

# Previsão de custos e preços de commodities agrícolas

Uma abordagem com Modelos Estatísticos e Aprendizado de Máquina

Fernanda Aiko Hamatsu

Prof. Dr. Ivan Carlos Alcântara de Oliveira

## Motivação e Objetivos

A agricultura no Brasil representa aproximadamente 25\% de seu Produto Interno Bruto (PIB), um mercado que cresce a cada ano e apresenta um grande potencial para soluções tecnológicas e inovadoras. Soluções utilizando dados podem auxiliar os agricultores em diversas áreas, sendo uma delas sua estratégia comercial para precificação de seus cultivos e posicionamento de mercado, visando preços mais justos e que o proporcionem grandes margens de lucro. [Centro de Estudos Avançados em Economia Aplicada (CEPEA) 2024]

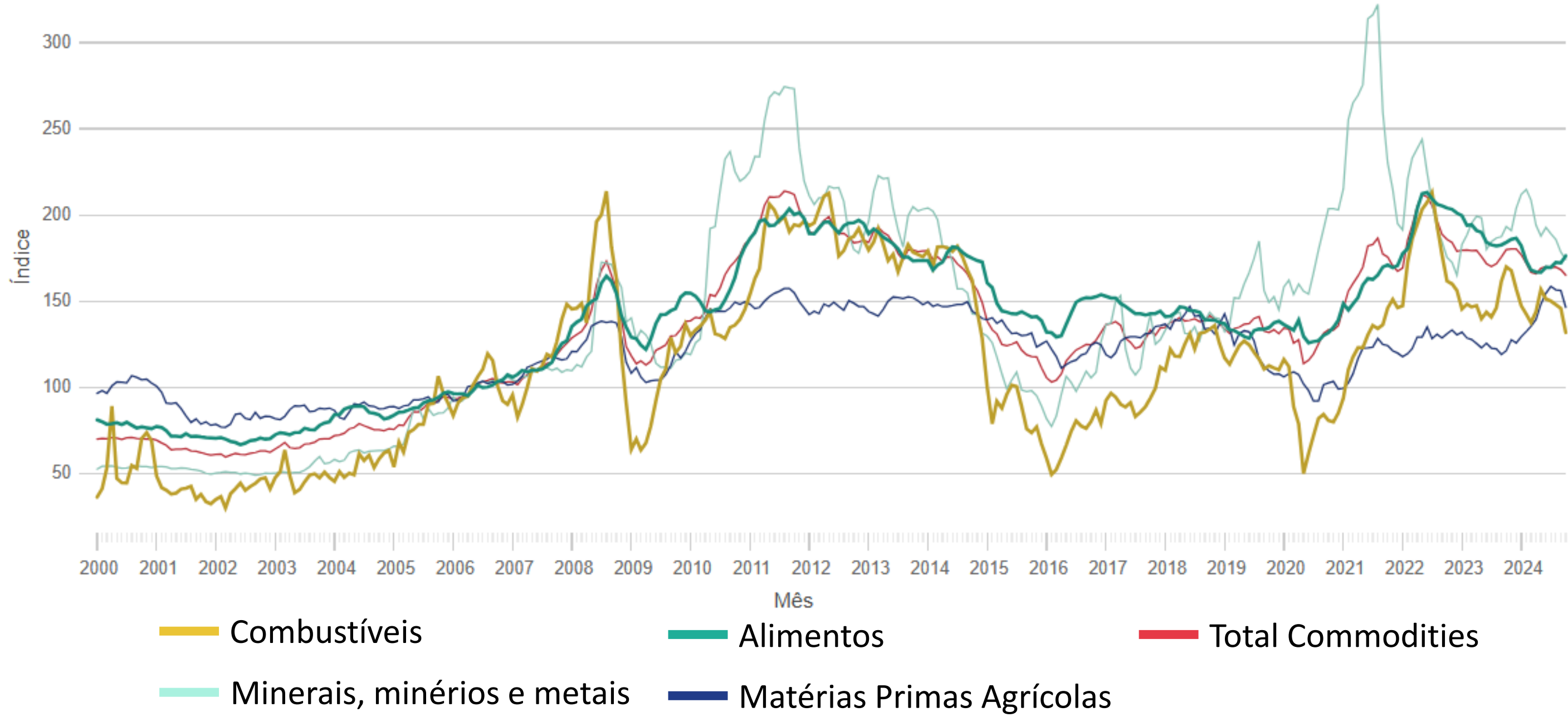


Figura 1: Índice de Preços das Exportações – Novembro de 2024

Fonte: Ministério do Desenvolvimento, Indústria, Comércio e Serviços

O objetivo deste trabalho está no desenvolvimento de modelos preditivos de Séries Temporais que retornem a previsão dos preços das commodities agrícolas Soja, Milho, Trigo e Algodão. Para isso serão utilizados: modelos estatísticos simples, modelos estatísticos avançados, e modelos de aprendizado de máquina.

## Metodologia

Pesquisadores da Universidade Agrícola de Sichuan estudaram aplicações de modelos preditivos para analisar e prever os comportamentos das commodities agrícolas na China (JIANG, 2023). Tendo este e outros estudos como base, foram utilizados dados históricos de cotações de cada cultivo selecionado do Brasil, obtidos através do Portal Agrolink que disponibiliza uma base atualizada mensalmente com valores históricos a partir de 2004, podendo ser filtrados e analisados por Estado ou País. Diferentes variáveis, como câmbio, cotação na Bolsa de Valores e Precipitação também foram exploradas através dos modelos multivariados.

A aplicação desenvolvida através da linguagem de programação Python, explorou os seguintes modelos: Auto Regressivo, ARMA & ARIMA, XGBoost e o Prophet (Meta). Métricas como MAE e RMSE foram aplicadas sobre os resultados gerados utilizando da partição de 80% dados de aprendizado e 20% dados de teste para validação.

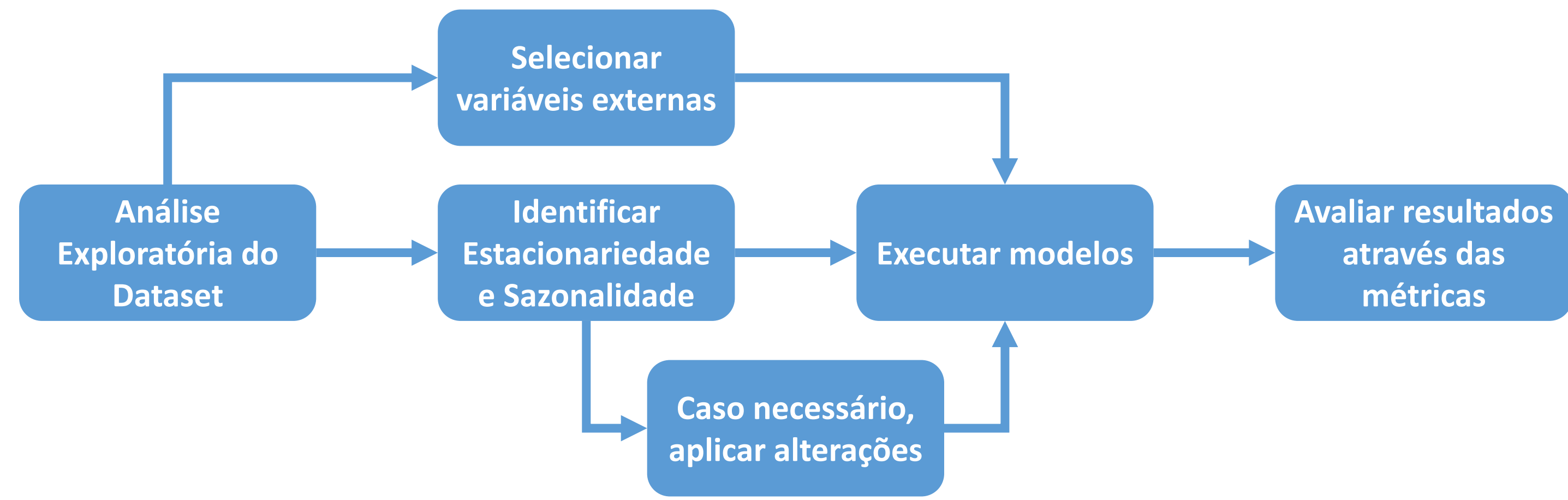


Figura 2: Fluxo parcial do desenvolvimento do projeto

Fonte: Elaborado pelo Autor

## Resultados Parciais

Experimentos foram realizados para todos os cultivos. Métricas e performances com o uso de modelos Multivariados para Soja são ilustrados no Quadro 1, todos avaliados com a separação de 20% dos dados para teste. A partir da análise das curvas previstas e dos valores das métricas de erro obtidas, os melhores modelos foram o ARIMA Multivariado e o Prophet Multivariado (Figura 3).

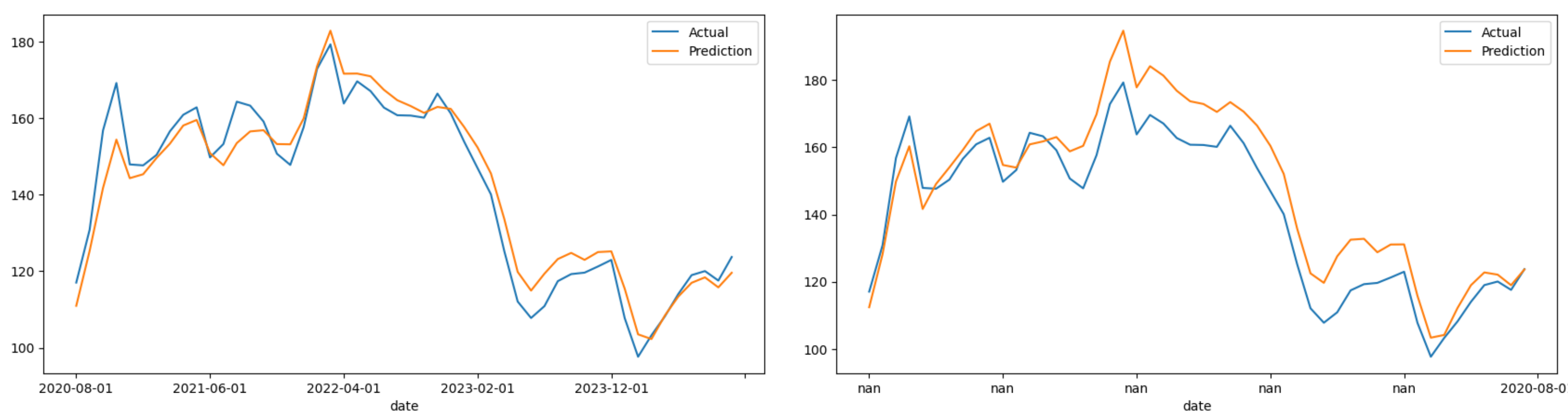


Figura 3: Resultados ARIMA (esq.) e Prophet (dir.)

Fonte: Elaborado pelo Autor

Modelo	MAE	RMSE	R-Squared
Auto-Regressivo	45.61	49.58	-3.78
ARIMA Univariado Não-Estacionário	43.99	48.91	-3.66
ARIMA Univariado Estacionário	0.28	0.38	-0.13
XGBoost Univariado	44.10	51.33	-4.13
Prophet Univariado	57.66	63.16	-6.77
ARIMA Multivariado	4.42	5.47	0.94
XGBoost Multivariado	43.54	48.73	-3.62
Prophet Multivariado	7.67	8.76	0.85

Quadro 1: Métricas dos modelos para soja

Fonte: Elaborado pelo Autor

## Considerações finais e próximos passos

A partir dos resultados parciais obtidos, devem ser analisadas as previsões e utilizar técnicas de detecção para *overfitting*, a validação das previsões futuras com possíveis usuários da solução (ex.: agricultores) e a aplicação dos modelos para todos os cultivos, selecionando os melhores para cada.

## Referências

JIANG, Sha; WANG, Liming; MIAO, Xiaoyu. A Study on Pricing and Replenishment of Vegetable Commodities Based on Goal Planning and Time Series Forecasting Models. Ya'an, China: IEEE, 2023.