

# MLOps na predição de séries temporais

Laís Tie Takaki<sup>1</sup>, Bruno Bogaz Zarpelão<sup>1</sup>

<sup>1</sup>Departamento de Computação – Universidade Estadual de Londrina (UEL)  
Caixa Postal 10.011 – CEP 86057-970 – Londrina – PR – Brasil

laistie.takaki@uel.br, brunozarpelao@uel.br

**Abstract.** *Advancements in the field of Artificial Intelligence (AI) have enabled the introduction of Machine Learning (ML) models into everyday life. Consequently, there is an urgent need to meet the demand for systems that integrate AI as a core component, automating activities that were previously time-consuming and exhaustive for humans. As a starting point, this work will focus on a case study involving an ARIMA model for agronomic purposes, due to the increasing demand for the use of smart solutions in agriculture. The theoretical foundation of the work encompasses three main pillars: ML, time series, and MLOps, which will support the project's development. Ultimately, the project aims to contribute to the introduction of MLOps practices not only in agriculture but also in other contexts, through discussions on the applicability and challenges associated with ML systems in practice.*

**Resumo.** *Os avanços no ramo da Inteligência Artificial (IA) viabilizaram a introdução de modelos de Aprendizado de Máquina, Machine Learning (ML), ao cotidiano. Sendo assim, urge-se suprir a demanda por sistemas que possuem a IA como principal componente e atuam como ferramentas na automatização de atividades que, até então, demandavam tempo e eram exaustivas para o ser humano. Como ponto de partida, optou-se por trabalhar com o estudo de caso de uma aplicação com modelo ARIMA para fins agrônômicos, devido à crescente demanda pela utilização de soluções inteligentes para a agricultura. O embasamento teórico do trabalho abrange três principais pilares: ML, séries temporais e MLOps, sobre os quais se apoiarão o desenvolvimento do projeto. Por fim, espera-se que o projeto contribua para a introdução de práticas MLOps não somente na agricultura, como também em outros contextos, por meio da discussão acerca da aplicabilidade e dos desafios associados a sistemas de ML na prática.*

## 1. Introdução

Os últimos anos foram marcados pela incorporação de modelos de Inteligência Artificial (IA) no cotidiano, especialmente após o lançamento do *ChatGPT* [8]. Não obstante, avanços significativos no ramo do Aprendizado de Máquina, *Machine Learning (ML)*, vêm transformando a maneira como realizamos tarefas, desde as mais simples - como buscar informações e estudar [9] [17] - até mesmo mais complexas - realizar o cultivo inteligente na agricultura [18].

Nesse contexto, o surgimento de aplicações que possuem ML como sua principal ferramenta aumentou. Entretanto, apesar dos esforços para lançar e manter esses sistemas, desenvolvedores apresentam dificuldades para obter sucesso na tarefa, pois os modelos

não são projetados para acompanhar mudanças no ambiente, sejam elas na natureza dos dados ou na própria proposta de aplicação do sistema [10].

Além disso, estudos mostraram que, entre os profissionais entrevistados em 2020 para avaliação da maturidade de implementação de IA na indústria, o setor da agronomia não apresenta contribuições significativas [2]. Nesse sentido, a agricultura de precisão, nos últimos anos, apresentou uma crescente demanda pela agregação de tecnologias aos cultivos, mas ainda se trata de um meio que necessita de estudos antes dessa demanda ser suprida, de modo a escolher a melhor abordagem nesse novo contexto [14].

Sendo assim, o ramo de desenvolvimento de softwares se deparou com desafios para incorporar o novo setor de IA ao seus serviços. Diante disso, o *Machine Learning Operations (MLOps)* surgiu como um conjunto de práticas que contemplam tarefas como o desenvolvimento, a realização de testes e de operações para manter essas aplicações [3].

Por esses motivos, o presente projeto propõe uma abordagem MLOps em um sistema para fins agronômicos, a fim de se identificar pontos-chave a serem considerados nessa implantação de conjunto de práticas em um cenário real. Espera-se, também, que o Trabalho de Conclusão de Curso (TCC) contribua para a introdução de soluções de ML em outros contextos.

Assim, o TCC se iniciará com a discussão e fundamentação dos princípios de MLOps, que será o principal pilar do projeto. Também serão abordados conceitos relacionados a ML e séries temporais, que servirão de base para o desenvolvimento da aplicação.

Posteriormente, será realizado o estudo de caso, no qual se utilizará MLOps para se desenvolver um sistema baseado no algoritmo estatístico ARIMA para prever séries temporais. Essa aplicação se trata de um dos projetos desenvolvidos pelo núcleo UEL do Centro de Inteligência Artificial no Agro (CIA-Agro), uma proposta dentro dos Novos Arranjos de Pesquisa e Inovação (NAPI) do Estado do Paraná e que tem como objetivo promover a incorporação de IA ao meio agronômico.

Essa seção do TCC servirá de base para discussão da eficiência da aplicação das práticas no meio agronômico, pois nesse estudo de caso serão testadas diferentes abordagens nas várias fases de desenvolvimento do software. Por fim, será feita uma coletânea dos resultados observados, a fim de se destacar prós e contras da introdução de MLOps a um contexto real.

## **2. Fundamentação Teórico-Metodológica e Estado da Arte**

### **2.1. Machine Learning**

O Aprendizado de Máquina, *Machine Learning (ML)*, trata-se de um ramo da Inteligência Artificial (IA), no qual o propósito principal se baseia em ensinar máquinas a trabalharem com dados de maneira eficiente [13], de modo a emular a inteligência humana. Nesse contexto, a aplicação de modelos estatísticos é imprescindível para o treinamento dessas ferramentas, que então aprendem a tomar decisões baseadas em probabilidades relacionadas ao conjunto de informações apresentado.

O processamento de dados em grande escala e o consequente estabelecimento de relações entre eles é uma tarefa exaustiva e demorada para o ser humano. Com efeito,

a possibilidade de se utilizar ML em resoluções de problemas complexos, que exigem amplo processamento de informações, foi cada vez mais explorada e aplicada em diversos contextos nos últimos anos.

## 2.2. Séries temporais

Entre os tipos de dados mais comuns que alimentam os modelos de ML, é possível encontrar os dados temporais. Eles armazenam informações organizadas por tempo e possibilitam a inferência de outras informações baseadas nas previamente conhecidas.

Assim, o conjunto de dados que apresentam dependência serial, também conhecidos por séries temporais, é de extrema importância para contextos como o financeiro - na predição de oscilação no mercado de ações -, o da saúde - na avaliação de tempo adequado para início da medicação de indivíduos [16] - e o agrônômico - na predição de séries temporais, que será o enfoque da aplicação deste projeto.

### 2.2.1. Dados para séries temporais

O ponto principal para prever séries temporais está na fase de coleta de dados, no qual são definidas quais informações presentes no conjunto de dados são efetivamente relevantes para realizar a previsão [6]. Isto é uma tarefa difícil, pois os intervalos dos dados nem sempre são uniformes e leituras que envolvem espaço e tempo necessitam de estratégias diferentes quando comparadas às de conjuntos de dados com instâncias independentes e igualmente distribuídas [4].

Nesse sentido, o comportamento dos dados dita não somente a quantidade de transformações que serão realizadas no conjunto de dados, como também a característica da própria série temporal, que pode ser [5]:

- **Estacionária:** As métricas estatísticas (como a média, a variância e a covariância) não apresentam variações significativas ao longo do período observado, como observado no gráfico em verde na Figura 1.
- **Não estacionária:** Série que apresenta tendência - padrão de crescimento ou decrescimento a longo prazo - ou sazonalidade - padrões nos dados que se repetem periodicamente, representados em vermelho na Figura 1.



Figura 1. Exemplos de séries temporais (Fonte: Medium [1])

## 2.3. MLOps

O desenvolvimento de aplicações de software depende de equipes formadas por analistas, desenvolvedores, engenheiros e outros profissionais que não necessariamente têm conhecimento e compreensão totais do que é desenvolvido pelos outros integrantes.

Haja vista a crescente demanda por um fluxo de trabalho contínuo com constantes entregas de atualizações de projeto, é preciso que práticas organizacionais sejam implantadas [12]. Sob essa demanda, o DevOps (*Development and Operations*), ao propor um conjunto de práticas que procura aproximar todos os profissionais envolvidos no desenvolvimento de uma aplicação, aumentou a capacidade de uma empresa de distribuir aplicativos e serviços em alta velocidade.

Nos últimos anos, surgiram aplicações que contemplam o ML como principal ferramenta, adicionando mais um setor especializado ao desenvolvimento desses projetos. Apesar dos avanços na área da IA, essas aplicações costumam falhar antes mesmo de chegar à fase de produção [3] e, caso obtenham sucesso, falham para se adaptar a mudanças do ambiente [10].

Por isso, o MLOps está para aplicações de ML, assim como o DevOps está para as que não se enquadram nessa classificação. Com a constante preocupação em entregar automação e escalabilidade em aplicações, a adoção de práticas MLOps, apesar de se tratar de algo novo na indústria, apresenta benefícios para aumento e melhora no lançamento de sistemas com ML [3].

### 2.3.1. Princípios MLOps

O MLOps pode ser entendido como um conjunto de processos padronizados e tecnologias capazes de desenvolver, entregar e operar sistemas ML rapidamente e de modo eficiente [10]. Nesse sentido, foram eleitos os seguintes pilares desse conjunto [3], que sustentam as práticas que compõem o ciclo de vida MLOps (Figura 2), o qual possui como núcleo a função de modelagem e gerenciamento de dados:

- **Automação:** Refere-se à capacidade de um time MLOps de automatizar entregas de modelos de ML em um sistema ou como um componente, sem que haja intervenção manual (em roxo na Figura 2). É importante que esta etapa esteja bem consolidada, pois ela é responsável por flexibilizar o sistema e possibilitar que ele se adapte a mudanças em seu contexto, motivo pelo qual muitos projetos são descontinuados [10].
- **Implantação contínua:** Fornece, por meio do acompanhamento das dependências do modelo e sua containerização, a orquestração dos componentes de sua infraestrutura (uma mistura das fases de *Model deployment* e *Prediction serving* na Figura 2).
- **Versionamento:** Preocupação também presente no DevOps, o versionamento consiste em organizar o código em versões, possibilitando a eventual recuperação e consulta de partes do projeto antigas (auxiliando na fase de *Training operationalization* do ciclo de vida MLOps).
- **Realização de testes:** A realização de testes também deve ser feita constantemente, a fim de se diminuir os custos de manutenção a longo prazo e aumentar



Figura 2. Ciclo de vida MLOps (Fonte: Google Cloud [10])

a confiabilidade do sistema (presente na fase de *Training operationalization* na Figura 2).

- **Reprodutibilidade:** A aplicação do modelo ML (fase em azul escuro acima) em certos conjuntos de dados deve produzir resultados similares, se não iguais, o que promove o desenvolvimento de um sistema menos enviesado e com acurácia relativamente constante.
- **Monitoramento:** Após o lançamento da aplicação, é importante que haja seu acompanhamento contínuo (em azul claro na Figura 2), para verificar se o modelo está performando como esperado. Caso contrário, é possível destrinchar e discutir os problemas do sistema.

#### 2.4. MLOps na predição de séries temporais

Apesar das séries temporais possuírem aplicações em diversos contextos, como mencionado na subseção acima, por se tratar um campo novo na área de desenvolvimento e de IA, as contribuições para o mercado ainda são recentes [19] [7] [15]. Além disso, muitas empresas ainda parecem estar amadurecendo suas abordagens MLOps, com cientistas de dados ainda gerenciando fluxos de trabalho ML em grande escala [11].

Por esse motivo, este trabalho procurará facilitar a introdução dos princípios MLOps a aplicações reais, em especial para o meio agrônômico, que apresenta crescente demanda pela agregação de tecnologias aos cultivos [14].

### 3. Objetivos

O presente trabalho tem como objetivo geral realizar um estudo de caso que envolve a aplicação de técnicas de MLOps ao desenvolvimento de um sistema que permite a análise

de séries temporais agrometeorológicas. Por se tratar de uma área de estudo recente e com crescente demanda, procura-se também definir e explorar as dificuldades de implantação das práticas associadas ao DevOps em uma sistema que possui um modelo ML como ferramenta principal. Mais especificamente, o projeto busca:

1. Elencar, para cada fase do ciclo de vida MLOps, tarefas que permitam analisar e compreender os benefícios do conjunto de práticas.
2. Discutir os desafios encontrados para colocar em prática esses princípios e os prós e contras associados à execução da tarefa, tendo em vista o contexto da aplicação.

#### 4. Procedimentos metodológicos/Métodos e técnicas

O primeiro passo consiste na investigação do problema, no qual será realizada uma revisão bibliográfica acerca dos dois pilares do projeto: MLOps e séries temporais, a fim de se elencar critérios e prioridades em se tratando do tema proposto. É importante também que se faça um estudo dos dados climáticos e do modelo ARIMA disponibilizados para o desenvolvimento da aplicação, para compreender a abordagem desse modelo de ML nesse contexto.

Em segundo lugar, o desenvolvimento de solução abordará um problema prático; nesse caso, o objeto de estudo será o software de ML do projeto desenvolvido no CIA-Agro. A seguir, ocorrerá a validação de projeto, na qual serão analisados os resultados obtidos ao fim do período de desenvolvimento da aplicação, que contribuirá para a conclusão do item 1 da Seção 3.

Por fim, será realizada uma avaliação de implementação, ao passo de desenvolvimento do item 2 os objetivos específicos, gerando conhecimento científicos sobre a pesquisa realizada; fechando, assim, o ciclo de *Design Science Research* de Wieringa [20].

#### 5. Cronograma de Execução

Atividades:

1. Revisão bibliográfica focada em MLOps e séries temporais;
2. Estudo dos dados climáticos e modelo ARIMA;
3. Desenvolvimento da aplicação de séries temporais;
4. Validação de projeto e avaliação de implementação;
5. Escrita do TCC.

**Tabela 1. Cronograma de Execução**

	jul	ago	set	out	nov	dez	jan
Atividade 1	x						
Atividade 2	x	x					
Atividade 3		x	x	x	x	x	x
Atividade 4					x	x	x
Atividade 5			x	x	x	x	x

## 6. Contribuições e/ou Resultados esperados

O principal resultado esperado deste projeto é identificar pontos-chave a serem considerados ao aplicar MLOps no desenvolvimento de sistemas, discutindo também desafios, prós e contras associados a uma aplicação real baseada em ML. Além disso, pretende-se propor soluções para as dificuldades encontradas e possíveis melhorias no modelo e no sistema, levando em consideração sua aplicabilidade em outros contextos além do agrônômico.

## 7. Espaço para assinaturas



Aluno

Londrina, 29 de julho de 2024.



Orientador

## Referências

- [1] Gisely Alves. Séries temporais: Parte 1, 2020. Acesso em: 20 de julho de 2024.
- [2] Fabian Schladitz Jerome Buvat Ramya Krishna Puttur Marie-caroline Baerd Ron Tolido Jerry Kurtz Subrahmanyam KVJ Anne-Laure Thieullent, Ashwin Yardi and Gaurav Aggarwal. The ai-powered enterprise: Unlocking the potential of ai at scale. Technical report, Capgemini Research Institute, 2020.
- [3] Gabriel Araujo, Marcos Kalinowski, Markus Endler, and Fabio Calefato. Professional insights into benefits and limitations of implementing mlops principles. In *Proceedings of ICEIS*, 04 2024.
- [4] Gowtham Atluri, Anuj Karpatne, and Vipin Kumar. Spatio-temporal data mining: A survey of problems and methods. *ACM Comput. Surv.*, 51(4), aug 2018.
- [5] Chris Chatfield. *Time-Series Forecasting*. Chapman and Hall/CRC, 2000.
- [6] Douglas C. Montgomery Cheryl L. Jennings, Murat Kulahci. *Introduction to Time Series Analysis and Forecasting*. John Wiley Sons. Inc, 2008.
- [7] Adam Ekblom. Monitoring with mlops for clinical decision support. Technical report, Lund University, 2024.
- [8] Irene S. Gabashvili. The impact and applications of chatgpt: a systematic review of literature reviews, 2023.
- [9] Radhika R Halde. Application of machine learning algorithms for betterment in education system. In *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)*, pages 1110–1114, 2016.
- [10] Jarek Kazmierczak Khalid Salama and Donna Schut. Practitioners guide to mlops: A framework for continuous delivery and automation of machine learning. Technical report, Google Cloud, 2021.

- [11] Dominik Kreuzberger, Niklas Kühn, and Sebastian Hirschl. Machine learning operations (mlops): Overview, definition, and architecture. *IEEE Access*, 11:31866–31879, 2023.
- [12] Leonardo Leite, Carla Rocha, Fabio Kon, Dejan Milojicic, and Paulo Meirelles. A survey of devops concepts and challenges. *ACM Comput. Surv.*, 52(6), nov 2019.
- [13] Batta Mahesh. Machine learning algorithms - a review. *International Journal of Science and Research (IJSR)*, 2019.
- [14] Subhrajit Mandal, Anamika Yadav, Florence A. Panme, Kshetrimayum Monika Devi, and Shravan Kumar S.M. Adaption of smart applications in agriculture to enhance production. *Smart Agricultural Technology*, 7:100431, 2024.
- [15] Saadin Oyucu and Ahmet Aksöz. Integrating machine learning and mlops for wind energy forecasting: A comparative analysis and optimization study on türkiye’s wind data. *Applied Sciences*, 14(9), 2024.
- [16] Robert B. Penfold and Fang Zhang. Use of interrupted time series analysis in evaluating health care quality improvements. *Academic Pediatrics*, 13(6, Supplement):S38–S44, 2013. Quality Improvement in Pediatric Health Care.
- [17] Abderahman Rejeb, Karim Rejeb, Andrea Appolloni, Horst Treiblmaier, and Mohammad Iranmanesh. Exploring the impact of chatgpt on education: A web mining and machine learning approach. *The International Journal of Management Education*, 22(1):100932, 2024.
- [18] Abhinav Sharma, Arpit Jain, Prateek Gupta, and Vinay Chowdary. Machine learning applications for precision agriculture: A comprehensive review. *IEEE Access*, 9:4843–4873, 2021.
- [19] Rakshith Subramanya, Seppo Sierla, and Valeriy Vyatkin. From devops to mlops: Overview and application to electricity market forecasting. *Applied Sciences*, 12(19), 2022.
- [20] Roel J. Wieringa. *Design Science Methodology for Information Systems and Software Engineering*. Springer, 2014.