

Análise estatística e modelos de *machine learning* na produção agrícola brasileira: Tendências temporais e eficiência produtiva ao longo de quatro décadas (1980-2019)

Eduardo Silva Vasconcelos

Doutor em Engenharia Elétrica
Instituto Federal Goiano Goiânia, Goiás – Brasil

Leandro Aureliano da Silva

Doutor em Engenharia Elétrica Universidade de Uberaba (UNIUBE)
Uberaba, Minas Gerais, Brasil

RESUMO

A pesquisa apresentada aborda a produção agrícola no Brasil ao longo de quatro décadas, destacando a importância do setor para a economia do país e a necessidade de técnicas avançadas para análise e previsão. Utilizando dados da Organização das Nações Unidas para a Alimentação e a Agricultura (FAO), o estudo avalia as tendências de longo prazo e a eficiência produtiva de diferentes culturas entre 1980 e 2019. Os objetivos primordiais incluem a análise da produção total de itens agrícolas, a avaliação da área colhida e a proporção de produção por área colhida, além de identificar mudanças significativas nos itens mais produzidos usando o Gráfico de Pareto e testar a normalidade dos dados com o teste Shapiro-Wilk. A metodologia aplicada envolve técnicas estatísticas e de Machine Learning, como Random Forest e Support Vector Machines (SVM), para prever a produção agrícola e identificar padrões complexos nos dados históricos. Os principais resultados indicam um crescimento robusto na produção de itens como cana-de-açúcar e oleaginosas, refletindo a expansão do agronegócio e a adoção de tecnologias modernas. Em contraste, a produção de cereais apresentou flutuações significativas. A análise da eficiência produtiva mostrou um aumento na proporção de produção por área colhida, especialmente após 2000, atribuível a melhorias nas práticas agrícolas e uso de tecnologias avançadas. No entanto, os modelos de Machine Learning apresentaram desempenho preditivo limitado, com altos valores de MSE e R^2 negativos, ressaltando a necessidade de abordagens mais robustas e integração de dados adicionais, como informações climáticas. Esses resultados destacam a importância de políticas agrícolas sustentáveis, investimentos contínuos em tecnologia e estratégias de gestão de riscos para promover a sustentabilidade e eficiência no setor agrícola brasileiro a longo prazo.

Palavras-chave: Estatística aplicada, Machine learning, Produção agrícola, Análise temporal.

1 INTRODUÇÃO

A produção agrícola no Brasil é crucial para a economia, contribuindo significativamente para o Produto Interno Bruto (PIB) e as exportações. Nas últimas quatro décadas, o setor agrícola brasileiro passou por transformações substanciais, impulsionadas por avanços tecnológicos, políticas agrícolas favoráveis e variações climáticas. A análise estatística e o uso de técnicas de Machine Learning surgem como ferramentas essenciais para compreender essas transformações e prever tendências futuras. Essas abordagens permitem não apenas a análise de grandes volumes de dados, mas também a extração de padrões e insights valiosos que podem orientar a tomada de decisões estratégicas.



Os dados utilizados neste estudo são provenientes da Organização das Nações Unidas para a Alimentação e a Agricultura (FAO), uma agência especializada das Nações Unidas focada na erradicação da fome, na segurança alimentar e na gestão sustentável dos recursos naturais. A base de dados da FAO fornece estatísticas detalhadas de colheita para 173 produtos em diversas regiões do mundo. Para esta pesquisa, foram selecionadas especificamente as informações relativas ao Brasil, extraídas do arquivo "Production_Crops_E_Americas.csv". A análise foi conduzida até o ano de 2019, o último ano disponível nos dados atualizados em janeiro de 2024.

O problema de pesquisa abordado é como as técnicas estatísticas e de Machine Learning podem ser utilizadas para detectar tendências de longo prazo na produção agrícola no Brasil e avaliar a eficiência produtiva ao longo do período de 1980 a 2019. O objetivo geral é avaliar essas tendências e a eficiência produtiva utilizando técnicas estatísticas e de Machine Learning.

Os objetivos específicos incluem analisar a produção total de diferentes itens agrícolas, avaliar a área colhida total e a proporção de produção por área colhida, e identificar mudanças significativas nos itens mais produzidos usando o Gráfico de Pareto, além de testar a normalidade dos dados com o teste Shapiro-Wilk.

A análise detalhada da produção agrícola e da eficiência produtiva é essencial para o planejamento estratégico e a formulação de políticas agrícolas eficazes. A aplicação de técnicas avançadas de estatística e Machine Learning permite uma avaliação precisa e a identificação de tendências e padrões que podem influenciar decisões futuras. Este estudo busca fornecer insights valiosos para pesquisadores, agricultores e formuladores de políticas, contribuindo para um entendimento aprofundado das mudanças na produção agrícola brasileira.

Este estudo utiliza dados históricos de produção agrícola no Brasil, extraídos do banco de dados de produção vegetal disponível no Kaggle. As técnicas estatísticas aplicadas incluem análise discriminante e métodos de previsão, enquanto as técnicas de Machine Learning envolvem a mineração de dados e a modelagem preditiva. As análises foram realizadas utilizando o software Python e suas bibliotecas estatísticas. A detecção de tendências de longo prazo na produção agrícola exige períodos mais longos de observação para assegurar que as variações observadas sejam estatisticamente significativas e representativas de mudanças reais, em vez de flutuações temporárias.

A literatura destaca a importância de períodos de observação suficientemente longos para a detecção eficaz de tendências de longo prazo na produção agrícola. Box, Jenkins e Reinsel (2015) argumentam que a análise de séries temporais é uma ferramenta poderosa para identificar padrões e prever comportamentos futuros com base em dados históricos. Cochran (1977) ressalta a importância de longos períodos de observação para identificar tendências persistentes. Anderson, Sweeney e Williams (2016) enfatizam que dados coletados ao longo de várias décadas são essenciais para a modelagem precisa de tendências em



fenômenos econômicos e agrícolas.

Diversos estudos analisaram a produção agrícola no Brasil, destacando suas transformações ao longo das décadas. A FAO (2019) forneceu uma classificação climática agrícola, enquanto Dolferus, Ji e Ricardo (2018) exploraram o impacto do estresse hídrico na produção de cereais. Machado e Santos (2019) demonstraram como técnicas avançadas podem melhorar a previsão da produtividade agrícola, e Carvalho e Fernandes (2021) investigaram os impactos das mudanças climáticas na agricultura.

O período de 1980 a 2019 foi marcado por significativas transformações no setor agrícola brasileiro. A produção de grãos, como soja e milho, aumentou substancialmente devido à introdução de novas tecnologias agrícolas e políticas de incentivo à produção. A adoção de tecnologias modernas, como sistemas de irrigação avançados e sementes geneticamente modificadas, foi crucial para o aumento da produtividade agrícola. As políticas governamentais desempenharam um papel significativo no desenvolvimento do setor agrícola. Programas de financiamento e subsídios incentivaram os agricultores a investir em tecnologias e práticas sustentáveis. O impacto das mudanças climáticas na produção agrícola tornou-se uma preocupação crescente, exigindo a implementação de práticas de manejo adaptativas para garantir a sustentabilidade do setor.

2 METODOLOGIA

A metodologia deste estudo visa garantir que os resultados sejam robustos, replicáveis e relevantes, sendo guiada pelos objetivos específicos e pela natureza dos dados disponíveis. O estudo é quantitativo e exploratório, analisando grandes conjuntos de dados históricos para identificar tendências de longo prazo e avaliar a eficiência produtiva. A abordagem quantitativa permite a aplicação de técnicas estatísticas rigorosas e algoritmos de Machine Learning para a análise de dados.

2.1 COLETA DE DADOS

Os dados foram extraídos do arquivo "Production_Crops_E_Americas.csv", parte do banco de dados de produção vegetal disponível no Kaggle, que fornece estatísticas de colheita para 173 produtos em diversas regiões do mundo. A pesquisa focou especificamente nas informações relativas ao Brasil, abrangendo produção, área colhida e rendimento de vários produtos agrícolas. A análise foi conduzida até 2019, utilizando os dados mais recentes disponíveis.

2.2 LIMPEZA E TRANSFORMAÇÃO DE DADOS

Registros incompletos ou inconsistentes foram removidos para garantir a precisão da análise. Os formatos de dados foram convertidos e as unidades de medida padronizadas, facilitando comparações diretas e cálculos precisos. As análises foram realizadas utilizando o software Python e suas bibliotecas estatísticas



e de Machine Learning, incluindo pandas para manipulação e análise de dados, numpy para operações matemáticas e estatísticas, scikit-learn para algoritmos de Machine Learning, statsmodels para modelagem estatística e testes de hipóteses, e matplotlib e seaborn para visualização de dados e criação de gráficos.

2.3 MINERAÇÃO DE DADOS

Informações relevantes foram extraídas e preparadas para análise utilizando técnicas estatísticas e de Machine Learning. A aplicação de algoritmos como Random Forest e Support Vector Machines (SVM) permitiu identificar padrões e prever a produção agrícola. Esses algoritmos foram treinados e testados com os dados históricos para avaliar sua precisão preditiva.

2.4 MÉTODOS ESTATÍSTICOS E TÉCNICAS DE MACHINE LEARNING

O teste t foi aplicado com o objetivo de verificar diferenças significativas entre médias, identificando variações significativas na produção ao longo do tempo. A análise discriminante foi utilizada para identificar variáveis que diferenciam anos de alta e baixa produção, classificando-os com base na produção agrícola. Algoritmos de Machine Learning, como Random Forest e Support Vector Machines (SVM), foram empregados para minerar dados e identificar padrões complexos, com a finalidade de prever a produção agrícola e avaliar sua precisão preditiva.

2.5 GENERALIZAÇÃO DOS RESULTADOS

Os resultados são específicos para o Brasil e para o período de 1980 a 2019. A generalização para outros países ou períodos deve ser feita com cautela.

2.6 TESTE DE NORMALIDADE (SHAPIRO-WILK) E ANÁLISE DE SIGNIFICÂNCIA (TESTE T DE UMA AMOSTRA)

O teste de Shapiro-Wilk foi aplicado para verificar a normalidade dos dados de produção. Para dados normalmente distribuídos, o teste t de uma amostra foi utilizado para avaliar a significância das mudanças na produção. Esse teste compara a média de uma amostra com uma média hipotética, verificando se as alterações observadas são estatisticamente significativas.

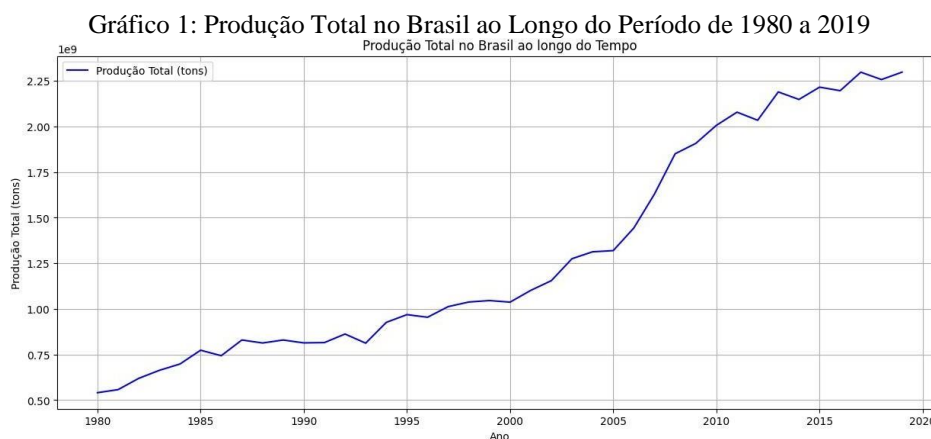
2.7 ANÁLISE DOS DADOS

A análise de dados incluiu a produção total, a área colhida, a eficiência produtiva e análises específicas como o Gráfico de Pareto, o teste de normalidade Shapiro-Wilk e o teste t de uma amostra. Medidas de tendência central (média, mediana) e de dispersão (desvio padrão, variância) foram utilizadas para fornecer uma visão geral da produção agrícola ao longo dos anos, destacando períodos de alta e baixa

produção. Gráficos de linha e os gráficos de Pareto ajudaram a visualizar a evolução da produção e a distribuição em diferentes décadas.

3 ANÁLISE DOS DADOS

3.1 PRODUÇÃO TOTAL NO BRASIL AO LONGO DO TEMPO (1980-2019)



Fonte: Elaboração Própria.

Observa-se no Gráfico 1 uma tendência de crescimento consistente na produção total ao longo dos anos. Esse aumento pode ser atribuído a avanços tecnológicos na agricultura, melhorias nas práticas agrícolas, políticas governamentais de incentivo ao agronegócio, expansão da área cultivada e aumento da eficiência produtiva. A média da produção total anual foi de aproximadamente 8709 milhões de toneladas, enquanto a mediana foi de 7116 milhões de toneladas. A diferença entre esses valores sugere a presença de alguns anos com produções excepcionalmente altas que elevaram a média.

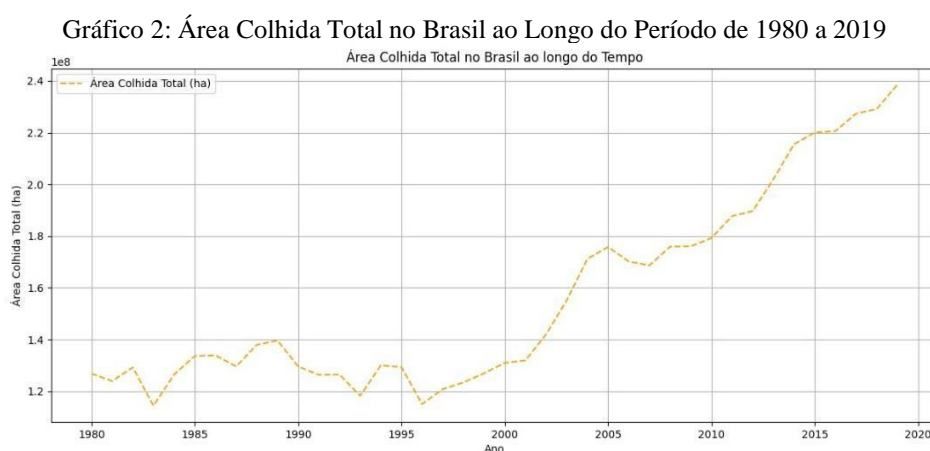
O desvio padrão elevado (379.862.884,82 toneladas) e a variância (1,44296E+17 toneladas²) indicam uma alta variabilidade na produção anual. Essa grande dispersão dos dados ao redor da média reflete flutuações significativas na produção ao longo dos anos.

O Gráfico 1 mostra períodos de aumento acentuado na produção, especialmente na década de 2000, correlacionados com melhorias nas práticas agrícolas, maior adoção de tecnologia e políticas governamentais favoráveis ao setor. Períodos de menor crescimento ou estabilização podem refletir condições climáticas adversas, crises econômicas ou outros fatores limitantes.

A análise dos dados de produção total agrícola no Brasil revela um crescimento robusto ao longo das últimas quatro décadas, com uma média anual elevada, mas também uma significativa variabilidade. Esse crescimento contínuo, aliado a uma alta dispersão, sugere que o setor agrícola brasileiro tem potencial para altas produtividades, mas também é suscetível a fatores de instabilidade que afetam a produção anual. A alta variabilidade observada ressalta a importância de estratégias de gestão de riscos e investimentos contínuos em tecnologia e infraestrutura para mitigar os impactos negativos e promover a sustentabilidade do crescimento produtivo no longo prazo.

3.2 ANÁLISE DA ÁREA COLHIDA TOTAL NO BRASIL (1980-2019)

Conforme observa-se no Gráfico 2, a seguir, a expansão da área colhida no Brasil entre 1980 e 2019 reflete as profundas transformações e a dinâmica do setor agrícola no país. Durante este período, o Brasil experimentou um aumento substancial na área dedicada à agricultura, impulsionado por políticas de desenvolvimento agrícola, avanços tecnológicos e a globalização dos mercados de commodities. A trajetória de crescimento da área colhida, evidenciada por estatísticas descritivas, mostra tanto a expansão quantitativa quanto as variações anuais significativas. A análise também destaca as implicações econômicas, sociais e ambientais associadas a esse crescimento, proporcionando uma compreensão abrangente das tendências ao longo de quatro décadas de desenvolvimento agrícola no Brasil.



Fonte: Elaboração Própria.

Conforme observa-se no Gráfico 2, a expansão da área colhida no Brasil entre 1980 e 2019 reflete as profundas transformações e a dinâmica do setor agrícola no país. Durante este período, o Brasil experimentou um aumento substancial na área dedicada à agricultura, impulsionado por políticas de desenvolvimento agrícola, avanços tecnológicos e a globalização dos mercados de commodities. A trajetória de crescimento da área colhida, evidenciada por estatísticas descritivas, mostra tanto a expansão quantitativa quanto as variações anuais significativas. A análise também destaca as implicações econômicas, sociais e ambientais associadas a esse crescimento, proporcionando uma compreensão abrangente das tendências ao longo de quatro décadas de desenvolvimento agrícola no Brasil.

A análise visual do Gráfico 2 revela uma tendência geral de crescimento na área colhida ao longo dos anos. A partir da década de 2000, observa-se um aumento mais acentuado, indicando uma possível expansão das fronteiras agrícolas e/ou intensificação das atividades agrícolas no Brasil. A média da área colhida total durante o período analisado foi de 150.177.982,425 hectares, enquanto a mediana foi de 131.900.050 hectares. A diferença entre a média e a mediana sugere que, ao longo do período, houve anos com áreas colhidas significativamente maiores que elevaram a média.



O desvio padrão, de 35.008.797,68 hectares, indica uma grande variabilidade na área colhida de um ano para o outro, influenciada por mudanças climáticas, políticas agrícolas, desenvolvimento tecnológico e expansão territorial. A variância, de 1,22562E+15 hectares², confirma essa alta dispersão dos dados ao longo do tempo.

Durante as décadas de 1980 e 1990, a área colhida manteve-se relativamente estável, com pequenas flutuações, refletindo uma fase de consolidação das práticas agrícolas e da ocupação territorial. A partir da década de 2000, nota-se um crescimento mais acentuado na área colhida, relacionado à intensificação da agricultura, políticas de incentivo ao agronegócio e avanços tecnológicos que permitiram uma maior exploração das terras agrícolas.

A análise da área colhida total ao longo dos anos revela uma trajetória de crescimento robusta, especialmente nas últimas duas décadas. A média e a mediana indicam que a área colhida típica aumentou significativamente ao longo do tempo. A grande variabilidade observada nas medidas de dispersão sugere que a área colhida foi influenciada por diversos fatores econômicos, ambientais e tecnológicos.

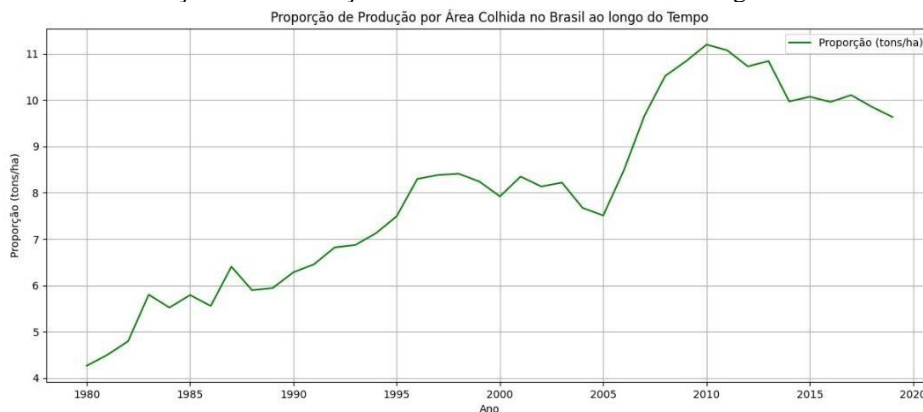
Esse crescimento contínuo na área colhida, aliado à alta variabilidade, enfatiza a importância de políticas agrícolas que promovam a sustentabilidade e a eficiência na utilização das terras agrícolas. Compreender essas dinâmicas é essencial para o planejamento de estratégias de desenvolvimento rural que garantam a produção sustentável e a segurança alimentar no longo prazo

3.3 PROPORÇÃO DE PRODUÇÃO POR ÁREA COLHIDA NO PERÍODO DE 1980 A 2019

A análise da proporção de produção por área colhida no Brasil entre 1980 e 2019 revela a evolução da eficiência agrícola ao longo de quatro décadas. Esse período é marcado por avanços significativos em práticas agrícolas, tecnologia e políticas de incentivo ao agronegócio. Observando os dados de produção e área colhida, é possível identificar tendências e variações que refletem a capacidade do setor agrícola brasileiro de otimizar a produção em relação à área cultivada.

A área colhida foi analisada para verificar o quanto de terra em hectares foi utilizado para a agricultura ao longo do período e para avaliar a proporção entre área colhida e produção. Medidas de tendência central e dispersão foram calculadas e gráficos de linha e de Pareto foram usados para visualizar a evolução da área colhida e a proporção entre área colhida e produção durante o período de 1980 a 2019. A eficiência produtiva foi calculada como a razão entre a produção total (toneladas) e a área colhida (hectares). Esse índice fornece uma medida da produtividade agrícola.

Gráfico 3: Relação entre Produção e Área Colhida no Brasil ao Longo de 1980 a 2019



Fonte: Elaboração Própria.

O gráfico mostra uma tendência geral de aumento na eficiência de produção agrícola no Brasil ao longo do tempo, especialmente notável após o ano 2000. Esse aumento na proporção de produção por área colhida reflete melhorias significativas nas práticas agrícolas, adoção de novas tecnologias e possivelmente melhores condições climáticas e de solo. A média da proporção foi de 5.582983955080975 toneladas por hectare, enquanto a mediana foi de 5.776893733676882 toneladas por hectare. A média ligeiramente inferior à mediana sugere que a maioria dos anos teve uma eficiência de produção relativamente alta, mas alguns anos com valores mais baixos puxaram a média para baixo.

O desvio padrão de 1,3105616642860847 toneladas por hectare indica uma variação significativa na eficiência de produção ao longo dos anos. A variância, sendo o quadrado do desvio padrão (1,717571875896312 toneladas por hectare²), reflete a dispersão dos dados em torno da média de maneira mais amplificada, confirmando a significativa variação na eficiência de produção por hectare ao longo do período.

Anos de crescimento acelerado, especialmente entre 2000 e 2010, mostram um aumento acentuado na proporção de produção por hectare. Esse período pode ser associado à intensificação do uso de tecnologias agrícolas avançadas, melhoramento genético das sementes e práticas de manejo agrícola mais eficientes. Após o pico em torno de 2010, há uma tendência de estabilização seguida de uma leve queda na eficiência de produção por hectare, possivelmente devido ao esgotamento do solo, mudanças climáticas adversas ou saturação do uso de tecnologias sem novos avanços significativos.

A análise da proporção de produção por área colhida ao longo dos anos revela um aumento significativo na eficiência de produção agrícola no Brasil, especialmente nas duas últimas décadas. A média e a mediana próximas indicam uma distribuição relativamente simétrica da eficiência de produção, embora a alta variabilidade observada nas medidas de dispersão sugira que essa eficiência tem sido influenciada por vários fatores ao longo do tempo. Esses resultados ressaltam a importância de investimentos contínuos em pesquisa e desenvolvimento agrícola, bem como a adoção de práticas sustentáveis que possam manter ou



melhorar a eficiência de produção no longo prazo. Estratégias para mitigar os impactos negativos de variações climáticas e para a regeneração de solos agrícolas são cruciais para assegurar a sustentabilidade e a segurança alimentar futura do Brasil.

3.4 ANÁLISE DE PARETO: IDENTIFICAÇÃO DOS ITENS MAIS PRODUZIDOS NO PERÍODO SELECIONADO

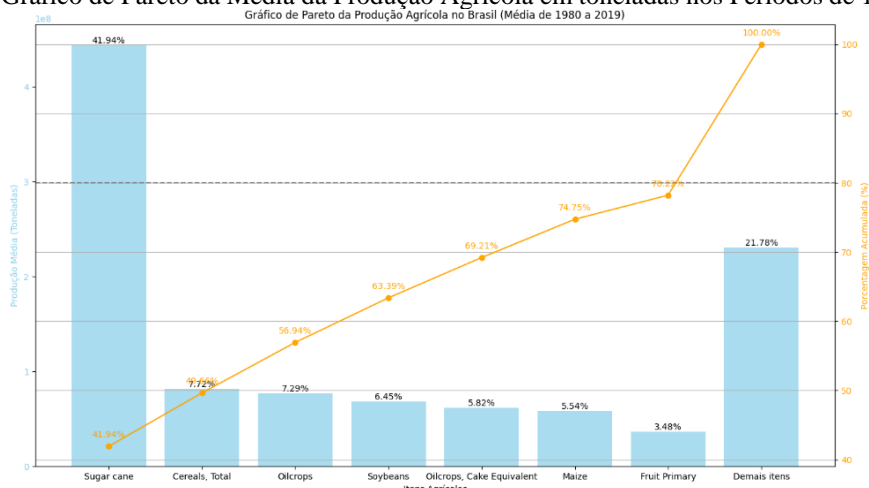
A análise de Pareto foi utilizada para identificar os itens agrícolas mais produzidos nos anos de 1980, 1990, 1995, 2000, 2005, 2010, 2015 e 2019. O Gráfico de Pareto visualiza a distribuição da produção, destacando os itens que representam a maior parte da produção total, sendo fundamental para identificar mudanças nos itens de destaque ao longo do tempo.

Os anos selecionados cobrem intervalos regulares ao longo das quatro décadas analisadas, capturando mudanças e tendências significativas nas práticas agrícolas e na produção. Cada ano escolhido reflete as condições e os desenvolvimentos agrícolas específicos daquele período, permitindo a análise de tendências de longo prazo sem a necessidade de examinar todos os anos individualmente, o que melhora a eficiência da análise.

A seleção de anos específicos reduz a redundância de dados sem comprometer a integridade da análise. A produção agrícola pode apresentar variações sazonais e anuais; no entanto, tendências de longo prazo são mais bem capturadas em intervalos de cinco a dez anos. A amostragem estratificada divide a população em subgrupos homogêneos (estratos) e seleciona amostras representativas de cada estrato. As décadas representam os estratos, e os anos escolhidos são amostras representativas dentro desses estratos, assegurando a consideração da variação dentro de cada década.

Os anos selecionados coincidem com períodos de mudanças significativas na agricultura brasileira, como a introdução de novas tecnologias, políticas agrícolas e eventos climáticos importantes. Por exemplo, os anos de 1980 representam a estabilização após a Revolução Verde; os anos de 1990 mostram a expansão das fronteiras agrícolas e modernização das práticas; os anos de 2000 marcam a adoção massiva de biotecnologia e expansão do agronegócio; e os anos de 2010 destacam a implementação de políticas sustentáveis e avanços em agricultura de precisão.

Gráfico 4: Gráfico de Pareto da Média da Produção Agrícola em toneladas nos Períodos de 1980 a 2019



Fonte: Elaboração Própria.

O gráfico de Pareto apresentado ilustra a produção agrícola média dos principais itens no Brasil ao longo do período de 1980 a 2019. A análise detalhada dos dados permite identificar os itens mais significativos em termos de produção e avaliar a concentração dessa produção entre os diferentes itens agrícolas.

A produção média de cana-de-açúcar representa 41,94% da produção total, destacando-se como o item agrícola mais produzido no Brasil. Os cereais, totalizam 9,72% da produção, ocupando o segundo lugar. Oleaginosas correspondem a 7,29% e soja a 6,45%. Oleaginosas, equivalentes a torta, representam 5,82% da produção média. O milho é responsável por 5,54%, e frutas primárias por 3,48%. Os demais itens agrícolas, agrupados, representam 21,78% da produção total.

A análise de Pareto indica que poucos itens agrícolas representam uma grande parcela da produção total. Os itens listados até "frutas primárias" representam aproximadamente 80% da produção agrícola no Brasil, seguindo o princípio de Pareto (80/20). Isso significa que focar em melhorar a eficiência e a produção desses itens pode trazer benefícios significativos ao setor agrícola brasileiro.

A concentração da produção agrícola no Brasil é altamente dominada por poucos itens, com a cana-de-açúcar sendo o mais dominante. Há uma necessidade potencial de diversificação para mitigar riscos associados à dependência de poucos produtos agrícolas. As políticas agrícolas podem ser orientadas para apoiar os itens que já mostram uma produção significativa, buscando aumentar a eficiência e a produtividade nesses setores.

O gráfico de Pareto e a análise estatística associada demonstram a concentração da produção agrícola em poucos itens principais. Este conhecimento pode ser utilizado para direcionar estratégias de melhoria na produção agrícola, políticas de apoio e investimentos em pesquisa e desenvolvimento para otimizar a produção dos itens mais significativos. Além disso, pode orientar a diversificação agrícola, reduzindo a vulnerabilidade a fatores externos e promovendo um crescimento sustentável no setor agrícola brasileiro.

3.5 ANÁLISE ESTATÍSTICA DETALHADA DOS DADOS PERCENTUAIS DE PRODUÇÃO AGRÍCOLA

Tabela 1: Percentual dos itens que mais contribuíram para a produção agrícola no Brasil (1980-2019) conforme Análise de Pareto.

	Y1980	Y1985	Y1990	Y1995	Y2000	Y2005	Y2010	Y2015	Y2019
Sugar cane	36,088	45,065	45,872	43,965	44,222	45,385	52,833	47,997	45,362
Oilcrops	7,4763	6,9364	6,8967	7,3099	8,7937	10,907	9,7308	11,865	13,267
Cereals, Total	8,064	6,565	5,674	7,186	6,309	5,974	5,535	6,783	7,304
Soybeans	-	-	-	3,718	4,451	5,492	5,063	6,235	6,885
Fruit Primary	4,807	4,671	5,293	5,062	5,155	4,152			
Maize	4,946	4,014	3,728	5,250	4,383	3,768			
Roots and Tubers, Total	6,387	4,752	4,786	4,192	3,575				
Cassava	5,697	4,216	4,247						
Demais Itens	26,535	23,780	23,503	23,316	23,112	24,322	26,838	27,120	27,182
SOMA	100	100	100	100	100	100	100	100	100

Fonte: Elaboração Própria.

A Tabela 1 fornece uma visão abrangente dos percentuais de produção dos principais itens agrícolas no Brasil ao longo dos anos de 1980 a 2019. Esses percentuais refletem a contribuição de cada item para a produção total, permitindo uma análise detalhada das tendências e mudanças na importância relativa de cada cultura conforme a análise de Pareto.

Cana-de-Açúcar: A produção de cana-de-açúcar mostra um aumento significativo, passando de 36,08% em 1980 para 45,36% em 2019. Este crescimento reflete a crescente demanda por açúcar e etanol, tornando a cana-de-açúcar uma cultura estratégica na matriz agrícola brasileira, atendendo tanto ao mercado interno quanto externo.

Oleaginosas: As oleaginosas, principalmente a soja, também apresentam um crescimento notável, aumentando de 7,47% em 1980 para 13,27% em 2019. A soja, em particular, tem sido fundamental para a economia agrícola brasileira devido à sua relevância para exportação e produção de biocombustíveis.

Cereais Totais: A produção de cereais apresenta flutuações ao longo dos anos, variando de 8,06% em 1980 para 7,30% em 2019. Apesar das variações, os cereais mantêm um papel essencial na segurança alimentar e na economia agrícola do Brasil.

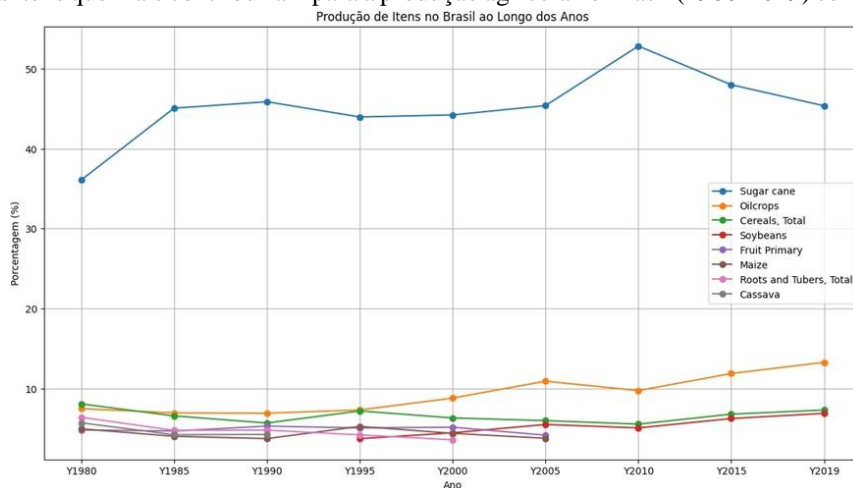
Os itens "cana-de-açúcar", "oleaginosas" e "cereais totais" mantiveram uma produção contínua e significativa de 1980 até 2019. Esta consistência temporal proporciona uma base sólida para uma análise estatística robusta, permitindo uma comparação justa e equilibrada das médias ao longo das quatro décadas.

O Gráfico 5, a seguir, ilustra visualmente a produção dos itens que mais contribuíram para a produção agrícola no Brasil entre 1980 e 2019. Os itens "cana-de-açúcar", "oleaginosas" e "cereais totais" foram os que mais se destacaram na análise de Pareto, justificando sua seleção para a aplicação do teste t.

Observa-se no Gráfico 5 que a Cana-de-Açúcar mostra um crescimento consistente, destacando-se como um dos itens mais produzidos ao longo de todo o período; as Oleaginosas apresentam um crescimento

constante, refletindo sua importância crescente no cenário agrícola brasileiro e os Cereais Totais mantêm uma produção estável, demonstrando a importância contínua desses itens na dieta e na economia brasileira.

Gráfico 5: Produção dos itens que mais contribuíram para a produção agrícola no Brasil (1980-2019) conforme Análise de Pareto.



Fonte: Elaboração Própria.

Portanto, o Gráfico 5 e a análise estatística indicam que "cana-de-açúcar", "oleaginosas" e "cereais totais" foram os itens que mais se destacaram em termos de produção ao longo dos anos de 1980 a 2019. A consistência na produção desses itens ao longo do tempo sublinha a importância dessas culturas na agricultura brasileira. A aplicação do teste t mostrou que não há diferenças significativas na produção dentro de cada item ao longo do período analisado, confirmando a estabilidade e a relevância contínua desses produtos na matriz agrícola do Brasil.

3.6 ANÁLISE ESTATÍSTICA DETALHADA E APLICAÇÃO DE MACHINE LEARNING NA PRODUÇÃO AGRÍCOLA BRASILEIRA

3.6.1 Testes estatísticos: Shapiro-Wilk e Teste t de uma amostra

O teste de Shapiro-Wilk foi utilizado para verificar a normalidade dos dados de produção agrícola. Os resultados indicaram que os dados de "Sugar cane", "Oilcrops" e "Cereals Total" são normalmente distribuídos, com valores p de 0,1313, 0,2289 e 0,8636, respectivamente, todos acima do nível de significância de 0,05. Isso permitiu a aplicação subsequente do teste t de uma amostra.

O teste t de uma amostra foi aplicado para comparar a média observada com valores de referência predefinidos. Os resultados mostraram que não há diferenças significativas entre as médias observadas e os valores de referência para "Sugar cane" ($p = 0,8945$), "Oilcrops" ($p = 0,1504$) e "Cereals Total" ($p = 0,7284$), indicando que as médias observadas são estatisticamente equivalentes aos valores de referência, sugerindo estabilidade na produção desses itens ao longo do período estudado.



3.6.2 Aplicação de Machine Learning na produção agrícola (1980-2019)

A aplicação de técnicas de Machine Learning foi realizada para prever a produção agrícola no Brasil, utilizando os algoritmos Random Forest e Support Vector Machines (SVM). A preparação dos dados incluiu limpeza, normalização e divisão em conjuntos de treinamento e teste.

- Random Forest: Utilizado por sua eficácia em manejar grandes volumes de dados e lidar com variabilidade anual. Este algoritmo combina múltiplas árvores de decisão para melhorar a precisão preditiva.
- SVM: Aplicado para identificar padrões complexos e não linearmente separáveis nos dados, utilizando variáveis como área colhida, condições climáticas e práticas agrícolas.

Os resultados indicaram que ambos os modelos enfrentaram dificuldades significativas na previsão precisa da produção agrícola. O Random Forest apresentou um Mean Squared Error (MSE) de $3.012042687858717e+15$ e um R^2 Score de -5.377521598510755 , enquanto o SVM apresentou um MSE de $5.005075822983964e+14$ e um R^2 Score de -0.05974524504353451 . Esses valores negativos de R^2 indicam que as previsões dos modelos são menos precisas do que uma simples média dos valores observados.

Os resultados dos modelos de Machine Learning destacam a necessidade de uma abordagem mais robusta. Fatores como a alta variabilidade dos dados agrícolas, a complexidade do problema e o ajuste insuficiente dos hiperparâmetros podem ter contribuído para o desempenho subótimo observado. Para melhorar a precisão preditiva, recomenda-se a inclusão de técnicas avançadas de pré-processamento de dados, ajuste fino de hiperparâmetros e a consideração de modelos híbridos que possam capturar melhor a variabilidade dos dados.

4 DISCUSSÃO DOS RESULTADOS E IMPLICAÇÕES PRÁTICAS PARA A PRODUÇÃO AGRÍCOLA BRASILEIRA

4.1 DISCUSSÃO DOS RESULTADOS

A interpretação dos resultados obtidos nas análises estatísticas e na aplicação de técnicas de Machine Learning na produção agrícola brasileira entre 1980 e 2019 revelou tendências importantes e implicações práticas significativas para o setor agrícola. A análise será abordada em duas partes principais: interpretação dos resultados e implicações práticas para a prática agrícola e formulação de políticas.

4.2 INTERPRETAÇÃO DOS RESULTADOS

A análise estatística confirmou a normalidade dos dados de produção de "Sugar cane", "Oilcrops" e "Cereals Total" através do teste de Shapiro-Wilk, permitindo a aplicação válida do teste t de uma amostra.



Os testes t indicaram que não há diferenças significativas entre as médias observadas e os valores de referência, sugerindo estabilidade na produção desses itens ao longo do período analisado.

Os modelos de Machine Learning (Random Forest e SVM) apresentaram baixo desempenho preditivo, conforme indicado pelos altos valores de MSE e R^2 negativos. A alta variabilidade dos dados de produção agrícola, a qualidade dos dados históricos e a complexidade do problema contribuíram para esses resultados, ressaltando a necessidade de abordagens mais robustas e da integração de dados adicionais.

4.3 TENDÊNCIAS OBSERVADAS

As análises revelaram tendências significativas na produção agrícola brasileira:

4.3.1 Crescimento da Produção

"Sugar cane" e "Oilcrops" mostraram um crescimento expressivo, refletindo a expansão do agronegócio e a adoção de tecnologias modernas. Em contraste, "Cereals Total" apresentou flutuações.

4.3.2 Eficiência Produtiva

A proporção de produção por área colhida aumentou, especialmente após 2000, devido a melhorias nas práticas agrícolas, uso de tecnologias avançadas e políticas de incentivo à agricultura sustentável.

4.4 IMPLICAÇÕES PARA A PRÁTICA AGRÍCOLA E FORMULAÇÃO DE POLÍTICAS

Os resultados deste estudo oferecem várias implicações práticas para o setor agrícola e a formulação de políticas:

4.4.1 Melhoria das Práticas Agrícolas

Identificar tendências e padrões pode ajudar os agricultores a adotar práticas mais eficientes e sustentáveis, aumentando a produtividade e a sustentabilidade.

4.4.2 Investimentos em Tecnologia

O aumento da eficiência produtiva destaca a importância de investimentos contínuos em tecnologia agrícola e pesquisa para manter e melhorar a produtividade.

4.4.3 Gestão de Riscos

A alta variabilidade na produção agrícola sugere a necessidade de estratégias de gestão de riscos, incluindo seguro agrícola e diversificação de culturas, para mitigar impactos climáticos e de mercado.



4.4.4 Incentivos à Agricultura Sustentável

Políticas que incentivam práticas agrícolas sustentáveis e o uso eficiente dos recursos podem contribuir para a sustentabilidade do setor no longo prazo.

4.4.5 Apoio à Pesquisa e Desenvolvimento

Investimentos em pesquisa e desenvolvimento de novas tecnologias agrícolas são cruciais para enfrentar desafios futuros e melhorar a eficiência produtiva.

4.4.6 Integração de Dados Climáticos

A inclusão de dados climáticos em análises futuras pode melhorar a precisão preditiva e ajudar na formulação de políticas adaptativas que respondam às mudanças climáticas.

5 CONCLUSÃO

Este estudo explorou a produção agrícola no Brasil ao longo de quatro décadas, aplicando técnicas estatísticas e de Machine Learning para identificar tendências, avaliar a eficiência produtiva e prever a produção futura. Os principais achados incluem:

1. **Tendências de Crescimento:** A produção de "Sugar cane" e "Oilcrops" cresceu significativamente, enquanto "Cereals Total" apresentou flutuações.
2. **Eficiência Produtiva:** A eficiência produtiva aumentou, refletindo melhorias nas práticas agrícolas e uso de tecnologias avançadas.
3. **Previsões de Machine Learning:** Os modelos Random Forest e SVM apresentaram baixo desempenho preditivo, indicando a necessidade de abordagens mais robustas e dados adicionais.

As contribuições do estudo incluem a identificação de tendências importantes, a aplicação de Machine Learning e insights para políticas agrícolas. Futuras pesquisas podem focar no aprimoramento dos modelos de Machine Learning, integração de dados climáticos e desenvolvimento de modelos híbridos para melhorar a precisão preditiva. A análise estatística e a aplicação de Machine Learning oferecem uma visão abrangente das tendências e desafios do setor agrícola brasileiro, destacando a importância de práticas sustentáveis, investimentos em tecnologia e políticas agrícolas eficazes para garantir a segurança alimentar e a sustentabilidade a longo prazo.



REFERÊNCIAS

ANDERSON, D. R.; SWEENEY, D. J.; WILLIAMS, T. A. Estatística Aplicada à Administração e Economia. 3. ed. São Paulo: Cengage Learning, 2016. 792 p.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. Time Series Analysis: Forecasting and Control. 5. ed. New Jersey: John Wiley & Sons, 2015. 712 p.

CARVALHO, M. P.; FERNANDES, E. C. M. Agricultura e mudanças climáticas: impactos e adaptação. In: Ciência Rural, v. 51, n. 6, p. 1-14, 2021. Disponível em: <https://www.scielo.br/j/cr/a/3gJ3LwQdQq3kZsmJ3mRmF7L/?lang=pt>. Acesso em: 25 maio 2024.

COCHRAN, W. G. Sampling Techniques. 3. ed. New York: John Wiley & Sons, 1977. 428 p.

DOLFERUS, R.; JI, X.; RICARDO, J. Stress signaling in cereals: the ‘drought tolerance’ model revisited. In: The Plant Journal, v. 93, n. 5, p. 693-707, 2018. Disponível em: <https://doi.org/10.1111/tpj.13705>. Acesso em: 25 maio 2024.

FAO. FAO Data. Disponível em: <https://www.kaggle.com/datasets/imtkaggleteam/crop-production>. Acesso em: 22 janeiro 2024.

FAO. Guide to agricultural climate classification. Rome: FAO, 2019. Disponível em: <http://www.fao.org/3/ca6039en/ca6039en.pdf>. Acesso em: 25 maio 2024.

MACHADO, D. R.; SANTOS, G. M. Análise temporal da produtividade agrícola utilizando séries temporais e redes neurais. In: Revista Brasileira de Meteorologia, v. 34, n. 1, p. 41-53, 2019. Disponível em: <http://rbmet.org.br/port/revista>. Acesso em: 25 maio 2024.