

Desvendando novos vínculos da Equação do Preço das Commodities Agrícolas no Brasil

Rafael Passos Domingues

I. RESUMO

Este trabalho busca explorar a capacidade preditiva de conjuntos de regressão, comparando diferentes conjuntos entre si e com abordagens que consideram apenas um modelo estatístico: a melhor combinação encontrada a partir dos modelos de referência [Ribeiro & Coelho]. O foco está na previsão de preços de commodities agrícolas para o curto prazo, especificamente um mês à frente. Este horizonte de previsão é escolhido devido à sua relevância para permitir que produtores rurais de pequeno a médio porte desenvolvam planejamentos estratégicos de curto prazo para atender às suas necessidades imediatas. Pela primeira vez, busca-se correlações entre variações nos preços dos insumos, flutuações cambiais e mudanças climáticas em conjunto com as variações isoladas nos preços do café, soja, milho e outras commodities agrícolas, pautados na metodologia “*dividir para conquistar*”, espera-se que tal cruzamento de dados possa revelar novos vínculos ocultos na equação de preço das commodities agrícolas brasileira.

II. INTRODUÇÃO

A operação intrínseca ao agronegócio constitui uma rede de atividades de impacto significativo na economia local, regional e nacional. Em momentos de crise, o agronegócio tem demonstrado ser uma área crucial, contribuindo para o crescimento de diversos indicadores econômicos, incluindo o Produto Interno Bruto (PIB). O contexto econômico atual no Brasil destaca-se pelo agronegócio representar 24,1% do PIB do país, impulsionando constantemente o crescimento nacional e controlando a inflação mesmo diante do caos dos conflitos na Europa e a pandemia do covid-19 [Cepea/CNA].

No âmbito estadual, no estado de Minas Gerais, em 2022 o PIB do agronegócio mineiro cresceu 6,3% no volume produzido, atingindo a maior participação no total do PIB do estado observada na série histórica desde 2010 [Fundação João Pinheiro]. Projeções do Ministério da Agricultura, Pecuária e Abastecimento (MAPA) indicam um aumento de 19,1% entre 2022/23 e 2032/33, passando de 77,5 milhões de hectares em 2022/23 para 92,3 milhões em 2032/33, o que corresponde a um acréscimo anual de 1,7% no melhor cenário. Este setor não apenas contribui para o desenvolvimento regional, mas também influencia a renda familiar das famílias mais carentes. Diante desse cenário, compreender as projeções de preços futuros torna-se crucial, pois isso impacta o planejamento econômico, principalmente de pequenos produtores.

Neste contexto, a previsão de séries temporais de preços de commodities agrícolas desempenha um papel essencial no

cenário econômico. A precisão dessas previsões é tão vital quanto os resultados esperados e está diretamente ligada à acurácia do modelo adotado. Uma abordagem que tem se destacado na literatura para melhorar a capacidade preditiva dos modelos de regressão é o uso de conjuntos de regressão [Ribeiro & Coelho].

Nas seções seguintes será apresentado quatro distintos estudos de caso, cada qual cruzando dados específicos de modo isolado, para entender a variação de preços das commodities agrícolas no mercado brasileiro. O Caso 1 investiga a relação entre os preços do café e do adubo, enquanto o Caso 2 explora a correlação entre os preços do café e a taxa de câmbio. Os Casos 3 e 4 ampliam a análise, pela primeira vez cruzando dados de preços do café com os preços da soja, milho e índice pluviométrico, respectivamente.

Este estudo desempenha um papel significativo na literatura de computação e aprendizado de máquina, ao aplicar, avaliar e discutir o desempenho de abordagens de bagging, boosting e stacking, representadas por RF, GBM e XGB, na previsão de preços de commodities agrícolas em curto prazo. A inserção de modelos de séries temporais utilizando conjuntos de regressão para antecipar fenômenos ligados ao agronegócio preenche uma lacuna importante em uma área vital para a construção de indicadores econômicos.

É notável a escassez de artigos de revisão neste contexto, o que pode ser interpretado como uma oportunidade valiosa de contribuição. A falta de foco generalizado sobre esse estudo no Brasil revela, de certa forma, uma perspectiva otimista: a nação possui um solo agrícola privilegiado em comparação com outros países, como os do continente Africano, onde a pesquisa em agricultura de precisão é mais desenvolvida.

A ausência de um volume significativo de estudos revisados pode ser interpretada como um indicativo de que o Brasil, mesmo com recursos agrícolas vastos, ainda não explorou totalmente o potencial da aplicação de tecnologias avançadas na agricultura. Isso sugere que, ao investir em pesquisa e adoção de abordagens inovadoras, o Brasil pode alavancar ainda mais seu setor agrícola, atingindo níveis de eficiência e produtividade sem precedentes.

Dessa forma, os resultados deste estudo têm o potencial não apenas de preencher uma lacuna no conhecimento existente, mas também de estimular pesquisadores, agricultores e stakeholders do mercado financeiro a reconhecerem a importância estratégica de investir em métodos avançados de previsão para impulsionar o agronegócio brasileiro a patamares superiores de sucesso e sustentabilidade, especialmente aos produtores de pequeno a médio porte que são os principais movimentadores ocultos da economia brasileira.

III. METODOLOGIA

A metodologia adotada neste estudo combina abordagens de aprendizado de máquina, especificamente Support Vector Machines (SVM) e XGBoost, para analisar e prever os preços das commodities agrícolas no Brasil. A escolha dessas técnicas baseia-se em sua eficácia comprovada em lidar com dados não lineares e na capacidade de modelar séries temporais.

Iniciamos utilizando o SVM para a extração de características relevantes das séries temporais univariadas dos preços das commodities agrícolas. O SVM destaca-se pela sua adaptabilidade a relações não lineares, fornecendo uma representação robusta dos padrões temporais presentes nos dados. A aplicação do SVM permite identificar características importantes, contribuindo para uma compreensão mais profunda dos fatores que afetam os preços.

Em seguida, implementamos o XGBoost para otimizar o modelo, levando em consideração a complexidade das relações presentes nos dados. O XGBoost é conhecido por sua capacidade de lidar com sobreajuste e relações complexas, tornando-o uma escolha adequada para aprimorar a precisão do modelo SVM.

A integração dessas duas abordagens visa capturar nuances temporais e não lineares, proporcionando uma representação mais abrangente da equação subjacente aos preços das commodities agrícolas no Brasil.

IV. 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.

V. 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.

VI. 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.

VII. CONSIDERAÇÕES FINAIS

Este estudo propôs uma abordagem inovadora para desvendar a equação dos preços das commodities agrícolas no Brasil, utilizando uma combinação de Support Vector Machines (SVM) e XGBoost. Os resultados obtidos demonstram uma melhoria significativa na precisão da previsão, destacando a eficácia dessa abordagem integrada.

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.

A análise das características extraídas revela a importância de fatores temporais não lineares e complexas relações entre variáveis que influenciam os preços das commodities agrícolas. Esses insights não apenas aprimoram a capacidade de previsão, mas também fornecem uma base sólida para o desenvolvimento de estratégias mais eficientes e resilientes no setor agrícola brasileiro.

O desvendamento da equação dos preços das commodities agrícolas contribui para uma compreensão mais profunda e precisa desse mercado dinâmico, beneficiando stakeholders, governos e entidades reguladoras na tomada de decisões informadas e estratégicas.

VIII. REFERÊNCIAS BIBLIOGRÁFICAS

- 1) Ribeiro, Matheus Henrique Dal Molin, and Leandro dos Santos Coelho. "Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series." *Applied soft computing* 86 (2020): 105837.
- 2) Odion, Divinefavour, et al. "Towards improving farmers livelihood in Nigeria using food price forecasting."
- 3) Banerjee, Saikat, and Abhoy Chand Mondal. "An ingenious method for estimating future crop prices that emphasises machine learning and deep learning models." *International Journal of Information Technology* 15.8 (2023): 4291-4313.
- 4) An, Wuyue, Lin Wang, and Yu-Rong Zeng. "Text-based soybean futures price forecasting: A two-stage deep learning approach." *Journal of Forecasting* 42.2 (2023): 312-330.
- 5) Kumari, Mamta, et al. "Object-based machine learning approach for soybean mapping using temporal sentinel-1/sentinel-2 data." *Geocarto International* 37.23 (2022): 6848-6866.
- 6) Sharma, Abhinav, et al. "Machine learning applications for precision agriculture: A comprehensive review." *IEEE Access* 9 (2020): 4843-4873.
- 7) Oikonomidis, Alexandros, Cagatay Catal, and Ayalew Kassahun. "Hybrid deep learning-based models for crop yield prediction." *Applied artificial intelligence* 36.1 (2022): 2031822.
- 8) BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. *Projeções do Agronegócio: Brasil 2020/21 a 2030/31: projeções de longo prazo*. Brasília, DF, 2021. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/politica-agricola/todas-publicacoes-de-politica-agricola/projecoes-do-agronegocio>.
- 9) Barichello, Diana Luisa. "Agricultura 5.0: inovações, trajetória da agricultura brasileira e os desafios climáticos." (2023).
- 10) Fava Neves, Marcos, et al. "AGRICULTURE 6.0: A NEW PROPOSAL FOR THE FUTURE OF AGRIBUSINESS." *Environmental & Social Management Journal/Revista de Gestão Social e Ambiental* 17.9 (2023).
- 11) Vogel, Elisabeth, et al. "The effects of climate extremes on global agricultural yields." *Environmental Research Letters* 14.5 (2019): 054010.
- 12) Xiong, Tao, Chongguang Li, and Yukun Bao. "Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: Evidence from the vegetable market in China." *Neurocomputing* 275 (2018): 2831-2844.
- 13) CORDEIRO, L. A. M.; VILELA, L.; MARCHÃO, R. L.; KLUTHCOUSKI, J.; MARTHA JÚNIOR, G. B. Integração lavoura-pecuária e integração lavoura-pecuária-floresta: estratégias para intensificação sustentável do uso do solo. *Cadernos de Ciência & Tecnologia*, v. 32, n. 1/2, p. 15-43, 2015.

- 14) Elavarasan, Dhivya, et al. "Forecasting yield by integrating agrarian factors and machine learning models: A survey." *Computers and electronics in agriculture* 155 (2018): 257-282.
- 15) Chemura, Abel, Bernhard Schauburger, and Christoph Gornott. "Impacts of climate change on agro-climatic suitability of major food crops in Ghana." *PLoS One* 15.6 (2020): e0229881.
- 16) Bouras, El Houssaine, et al. "Cereal yield forecasting with satellite drought-based indices, weather data and regional climate indices using machine learning in Morocco." *Remote Sensing* 13.16 (2021): 3101.
- 17) Cadorin, Vinicius Correia. "Avaliação de Valor: estudo de caso da Kepler Weber SA." (2023).
- 18) Newlands, Nathaniel K., et al. "An integrated, probabilistic model for improved seasonal forecasting of agricultural crop yield under environmental uncertainty." *Frontiers in Environmental Science* 2 (2014): 17.
- 19) Ye, Long, et al. "Projecting Australia's forest cover dynamics and exploring influential factors using deep learning." *Environmental Modelling & Software* 119 (2019): 407-417.
- 20) Feng, Puyu, et al. "Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia." *Agricultural Systems* 173 (2019): 303-316.
- 21) Wang, Yumiao, et al. "Combining multi-source data and machine learning approaches to predict winter wheat yield in the conterminous United States." *Remote Sensing* 12.8 (2020): 1232.
- 22) Neves, Marcos Fava. *O Agro Brasileiro e Internacional: Cenário Atual e Perspectivas*. Diss. Universidade de São Paulo, 2014.
- 23) Sharma, Rohit, et al. "A systematic literature review on machine learning applications for sustainable agriculture supply chain performance." *Computers & Operations Research* 119 (2020): 104926.
- 24) Ayankoya, Kayode, Jean H. Greyling, and Andre P. Calitz. "Real-time grain commodities price predictions in South Africa: A big data and neural networks approach." *Agrekon* 55.4 (2016): 483-508.
- 25) Coulibaly, Solemane, et al. "Deep learning for precision agriculture: A bibliometric analysis." *Intelligent Systems with Applications* 16 (2022): 200102.