

XGBoost e a prática orçamentária do MAPA: influência no IDH e PIB do Agronegócio

JORGE LUIZ DE CASTRO¹; Prof. Miguel Ângelo Lellis Moreira ²

¹Faculdade de Humanidades Pedro Segundo. Bacharel em Matemática. SGAN 911 GREEN PARK, Bloco D – 102, Brasília, Distrito Federal - Brasil

² Universidade Federal Fluminense. Doutorado em Engenharia de Produção – Pesquisa Operacional. R. Passo da Pátria, São Domingos, 24210-240, Niterói - RJ, Brasil.

*autor correspondente: jorluc@me.com

XGBoost e a prática orçamentária do MAPA: influência no IDH e PIB do Agronegócio

Resumo

A gestão eficiente de recursos públicos é um desafio governamental global. O avanço em análise de dados e aprendizado de máquina oferece novas oportunidades para melhorar o planejamento orçamentário. O XGBoost surge como técnica promissora para analisar grandes volumes de dados e extrair insights para decisões no setor público. Este estudo investigou a aplicação do XGBoost na análise de despesas e identificação dos diversos fatores que compõem o planejamento e execução orçamentária. A pesquisa se justificou pela necessidade de aprimorar a gestão de recursos públicos, buscando maior eficiência e transparência nos processos de tomada de decisão. O objetivo geral do trabalho foi avaliar a viabilidade dessa técnica para a previsão de resultados e análise de elementos impactantes no contexto do orçamento público. A metodologia empregada consistiu na aplicação do algoritmo a dados abertos governamentais de fontes públicas entre 2017 e 2022, com o Índice de Desenvolvimento Humano e a variação percentual do PIB do Agronegócio como variáveis dependentes do modelo. A análise incluiu otimização de hiperparâmetros via GridSearchCV e avaliação da importância das features. Os resultados demonstraram o potencial do XGBoost em aprimorar decisões baseadas em dados sobre o manuseio de recursos, revelando a influência de variáveis específicas sobre as despesas e fornecendo insights valiosos para a otimização do planejamento orçamentário. As conclusões apontam para a importância da integração de técnicas de aprendizado de máquina na gestão, contribuindo para a eficiência, transparência e embasamento das práticas orçamentárias.

Abstract

Efficient management of public resources is a global governmental challenge. Advancements in data analysis and machine learning offer new opportunities to improve budget planning. XGBoost emerges as a promising technique to analyze large volumes of data and extract insights for public sector decisions. This study investigated the application of XGBoost, an advanced machine learning technique, in expense analysis and identification of various factors that compose budget planning and execution. The research was justified by the need to enhance public resource management, seeking greater efficiency and transparency in decision-making processes. The general objective was to evaluate this technique's viability for result prediction and analysis of impactful elements in the public budget context. The methodology involved applying the algorithm to open government data from public sources between 2017 and 2022, with the Human Development Index and Agribusiness GDP percentage variation as the model's dependent variables. The analysis included hyperparameter optimization via GridSearchCV and feature importance evaluation. Results demonstrated XGBoost's potential to improve data-driven decisions on resource handling, revealing specific variables' influence on expenses and providing valuable insights for budget planning optimization. Conclusions point to the importance of integrating machine learning techniques in management, contributing to efficiency, transparency, and substantiation of budgetary practices.

Keywords: Budget, Machine learning, Extreme Gradient Boosting, XGBoost, Agricultural sector, Economic development, HDI, GDP Agribusiness.

Introdução

A gestão empresarial moderna exige práticas que garantam eficiência operacional e sustentabilidade financeira. Neste contexto, o emprego de técnicas de inteligência artificial na interpretação de informações emerge como ferramenta crucial para antecipar e administrar despesas (Minenhle, 2022). Esta pesquisa investiga a implementação de técnicas de aprendizado de máquina avançados na estimativa de gastos em iniciativas diversas, bem como na detecção de elementos que impactam o planejamento e a execução tanto na esfera governamental quanto no âmbito empresarial privado (Haytham & Elmousalami, 2019).

O XGBoost, algoritmo baseado em árvore de decisão, oferece uma abordagem robusta para previsão e otimização orçamentária (Capone, 2024). Capone também demonstra que seu uso melhora significativamente a precisão das estimativas, destacando-se pela capacidade de processar grandes volumes de dados e identificar padrões complexos.

Nesse contexto, a pesquisa de Zainab et al. (2022) destaca a relevância dessas tecnologias na projeção de custos e modelagem de fatores-chave em projetos. O "gradient boosting" permite examinar variáveis influentes, como dotação inicial, despesas liquidadas e aspectos socioeconômicos, gerando insights valiosos para gestão financeira e execução de iniciativas (Puhach, 2023).

No setor público, a utilização dessas tecnologias pode levar a uma alocação mais eficiente de recursos, melhorando a qualidade dos serviços prestados à população. Tahu et al. (2024) destacam a influência da receita regional e dos fundos de alocação geral no crescimento econômico, enfatizando o papel das despesas de capital como variável interveniente. No setor privado, as empresas podem otimizar seus processos de planejamento financeiro, reduzindo custos e aumentando a competitividade.

Várias pesquisas realizadas demonstram que a produtividade agrícola é fundamental para a economia. No setor agropecuário, seu aumento alivia a pressão sobre recursos naturais escassos (Torres, 2023). Torres também indica que os investimentos em pesquisa e desenvolvimento são vitais para elevar a produtividade agrícola e impulsionar a inovação, contribuindo significativamente para o crescimento do Produto Interno Bruto.

A relação entre o desenvolvimento do setor agrícola e o Índice de Desenvolvimento Humano [IDH] também tem sido estudada. No contexto brasileiro, Piras, Wesz e Ghinoi (2021) observaram que a expansão da soja está associada a melhorias no IDH em municípios, embora os benefícios sejam mais evidentes na dimensão de renda do que em educação e longevidade.

A distribuição desigual de terras no Brasil, um problema histórico, tem consequências duradouras que impactam negativamente o IDH, com efeitos mais pronunciados na longevidade e nos níveis de renda da população, conforme apontado por Ondetti (2020). Esta

constatação evidencia a importância de implementar políticas públicas que visem o fortalecimento e apoio aos pequenos produtores e a busca por um equilíbrio entre aumento da produtividade, preservação ambiental e progresso social.

A análise do setor agrícola brasileiro ilustra a importância de uma gestão orçamentária eficaz. O investimento em áreas como pesquisa e desenvolvimento pode impulsionar a produtividade e, conseqüentemente, o crescimento econômico. Além disso, a relação entre o progresso agrícola e o IDH enfatiza a complexidade da relação entre desenvolvimento agrícola e bem-estar social, sugerindo a necessidade de análises mais apurada para definir políticas públicas eficazes (Soleman et al., 2022).

O uso de machine learning representa uma evolução significativa em relação aos métodos tradicionais, proporcionando base científica para decisões estratégicas e alocação eficiente de recursos. Sua relevância é amplificada pela crescente disponibilidade de dados abertos e necessidade de transparência na administração pública (Tores, 2023).

Este trabalho busca contribuir para o avanço do conhecimento nesta área crucial, oferecendo uma visão abrangente dos desafios e oportunidades na gestão financeira e orçamentária. A aplicação de algoritmos como o XGBoost na previsão orçamentária tem demonstrado resultados promissores, permitindo uma análise mais profunda e precisa dos fatores que influenciam as despesas.

À medida que governos e empresas buscam maneiras de otimizar seus processos de gestão financeira, a integração de tecnologias de análise de dados se torna cada vez mais relevante. Essas ferramentas não apenas melhoram a precisão das previsões, mas também fornecem insights valiosos para a tomada de decisões estratégicas. A aplicação de técnicas avançadas na gestão orçamentária representa uma oportunidade significativa para melhorar a eficiência e eficácia tanto no setor público quanto no privado. Ao fornecer uma base mais sólida para o planejamento financeiro e a alocação de recursos, essas tecnologias têm o potencial de impulsionar o desenvolvimento econômico e social, promovendo uma gestão mais transparente e responsável dos recursos disponíveis.

Material e Métodos

O estudo adotou uma abordagem metodológica concisa para examinar o orçamento do setor agrícola e de desenvolvimento rural no Brasil, concentrando-se em três aspectos principais: a analogia entre tipos de ações orçamentárias e conceitos financeiros privados, a análise do desempenho de subfunções críticas e a avaliação da diversidade temática do orçamento.

Tahu et al. (2024) e Puhach (2023) fornecem insights para estabelecer, no Orçamento Fiscal brasileiro, uma analogia entre ações orçamentárias e conceitos da gestão financeira privada. Na Lei Orçamentária Anual, Atividades foram equiparadas aos gastos operacionais recorrentes (“Operational Expenditure [OPEX]”), enquanto Projetos foram comparados aos investimentos em bens de capital para expansão ou melhoria operacional (“Capital Expenditure [CAPEX]”). Operações Especiais formaram uma categoria única sem paralelo direto no setor privado. Essa comparação, embora não perfeita devido às particularidades do setor público, ofereceu uma perspectiva valiosa para análises, permitindo uma melhor compreensão da alocação de recursos e seus impactos no contexto orçamentário brasileiro.

A coleta dos dados foi realizada a partir de bases abertas de entidades governamentais, para o período de 2018 a 2022. São elas: Acompanhamento Físico-Financeiro das Ações Orçamentárias, Orçamento Fiscal e da Seguridade Social das bases de dados do Orçamento Federal, Censo Agropecuário de 2017 do IBGE, a série histórica do Produto Interno Bruto [PIB] do agronegócio brasileiro e do dólar americano [USD] das bases do Cepea/Esalq/USP-CNA. Um rigoroso processo de limpeza, clusterização e classificação dos dados foi realizado. Observações com valores faltantes (“null” e “NA”) foram tratadas ou removidas e aplicou-se “one-hot encoding” para atributos categóricos selecionados.

A análise descritiva das variáveis categóricas e numéricas está resumida nas Tabelas 1 e 2, respectivamente, e visa fornecer uma visão estatística de negócios sobre o setor agropecuário no Brasil. O dicionário de dados detalhado encontra-se na Tabela 8 no Apêndice. Os resultados dessa análise embasam discussões sobre a distribuição regional de recursos, centralização de decisões no MAPA e foco em projetos específicos. Esses aspectos são essenciais para entender a eficácia e o direcionamento das políticas agropecuárias no Brasil.

Tabela 1 – Análise descritiva das variáveis categóricas (qualitativas)

Variável	Categorias	Moda
regiao	Sul, Sudeste, Norte, Nordeste, Centro-Oeste e BR (abrangência nacional)	Nordeste
uf	as 27 unidades federativas e BR (abrangência nacional)	BR
unidadeorçamentaria	CONAB, EMBRAPA, INCRA, SFB e MAPA (Administração Direta)	MAPA
subfuncao	são 22 subfunções distintas	Promoção da Produção Agropecuária
tipoacao	Atividade, Projeto e Operação Especial	Atividade

Fontes: GOVERNO DO BRASIL, IBGE, CEPEA e CNA no período de 2018 a 2022

Tabela 2 – Análise descritivas das variáveis numéricas (quantitativas)

Variável	Min.	Max.	Média	1º quartil	Mediana	3º quartil
Autorizado	0	2,41E+09	63757803	350000	1555000	17742156
Pago	0	1,18E+09	24929729	0	23750	983903,3
Autorizado_USD	0	4,68E+08	13687846	74074,07	355329,9	3856394
Pago_USD	0	2,19E+08	5345220	0	5071,386	206511,2
IDH	0,676	0,859	0,756	0,726	0,757	0,774
Populacao	513466	2,14E+08	55513739	3471779	9558567	46282864
PIB	13369988	2,72E+09	3,57E+08	1E+08	2,74E+08	4,18E+08
Valor_PIB_Agronegocio	2675417	7,24E+08	85748300	21345487	66233679	1,06E+08
Valor_PIB_Agricola	1991742	5,37E+08	61930409	15688058	46960474	76308447
Valor_PIB_Pecuario	683675,5	1,91E+08	23817891	5760280	17074858	29336094
%_PIB_Agronegocio	0,2	0,266	0,24	0,205	0,253	0,259
%_PIB_Agricola	0,145	0,197	0,173	0,149	0,182	0,183
%_PIB_Pecuario	0,051	0,077	0,067	0,06	0,069	0,07
Total_Propriedade_Agricola	5246	5073324	1336178	80959	281688	762848
Populacao_Por_Propriedade	13,289	597,703	75,161	30,702	38,001	42,27
Log_Pop_Por_Propriedade	1,123	2,776	1,652	1,487	1,58	1,626
Norm_Pop_Por_Propriedade	-0,515	4,351	0	-0,37	-0,309	-0,274
Nao_Recebe_Orientacao_Tecnica	1211	4047881	1064868	71267	182811	704292
%_Nao_Recebe_Orientacao_Tecnica	0,231	0,965	0,773	0,754	0,798	0,886
Recebe_Orientacao_Tecnica	1293	1025443	271309,8	10205	33419	182283
%_Receber_Orientacao_Tecnica	0,035	0,769	0,227	0,114	0,202	0,246
Governo_Orientacao_Tecnica	1125	388077	102597,8	5484	18281	58551
Propria_Orientacao_Tecnica	132	316394	83437,05	3343	12592	64141
Cooperativas_Orientacao_Tecnica	14	251520	66784,2	522	3471	63410
Emp_Integradoras_Orientacao_Tecnica	8	134950	36110,7	172	1625	52534
Emp_Privadas_Orientacao_Tecnica	12	28302	7536,082	102	835	8968
ONG_Orientacao_Tecnica	1	8662	2317,912	73	315	3544
Sistema_S_Orientacao_Tecnica	20	7680	2010,234	104	354	1457
Outra_Origem__Orientacao_Tecnica	35	52117	13779,73	750	2089	8611

Fontes: GOVERNO DO BRASIL, IBGE, CEPEA e CNA no período de 2018 a 2022

A análise exploratória dos dados foi conduzida utilizando métodos visuais e por meio do desenvolvimento de matriz de correlação (disponível na Figura 4 do Apêndice), visando identificar relações significativas entre as variáveis quantitativas e as variáveis categóricas, que foram codificadas em variáveis “dummy” utilizando o método de codificação “one-hot”. Além disso, uma análise dos coeficientes de correlação foi realizada, considerando valores absolutos acima de 0,7 na matriz de correlação como indicativos de alta correlação e potencial multicolinearidade.

Para a seleção de “features”, foi implementada a técnica de “Recursive Feature Elimination [RFE]”, utilizando um modelo de regressão linear como base. A seleção considerou a relevância para os alvos (IDH e variação percentual do Produto Interno Bruto do Agronegócio [PIB-AGRO]), evitou multicolinearidade e buscou incluir uma variedade de tipos de dados que capturassem diferentes aspectos do problema econômico. Técnicas de normalização foram aplicadas aos dados para garantir que todas as variáveis estivessem na mesma escala, facilitando a comparação e melhorando o desempenho dos modelos subsequentes.

Como principal ferramenta de modelagem, utilizou-se o XGBoost, dada sua eficácia em problemas de regressão complexos. Para otimizar o desempenho do modelo, a técnica de GridSearchCV foi empregada para ajuste de hiperparâmetros. Definiu-se uma grade de hiperparâmetros a serem otimizados, incluindo número de estimadores, profundidade máxima das árvores, taxa de aprendizado e fração de sub amostras. O GridSearchCV foi configurado para realizar uma busca exaustiva através dessa grade, utilizando validação cruzada para avaliar o desempenho de cada combinação de hiperparâmetros. Para a avaliação do modelo, os dados foram divididos em conjuntos de teste (20%) e treinamento (80%). As métricas “Coefficient of Determination” [R^2], “Mean Squared Error [MSE]”, “Root Mean Squared Error [RMSE]” e “Mean Absolute Error [MAE]” foram utilizadas para avaliar o desempenho do modelo.

Após o treinamento, a importância relativa de cada “feature” utilizada no XGBoost foi extraída e analisada, permitindo identificar quais variáveis tiveram maior impacto nas previsões do modelo. Uma análise geográfica e temática do orçamento foi realizada, examinando sua distribuição em todas as regiões do Brasil e analisando os programas e ações descritos, com especial atenção às perspectivas do Censo Agropecuário de 2017. Esta abordagem heterogênea permitiu identificar padrões e fatores relevantes na gestão financeira pública do setor agrícola e de desenvolvimento rural, além de oferecer insights valiosos para aprimorar futuras decisões (“data-driven decision”) e alocação de recursos.

Utilizar a combinação de técnicas estatísticas clássicas com modelos preditivos avançados proporcionou uma compreensão mais profunda e dinâmica dos fatores que influenciam os gastos públicos neste setor crucial para a economia brasileira. Todo o código-fonte utilizado neste estudo, bem como os conjuntos de dados (respeitando políticas de privacidade), foram disponibilizados em repositório público (Castro, 2024), garantindo a transparência e replicabilidade da pesquisa. Esta metodologia visou promover uma abordagem orientada a dados para a gestão financeira pública, alinhando-se com as tendências mais recentes em econometria e ciência de dados aplicadas ao setor público, de forma comparativa ao setor privado.

Resultados e Discussão

A etapa inicial de preparação dos dados revelou um cenário ideal para o estudo. Todas as 465 observações do “dataset” original foram aproveitadas na análise, demonstrando uma coleta de informações excepcionalmente bem executada, livre de valores ausentes ou inválidos. Esta base sólida proporcionou um alicerce confiável para as etapas subsequentes da pesquisa.

Na análise da matriz de correlação, buscando identificar multicolinearidade na previsão de %_PIB_Agronegocio, o estudo observou correlações significativas. As mais fortes, com valores absolutos superiores a 0,7, incluíram o PIB e seus componentes setoriais, variáveis orçamentárias em diferentes moedas, e indicadores específicos do agronegócio. Notavelmente, a relação entre %_PIB_Agronegocio e %_PIB_Agricola apresentou uma correlação de 0,980, indicando uma forte interdependência.

A seleção de “features” para o modelo XGBoost considerou a robustez do algoritmo à multicolinearidade, permitindo a inclusão de variáveis potencialmente relevantes, mesmo com correlações moderadas entre si. O estudo selecionou um conjunto abrangente de indicadores, incluindo o PIB geral, o valor do PIB-AGRO, o IDH, a distribuição populacional por propriedade, o acesso à orientação técnica, variáveis regionais, unidades orçamentárias, tipos de ações de Projetos (análogo ao CAPEX) e de Atividades (análogo ao OPEX), e valores autorizados e pagos.

A justificativa para esta seleção baseou-se na inclusão de variáveis que, apesar de baixas correlações lineares, poderiam apresentar relações não-lineares importantes com %_PIB_Agronegocio. O estudo buscou abranger aspectos econômicos, sociais, regionais e de políticas públicas para fornecer um contexto abrangente, evitando a inclusão de variáveis com altíssima multicolinearidade entre si.

A análise inicial desse conjunto de variáveis revela desafios específicos na gestão orçamentária das subfunções científicas e tecnológicas, como a tendência de declínio nos investimentos e a diminuição da sua participação no orçamento total do ministério. Isso pode indicar uma mudança nas prioridades ou restrições orçamentárias específicas para a área tecnológica. A melhoria na eficiência da execução, com uma relação entre o que é empenhado e pago mais equilibrada, é um ponto positivo, mas ocorre em um contexto de redução geral dos recursos. Isso pode sugerir uma otimização forçada devido à limitação de recursos. A aplicação de técnicas de aprendizado de máquina na análise de dados é fundamental para uma gestão orçamentária eficaz, elemento crítico para superar os desafios vindouros nos setores de agricultura e pecuária.

Na Figura 1, é possível observar uma tendência de crescimento nos valores de 2018 a 2021, seguida por uma queda em 2022. O pico ocorreu em 2021 para a maioria dos estágios, tais como: dotação, autorizado e empenhado. Em todos os anos da amostra, o valor autorizado é maior que a dotação inicial, sendo a diferença mais significativa em 2019, com um aumento de cerca de R\$ 0,89 bilhões.

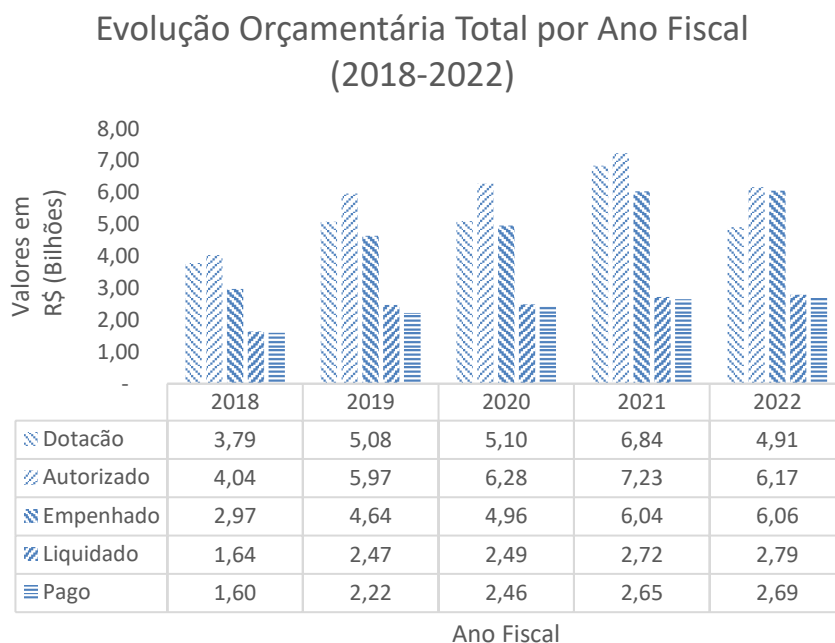


Figura 1. Cenário de avaliação entre os estágios do orçamento total do MAPA no período.

Fonte: Dados do Orçamento Federal - Orçamento Fiscal e da Seguridade Social (GOVERNO DO BRASIL, 2018 a 2022)

A execução orçamentária revela uma diferença consistente entre os valores empenhados e os valores efetivamente pagos. Em média, apenas cerca de 50% do valor empenhado é pago. No entanto, a razão entre pago e empenhado melhorou ao longo dos anos, indo de 53,9% em 2018 para 44,4% em 2022.

A variação ano a ano destaca o maior salto de 2020 para 2021, com um aumento significativo na dotação, de R\$ 5,10 bilhões para R\$ 6,84 bilhões. Em 2022, houve uma queda na dotação, mas os níveis de empenho e pagamento mantiveram-se altos. O padrão de execução orçamentária segue uma lógica consistente, onde a dotação é menor que o autorizado, que por sua vez é maior que o empenhado, que é maior que o liquidado e é aproximadamente igual ao que é pago.

A eficácia na utilização do orçamento mostra que a porcentagem do orçamento autorizado e o que é efetivamente pago variou de 39,6% em 2018 a 43,6% em 2022. A análise dos dados evidencia discrepância entre o planejado e o executado, bem como a dificuldade em converter empenhos em pagamentos efetivos. Além disso, demonstram uma tendência

de crescimento do orçamento ao longo dos anos, com exceção de 2022, o que pode indicar mudanças nas prioridades governamentais ou restrições orçamentárias.

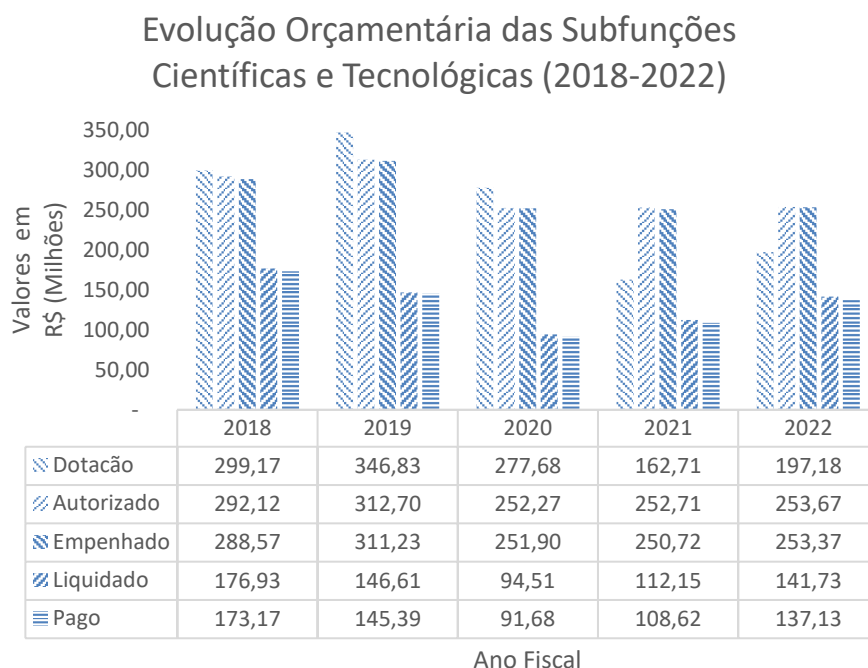


Figura 2. Cenário de avaliação entre os estágios do orçamento dos valores das subfunções da amostra (científica/tecnológica) do orçamento do MAPA no período.
Fonte: Dados do Orçamento Federal - Orçamento Fiscal e da Seguridade Social (GOVERNO DO BRASIL, 2018 a 2022)

A Figura 2 e a Figura 3 mostram uma segunda análise, de relações e tendências orçamentárias, das subfunções de aspectos científicos e tecnológicos, nomeadamente: Desenvolvimento Científico [DC], Desenvolvimento Tecnológico e Engenharia [DTE] e Difusão do Conhecimento Científico e Tecnológico [DCCT] do MAPA em suas unidades orçamentárias.

A tendência geral dos valores orçamentários para as subfunções tecnológicas entre 2018 e 2022 mostra um declínio. O pico é observado em 2019 para a dotação, com R\$ 346,83 milhões, mas os valores autorizados e empenhados reduziram a partir desse ano.

Em relação à dotação e ao autorizado, diferentemente do orçamento total, nas subfunções tecnológicas, o valor autorizado é geralmente menor ou igual à dotação inicial. A partir de 2021, o valor autorizado se estabilizou em torno de R\$ 250 milhões, mesmo com variações na dotação.

A execução orçamentária revela uma diferença consistente entre os valores empenhados e os valores que são efetivamente pagos, porém essa diferença diminuiu ao longo dos anos. Em 2022, o valor empenhado é igual ao autorizado, ambos em torno de R\$ 253 milhões, indicando uma execução completa do orçamento autorizado.

Variação Percentual dos Valores das Subfunções Científicas e Tecnológicas (2018-2022)

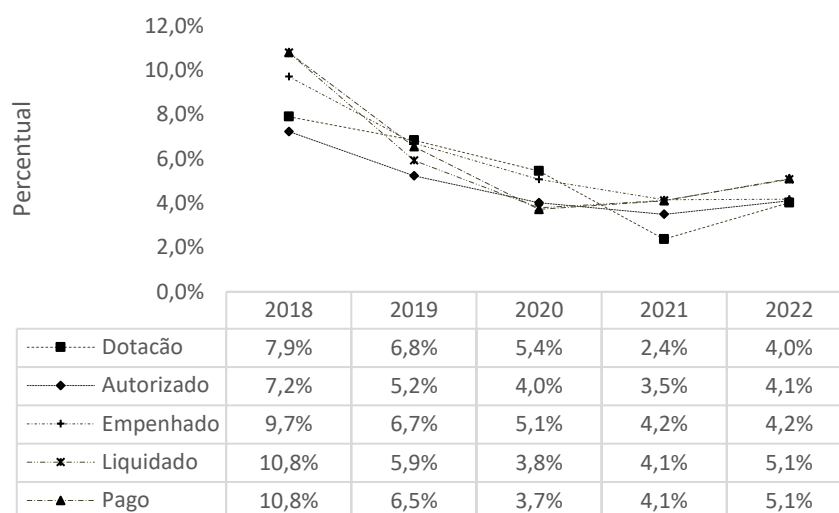


Figura 3. Cenário de variação percentual entre os estágios do orçamento dos valores das subfunções da amostra (científica/tecnológica) em relação ao orçamento total do MAPA no período 2018/2022.

Fonte: Dados do Orçamento Federal - Orçamento Fiscal e da Seguridade Social (GOVERNO DO BRASIL, 2018 a 2022)

A eficiência na execução melhorou significativamente, com a razão entre pago e empenhado passando de 58,6% em 2018 para 56% em 2022. A variação ano a ano destaca o maior declínio de 2019 para 2020, com uma queda significativa em todos os valores. Houve uma leve recuperação nos valores liquidados e pagos de 2020 para 2022.

O padrão de execução segue uma lógica onde a dotação é maior ou igual ao autorizado, que é igual ao empenhado, que por sua vez é maior que o liquidado, que é igual ao valor pago. Nos últimos anos, de 2020 a 2022, os valores autorizados e empenhados foram idênticos.

Comparando com o orçamento total, enquanto este mostrou uma tendência de crescimento até 2021, a amostra analisada subfunções de aspectos científicos e tecnológicos (DC, DTE e DCCT), apresentou uma evidente tendência de declínio. A proporção destas subfunções no orçamento total diminuiu significativamente, de cerca de 7,2% em 2018 para cerca de 4,1% em 2022, considerando os valores autorizados (planejado que foi aprovado). Se for considerado os valores que foram pagos, ou seja, o planejamento efetivamente realizado, esse declínio percentual é ainda maior, de 10,6% em 2018 para 5,2% em 2022.

As observações revelam desafios na gestão orçamentária das subfunções analisadas, como declínio nos investimentos e redução de sua participação no orçamento total do

ministério. Isso pode indicar mudanças nas prioridades ou restrições orçamentárias na área científica e tecnológica. A melhoria na eficiência da execução, com uma relação entre empenhado e pago mais equilibrada, é um ponto positivo. Isso ocorre em um contexto de redução geral dos recursos, onde a pandemia certamente teve impacto, mas não é o único motivo para a redução do orçamento do MAPA, pois conforme Monteiro et al. (2021) “houve uma combinação de fatores econômicos, políticos e estratégicos”. Esta situação sugere uma otimização forçada devido à limitação de recursos. Tais tendências levantam questões sobre o impacto potencial na capacidade do MAPA de promover a inovação e modernização tecnológica.

Ampliando um pouco as fronteiras de análise é possível observar quanto o investimento em agricultura e pecuária tem se mostrado um catalisador fundamental para o desenvolvimento econômico e social em economias emergentes. De acordo com Diallo e Wouterse (2022), o investimento do governo em setores como infraestrutura, pesquisa e desenvolvimento agrícola tem sido crucial para aumentar o PIB agrícola e reduzir a pobreza rural na África. O estudo destaca que os investimentos públicos direcionados a essas áreas em estados menos desenvolvidos e dependentes da agricultura podem levar à melhoria da igualdade de renda regional, ao aumento da produtividade agrícola e ao alívio da pobreza rural.

Além do impacto direto no PIB, os investimentos em agricultura e pecuária também influenciam positivamente o IDH. Estudos em vários países mostram que o IDH tem um impacto positivo e significativo sobre o PIB, sugerindo uma relação recíproca entre desenvolvimento humano e crescimento econômico (Soleman et al., 2022). Isso indica que investimentos que melhoram a saúde, educação e renda não apenas beneficiam diretamente a população, mas também impulsionam o crescimento econômico geral.

Em suma, os investimentos em agricultura e pecuária têm um impacto positivo significativo tanto no PIB quanto no IDH, especialmente quando combinados com políticas direcionadas que abordem desigualdades regionais e setoriais, promovendo assim um desenvolvimento econômico e social mais equitativo e sustentável.

No Brasil o setor agrícola, especialmente nas cadeias produtivas de soja, milho e gado, tem sido um importante motor de crescimento econômico. No entanto, as políticas agrícolas têm beneficiado principalmente grandes produtores, levantando preocupações sobre a distribuição equitativa dos benefícios. Corcioli, Medina e Arrais (2022) revelam que 75% do crédito operacional do governo vai para produtores de soja, milho e gado, com a maior parte destinada aos produtores de soja, que são predominantemente grandes agricultores. Além disso, 76,1% da cadeia produtiva da soja no Brasil é controlada por empresas multinacionais

estrangeiras, sugerindo que recursos públicos estão indiretamente financiando grandes empresas em detrimento de pequenos agricultores locais e do agronegócio nacional.

O último Censo Agropecuário, (IBGE, 2017), revela uma significativa lacuna na assistência técnica ao setor agrícola brasileiro, com a maioria das propriedades não recebendo orientação. Há também notáveis disparidades regionais, com o Sul tendo uma cobertura muito melhor de orientação técnica que o Nordeste, proporcionalmente. O governo desempenha um papel crucial na provisão de assistência técnica, mas há espaço substancial para aumentar a participação de outros atores; como mostra a Tabela 3.

Tabela 3 - Propriedades Rurais com Acesso à Orientação Técnica por Região

	Brasil	%	Centro-Oeste	Norte	Sul	Sudeste	Nordeste
Total de Propriedades	5.073.324	100,0%	347.263	580.613	853.314	969.415	2.322.719
Total que Não Recebe Orientação Técnica	4.047.881	79,8%	265.213	520.262	438.669	691.822	2.131.915
Total que Recebe Orientação Técnica	1.025.443	20,2%	82.050	60.351	414.645	277.593	190.804
Por tipo de Orientação Técnica							
Governo	388.077	7,6%	20.508	40.224	124.015	88.905	114.425
Própria	316.394	6,2%	43.610	14.637	98.051	115.266	44.830
Cooperativas	251.520	5,0%	13.193	2.223	155.171	66.319	14.614
Empresas Integradoras	134.950	2,7%	4.884	1.625	110.162	14.229	4.050
Empresas Privadas	28.302	0,6%	3.037	822	17.735	4.692	2.016
ONG	8.662	0,2%	246	797	850	1.012	5.757
Sistema S	7.680	0,2%	1.457	719	1.694	1.848	1.962
Outra Origem	52.117	1,0%	3.231	2.725	15.085	17.901	13.175

Fonte: Censo Agropecuário 2017, IBGE

Apesar dos crescentes desafios, o setor agrícola tem demonstrado potencial para impulsionar o desenvolvimento regional. Piras, Wesz e Ghinoi (2021) observam que a expansão da soja está associada a melhorias no IDH em municípios brasileiros, embora os benefícios sejam mais evidentes na dimensão de renda do que em educação e longevidade.

Do ponto de vista ambiental, há iniciativas promissoras para conciliar produção e sustentabilidade. Silva et al. (2023) demonstram que a recuperação de pastagens degradadas no Cerrado pode aumentar a produtividade pecuária e reduzir as emissões de carbono. Reis et al. (2020) destacam o potencial dos sistemas integrados de lavoura-pecuária-floresta em promover a intensificação sustentável da agricultura no Cerrado. Essas abordagens oferecem caminhos para melhorar tanto o desempenho econômico quanto ambiental do setor agropecuário brasileiro.

Ondetti (2020) apontam que a desigualdade histórica na distribuição de terras no Brasil tem efeitos negativos de longo prazo sobre o IDH, afetando principalmente a expectativa de

vida e a renda. Isso ressalta a necessidade de políticas que promovam uma distribuição mais equitativa dos benefícios do crescimento agrícola, apoiem pequenos agricultores e agricultores familiares, e busquem um equilíbrio entre produtividade, sustentabilidade ambiental e desenvolvimento social. A reestruturação da política agrícola brasileira deve, portanto, visar não apenas o crescimento econômico, mas também o desenvolvimento humano e a sustentabilidade ambiental em longo prazo.

Na modelagem preditiva do percentual do PIB-AGRO com XGBoost Regressor, buscou-se um equilíbrio entre concisão e abrangência. Segundo Darmawan, Yuliana e Hadi, (2023) a seleção de variáveis e a mitigação de multicolinearidade e “overfitting” são cruciais para garantir a robustez na aplicação do modelo, visando otimizar a gestão das cadeias de fatores que influenciam o agronegócio.

Tabela 4 – Hiperparâmetros e Resultados do Modelo para % do PIB-AGRO

Melhores hiperparâmetros:

`{'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 200, 'reg_alpha': 0, 'reg_lambda': 0.1}`

Resultados para Previsão do modelo para '% PIB Agronegocio':

MSE: 9.142347787437232e-05

RMSE: 0.009561562522640968

MAE: 0.0035760367636491486

R2 Score: 0.8764343967963025

Fonte: Saida da execução do Modelo GXBoost Regressor para % do PIB-AGRO

O processo de otimização do modelo XGBoost, realizado através do GridSearchCV, resultou em um conjunto de hiperparâmetros que sugerem uma abordagem equilibrada entre complexidade e generalização (Darmawan, Yuliana e Hadi, 2023). A taxa de aprendizagem de 0,1, profundidade máxima de 5 e 200 estimadores, indicam um modelo capaz de capturar nuances importantes dos dados sem incorrer em “overfitting”.

A escolha dos parâmetros de regularização, com alpha igual a zero e lambda igual a 0,1, indica a utilização da Regularização L2 (Ridge) de forma moderada. Essa decisão é justificada pelo grande número de variáveis preditoras, 17 no total, e pela possibilidade de algumas delas serem redundantes ou pouco relevantes para o modelo. A aplicação da Regularização L2 contribui para a robustez do modelo, tornando-o menos suscetível a pequenas flutuações nos dados e prevenindo também o “overfitting”.

A avaliação de desempenho do modelo apresentou bons resultados. O coeficiente de determinação (R^2) de 0,8764 indica que aproximadamente 87,64% da variabilidade na contribuição para o % do PIB_AGRO é explicada pelas variáveis selecionadas. Este elevado poder explicativo confere credibilidade às conclusões derivadas do modelo. Adicionalmente,

os baixos valores de MSE, RMSE e MAE colaboram para a precisão das previsões, reforçando a confiabilidade dos resultados obtidos.

Tabela 5 - Importância relativa das “features” para o modelo % do PIB_AGRO

Seq.	Feature	Importância
1	%_Receber_Orientacao_Tecnica	0,281825
2	tipoacao_Projeto	0,215948
3	IDH	0,117489
4	Log_Pop_Por_Propriedade	0,105853
5	PIB	0,089093
6	unidadeorcamentaria_EMBRAPA	0,051075
7	regiao_Sudeste	0,039999
8	regiao_Nordeste	0,038909
9	Pago_USD	0,021143
10	unidadeorcamentaria_MAPA	0,010413
11	regiao_Norte	0,009173
12	regiao_Sul	0,008256
13	unidadeorcamentaria_INCRA	0,005694
14	regiao_Centro-Oeste	0,003296
15	unidadeorcamentaria_CONAB	0,001835
16	regiao_BR	0,000000
17	unidadeorcamentaria_SFB	0,000000

Fonte: Saída da execução do Modelo GXBoost Regressor para % do PIB-AGRO

A análise da importância relativa das “features” revelou insights valiosos sobre os fatores que mais influenciam o desempenho do agronegócio no contexto econômico brasileiro. A variável relacionada ao recebimento de orientação técnica emergiu como o fator mais crucial, respondendo por 28,18% da importância total. Este resultado enfatiza o papel fundamental da disseminação de conhecimento e suporte técnico no setor agrícola, sugerindo que investimentos nesse aspecto podem ter um impacto significativo no crescimento do agronegócio.

O segundo fator mais relevante, com 21,59% de importância, foi identificado como os investimentos em Projetos (análogo ao CAPEX). Esta constatação ressalta e reforça a necessidade de alocação de recursos para infraestrutura, ciência e tecnologia como elementos-chave para impulsionar a produtividade e, consequentemente, a contribuição do setor para o PIB nacional.

O Índice de Desenvolvimento Humano (IDH) emergiu como a terceira variável mais influente, seguido pela distribuição logarítmica da população por propriedade e pelo PIB geral. A presença do PIB e do IDH entre os principais fatores sugere que o sucesso do setor está intimamente ligado ao panorama econômico e social mais amplo do país.

A relevância da variável logarítmica da população por propriedades agrícolas indica que a estrutura fundiária e a distribuição demográfica nas áreas rurais têm um impacto considerável na produtividade e no valor gerado pelo agronegócio. Esta constatação pode ter implicações importantes para políticas de reforma agrária e planejamento territorial.

O estudo notou que algumas variáveis, como as relacionadas a regiões específicas do Brasil ou a certas unidades orçamentárias, demonstraram baixa ou nenhuma importância no modelo. Isto sugere que fatores locais ou institucionais específicos podem ter menos influência no desempenho do agronegócio quando comparados a indicadores socioeconômicos e técnicos mais abrangentes.

A construção do outro modelo XGBoost Regressor para o IDH também exigiu a exploração diligente das variáveis disponíveis, com o zelo de evitar multicolinearidade e “overfitting” (Darmawan, Yuliana e Hadi, 2023).

O modelo demonstrou um desempenho expressivo na previsão do IDH, evidenciado por um R^2 de 0,9632, indicando que 96,32% da variabilidade do IDH é explicada pelas variáveis independentes selecionadas. O erro quadrático médio (MSE) de 5,07e-05 e o erro absoluto médio (MAE) de 0,0037 corroboram a alta precisão do modelo, sugerindo que as previsões estão muito próximas dos valores reais do IDH, conforme demonstra a Tabela 6.

Tabela 6 – Hiperparâmetros e Resultados do Modelo para IDH

Melhores hiperparâmetros:
<code>{'learning_rate': 0.2, 'max_depth': 5, 'n_estimators': 100, 'reg_alpha': 0, 'reg_lambda': 0}</code>
Resultados para Previsão do modelo para '% PIB Agronegocio':
MSE: 5.072378460789067e-05
RMSE: 0.007122063226895046
MAE: 0.0037436445246460644
R2 Score: 0.9631932887070622

Fonte: Saída da execução do Modelo GXBoost Regressor para IDH

A análise da Tabela 7 destaca o acesso à orientação técnica como fator crucial para o IDH, representando 52,11% de importância relativa. Esse dado indica que políticas públicas direcionadas à disseminação de conhecimento e suporte aos produtores rurais podem impactar significativamente o desenvolvimento humano. A implementação estratégica dessas iniciativas pode ocasionar avanços socioeconômicos significativos em comunidades rurais.

Tabela 7 - Importância relativa das “features” para o modelo IDH

Seq.	Feature	Importância
1	%_Receber_Orientacao_Tecnica	0,521084
2	regiao_Sudeste	0,318894
3	regiao_Centro-Oeste	0,106709
4	Log_Pop_Por_Propriedade	0,017870
5	PIB	0,011041
6	regiao_BR	0,010882
7	tipoacao_Projeto	0,004801
8	regiao_Norte	0,003557
9	Pago_USD	0,003508
10	unidadeorcamentaria_EMBRAPA	0,000697
11	unidadeorcamentaria_INCRA	0,000426
12	regiao_Sul	0,000366
13	unidadeorcamentaria_MAPA	0,000165
14	unidadeorcamentaria_SFB	0,000000

Fonte: Saída da execução do Modelo GXBoost Regressor para IDH

A localização geográfica emerge como o segundo e terceiros fatores mais importantes, com a região Sudeste contribuindo com 31,89% e a região Centro-Oeste com 10,67% para a explicação do IDH. Yuliana, Prihatiningsih e Ispa (2024) propõe que essa distribuição desigual da importância regional indica disparidades significativas no desenvolvimento entre as diferentes áreas do país, destacando a necessidade de políticas regionais específicas para promover um desenvolvimento mais equitativo.

As variáveis econômicas tradicionais, como o PIB e os pagamentos em USD, mostraram uma importância relativamente baixa (1,10% e 0,35%, respectivamente) na determinação do IDH. Essa constatação desafia a noção convencional de que o crescimento econômico possui relação direta com desenvolvimento humano, sugerindo a existência de outros fatores.

O logaritmo da população por propriedade apresentou uma importância moderada (1,79%), indicando que a estrutura fundiária tem alguma influência no IDH, possivelmente relacionada à eficiência na produção agrícola e à distribuição de recursos.

As unidades orçamentárias, como EMBRAPA, INCRA, MAPA e SFB, mostraram uma importância marginal no modelo, o que pode indicar que a mera existência dessas instituições não é suficiente para impactar significativamente o IDH. Isso sugere a necessidade de uma avaliação mais aprofundada da eficácia e do alcance dos programas implementados por essas entidades.

A otimização dos hiperparâmetros do modelo XGBoost resultou em uma configuração com 100 estimadores, profundidade máxima de 5 e taxa de aprendizado de 0,2, sem

regularização adicional. Essa configuração equilibra a complexidade do modelo com sua capacidade de generalização, evitando “overfitting”.

Os resultados têm implicações importantes para a formulação de políticas públicas. Primeiramente, o foco no fornecimento de orientação técnica e na melhoria do acesso à informação para produtores rurais deve ser uma prioridade. Em segundo lugar, estratégias de desenvolvimento regional devem ser implementadas para reduzir as disparidades entre as regiões, com atenção especial às áreas fora do eixo Sudeste-Centro-Oeste.

Adicionalmente, o estudo sugere que Projetos, investimentos diretos em capital (análogo ao CAPEX), têm um impacto limitado no IDH, indicando que a qualidade e a direção dos investimentos podem ser mais importantes que o volume. Políticas que promovam a eficiência no uso de recursos e a transferência de tecnologia podem ser mais efetivas na promoção do desenvolvimento humano.

É importante notar que o modelo utilizou todas as 465 observações disponíveis, o que confere robustez aos resultados. No entanto, futuras pesquisas poderiam beneficiar-se da inclusão de variáveis adicionais relacionadas à saúde e educação, componentes diretos do IDH que não fizeram parte do escopo da amostragem desse estudo, com isso não representados no conjunto de “features”.

Considerações Finais

O estudo permitiu uma compreensão profunda dos fatores que influenciam o desempenho do agronegócio no Brasil. A preparação inicial dos dados mostrou um cenário ideal, sem valores ausentes ou inválidos, o que garantiu a confiabilidade da análise. As conclusões destacam a importância do suporte técnico e da gestão eficiente dos recursos. Observou-se que o agronegócio depende fortemente de variáveis socioeconômicas, regionais e políticas públicas.

A análise orçamentária do MAPA revelou uma tendência de declínio nos investimentos em áreas científicas e tecnológicas. Este fato aponta para a necessidade de reavaliar as prioridades de alocação de recursos. A melhoria na eficiência da execução orçamentária é um ponto positivo, mas ocorre em um contexto de redução geral dos recursos.

O resultado das análises realizadas enfatiza a importância de programas de assistência técnica para os produtores rurais. A estrutura fundiária e a distribuição demográfica nas áreas rurais também têm um impacto significativo no desempenho do agronegócio.

A utilização de técnicas de modelagem preditiva mostrou-se eficaz na captura das complexidades do setor agrícola. Os resultados reforçam a necessidade de políticas públicas direcionadas que promovam o desenvolvimento sustentável e equitativo do agronegócio.

A modelagem preditiva do percentual do PIB do agronegócio utilizando o XGBoost Regressor demonstrou um bom poder explicativo, com um R^2 de 0,8764. Este resultado confere credibilidade às conclusões derivadas do modelo e oferece uma base sólida para a formulação de políticas públicas direcionadas ao setor.

A variável relacionada ao recebimento de orientação técnica emergiu como o fator mais crucial na determinação do desempenho do agronegócio, respondendo por 28,18% da importância total no modelo. Esta constatação enfatiza o papel fundamental da disseminação de conhecimento e suporte técnico no setor agrícola.

Os investimentos em projetos (análogo ao CAPEX) foram identificados como o segundo fator mais relevante, com 21,59% de importância. Este resultado ressalta a necessidade de alocação de recursos para infraestrutura e desenvolvimento tecnológico como elementos-chave para impulsionar a produtividade e, consequentemente, a contribuição do setor para o PIB nacional.

O Índice de Desenvolvimento Humano (IDH) emergiu como a terceira variável mais influente no modelo do agronegócio, seguido pela distribuição logarítmica da população por propriedade e pelo PIB geral. Esta configuração de importância destaca a intrincada relação entre o desenvolvimento socioeconômico abrangente e o desempenho específico do agronegócio.

Por sua vez, a análise da modelagem preditiva do IDH também revelou que o acesso à orientação técnica no setor do agronegócio é o fator mais crítico na determinação do índice, com uma importância relativa de 52,11%. Este resultado reforça a necessidade de políticas públicas focadas na disseminação de conhecimento técnico e apoio aos produtores rurais podem ter um impacto substancial no desenvolvimento humano. A inclusão de variáveis adicionais relacionadas à saúde e educação em futuras pesquisas poderia enriquecer ainda mais essa análise, proporcionando uma compreensão mais detalhada do quanto influencia o desempenho do agronegócio nesse índice.

A localização geográfica emergiu como o segundo e terceiro dos fatores mais importantes na determinação do IDH, com a região Sudeste contribuindo com 31,89% e a região Centro-Oeste com 10,67%. Esta distribuição desigual da importância regional indica disparidades significativas no desenvolvimento entre as diferentes áreas do país, destacando a necessidade de políticas regionais específicas para promover um desenvolvimento mais equitativo.

As variáveis como o PIB e os valores orçamentários de pagamentos, efetivamente efetuadas e convertidos para USD, mostraram uma importância relativamente baixa na determinação do IDH. Esta constatação desafia a noção convencional de que o crescimento econômico através de investimentos é o principal motor do desenvolvimento humano, sugerindo que outros fatores, como a disseminação de conhecimento e suporte técnico, podem ter um impacto mais direto no bem-estar da população.

O estudo identificou desafios na gestão orçamentária das subfunções científicas e tecnológicas, como a tendência de declínio nos investimentos e a diminuição de sua participação no orçamento total do ministério. No entanto, houve uma melhoria na eficiência da execução orçamentária, sugerindo uma otimização forçada devido à limitação de recursos.

A robustez e a capacidade preditiva dos modelos desenvolvidos oferecem uma base sólida para a formulação de recomendações de políticas públicas e estratégias de desenvolvimento para o setor agropecuário. Os resultados apontam para a necessidade de uma abordagem heterogênea que priorize o fortalecimento e expansão de programas de assistência técnica e extensão rural, visando ampliar o acesso dos produtores a conhecimentos e tecnologias atualizadas.

O trabalho enfatiza a necessidade de atenção à estrutura fundiária e às políticas de distribuição de terras, buscando um equilíbrio que maximize a produtividade e a eficiência do uso do solo. Adicionalmente, propõe a adoção de uma visão mais disruptiva do desenvolvimento econômico, reconhecendo que o crescimento do agronegócio está intrinsecamente ligado ao progresso socioeconômico geral do país.

A abordagem metodológica adotada, combinando técnicas de machine learning com uma seleção criteriosa de variáveis, contribui significativamente para o entendimento dos determinantes do sucesso do agronegócio no Brasil, oferecendo uma ferramenta valiosa para o planejamento estratégico e a formulação de políticas baseadas em evidências. A aplicação prática destes conhecimentos tem o potencial de catalisar o crescimento sustentável do setor, reforçando sua posição como um pilar fundamental da economia brasileira.

Futuras pesquisas poderiam explorar a incorporação de variáveis climáticas e ambientais nos modelos, dada a crescente importância da sustentabilidade no setor agrícola. Além disso, análises comparativas com outros países de perfil agrícola semelhante poderiam identificar melhores práticas e oportunidades de aprendizado mútuo.

É crucial manter os modelos atualizados com dados recentes, permitindo ajustes em tempo real às mudanças nas condições econômicas e setoriais, assegurando sua contínua relevância e precisão na orientação de políticas e estratégias para o agronegócio brasileiro.

Agradecimento

O nome de DEUS seja louvado!

Referências

Capone, C.; Talgat, S.; Hazır, Ö.; Abdrasheva, K.; Kozhakhmetova, A. 2024. Artificial Intelligence Models for Predicting Budget Expenditures. Eurasian Journal of Economic and Business Studies. Disponível em: <<https://typeset.io/pdf/artificial-intelligence-models-for-predicting-budget-546nhxoizr.pdf>>. Acesso em: 14 jul. 2024.

Castro, J. (2024). MBA_USP. GitHub. https://github.com/Jotta-se/MBA_USP.

Centro de Estudos Avançados em Economia Aplicada [CEPEA]; Confederação Nacional da Agricultura e Pecuária [CNA]. 2024. PIB do agronegócio brasileiro de 1996 a 2024. Disponível em: <[https://www.cepea.esalq.usp.br/upload/kceditor/files/Planilha_PIB_Cepea_Portugues_Site\(5\).xlsx](https://www.cepea.esalq.usp.br/upload/kceditor/files/Planilha_PIB_Cepea_Portugues_Site(5).xlsx)>. Acesso em: 15 jul. 2024.

Corcioli, G.; Medina, G.S.; Arrais, C.A. 2022. Missing the Target: Brazil's Agricultural Policy Indirectly Subsidizes Foreign Investments to the Detriment of Smallholder Farmers and Local Agribusiness. *Frontiers in sustainable food systems*. Disponível em: <<https://typeset.io/pdf/missing-the-target-brazil-s-agricultural-policy-indirectly-2xw2y1in.pdf>>. Acesso em: 21 abr. 2024.

Darmawan, H.; Yuliana, M.; Hadi, M.Z.S. 2023. GRU and XGBoost Performance with Hyperparameter Tuning Using GridSearchCV and Bayesian Optimization on an IoT-Based Weather Prediction System. *International Journal on Advanced Science, Engineering and Information Technology*. Disponível em: <https://ijaseit.insightsociety.org/index.php/ijaseit/article/view/18377/pdf_2425>. Acesso em: 01 mai. 2024.

Diallo, M.; Wouterse, F. 2022. Agricultural development promises more growth and less poverty in Africa: Modelling the potential impact of implementing the Comprehensive Africa Agriculture Development Programme in six countries. *Development Policy Review*. Disponível em: <<https://onlinelibrary.wiley.com/doi/pdfdirect/10.1111/dpr.12669>>. Acesso em: 5 fev. 2024.

Governo do Brasil. 2024a. Dados do Orçamento Federal - Acompanhamento Físico-Financeiro das Ações Orçamentárias. Disponível em: <<https://dados.gov.br/dados/conjuntos-dados/orcamento-federal---acompanhamento-fisico-financeiro-das-acoes-orcamentarias>>. Acesso em: 15 fev. 2024.

Governo do Brasil. 2024b. Dados do Orçamento Federal - Orçamento Fiscal e da Seguridade Social. Disponível em: <<https://dados.gov.br/dados/conjuntos-dados/orcamento-federal---orcamento-fiscal-e-da-seguridade-social>>. Acesso em: 15 fev. 2024.

Haytham, H., Elmousalami. (2019). 3. Comparison of Artificial Intelligence Techniques for Project Conceptual Cost Prediction. arXiv: Learning,

Instituto Brasileiro de Geografia e Estatística [IBGE]. 2019. Censo Agropecuário 2017: Resultados definitivos. IBGE, Rio de Janeiro, RJ, Brasil. Disponível em: <<https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/21814-2017-censo-agropecuario.html>>. Acesso em: 21 abr. 2024.

Monteiro, J.J.; Rengel, R.; Rosa, F.S.; Lunkes, R.J.; Haskel, A.P. 2021. Reflexos da COVID-19 no orçamento público do governo federal. *Revista de Gestão e Organizações* 14: 1. Disponível em: <<https://bell.unochapeco.edu.br/revistas/index.php/rgo/article/view/5560/3249>>. Acesso em: 12 jul. 2024.

Minenhle, Hlongwane. (2022). 1. AI in Finance: Challenges, Techniques, and Opportunities. *ACM Computing Surveys*, doi: 10.1145/3502289

Piras, S.; Wesz, V.J.; Ghinoi, S. 2021. Soy Expansion, Environment, and Human Development: An Analysis across Brazilian Municipalities. Disponível em: <<https://www.mdpi.com/2071-1050/13/13/7246/pdf>>. Acesso em: 24 abr. 2024.

Puhach, Y.V. 2023. Modern econometric approaches of machine learning in the simulation of factors influencing local budget expenditures. *Actual Problems of Economics*. Disponível em: <https://eco-science.net/wp-content/uploads/2023/03/3.23._topic_-Yurii-V.-Puhach-67-74.pdf>. Acessado em: 01 mai. 2024.

Ondetti, G. 2020. Ideational Bases of Land Reform in Brazil: 1910 to the Present. Disponível em: <<https://typeset.io/pdf/ideational-bases-of-land-reform-in-brazil-1910-to-the-1ifeyp7e4q.pdf>>. Acesso em: 13 jul. 2024.

Silva, L.A.P.; Sena-Souza, J.P.; Souza, C.M.P.; Silva, C.R.; Bolfe, E.L.; Chagas-Reis, C.C.; Leite, M.E. 2023. Vetores de degradação das pastagens no Cerrado norte mineiro – Minas Gerais - BR. Raega - *O Espaço Geográfico em Análise* 57: 66-80. Disponível em: <<https://revistas.ufpr.br/raega/article/viewFile/89342/50203>>. Acessado em: 10 jun. 2024.

Soleman, R.; Ebiyeska, H.; Wibowo, M.G.; Nainggolan, B. 2022. The Effect of HDI and Macroeconomic Variables on Economic Growth in Indonesia 2015-2020. *Jurnal Ilmu Ekonomi Terapan*. Disponível em: <<https://e-journal.unair.ac.id/JIET/article/view/34641/21483>>. Acesso em: 01 mai. 2024.

Tahu, G.P.; Yuesti, A.; Bhegawati, D.A.S.; Menes, C.C.; Orbaningsih, D. 2024. The Influence of Regional Original Income and General Allocation Funds on Economic Growth with Capital Expenditures as an Intervening Variable. *Revista De Gestão Social e Ambiental* 18: e05795. Disponível em: <<https://rgsa.openaccesspublications.org/rgsa/article/view/5795/2175>>. Acesso em: 01 jan. 2024.

Torres, D.A.P. 2023. Investimento em pesquisa e a produtividade agropecuária: uma revisão da literatura recente. *Revista de Política Agrícola* 32: 106-120. Disponível em: <<https://ainfo.cnptia.embrapa.br/digital/bitstream/doc/1156720/1/Investimento-em-pesquisa.pdf>>. Acesso em: 24 jul. 2024.

Yuliana; Prihatiningsih, A.; Ispa, A.Y. 2024. Dynamics Of Economic Development Disparities A New Perspective In Regional Analysis. Disponível em: <<https://www.putrapublisher.org/ojs/index.php/jamis/article/view/511/608>>. Acesso em: 19 jul. 2024.

Zainab, H.A.; Burhan, A.M.; Kassim, M.; Al-Khafaji, Z.H.K. 2022. Developing an Integrative Data Intelligence Model for Construction Cost Estimation. Complexity. Disponível em: <<https://onlinelibrary.wiley.com/doi/epdf/10.1155/2022/4285328>>. Acesso em: 01 mai. 2024.

Apêndice

Tabela 1 – Dicionário de Dados das variáveis

(continua)

Nome do Atributo	Tipo	Descrição
exerciciofiscal	N	O exercício FISCAL é o período de tempo ao qual se referem a previsão das receitas e a fixação das despesas registradas na LOA. Este princípio é mencionado no caput do art. 2º da Lei nº 4.320, de 1964. Segundo o art. 34 dessa lei, o exercício financeiro coincidirá com o ano civil (1º de janeiro a 31 de dezembro). Fonte: MTO do respectivo exercício.
regiao		Região brasileira correspondem às divisões do território nacional.
uf	C	Significa Unidade Federativa, ou seja, cada Estado da Federação é uma unidade.
unidadeorcametaria	C	A classificação institucional reflete as estruturas organizacional e administrativa e compreende: Empresa Brasileira de Pesquisa Agropecuária [EMBRAPA], Instituto Nacional de Colonização e Reforma Agrária [INCRA], Serviço Florestal Brasileiro [SFB], Companhia Nacional de Abastecimento [CONAB], e Administração Direta do MAPA. Fonte: MTO do respectivo exercício.
subfuncao	C	Representa um nível de agregação imediatamente inferior à função e deve evidenciar a natureza da atuação governamental. A classificação funcional é representada por cinco dígitos sendo os dois primeiros relativos às funções e os três últimos às subfunções. Fonte: MTO do respectivo exercício.
tipoacao	C	Operação da qual resultam produtos (bens ou serviços) que contribuem para atender ao objetivo de um programa. Incluem-se também no conceito de ação as transferências obrigatórias ou voluntárias a outros entes da Federação e a pessoas físicas e jurídicas, na forma de subsídios, subvenções, auxílios, contribuições, entre outros, e os financiamentos. A ação é identificada por um código alfanumérico de quatro dígitos, acrescido de quatro dígitos do localizador. Ao observar o 1º dígito do código, pode-se identificar por Projetos, Atividades e Operação Especial.

(continua)

Nome do Atributo	Tipo	Descrição
Autorizado	N	Representa o valor do Localizador na Lei Orçamentária do exercício selecionado mais os acréscimos dos créditos adicionais, bem como cancelamentos e remanejamentos de dotações.
Autorizado_USD	N	Representa o valor na Lei Orçamentária do exercício selecionado mais os acréscimos dos créditos adicionais, bem como cancelamentos e remanejamentos de dotações. Valor convertido em USD no valor do último dia do ano do respectivo exercício.
Pago_USD	N	Representa o total de despesas efetivamente pagas. Valor convertido em USD no valor do último dia do ano do respectivo exercício.
IDH	N	Índice de Desenvolvido Humano.
Populacao	N	Corresponde ao conjunto de pessoas que vive na respectiva unidade federativa ou região brasileira.
PIB	N	Produto Interno Bruto. Valor total de bens e serviços produzidos no país durante um determinado período.
Valor_PIB_Agronegocio	N	Representa o valor monetário total gerado por todas as atividades relacionadas à agricultura e pecuária, incluindo não apenas a produção primária, mas também o processamento, distribuição e serviços associados.
Valor_PIB_Agricola	N	É a parcela do PIB gerada especificamente pelas atividades de cultivo de plantas, incluindo lavouras temporárias e permanentes.
Valor_PIB_Pecuario	N	Refere-se à parte do PIB gerada pela criação de animais, incluindo bovinos, suínos, aves e outros.
%_PIB_Agronegocio	N	Percentual do PIB total que corresponde ao agronegócio.
%_PIB_Agricola	N	Percentual do PIB total que corresponde especificamente à agricultura.
%_PIB_Pecuario	N	Percentual do PIB total que corresponde à pecuária.
Total_Propriedade_Agricola	N	No contexto do Censo Agropecuário 2017, é a unidade de produção dedicada, total ou parcialmente, a atividades agropecuárias, florestais e aquícolas com o objetivo de produção para subsistência e/ou comercialização.
Populacao_Por_Propriedade	N	Número médio de pessoas dividido pela quantidade de propriedades agrícola de uma determinada unidade federativa ou região brasileira.

(conclusão)

Nome do Atributo	Tipo	Descrição
Log_Pop_Por_Propriedade	N	$\text{Log}(\text{Populacao_Por_Propriedade})$ - para capturar melhor as nuances da relação entre população e propriedades agrícolas, potencialmente melhorando o desempenho do modelo.
Norm_Pop_Por_Propriedade	N	$(\text{Populacao_Por_Propriedade} - \text{média}) / \text{desvio padrão}$ - para capturar melhor as nuances da relação entre população e propriedades agrícolas, potencialmente melhorando o desempenho do modelo.
Nao_Recebe_Orientacao_Tecnica	N	Número de propriedades que não recebem orientação técnica.
%_Nao_Recebe_Orientacao_Tecnica	N	Percentual de propriedades que não recebem orientação técnica em relação ao total de propriedades.
Recebe_Orientacao_Tecnica	N	Número de propriedades que recebem orientação técnica.
%_Receber_Orientacao_Tecnica	N	Percentual de propriedades que recebem orientação técnica em relação ao total de propriedades.
Governo_Orientacao_Tecnica	N	Número de propriedades que recebem orientação técnica do governo.
Propria_Orientacao_Tecnica	N	Número de propriedades que possuem orientação técnica própria.
Cooperativas_Orientacao_Tecnica	N	Número de propriedades que recebem orientação técnica de cooperativas.
Empresas_Integradoras_Orientacao_Tecnica	N	Número de propriedades que recebem orientação técnica de empresas integradoras.
Empresas_Privadas_Orientacao_Tecnica	N	Número de propriedades que recebem orientação técnica de empresas privadas.
ONG_Orientacao_Tecnica	N	Número de propriedades que recebem orientação técnica de ONGs.
Sistema_S_Orientacao_Tecnica	N	Número de propriedades que recebem orientação técnica do Sistema S (SEBRAE, SENAR, etc.).
Outra_Origem__Orientacao_Tecnica	N	Número de propriedades que recebem orientação técnica de outras fontes não especificadas anteriormente.

Fontes: GOVERNO DO BRASIL, IBGE, CEPEA e CNA no período de 2018 a 2022

Figura 4 – Matriz de Correlação das “features” relevantes para o modelo XGBoost

	Pago_USD	PIB	Log_Pop_Por_Propriedade	%_Receber_Orientacao_Tecnica	tipoacao_CAPEX	regiao_BR	regiao_Centro-Oeste	regiao_Norte	regiao_Sudeste	regiao_Sul	unidadeorcametaria_EMBRAPA	unidadeorcametaria_INCRA	unidadeorcametaria_MAPA	unidadeorcametaria_SFB
Pago_USD	1,00	0,01	-0,02	-0,04	0,02	0,44	-0,08	-0,12	-0,09	-0,08	-0,08	-0,07	0,10	-0,01
PIB	0,01	1,00	0,49	0,36	-0,02	0,03	-0,11	-0,31	0,66	0,09	0,03	-0,01	-0,02	0,01
Log_Pop_Por_Propriedade	-0,02	0,49	1,00	0,63	0,05	-0,05	0,37	-0,15	0,41	-0,10	0,08	0,00	-0,08	-0,01
%_Receber_Orientacao_Tecnica	-0,04	0,36	0,63	1,00	0,03	-0,08	0,45	-0,32	0,15	0,53	0,00	0,01	-0,02	-0,01
tipoacao_CAPEX	0,02	-0,02	0,05	0,03	1,00	0,21	0,01	-0,10	-0,04	-0,06	0,08	-0,07	-0,01	-0,02
regiao_BR	0,44	0,03	-0,05	-0,08	0,21	1,00	-0,19	-0,27	-0,21	-0,18	-0,05	0,04	-0,04	0,17
regiao_Centro-Oeste	-0,08	-0,11	0,37	0,45	0,01	-0,19	1,00	-0,17	-0,14	-0,12	0,01	-0,03	0,00	-0,03
regiao_Norte	-0,12	-0,31	-0,15	-0,32	-0,10	-0,27	-0,17	1,00	-0,20	-0,17	0,06	0,07	-0,05	-0,05
regiao_Sudeste	-0,09	0,66	0,41	0,15	-0,04	-0,21	-0,14	-0,20	1,00	-0,13	0,02	-0,02	0,02	-0,04
regiao_Sul	-0,08	0,09	-0,10	0,53	-0,06	-0,18	-0,12	-0,17	-0,13	1,00	-0,04	0,00	0,04	-0,03
unidadeorcametaria_EMBRAPA	-0,08	0,03	0,08	0,00	0,08	-0,05	0,01	0,06	0,02	-0,04	1,00	-0,22	-0,68	-0,05
unidadeorcametaria_INCRA	-0,07	-0,01	0,00	0,01	-0,07	0,04	-0,03	0,07	-0,02	0,00	-0,22	1,00	-0,45	-0,04
unidadeorcametaria_MAPA	0,10	-0,02	-0,08	-0,02	-0,01	-0,04	0,00	-0,05	0,02	0,04	-0,68	-0,45	1,00	-0,11
unidadeorcametaria_SFB	-0,01	0,01	-0,01	-0,01	-0,02	0,17	-0,03	-0,05	-0,04	-0,03	-0,05	-0,04	-0,11	1,00

-1,00
0,00
1,00

Fontes: GOVERNO DO BRASIL, IBGE, CEPEA e CNA no período de 2018 a 2022

Nota: Desconsiderados os valores absolutos acima de 0,7 na matriz de correlação, pois foram entendidos como indicativos de alta correlação e potencial multicolinearidade.