

# Desvendando a Equação dos Preços das Commodities Agrícolas no Brasil

Rafael Passos Domingues

## I. RESUMO

Este artigo aborda a análise e previsão dos preços das commodities agrícolas no Brasil, reconhecendo sua importância crítica para a economia nacional. Utilizando uma abordagem inovadora, combinamos as técnicas de Support Vector Machines (SVM) e XGBoost para a extração de características em séries temporais univariadas. A escolha dessas metodologias baseia-se na capacidade comprovada do SVM em lidar com dados não lineares e em sua eficácia na modelagem de séries temporais, enquanto o XGBoost é conhecido por sua habilidade em tratar relações complexas. A integração dessas técnicas busca capturar nuances temporais e não lineares, melhorando a precisão na previsão dos preços das commodities agrícolas. O estudo não apenas aprimora a capacidade de previsão, mas também oferece insights valiosos para stakeholders, governos e entidades reguladoras, contribuindo para o desenvolvimento de estratégias mais eficientes e resilientes no setor agrícola brasileiro.

## II. INTRODUÇÃO

No cenário econômico global, os preços das commodities agrícolas desempenham um papel crucial, influenciando não apenas o setor agrícola, mas também a estabilidade econômica de diversos países. No contexto brasileiro, onde a agricultura representa uma parte substancial da economia, compreender e prever os padrões dos preços das commodities agrícolas torna-se imperativo para planejamento, tomada de decisões e políticas agrícolas eficazes.

Este artigo busca desvendar a complexa equação subjacente aos preços das commodities agrícolas no Brasil, utilizando uma abordagem inovadora baseada na combinação das técnicas de Support Vector Machines (SVM) e XGBoost para a extração de features em séries temporais univariadas. A escolha dessas metodologias se fundamenta na capacidade comprovada de SVM em lidar com dados não lineares e em sua eficácia na modelagem de séries temporais, enquanto o XGBoost destaca-se por sua habilidade em lidar com relações complexas e em lidar com sobreajuste.

A abordagem integrada proposta visa capturar de forma mais precisa as nuances temporais e não lineares inerentes aos movimentos dos preços das commodities agrícolas. Ao utilizar SVM para extrair características relevantes das séries temporais e em seguida aplicar o XGBoost para otimizar o modelo, espera-se alcançar uma maior acurácia na previsão dos preços, contribuindo assim para uma compreensão mais profunda e precisa das variáveis que influenciam esse mercado dinâmico.

Ao desvendar a equação dos preços das commodities agrícolas no Brasil, este estudo não apenas aprimora a capacidade de previsão, mas também fornece insights valiosos para stakeholders, governos e entidades reguladoras, contribuindo para o desenvolvimento de estratégias mais eficientes e resilientes no setor agrícola nacional.

## III. REFERENCIAL TEÓRICO

### A. Support Vector Machines (SVM)

As Support Vector Machines (SVM) são uma classe de algoritmos de aprendizado de máquina eficazes para classificação e regressão. Na análise de séries temporais, as SVM destacam-se por sua capacidade de lidar com dados não lineares e complexos. O princípio central das SVM é encontrar o hiperplano que melhor separa as diferentes classes no espaço de características. Em séries temporais, as SVM podem ser utilizadas para extrair padrões não lineares, identificar tendências e capturar a dinâmica temporal dos dados.

No contexto da modelagem de preços de commodities agrícolas, as SVM podem ser aplicadas para identificar correlações não lineares entre variáveis, considerando a sazonalidade e outros fatores temporais que influenciam os preços. A capacidade das SVM em lidar com dados de alta dimensionalidade é crucial ao lidar com séries temporais, onde múltiplos fatores podem influenciar os movimentos dos preços ao longo do tempo.

### B. XGBoost

XGBoost, ou Extreme Gradient Boosting, é um algoritmo de aprendizado de máquina baseado em árvores de decisão que se destaca por sua eficácia na modelagem de relações complexas entre variáveis. É uma extensão do método de boosting, onde modelos mais fracos são combinados para formar um modelo mais robusto. No contexto de séries temporais, o XGBoost oferece vantagens significativas, sendo capaz de lidar com sobreajuste e capturar padrões temporais complexos.

O XGBoost é particularmente adequado para a análise de séries temporais univariadas, onde a dependência temporal das observações é uma consideração essencial. Ao incorporar técnicas de regularização e otimização, o XGBoost pode lidar eficazmente com a variabilidade temporal, permitindo uma modelagem mais precisa e generalizável dos preços das commodities agrícolas.

### C. Integração de SVM e XGBoost

A combinação de SVM e XGBoost neste estudo busca aproveitar as vantagens individuais dessas duas técnicas. Enquanto as SVM contribuem para a extração de características

não lineares e a consideração de padrões temporais, o XGBoost aprimora a capacidade de modelagem e previsão, proporcionando um equilíbrio entre interpretabilidade e complexidade.

Essa abordagem integrada visa superar desafios comuns na modelagem de séries temporais agrícolas, proporcionando uma análise mais abrangente e precisa dos fatores que influenciam os preços das commodities no Brasil. A compreensão detalhada desses algoritmos é fundamental para o desenvolvimento e aplicação eficaz dessa metodologia no contexto agrícola.

#### IV. DESAFIOS TÉCNICOS

##### 1) Padronização de Dados

- SVM e XGBoost podem ter sensibilidades diferentes em relação à escala e distribuição dos dados. A integração requer uma padronização cuidadosa para garantir que ambas as técnicas interpretem as características temporais de maneira consistente.

##### 2) Sintonização de Parâmetros

- Cada algoritmo tem seus próprios parâmetros que afetam o desempenho do modelo. Encontrar uma combinação ideal de parâmetros para ambas as técnicas pode ser desafiador e requer experimentação rigorosa.

##### 3) Interpretabilidade do Modelo Integrado

- Integrar SVM e XGBoost pode resultar em modelos mais complexos. Manter a interpretabilidade do modelo, especialmente em um contexto agrícola onde insights compreensíveis são valiosos, é um desafio.

##### 4) Computacionalmente Intensivo

- A combinação de SVM e XGBoost pode exigir recursos computacionais significativos. Gerenciar eficientemente a carga computacional durante a integração é essencial para garantir a viabilidade prática.

#### A. Desafios para a Prevenção de Overfitting

*Overfitting* ocorre quando um modelo se ajusta excessivamente aos dados de treinamento, capturando não apenas os padrões subjacentes, mas também o ruído aleatório. Isso leva a um desempenho insatisfatório ao lidar com novos dados, pois o modelo está muito adaptado às peculiaridades específicas do conjunto de treinamento.

##### Desafios e Estratégias de Prevenção

- Complexidade do Modelo
  - Modelos muito complexos, como árvores de decisão profundas, são propensos a overfitting. Estratégias incluem limitar a profundidade da árvore no XGBoost e ajustar o parâmetro de regularização  $C$  nas SVM.
- Tamanho do Conjunto de Treinamento
  - Com conjuntos de treinamento pequenos, o modelo pode se ajustar facilmente a padrões irrelevantes. Aumentar o tamanho do conjunto de treinamento pode ajudar a mitigar esse problema.

- Validação Cruzada

- A validação cruzada divide o conjunto de dados em partes para treinar e testar o modelo em diferentes subconjuntos. Isso ajuda a avaliar a capacidade do modelo de generalização e a identificar sinais de overfitting.

- Regularização

- Técnicas como a regularização L1 e L2 podem ser aplicadas em SVM e XGBoost para penalizar coeficientes excessivamente grandes. Isso ajuda a controlar a complexidade do modelo e evitar overfitting.

- Seleção Adequada de Atributos

- Selecionar apenas os atributos mais relevantes pode reduzir a complexidade do modelo e, portanto, mitigar o overfitting. A SVM, em particular, pode se beneficiar de uma seleção cuidadosa de features para evitar o excesso de ajuste aos dados de treinamento.

## V. RESULTADOS ESPERADOS

Espera-se que a integração das técnicas SVM e XGBoost resulte em um modelo mais preciso na previsão dos preços das commodities agrícolas no Brasil. A combinação das capacidades de extração de características não lineares da SVM com a habilidade do XGBoost em lidar com relações complexas tem o potencial de capturar nuances temporais e padrões ocultos, melhorando a precisão geral das previsões.

A expectativa é que a abordagem integrada seja eficaz na identificação de padrões temporais e não lineares nos dados das séries temporais agrícolas. Isso proporcionará uma compreensão mais profunda das variáveis que influenciam os preços das commodities ao longo do tempo, permitindo uma análise mais precisa e informada.

A integração de SVM e XGBoost visa melhorar a resiliência do modelo em relação a sazonalidades e variações temporais, comumente observadas nos preços das commodities agrícolas. Ao capturar de maneira mais eficaz as tendências a longo prazo e os movimentos cíclicos, o modelo resultante deve ser capaz de lidar adequadamente com a dinâmica temporal inerente a esse mercado.

Embora a combinação de SVM e XGBoost possa levar a modelos mais complexos, espera-se que o processo de integração inclua estratégias para manter a interpretabilidade do modelo. A capacidade de interpretar e explicar as relações descobertas entre as variáveis é crucial, especialmente em contextos como o agrícola, onde a compreensão dos fatores subjacentes é essencial para a tomada de decisões informadas.

Antecipa-se que os insights derivados do modelo aprimorado terão implicações práticas para stakeholders, governos e entidades reguladoras. O entendimento aprimorado dos fatores que influenciam os preços das commodities agrícolas pode informar o desenvolvimento de estratégias mais eficientes e adaptáveis, contribuindo para a estabilidade e sustentabilidade do setor agrícola no Brasil.

## VI. CONSIDERAÇÕES FINAIS

A utilização de dados da pandemia do COVID-19 como um período específico na análise dos preços das commodities agrícolas no Brasil se revela uma oportunidade única e valiosa. Este período excepcional não apenas serve como um termômetro que se desligou temporariamente, permitindo uma análise isolada e aprofundada, mas também destaca a relevância de outros fatores significativos que podem ter sido eclipsados por eventos extraordinários.

Durante a pandemia, observamos mudanças drásticas nos padrões de demanda e oferta, efeitos nos canais de distribuição e impactos nas cadeias de suprimentos. Essas perturbações podem oferecer insights cruciais sobre a resiliência do setor agrícola a choques externos e, ao mesmo tempo, realçar a influência de fatores climáticos e ambientais que podem ter sido anteriormente subestimados.

Ao considerar o período da pandemia como um divisor de águas, abre-se a possibilidade de identificar características específicas desse período que podem ter impactos duradouros. Além disso, a análise durante a pandemia pode funcionar como um teste de estresse para o modelo integrado SVM-XGBoost,

fornecendo uma avaliação crítica de sua robustez e capacidade de adaptação a cenários adversos.

A inclusão de fatores climáticos, muitas vezes subestimados na análise tradicional, ganha destaque, pois condições meteorológicas extremas podem desempenhar um papel determinante nos preços das commodities agrícolas. A integração desses fatores climáticos com as técnicas de SVM e XGBoost permite uma compreensão mais holística e precisa das variáveis que influenciam os preços, contribuindo para uma modelagem mais resiliente e adaptável.

Portanto, ao encarar a pandemia como uma oportunidade para destacar e compreender melhor outros fatores significativos, como condições climáticas e ambientais, esta pesquisa não apenas aprimora a capacidade de previsão dos preços das commodities agrícolas no Brasil, mas também proporciona uma base mais sólida para o desenvolvimento de estratégias agrícolas robustas e resilientes em face de desafios futuros.

## VII. REFERÊNCIAS BIBLIOGRÁFICAS

- 1) Bucene, L. C., RODRIGUES, L. H., & Meira, C. A. A. (2002). Mineração de dados climáticos para previsão de geada e deficiência hídrica para as culturas do café e da cana-de-açúcar para o Estado de São Paulo.
- 2) Volpato, M. M., Alves, H. M. R., Vieira, T. G., Boell, V. G., Souza, J. C. D., Boell, M. G., & Soares, W. L. (2015). Mineração de dados espectrais para modelagem de ocorrência da broca do café.
- 3) Groh, J., Slawitsch, V., Herndl, M., Graf, A., Vereecken, H., & Pütz, T. (2018). Determining dew and hoar frost formation for a low mountain range and alpine grassland site by weighable lysimeter. *Journal of Hydrology*, 563, 372-381.
- 4) Mironov, D., Heise, E., Kourzeneva, E., Ritter, B., Schneider, N., & Terzhevik, A. (2010). Implementation of the lake parameterisation scheme FLake into the numerical weather prediction model COSMO.
- 5) Ho, E., & Gough, W. A. (2006). Freeze thaw cycles in Toronto, Canada in a changing climate. *Theoretical and Applied Climatology*, 83, 203-210.
- 6) Akinoglu, B. G. (2008). Recent advances in the relations between bright sunshine hours and solar irradiation. *Modeling Solar Radiation at the Earth's Surface: Recent Advances*, 115-143.
- 7) Xu, Q., Yan, X., Grantz, D. A., Xue, X., Sun, Y., Lammers, P. S., & Cheng, Q. (2020). Improving estimation of evapotranspiration during soil freeze-thaw cycles by incorporating a freezing stress index and a coupled heat and water transfer model into the FAO Penman-Monteith model. *Agricultural and Forest Meteorology*, 281, 107847.
- 8) Reisner, J., Rasmussen, R. M., & Bruintjes, R. T. (1998). Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model. *Quarterly Journal of the Royal Meteorological Society*, 124(548), 1071-1107.
- 9) Thompson, G., Nygaard, B. E., Makkonen, L., & Dierer, S. (2009, September). Using the Weather Research and Forecasting (WRF) model to predict ground/structural icing. In 13th International Workshop on Atmospheric Icing on Structures, METEOTEST, Andermatt, Switzerland.
- 10) Liston, G. E., & Pielke, R. A. (2001). A climate version of the regional atmospheric modeling system. *Theoretical and Applied Climatology*, 68, 155-173.
- 11) da Mata Ribeiro, M., & Quimarães, S. S. (2018). Redes neurais utilizando tensorflow e keras. *RE3C-Revista Eletrônica Científica de Ciência da Computação*, 13(1).
- 12) Falcão, J. V. R., de Ávila Moreira, V., de Ávila Ramos, C., & de Oliveira Santos, F. A. (2019). Redes neurais deep learning com tensorflow. *RE3C-Revista Eletrônica Científica de Ciência da Computação*, 14(1).
- 13) Meyer, D. (2022). Machine Learning Emulators for Numerical Weather Prediction—Applications to Parametrization Schemes (Doctoral dissertation, University of Reading).
- 14) Chen, M., Qian, Z., Boers, N., Jakeman, A. J., Kettner, A. J., Brandt, M., & Lü, G. (2023). Iterative integration of deep learning in hybrid Earth surface system modelling. *Nature Reviews Earth & Environment*, 4(8), 568-581.
- 15) Dados abertos INPE. Disponível em: [https://www.gov.br/inpe/pt-br/aceso-a-informacao/dados-abertos/repositorio-de-arquivos/pda\\_inpe\\_2021-2023-versao\\_1-0\\_jun\\_2022.pdf](https://www.gov.br/inpe/pt-br/aceso-a-informacao/dados-abertos/repositorio-de-arquivos/pda_inpe_2021-2023-versao_1-0_jun_2022.pdf) Acesso em 20 de Outubro de 2023.
- 16) SIGMA CPTEC INPE. Disponível em: <https://sigma.cptec.inpe.br/> Acesso em 20 de Outubro de 2023.
- 17) ACERVO CPTEC INPE. Disponível em: <https://satelite.cptec.inpe.br/acervo/goes16.formulario.logic;jsessionid=4A0D2C9D2C899557D37310F3DFDE305F> Acesso em 20 de Outubro de 2023.