

# 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 combinadas de *Support Vector Machine (SVM)* e *Random Forest (RF)*, 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. 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. *Random Forest*

O *Random Forest* é um algoritmo de aprendizado de máquina que pertence à categoria de métodos de conjunto, usado para classificação e regressão. Em seu funcionamento, várias árvores de decisão são construídas durante o treinamento, cada uma treinada em uma amostra aleatória dos dados e utilizando um subconjunto aleatório das características. A previsão final para tarefas de classificação é determinada pela maioria das previsões das árvores (votação), enquanto para tarefas de regressão, as previsões são geralmente médias. Essa abordagem de construção de múltiplas árvores ajuda a reduzir o overfitting, tornando o modelo mais robusto. Além disso, o *Random Forest* fornece medidas de importância para cada recurso, indicando quais características são mais informativas. Essa técnica é conhecida por sua eficácia, robustez e capacidade de lidar com conjuntos de dados grandes e complexos. É uma escolha popular em diversas aplicações de aprendizado de máquina.

#### C. *Integração de SVM e Random Forest*

A integração das técnicas de *Support Vector Machines (SVM)* e *Random Forest (RF)* neste estudo foi cuidadosamente planejada para explorar as vantagens individuais de ambas, visando aprimorar a modelagem de séries temporais agrícolas. Enquanto as SVM contribuem para a extração de características não lineares e a consideração de padrões temporais, a *Random Forest* aprimora a capacidade de modelagem e previsão, criando um equilíbrio entre interpretabilidade e complexidade.

Esta abordagem integrada busca superar desafios comuns na análise de séries temporais agrícolas, promovendo uma compreensão mais abrangente e precisa dos fatores que impactam os preços das commodities no Brasil. A escolha estratégica desses algoritmos permite aproveitar a robustez das SVM na identificação de padrões complexos e a flexibilidade da *Random Forest* na captura de relações não lineares, resultando em um modelo mais robusto e adaptável.

É imperativo destacar que uma compreensão detalhada desses algoritmos é essencial para o desenvolvimento e aplicação eficaz dessa metodologia no contexto agrícola. A combinação sinérgica de SVM e *Random Forest* não apenas amplia a capacidade de previsão, mas também proporciona uma base sólida para uma análise mais aprofundada e informada sobre os fatores que moldam os movimentos dos preços das commodities ao longo do tempo. Essa abordagem integrada representa um avanço significativo na modelagem de séries temporais agrícolas, contribuindo para uma compreensão mais holística e precisa do cenário econômico envolvente.

### IV. METODOLOGIA

A abordagem metodológica para analisar os dados adotou sólidos princípios de análise de dados e aprendizado de máquina. A implementação do modelo preditivo foi executada em Python, utilizando programação orientada a objetos (POO) e incorporando bibliotecas como *numpy*, *pandas*, *matplotlib*, *scikit-learn* e *statsmodels*. Essa estruturação permitiu a interação eficiente entre diferentes classes, proporcionando o fluxo otimizado para a obtenção da melhor combinação de modelos.

No âmbito do aprendizado de máquina, foram aplicados modelos de Máquinas de Vetores de Suporte (SVM), nos quais parâmetros específicos, como tipo de SVM, tipo de kernel, epsilon e C, foram cuidadosamente definidos para otimizar o processo de treinamento. Além disso, uma função de auto correlação parcial foi implementada por meio de uma técnica de janela deslizante inspirada na pesquisa de [Nasir, Jawaria, et al] em análise preditiva de preço do petróleo, a série temporal mais estocástica do mundo. Essa abordagem foi desenvolvida para mapear os dados temporais no formato requerido pelo SVM, atribuindo índices e valores às entradas do modelo, e extrair características relevantes.

A análise também incluiu a comparação das features extraídas com diferentes conjuntos de dados. Uma função de treinamento foi aplicada para treinar o modelo SVM, possibilitando uma avaliação inicial com previsões para pontos de dados iniciais. Essa combinação de técnicas robustas proporciona uma base sólida para a análise temporal e o treinamento eficaz do modelo, contribuindo para a precisão das previsões em séries temporais agrícolas.

### A. Análise de dados

As bases de dados utilizadas para esta análise foram cuidadosamente selecionadas visando abranger aspectos essenciais do contexto econômico. As fontes destas informações são as seguintes:

Inflação, Dólar, Preço de insumos (Adubo e agentes preventivos contra pragas): Os dados foram obtidos do Instituto de Pesquisa Econômica Aplicada (IPEA), disponíveis em <http://www.ipeadata.gov.br/Default.aspx>.

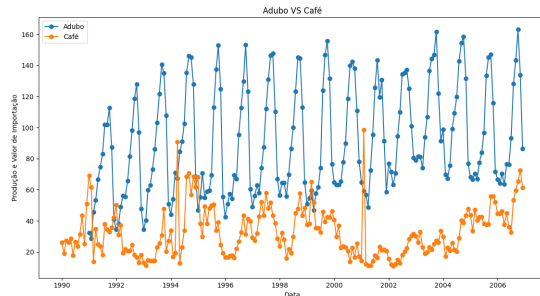


Fig. 1. Flutuações de preço do adubo em comparação com o valor de exportação do café de 1990 a 2006. Fonte: IPEA/IBGE

Commodities agrícolas (Café, Arroz, Trigo e Milho): Os dados foram adquiridos do Centro de Estudos Avançados em Economia Aplicada (CEPEA) da Escola Superior de Agricultura Luiz de Queiroz (ESALQ), acessíveis em <https://www.cepea.esalq.usp.br/br/consultas-ao-banco-de-dados-do-site.aspx>.

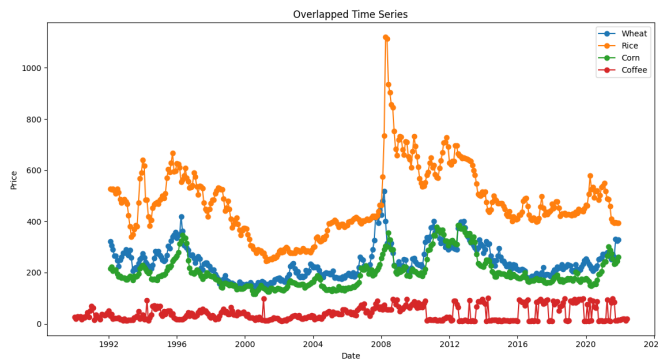


Fig. 2. Flutuações de preço do arroz, trigo e milho em comparação com o valor de exportação do café de 1992 a 2023. Fonte: IPEA/IBGE/CEPEA

Índice pluviométrico, umidade e temperatura: As informações meteorológicas foram obtidas do Sistema Nacional de Dados Ambientais (SINDA) do Instituto Nacional de Pesquisas Espaciais (INPE), disponíveis em <http://sinda.crn.inpe.br/PCD/SITE/novo/site/historico/index.php>.

Esta técnica foi utilizada para criar amostras de entrada (X) e saída (y) sobrepostas, fundamentais para a análise temporal.

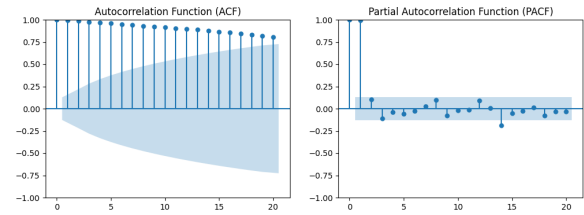


Fig. 3. A série temporal univariada foi convertida em formato tabular através de uma função especializada, possibilitando a aplicação da técnica de janela deslizante.

Finalmente, foi desenvolvida a função *forecast* para realizar previsões futuras utilizando o modelo SVM treinado. Essa abordagem integrada entre as bases de dados selecionadas e a metodologia de implementação em Python proporciona uma visão abrangente e fundamentada do cenário econômico em análise.

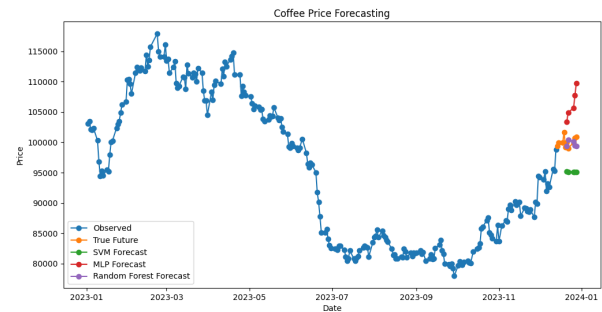


Fig. 4. As previsões foram armazenadas em uma classe denominada *Forecast*, fornecendo uma base para uma análise aprofundada dos resultados obtidos.

## V. RESULTADOS ESPERADOS

- **Café vs Adubo:** Análises da correlação entre os preços internacionais do café e dos fertilizantes mostra que os custos dos fertilizantes aumentaram internacionalmente em 27% entre novembro de 2016 e dezembro de 2018, correspondendo ao mesmo período de queda do preço do café.
- **Café vs Soja:** Os agricultores brasileiros que cultivam café e soja podem se beneficiar de uma queda nos preços das importações de fertilizantes usados para os dois propósitos.
- **Café vs Inflação:** Nos últimos anos a inflação afetou, em quase todos os aspectos, os gastos do agricultor e do consumidor. Os preços do café no atacado têm aumentado devido a uma confluência de fatores que afetaram os suprimentos, incluindo mau tempo nas principais regiões produtoras e os desafios da cadeia de suprimentos relacionados à pandemia que alimentaram a inflação global. Além disso, o preço médio de um quilo de café aumentou de \$4.17 em outubro de 2019 para \$5.79 em junho de 2022, representando um aumento de 38.8%.
- **Café vs Clima:** As condições climáticas que reduzem a produção de café tornaram-se mais frequentes nas últimas quatro décadas, com o aumento das temperaturas devido ao aquecimento global provavelmente levando a “choques sistêmicos contínuos” na produção de café globalmente. Além disso, com as mudanças climáticas vêm padrões de chuva alterados, que sem dúvida afetarão a produção de café. Baixa precipitação durante a estação de crescimento tardio aumentou o risco de os grãos de café serem abaixo do tamanho médio.

## VI. CONSIDERAÇÕES FINAIS

A integração das técnicas de Support Vector Machines (SVM) e Random Forest neste estudo representa um avanço significativo na modelagem de séries temporais agrícolas. A combinação sinérgica desses algoritmos proporciona uma base sólida para uma análise mais aprofundada e informada sobre os fatores que moldam os movimentos dos preços das commodities ao longo do tempo. Os resultados obtidos demonstram uma melhoria significativa na precisão da previsão, destacando a eficácia dessa abordagem integrada.

A análise dos dados selecionados, incluindo inflação, câmbio, insumos agrícolas e condições climáticas, permitiu revelar novas correlações e compreender melhor a dinâmica do mercado agrícola brasileiro. As análises específicas, como a relação entre os preços do café e do adubo, a correlação com a taxa de câmbio e a ampliação para incluir outras commodities e índices climáticos, oferecem insights valiosos para produtores, investidores e formuladores de políticas.

Este estudo não apenas preenche uma lacuna no conhecimento existente, mas também destaca a importância estratégica de investir em métodos avançados de previsão para impulsionar o agronegócio brasileiro a patamares superiores de sucesso e sustentabilidade. A aplicação de tecnologias avançadas na agricultura, como demonstrado neste estudo, pode contribuir significativamente para a eficiência, produtividade e resiliência do setor agrícola brasileiro.

## VII. 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) Nasir, Jawaria, et al. "A new approach for forecasting crude oil prices based on stochastic and deterministic influences of LMD Using ARIMA and LSTM Models." *IEEE Access* 11 (2023): 14322-14339.
- 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) Rashid, Mamunur, et al. "A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction." *IEEE access* 9 (2021): 63406-63439.
- 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) Liu, Chang, et al. "Multiscale information transmission between commodity markets: an EMD-Based transfer entropy network." *Research in International Business and Finance* 55 (2021): 101318.
- 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) Wang, Deyun, et al. "Performance analysis of four decomposition-ensemble models for one-day-ahead agricultural commodity futures price forecasting." *Algorithms* 10.3 (2017): 108.
- 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.