprints das rotas da API
│   ├── services/ ..... Lógica de negócio
│   ├── utils/ ..... Utilitários e validadores
│   └── exceptions.py ..... Tratamento de erros customizados
├── config/ ..... Configurações da aplicação
├── database/ ..... Cliente de banco de dados
└── requirements.txt ..... Dependências Python

```

4.8.4.1 Processamento de Dados e Ingestão

O mecanismo de processamento de arquivos CSV, implementado nas classes *DataService* e *TursoDataService*, realiza o tratamento dos dados brutos através das seguintes etapas:

- **Detecção de Codificação:** Implementou-se um algoritmo de busca iterativa para identificar o *encoding* correto (UTF-8, Latin-1, etc.) e o delimitador utilizado no arquivo, recorrendo à biblioteca *pandas*¹¹ como *fallback*.
- **Validação de Integridade:** O sistema verifica a presença das colunas obrigatórias (*Unidade*, *Taxa de Evasão*, *Semestre*), remove registros inconsistentes e filtra unidades identificadas como temporárias.
- **Normalização:** Taxas percentuais são convertidas para o formato decimal e os identifica-

¹¹ Documentação oficial do *pandas* disponível em: <https://pandas.pydata.org/docs/>

dores de semestre são padronizados para garantir a continuidade da série temporal.

- **Persistência Otimizada:** As inserções no banco de dados ocorrem em lotes de 200 registros, suportando tanto o SQLite local quanto o Turso.tech para ambientes distribuídos.

4.8.4.2 Modelo de Dados

A persistência de dados fundamenta-se em três modelos principais:

1. **FileUpload:** Responsável pelos metadados do arquivo e pelo rastreamento do status de processamento.
2. **DataRow:** Armazena as linhas normalizadas, incluindo campos temporais e índices otimizados para consultas por unidade.
3. **Analysis:** Persiste os resultados das previsões e decomposições em formato JSON, permitindo a recuperação rápida de análises anteriores.

4.8.5 Serviço de Análise Preditiva

O motor de análise, centralizado no *AnalysisService*, executa um fluxo de trabalho estatístico rigoroso para garantir a confiabilidade das previsões.

4.8.5.1 Decomposição e Modelagem ARIMA

A análise inicia-se com a decomposição sazonal aditiva, considerando um ciclo de dois semestres. Posteriormente, utiliza-se o algoritmo *auto-ARIMA* da biblioteca *sktime* com as seguintes restrições paramétricas:

- Parâmetros não sazonais: $max_p = 3, max_q = 3, max_d = 2$;
- Parâmetros sazonais: $max_P = 2, max_Q = 2, max_D = 1$;
- Frequência sazonal: $sp = 2$.

O sistema gera previsões para um horizonte de quatro semestres, acompanhadas de intervalos de confiança de 95%. A performance é mensurada através da métrica MAPE (*Mean Absolute Percentage Error*).

4.8.5.2 Estratégia de Validação

Para validar os modelos, o sistema permite dois modos de operação:

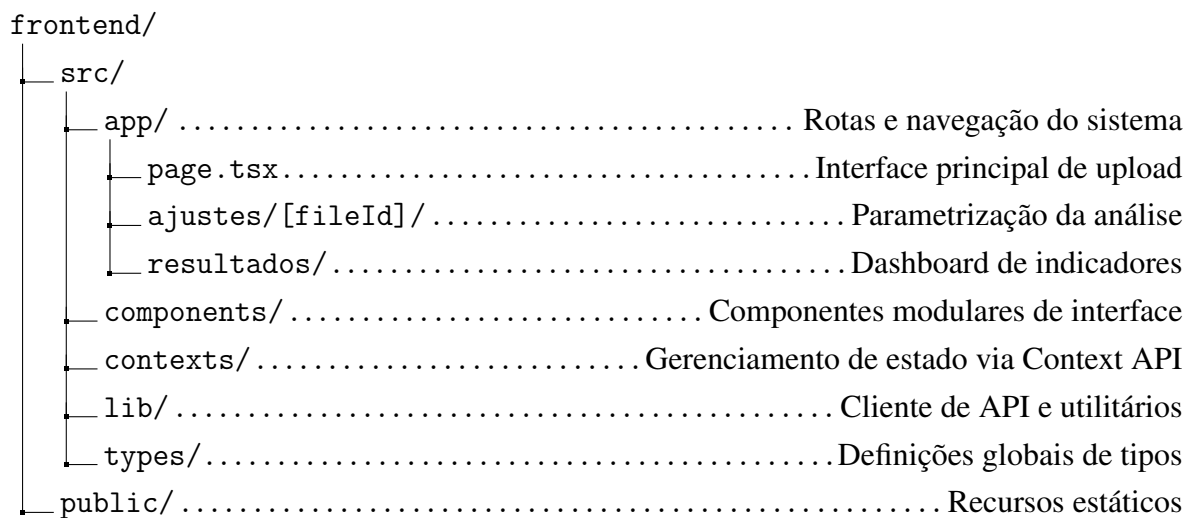
- **Validação por Semestre de Corte:** Os dados são particionados em conjuntos de treino e

teste, permitindo comparar os valores previstos com os reais observados após a data de corte.

- **Predição de Horizonte Futuro:** Utiliza-se a base histórica completa para projetar cenários futuros, auxiliando na tomada de decisão institucional.

4.8.6 Implementação do Frontend

A interface do usuário foi desenvolvida utilizando o *framework* Next.js 15, fundamentado na biblioteca React 19. Optou-se pelo uso de TypeScript para garantir a segurança de tipos e reduzir erros em tempo de execução. A arquitetura segue o padrão *App Router*, conforme detalhado na estrutura de diretórios a seguir:



4.8.6.1 Fluxo Lógico e Interfaces

O sistema foi projetado para guiar o usuário através de um fluxo lógico de três etapas principais:

1. **Módulo de Ingestão (Root):** Responsável pelo *upload* de arquivos CSV via interface *drag-and-drop*. Implementou-se uma validação rigorosa no lado do cliente utilizando esquemas da biblioteca *Zod*¹², limitando o tamanho dos arquivos a 16MB e assegurando a integridade do formato antes do envio ao servidor.
2. **Configuração e Parametrização:** Localizado na rota dinâmica `/ajustes/[fileId]`, este módulo permite ao pesquisador selecionar múltiplas unidades acadêmicas e definir o semestre de corte para a validação do modelo. A interface dispõe de uma tabela paginada para pré-visualização dos dados, permitindo a exclusão de registros e a edição de valores

¹² Documentação oficial do Zod disponível em: <https://zod.dev/>

inconsistentes antes do processamento estatístico.

3. **Dashboard de Resultados:** A visualização final consolida os indicadores de desempenho e as predições. O painel exibe KPIs fundamentais, como a acurácia baseada no MAPE e a tendência geral de evasão. Para o componente de visualização de dados, utilizou-se a biblioteca *Plotly.js*, permitindo interações dinâmicas como *zoom* e filtros sobre o gráfico de predição e intervalos de confiança.

4.8.6.2 Gerenciamento de Estado e Persistência Local

Para a gestão do estado global da aplicação, utilizou-se a *React Context API* através do *AnalysisContext*. Este mecanismo é responsável por orquestrar os resultados provenientes da API e garantir a persistência dos dados no *localStorage* do navegador. Essa abordagem previne a perda de informações durante a navegação entre as páginas de parametrização e resultados, além de permitir o carregamento assíncrono de análises previamente realizadas.

4.8.6.3 Análise Avançada e Visualização Temporal

A camada de visualização avançada decompõe a série temporal em seus elementos constituintes (tendência, sazonalidade e resíduos) através de componentes especializados (*DecompositionChart*). Adicionalmente, o sistema renderiza as funções de autocorrelação (ACF e PACF), fundamentais para a interpretação técnica do comportamento da evasão por parte de especialistas em estatística. Toda a interface foi concebida sob os princípios de *design* responsivo, assegurando a usabilidade em diferentes dispositivos e resoluções.

4.8.7 Tecnologias e Bibliotecas Utilizadas

Para a implementação da ferramenta Oráculo, selecionaram-se tecnologias consagradas no ecossistema de Desenvolvimento Web e Ciência de Dados, priorizando a interoperabilidade e o desempenho acadêmico (Sommerville, 2011; Hyndman; Athanasopoulos, 2018). A seguir, detalham-se os recursos utilizados em cada camada do sistema.

4.8.7.1 Tecnologias de Backend

O *backend* foi estruturado para suportar processamento estatístico intensivo e persistência resiliente de dados (Pressman; Maxim, 2020; Shmueli *et al.*, 2017). As bibliotecas e suas

respectivas finalidades estão consolidadas no 3.

Quadro 3 – Ecossistema de Tecnologias do Backend.

Categoria	Tecnologia	Finalidade e Versão
Framework Core	Flask 3.0.3	Micro-framework para criação de rotas e API RESTful.
Banco de Dados	SQLAlchemy 2.0.35	ORM para abstração de banco de dados e integração com Turso.tech.
Análise de Dados	pandas 2.2.2	Manipulação e limpeza de estruturas de dados (DataFrames).
Estatística	statsmodels 0.14.2	Execução de decomposição sazonal e cálculos de autocorrelação.
Predição	sktime 0.32.1	Framework para modelagem de séries temporais e auto-ARIMA.
Validação	pydantic 2.9.2	Garantia de integridade dos dados e tipos no servidor.
Produção	gunicorn 21.2.0	Servidor HTTP WSGI para ambiente de implantação real.

Fonte: Elaborado pelo autor.

4.8.7.2 Tecnologias de Frontend

A interface do usuário foi projetada sob os princípios de reatividade e acessibilidade (Nielsen, 1994), utilizando a *stack* descrita no Quadro 4.

Quadro 4 – Ecossistema de Tecnologias do Frontend.

Categoria	Tecnologia	Finalidade e Versão
Framework	Next.js 15.3.4	Gerenciamento de rotas, renderização híbrida e otimização.
Linguagem	TypeScript 5.8.3 ¹³	Tipagem estática para redução de erros em tempo de execução.
Estilização	Tailwind CSS 4.1.12 ¹⁴	Estilização baseada em utilitários para interface responsiva.
Visualização	Plotly.js 3.1.0	Geração de gráficos estatísticos interativos e dinâmicos.
Componentes	Radix UI ¹⁵	Primitivas de componentes acessíveis e customizáveis.
Validação	Zod 3.25.76	Validação de esquemas e formulários no lado do cliente.
Utilitários	html2pdf.js 0.10.3 ¹⁶	Funcionalidade de exportação de relatórios para formato PDF.

Fonte: Elaborado pelo autor.

4.8.8 Automação e Entrega Contínua (CI/CD)

A gestão do ciclo de vida da ferramenta Oráculo apoiou-se em práticas de Integração e Entrega Contínua (CI/CD). A utilização do *GitHub*¹⁷ como repositório central permitiu a orquestração de gatilhos automáticos: a cada *push* na ramificação principal, as plataformas *Vercel*¹⁸ (para o *frontend*) e *Render*¹⁹ (para o *backend*) realizam o *build* e a implantação imediata,

¹⁷ Documentação oficial do GitHub disponível em: <https://docs.github.com/>

¹⁸ Documentação oficial da Vercel disponível em: <https://vercel.com/docs>

¹⁹ Documentação oficial do Render disponível em: <https://docs.render.com/>

assegurando que a versão pública da ferramenta reflita o estado mais estável do código.

4.8.9 Custo-Eficiência e Acessibilidade

A viabilidade operacional do projeto foi assegurada pela seleção de infraestruturas que oferecem planos de uso gratuito para fins acadêmicos. Esta escolha estratégica elimina barreiras orçamentárias para a manutenção da ferramenta e garante que instituições ou estudantes com recursos limitados possam replicar a análise sem custos adicionais de servidor, promovendo a democratização do acesso a tecnologias de análise preditiva.

5 RESULTADOS

Este capítulo apresenta os resultados do trabalho, organizados na seguinte ordem: Quixadá, CT (Centro de Tecnologia), CH (Centro de Humanidades), FEAAC (Faculdade de Economia, Administração, Atuária e Contabilidade) e o consolidado da UFC.

5.1 Sazonalidade

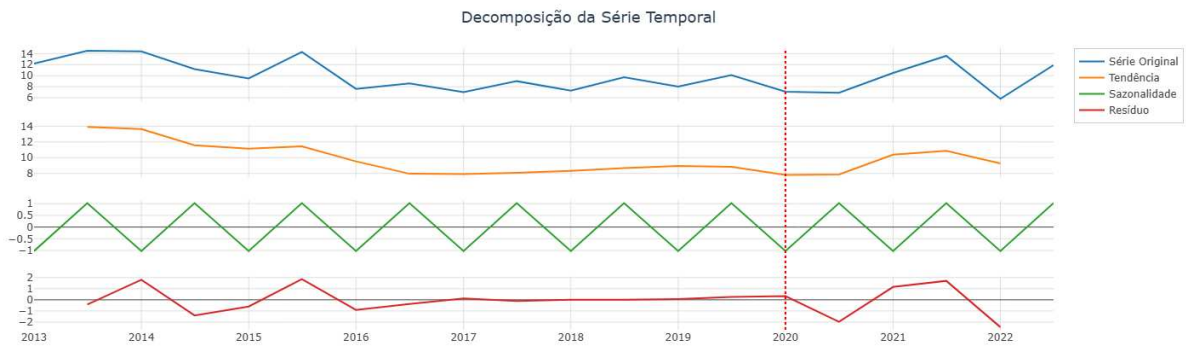
Durante a análise, observou-se que a sazonalidade exhibe comportamento uniforme em todas as unidades acadêmicas. Este padrão indica que a série temporal, de modo geral, possui um ciclo característico, com picos de evasão no segundo semestre de cada ano e reduções no início de cada semestre. Tal fenômeno é atribuído ao fato de que o ingresso de novos alunos supera o número de discentes que deixam a instituição, resultando em um gráfico de sazonalidade com formato triangular.

5.2 Modelagem Preditiva por Unidade e de toda UFC

5.2.1 Quixadá

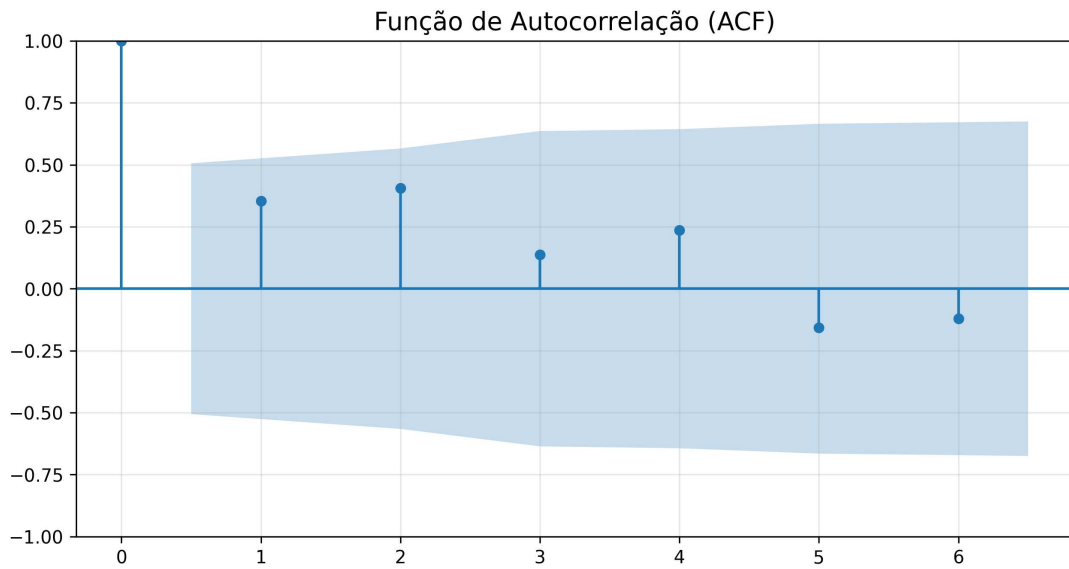
Conforme ilustrado no gráfico (Figura 7), a tendência no período pré-pandemia (à esquerda da linha vertical) demonstra uma queda constante, seguida por um leve aumento e uma nova redução antes da pandemia. Este comportamento indica uma tendência decrescente na série temporal durante esse intervalo. Contudo, após o início da pandemia de COVID-19, a linha de tendência começou a apresentar crescimento nos primeiros semestres, mas, ao final do período analisado, registrou novamente uma queda. A análise das Figuras 8 e 9 (ACF e PACF) não houve correlação significativa em nem um lag. Diante dessa estrutura, o algoritmo *auto-arima* convergiu para o modelo **ARIMA(1,1,2)**. A inclusão do parâmetro de diferenciação ($d = 1$) justifica-se pela instabilidade da tendência observada no campus. Conforme a Figura 10, as previsões acompanharam a tendência real, embora a volatilidade local tenha resultado em intervalos de confiança mais amplos.

Figura 7 – Gráfico de decomposição da série temporal do campus da UFC de Quixadá.



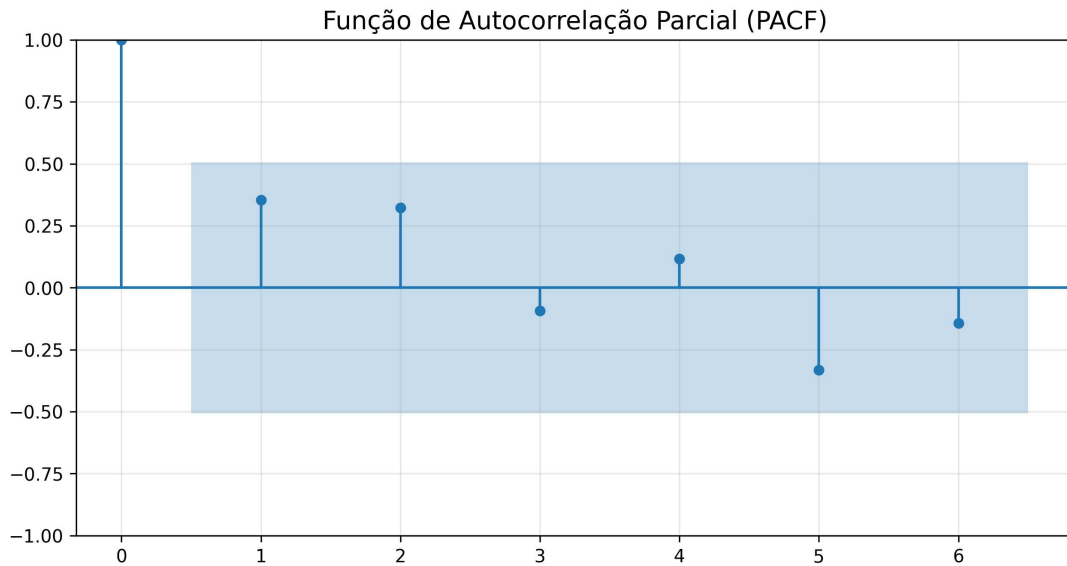
Fonte: Elaborado pelo autor.

Figura 8 – Gráfico de autocorrelação (ACF) da UFC de Quixadá.



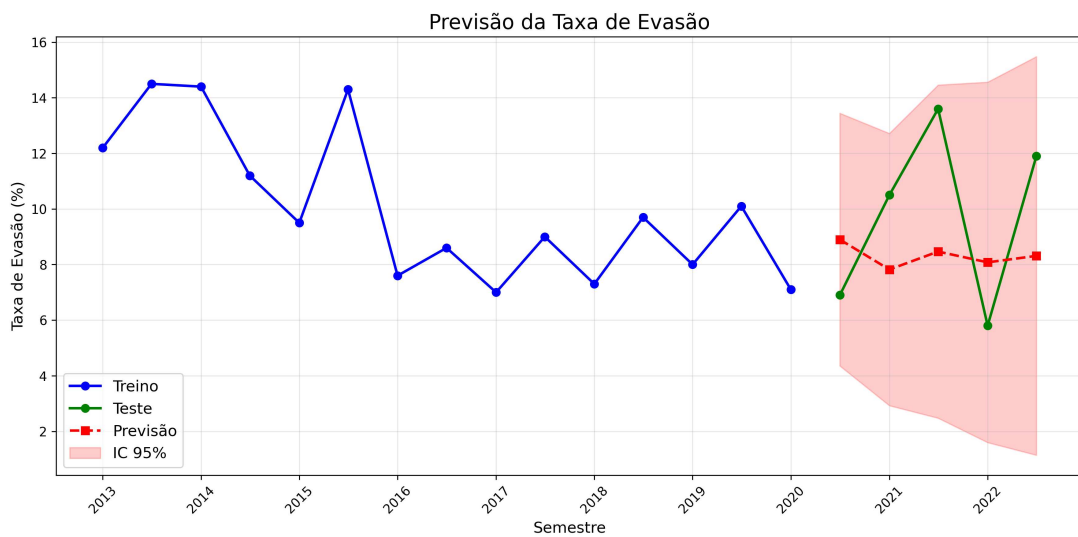
Fonte: Elaborado pelo autor.

Figura 9 – Gráfico de autocorrelação parcial (PACF) da UFC de Quixadá.



Fonte: Elaborado pelo autor.

Figura 10 – Gráfico de previsão da UFC de Quixadá.



Fonte: Elaborado pelo autor.

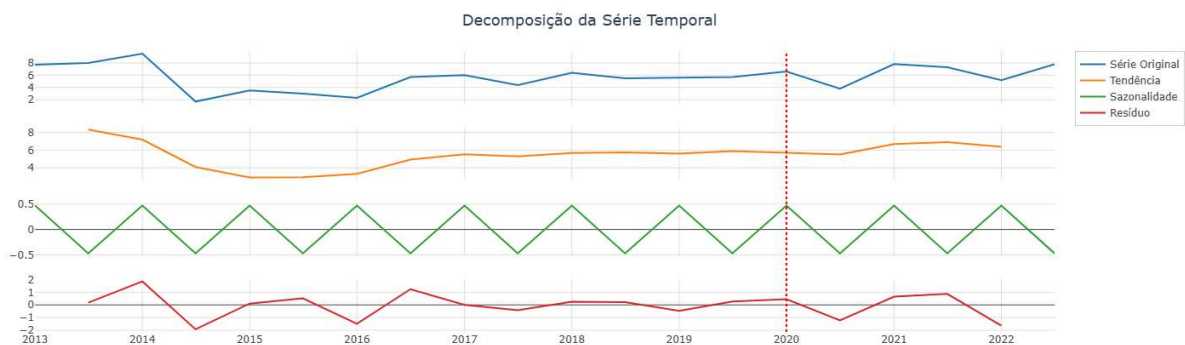
A Figura 10 apresentou a previsão da taxa de evasão utilizando o modelo ARIMA. A série temporal foi dividida em conjunto de treino e teste, em que os dados de treino abrangeram o período de 2013 a 2019, enquanto os dados de teste corresponderam aos anos de 2020 a 2022. A linha azul representou os valores observados no conjunto de treino, enquanto a linha verde indicou os valores reais do conjunto de teste. A linha vermelha demonstrou as previsões realizadas pelo modelo ARIMA, acompanhadas por uma faixa sombreada em rosa que indicou o intervalo de confiança de 95%.

Observou-se que as previsões seguiram a tendência geral dos dados reais, embora com algumas discrepâncias, principalmente nos extremos do período de teste. A amplitude do intervalo de confiança foi relativamente ampla, refletindo a incerteza inerente ao processo de previsão em séries temporais. De modo geral, o modelo ARIMA apresentou um desempenho satisfatório ao capturar a variação da taxa de evasão ao longo do tempo, demonstrando-se uma ferramenta útil para a análise e planejamento de políticas educacionais baseadas em dados históricos.

5.2.2 CT - Centro de Tecnologia

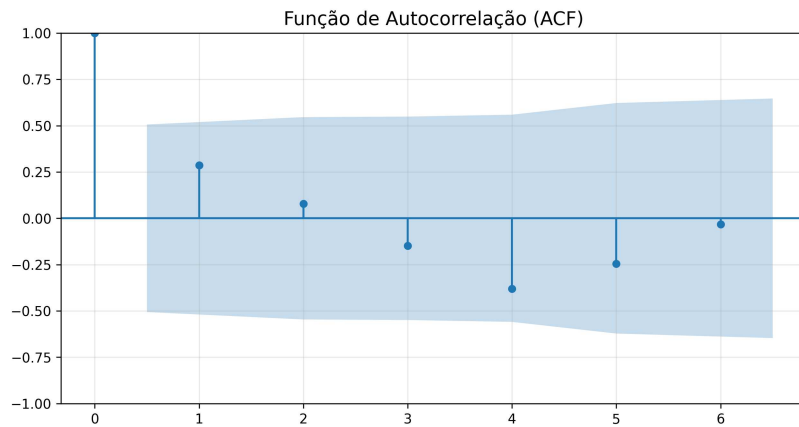
No período pré-pandemia, destacado no gráfico à esquerda da linha vertical vermelha, podemos perceber que a tendência apresenta uma queda acentuada. Após essa queda, nota-se um aumento que se manteve oscilante, com uma tendência de crescimento. Durante a pandemia, é possível observar uma leve queda inicial na tendência, seguida por um aumento mais expressivo, quase alcançando o valor de 7,5. Posteriormente, a tendência apresentou uma leve queda novamente, indicando possíveis ajustes ou variações decorrentes do impacto da pandemia. Para esta unidade, os gráficos de autocorrelação indicaram uma estrutura que levou o algoritmo *auto-arima* a selecionar o modelo **ARIMA(1,0,2)**. Este ajuste sugere que a série é estacionária em nível ($d = 0$). Conforme observado na Figura 14, as previsões mantiveram-se constantes, o que, estatisticamente, indica que o modelo identificou a média histórica como o melhor estimador diante da volatilidade residual do período pandêmico.

Figura 11 – Gráfico de decomposição da série temporal da unidade CT (Centro de Tecnologia)



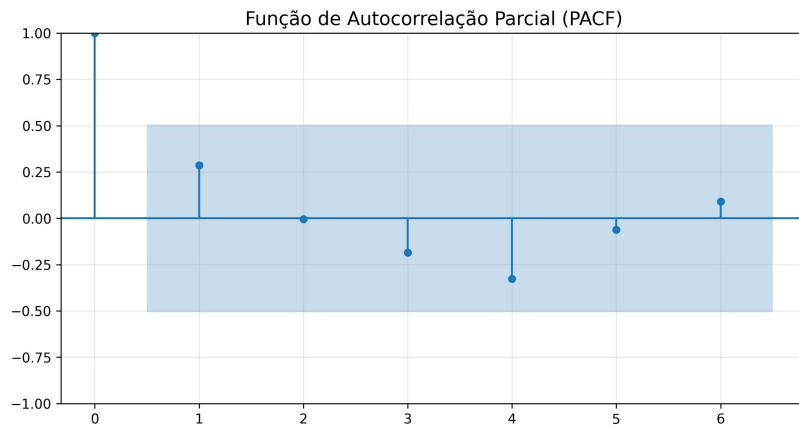
Fonte: Elaborada pelo autor.

Figura 12 – Gráficos de autocorrelação (ACF) da unidade CT



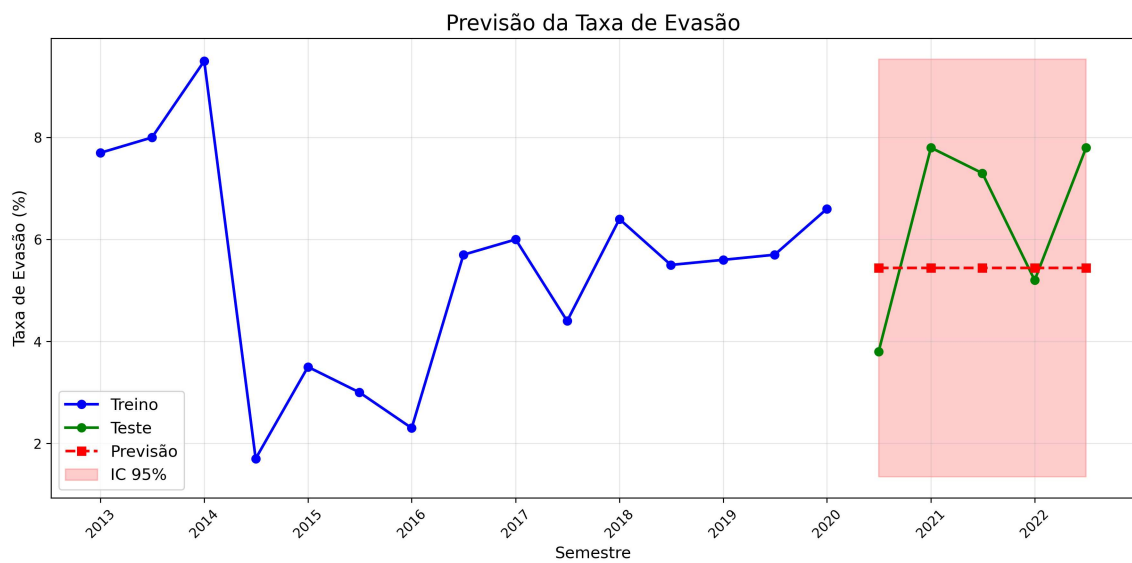
Fonte: Elaborada pelo autor.

Figura 13 – Gráficos de autocorrelação parcial (PACF) da unidade CT



Fonte: Elaborada pelo autor.

Figura 14 – Gráficos da previsão da unidade CT



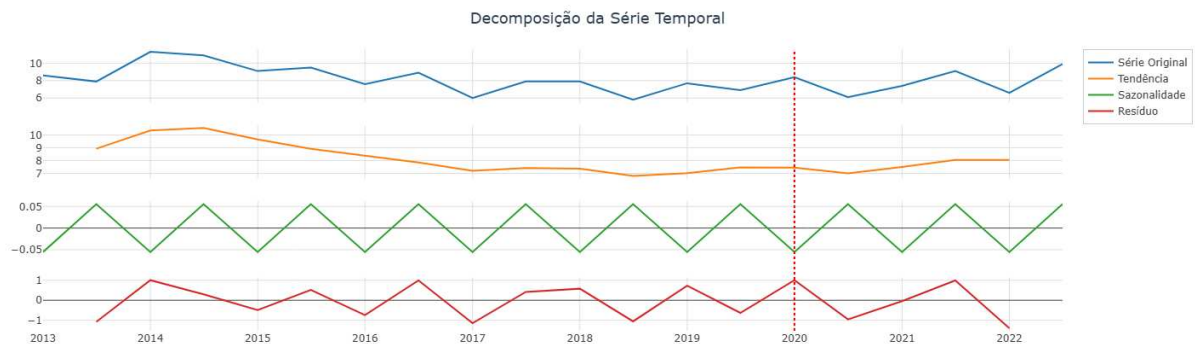
Fonte: Elaborada pelo autor.

5.2.3 CH - Centro de Humanidades

O gráfico revela que, no período pré-pandemia, representado à esquerda da linha vertical vermelha, a tendência apresenta um pico em 2014. Logo em seguida, há uma queda bastante acentuada. Entre 2017 e o segundo semestre de 2019, a tendência demonstra um comportamento de oscilação, sem mudanças significativas. No entanto, durante a pandemia de COVID-19, esse padrão de oscilação foi interrompido, e a tendência passou a apresentar um crescimento, indicando um impacto considerável desse período nos dados analisados.

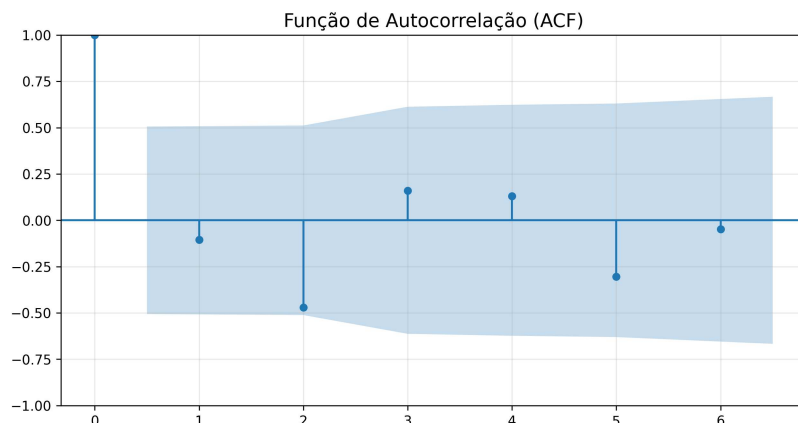
A modelagem estatística para o CH também convergiu para o **ARIMA(1,0,2)**. A escolha deste modelo reflete a tentativa do algoritmo de capturar a dinâmica de curto prazo (AR1) e suavizar ruídos aleatórios através dos componentes de média móvel (MA2). Assim como no CT, a estabilidade das previsões (Figura 18) evidencia que, apesar do crescimento recente, a série tende a retornar à sua média de longo prazo na ausência de novos choques estruturais.

Figura 15 – Gráfico de decomposição da série temporal da unidade CH (Centro de Humanidades) da UFC



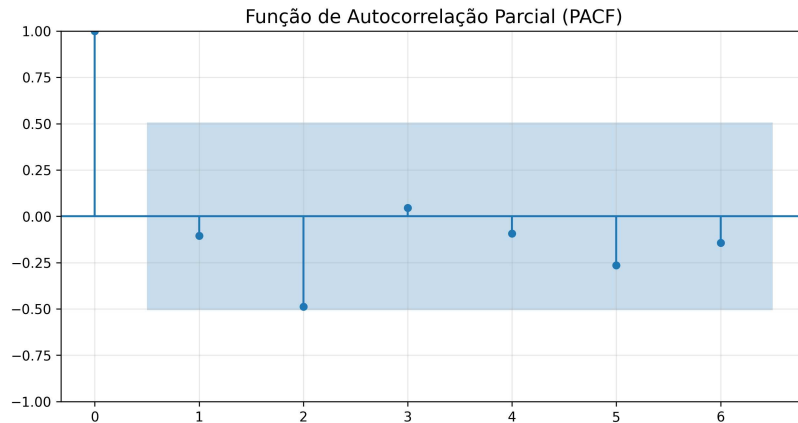
Fonte: Elaborada pelo autor.

Figura 16 – Gráficos de autocorrelação (ACF) da unidade CH



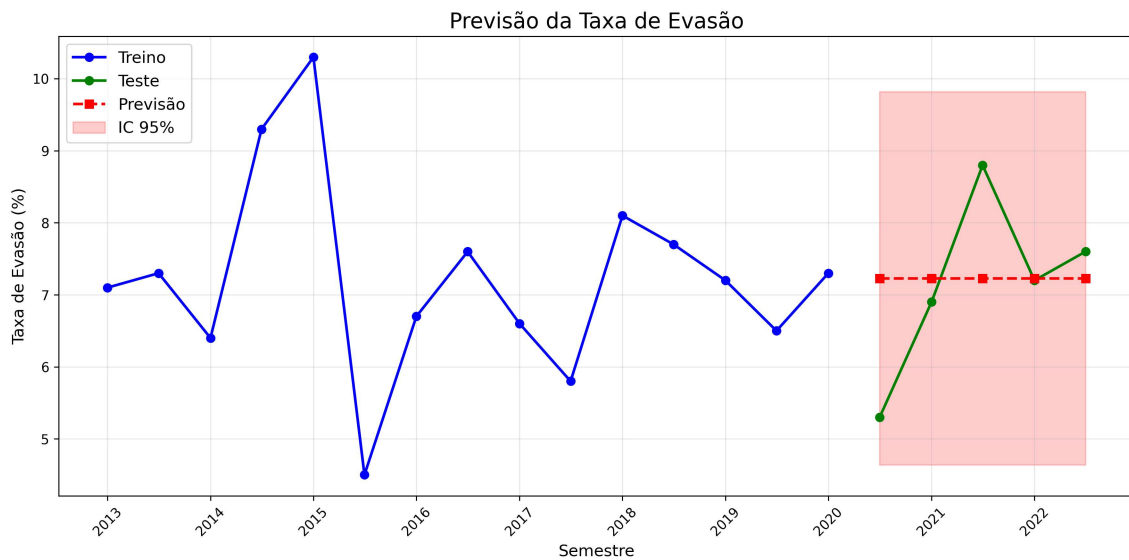
Fonte: Elaborada pelo autor.

Figura 17 – Gráficos de autocorrelação parcial (PACF) da unidade CH



Fonte: Elaborada pelo autor.

Figura 18 – Gráficos da previsão da unidade CH



Fonte: Elaborada pelo autor.

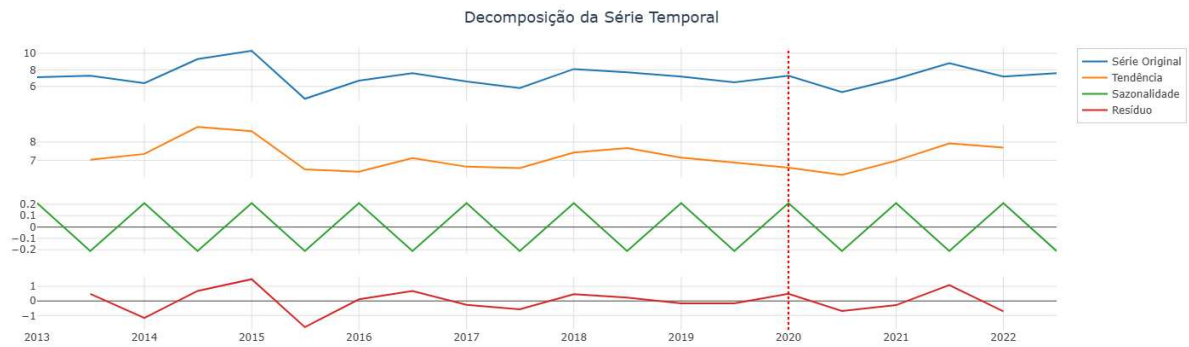
5.2.4 FEAAC - Faculdade de Economia, Administração, Atuária e Contabilidade

Observando a área do gráfico correspondente ao período pré-pandemia (à esquerda da linha vertical vermelha), nota-se que a tendência apresenta uma queda constante, com um leve aumento seguido de uma nova redução antes da pandemia. Isso indica uma tendência decrescente na série temporal nesse período. Contudo, durante a pandemia do COVID-19, vemos que a linha de tendência começou a crescer nos primeiros semestres, mas, ao final do período analisado, apresentou novamente uma queda.

O modelo identificado foi o **ARIMA(1,0,2)**. É importante destacar que a FEAAC apresentou o melhor desempenho preditivo do estudo (MAPE de 12,85%), o que indica que

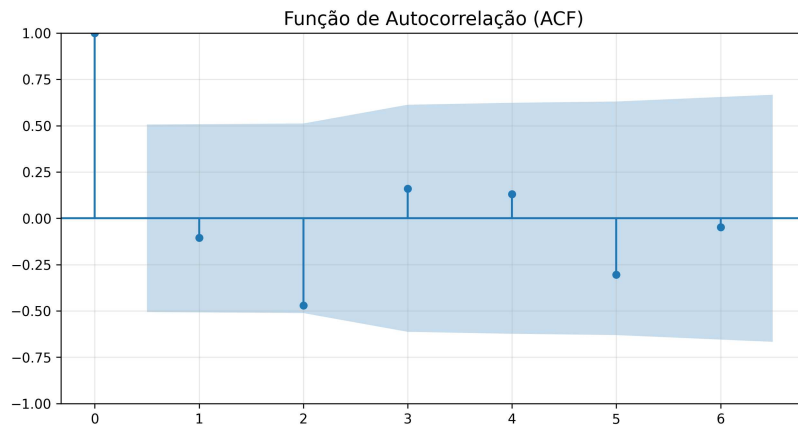
os parâmetros $(1, 0, 2)$ foram particularmente eficazes em mapear a realidade desta unidade. A baixa magnitude dos erros (Tabela 1) confirma que a estrutura autorregressiva de primeira ordem é suficiente para descrever a dinâmica de evasão deste centro acadêmico.

Figura 19 – Gráfico de decomposição da série temporal do campus FEAAC (Faculdade de Economia, Administração, Atuária e Contabilidade) da UFC



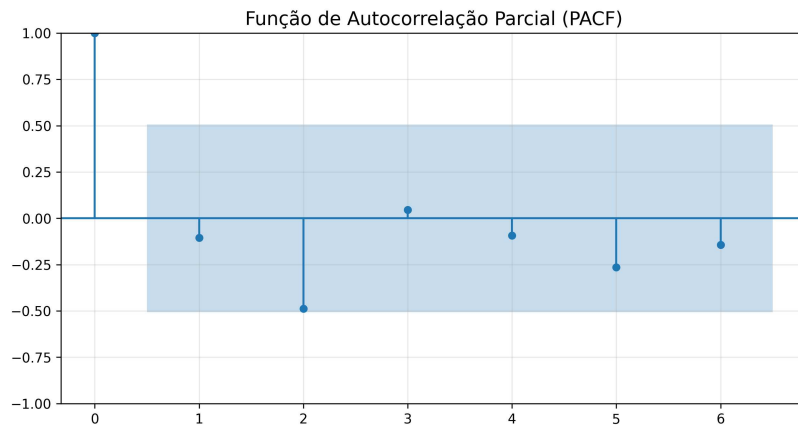
Fonte: Elaborada pelo autor.

Figura 20 – Gráficos de autocorrelação (ACF) da unidade FEAAC



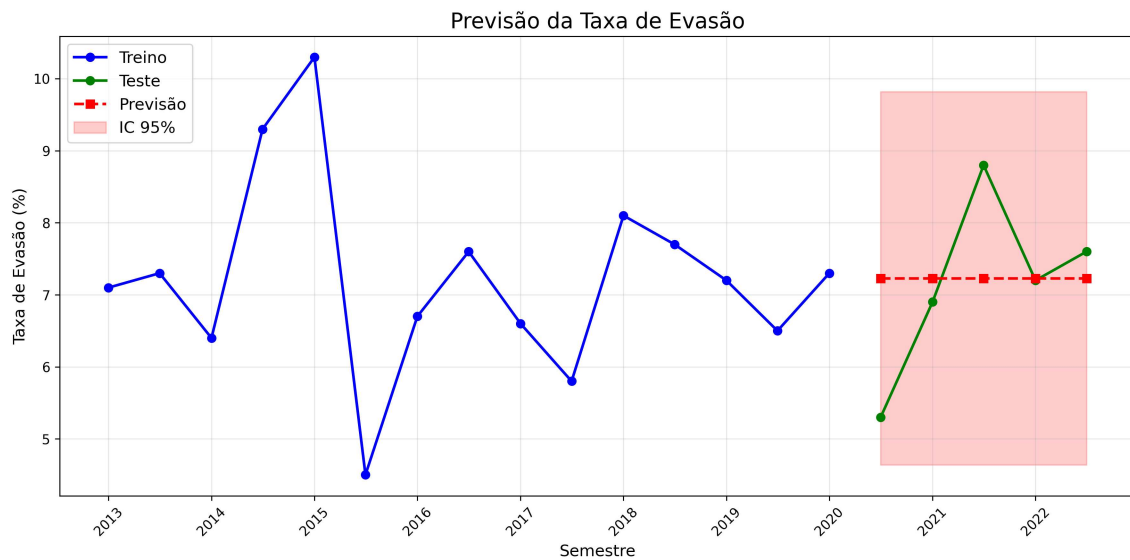
Fonte: Elaborada pelo autor.

Figura 21 – Gráficos de autocorrelação parcial (PACF) da unidade FEAAC



Fonte: Elaborada pelo autor.

Figura 22 – Gráficos da previsão da unidade FEAAC



Fonte: Elaborada pelo autor.

5.2.5 UFC

Observando a linha da tendência, vemos que, em toda a UFC, a tendência vinha em uma queda constante até 2017. Logo após essa data, a tendência apresentou uma leve crescente até a pandemia, representada no gráfico pela linha vertical vermelha. Durante a pandemia, após a linha vertical vermelha, observamos que a tendência apresentou um comportamento crescente constante.

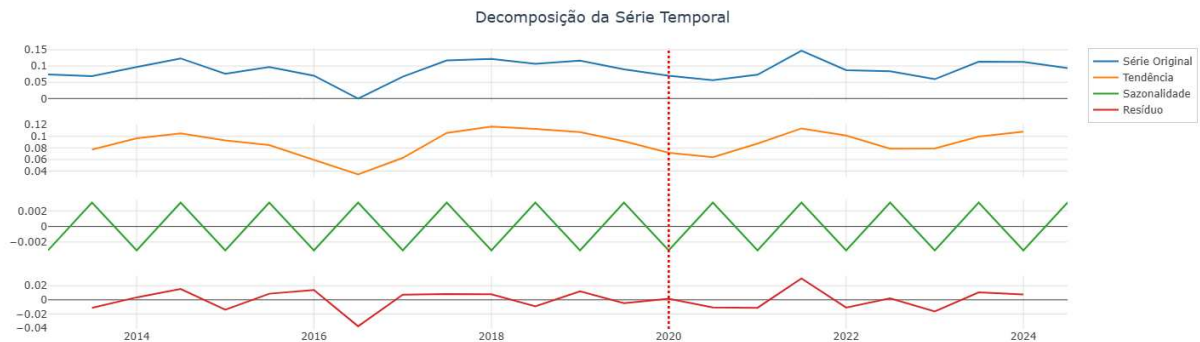
A Figura 23 apresenta o gráfico de decomposição da série temporal de toda a UFC, destacando a tendência, sazonalidade e o componente residual. Já as Figuras 24, 25 e 26 mostram, respectivamente, os gráficos de autocorrelação (ACF e PACF) e os resultados da previsão com o

modelo ARIMA.

Os gráficos ACF e PACF de toda a UFC indicam ausência de correlações estatisticamente significativas para a maioria dos lags, com exceção do lag 1, que apresentou correlação parcial significativa.

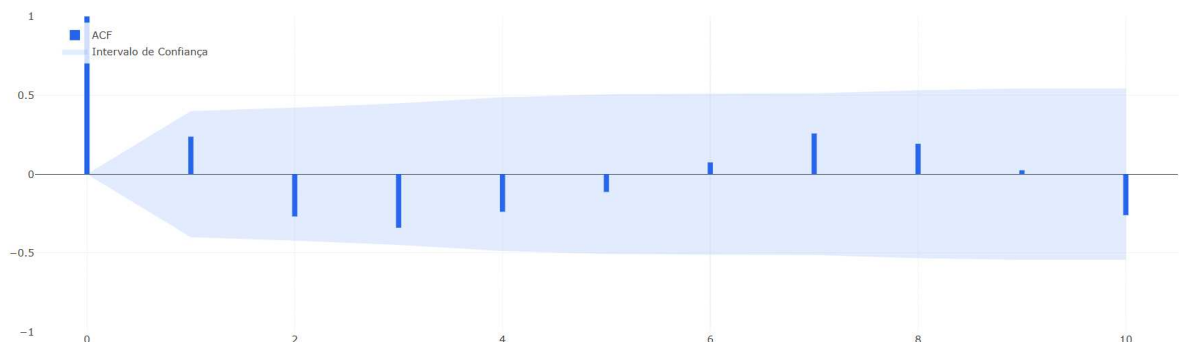
Para o consolidado da instituição, o modelo selecionado foi o **ARIMA(1,0,2)**. A análise da ACF e PACF para a UFC geral corroborou a predominância de correlação significativa no lag 1. Embora tenha havido divergências pontuais entre os valores reais e preditos durante a pandemia, a maioria das observações situou-se dentro do intervalo de confiança de 95% (Figura 26). Isso demonstra que o modelo, mesmo simplificado, possui robustez para servir como ferramenta de planejamento institucional, capturando a inércia do fenômeno da evasão em nível sistêmico.

Figura 23 – Gráfico de decomposição da série temporal de toda a UFC



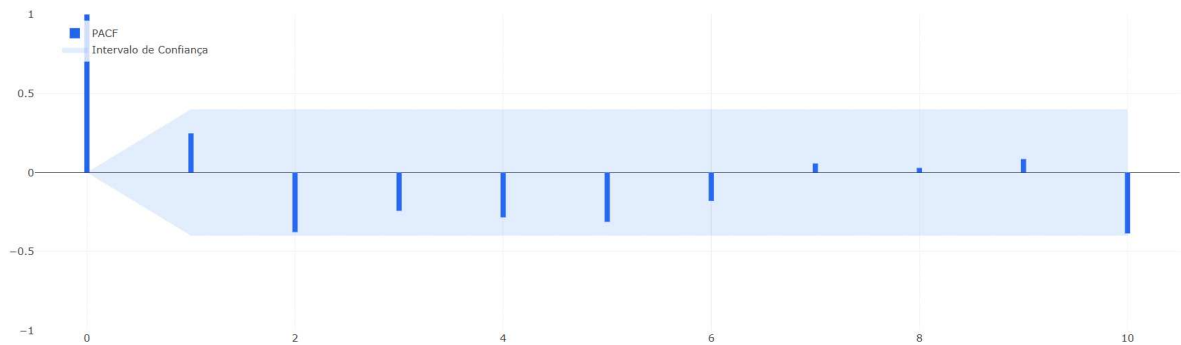
Fonte: Elaborada pelo autor.

Figura 24 – Gráficos de autocorrelação (ACF) de toda UFC



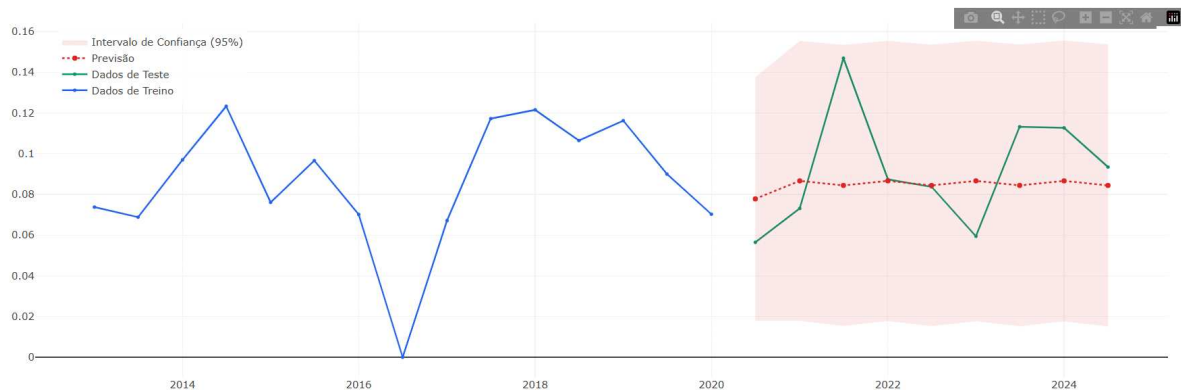
Fonte: Elaborada pelo autor.

Figura 25 – Gráficos de autocorrelação parcial (PACF) de toda UFC



Fonte: Elaborada pelo autor.

Figura 26 – Gráficos da previsão de toda UFC



Fonte: Elaborada pelo autor.

5.2.6 Comportamento das Previsões Constantes no Modelo ARIMA

Durante a etapa de modelagem, observou-se que, em alguns cenários, as previsões geradas pelo modelo *ARIMA* como o algoritmo *auto-ARIMA* apresentaram valores constantes ao longo do horizonte de previsão. Esse comportamento pode ser explicado pelo tipo de modelo ajustado automaticamente e pelas características estatísticas da série temporal analisada.

Em casos em que o *auto-ARIMA* identifica pouca variabilidade nos dados, quando a série apresenta valores quase constantes ou sem tendência aparente, o algoritmo tende a selecionar modelos extremamente simples, como o *ARIMA*(0,1,0) ou o *ARIMA*(0,0,0).

O modelo *ARIMA*(0,1,0) representa um *passaio aleatório* (*random walk*), em que a melhor previsão para o próximo valor é simplesmente o último valor observado. Já o modelo *ARIMA*(0,0,0) corresponde a uma *média constante*, em que todas as previsões são iguais à média dos dados históricos. Em ambos os casos, a ausência de componentes autorregressivos (AR) e de médias móveis (MA) faz com que o modelo não capture variações dinâmicas, resultando em

previsões constantes e intervalos de confiança invariáveis.

Esse comportamento foi observado nos casos da FEAAC, CH e CT, esses conjuntos de previsões analisados, nos quais tanto as médias previstas quanto os limites inferior e superior do intervalo de confiança permaneceram idênticos ao longo dos períodos futuros. Tais resultados indicam que o modelo não identificou padrões temporais significativos ou relações de dependência serial relevantes nos dados.

5.3 Avaliação Quantitativa do Desempenho Preditivo

Esta seção apresenta a avaliação quantitativa do desempenho do modelo ARIMA na previsão das taxas de evasão para cada unidade acadêmica e para a Universidade Federal do Ceará (UFC) como um todo. A acurácia das previsões foi mensurada utilizando o Erro Absoluto Médio (MAE), o Erro Percentual Absoluto Médio (MAPE) e a Raiz do Erro Quadrático Médio (RMSE). O MAE fornece a média da magnitude dos erros em pontos percentuais da taxa de evasão, sendo uma métrica de fácil interpretação. O RMSE, por sua vez, penaliza erros maiores de forma mais acentuada, indicando a presença de desvios mais significativos. O MAPE expressa o erro em termos percentuais relativos ao valor real, sendo útil para comparações entre séries com diferentes escalas.

A Tabela 1 resume os resultados obtidos:

Tabela 1 – Relatório comparativo de desempenho do modelo ARIMA por unidade acadêmica

Unidade	MAPE (%)	MAE	RMSE
Quixadá	32,34	3,14	3,33
CT	26,75	1,69	1,86
CH	19,76	1,44	1,53
FEAAC	12,85	0,85	1,13
Geral	25,25	1,45	1,73

Fonte: Elaborado pelo autor.

5.3.1 Análise por Unidade Acadêmica

Ao analisar os resultados por unidade acadêmica, observa-se que o modelo ARIMA apresentou desempenhos variados, mas com padrões consistentes.

A **FEAAC** (Faculdade de Economia, Administração, Atuária e Contabilidade) se destacou com o melhor desempenho preditivo entre as unidades individuais. Apresentou o menor MAPE de 12,85%, um MAE de 0,85 e um RMSE de 1,13. Isso indica que, em média, as

previsões para a taxa de evasão da FEAAC desviaram-se pouco dos valores reais, demonstrando uma alta precisão do modelo para esta unidade. A pequena diferença entre o MAE e o RMSE sugere que não houve uma ocorrência significativa de erros de grande magnitude.

O **CH** (Centro de Humanidades) também demonstrou uma boa performance do modelo, registrando um MAPE de 19,76%, MAE de 1,44 e RMSE de 1,53. Esses valores são ligeiramente superiores aos da FEAAC, mas ainda indicam uma capacidade preditiva robusta, com desvios médios controlados e uma boa consistência ao modelo.

O **CT** (Centro de Tecnologia) apresentou resultados intermediários, com MAPE de 26,75%, MAE de 1,69 e RMSE de 1,86, enquanto o **Campus de Quixadá** registrou os maiores erros entre as unidades individuais: MAPE de 32,34%, MAE de 3,14 e RMSE de 3,33. O valor elevado do MAE em Quixadá indica que, em média, as previsões se afastaram significativamente dos valores observados, possivelmente devido a maior volatilidade da série, comportamento irregular ao longo do tempo ou presença de eventos atípicos que dificultam o ajuste do modelo.

Em todas as unidades analisadas individualmente, a proximidade entre os valores de MAE e RMSE sugere que o modelo não gerou um grande número de valores discrepantes de previsão, ou seja, erros pontuais de valores muito altos não foram predominantes.

5.3.2 *Análise Geral (Agregada)*

Para a série temporal agregada da UFC, o modelo apresentou MAPE de 25,25%, MAE de 1,45 e RMSE de 1,73. Diferentemente do que pode ocorrer em séries agregadas com valores absolutos, aqui os erros permanecem compatíveis com a mesma ordem de grandeza das unidades individuais, indicando que a métrica empregada (taxa de evasão) mantém consistência entre os níveis agregado e desagregado.

Ainda assim, o desempenho da série geral é inferior ao da FEAAC e CH, mas superior ao de Quixadá, sugerindo que a heterogeneidade das diversas unidades contribui para uma maior complexidade na previsão da evasão total da universidade.

5.4 Óráculo

Prezando pela transparência científica e reprodutibilidade do estudo, os repositórios públicos estão disponíveis nos endereços:

- **Repositório Backend:** <https://github.com/IagoLSJ/oraculo-backend>

- **Repositório Frontend:** <https://github.com/IagoLSJ/oraculo>
- **Ferramenta Online:** <https://oraculo-rose.vercel.app/>

6 CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho cumpriu seu objetivo principal ao desenvolver e validar uma ferramenta de software denominada Oráculo, projetada para auxiliar a análise de evasão universitária por meio de intervalos de predição de séries temporais. Através da implementação de uma arquitetura que integra o modelo estatístico ARIMA auxiliado pelo algoritmo auto-ARIMA e a uma interface web intuitiva, foi possível automatizar o processo de identificação de anomalias e tendências em dados históricos da Universidade Federal do Ceará (UFC) entre 2013 e 2022.

A eficácia da ferramenta foi validada por meio de um estudo de caso focado no impacto da pandemia da COVID-19. O Oráculo permitiu identificar com clareza que, embora a evasão apresentasse uma tendência decrescente em diversas unidades acadêmicas antes da crise sanitária, o período pandêmico causou desvios significativos que em alguns casos ultrapassaram os limites estatísticos previstos. Do ponto de vista técnico, a modelagem estatística mostrou-se robusta, com os gráficos de autocorrelação (ACF) e autocorrelação parcial (PACF) confirmando que o resíduo do modelo se comporta como ruído branco na maioria das unidades analisadas, validando a escolha do modelo ARIMA para esta finalidade.

A arquitetura desenvolvida, baseada em um backend em Flask e frontend em Next.js, demonstrou ser uma solução eficiente para transformar cálculos estatísticos complexos em visualizações acessíveis. Assim, o Oráculo deixa de ser apenas um experimento acadêmico para se tornar um recurso valioso para gestores educacionais, permitindo o monitoramento proativo da evasão sem a necessidade de intervenção técnica constante em scripts de dados.

Como trabalhos futuros, propõe-se:

- O aperfeiçoamento da ferramenta Oráculo, integrando novos filtros por curso, exportação de relatórios automatizados e módulos de autenticação para diferentes níveis de gestão;
- Otimização da performance do backend, através da implementação de mecanismos de caching para resultados de modelos já computados e o uso de processamento assíncrono (ex: Celery ou Redis) para o treinamento de modelos mais complexos, reduzindo o tempo de resposta da API;
- A integração direta da ferramenta com os sistemas acadêmicos da instituição, permitindo a atualização dos dados em tempo real;
- A inclusão de variáveis exógenas no modelo, como indicadores socioeconômicos dos alunos, para uma análise multivariada da evasão.

Conclui-se que o desenvolvimento de ferramentas baseadas em dados e análise pre-

ditiva é fundamental para a Engenharia de Software aplicada à educação. O Oráculo demonstrou que a tecnologia pode mitigar a complexidade da análise de dados educacionais, fornecendo subsídios concretos para a formulação de políticas de permanência estudantil mais eficazes e resilientes.

REFERÊNCIAS

- ARMSTRONG, J. S.; COLLOPY, F. Error measures for generalizing about forecasting methods: Empirical comparisons. **International Journal of Forecasting**, v. 8, n. 1, p. 69–80, 1992.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. **Time Series Analysis: Forecasting and control**. 5. ed. Hoboken: John Wiley & Sons, 2016. ISBN 978-1118675021.
- BRASIL. Ministério da Educação. **Dados sobre a evasão universitária**. 2020. Acesso em: 01 dez. 2025. Disponível em: <https://www.gov.br/mec/pt-br>. Acesso em: 01 dez. 2025.
- CHACON, S.; STRAUB, B. **Pro Git**. 2. ed. Apress, 2014. Disponível em: <https://git-scm.com/book/en/v2>.
- CLEVELAND, R. B.; CLEVELAND, W. S.; MCRAE, J. E.; TERPENNING, I. STL: A seasonal-trend decomposition procedure based on loess. **Journal of Official Statistics**, v. 6, n. 1, p. 3–73, 1990.
- FIELDING, R. T. **Architectural Styles and the Design of Network-based Software Architectures**. Tese (Doutorado) – University of California, Irvine, 2000.
- FILHO, F. W. B. H.; LEAL, B. d. C.; VINUTO, T. d. S. Uma análise de modelos de séries temporais para predição de evasão discente: estudo de caso do ifce. In: CONGRESSO NACIONAL DE EDUCAÇÃO, 7., 2021, Campina Grande. **Anais [...]**. [S. l.], 2021.
- FOWLER, M. **Continuous Integration**. 2006. Acesso em: 15 jan. 2026. Disponível em: <https://martinfowler.com/articles/continuousIntegration.html>.
- GARCIA, L.; LARA, D.; GOMES, R.; CAZELLA, S. Ferramenta para predição do desempenho acadêmico no ensino superior. In: NOME DO EVENTO, numero., local. **Anais [...]**. [S. l.], ano.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. 2. ed. Melbourne: OTexts, 2018. Acesso em: 15 jan. 2026. Disponível em: <https://otexts.com/fpp2/>.
- HYNDMAN, R. J.; KHANDAKAR, Y. Automatic time series forecasting: The forecast package for R. **Journal of Statistical Software**, v. 27, n. 3, p. 1–22, 2008.
- LIMA, J. A.; SILVA, M. R.; SANTOS, T. P. Impactos da pandemia na evasão universitária: Um estudo de caso. **Revista Brasileira de Educação**, v. 26, n. 2, p. 123–145, 2021.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. 2. ed. São Paulo: Editora Edgard Blücher, 2006.
- NIELSEN, J. Usability inspection methods. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. [S. l.: s. n.], 1994. p. 413–414.
- NUANMEESRI, S.; POOMHIRAN, L.; CHOPVITAYAKUN, S.; KADMATEEKARUN, P. Improving dropout forecasting during the COVID-19 pandemic through feature selection and multilayer perceptron neural network. **International Journal of Information and Education Technology**, v. 12, n. 9, p. 851–857, 2022.
- PRESSMAN, R. S.; MAXIM, B. R. **Engenharia de Software: Uma abordagem profissional**. 9. ed. Porto Alegre: McGraw Hill Brasil, 2020.

RUBIN, D. B. Inference and missing data. **Biometrika**, v. 63, n. 3, p. 581–592, 1976.

SHMUELI, G.; BRUCE, P. C.; GENG, X.; PATEL, N. R. **Data Mining for Business Analytics: Concepts, techniques, and applications in R**. Hoboken: Wiley, 2017.

SOMMERVILLE, I. **Engenharia de Software**. 9. ed. São Paulo: Pearson, 2011.

SOUSA, L. M.; ALMEIDA, R. J. A transição para o ensino remoto: Desafios e oportunidades. **Revista de Educação e Tecnologia**, v. 15, n. 1, p. 45–60, 2022.

TANAKA, M. M.; FURLAN, L. L.; BRANCO, L. M.; VALERIO, N. I. Adaptação de alunos de medicina em anos iniciais da formação. **Revista Brasileira de Educação Médica**, v. 40, n. 4, p. 663–668, 2016.

UFC. Pró-Reitoria de Graduação. **Painel de Indicadores Acadêmicos**. 2025. Disponível em: <https://tinyurl.com/painel-prograd-ufc>. Acesso em: 15 jan. 2026.