

# Avaliação de modelos de *machine learning* na previsão de exportações brasileiras de amendoim (*arachis hypogea L.*)

MATEUS BONACINA ZANGUETTIN <sup>1</sup>; RAFAEL JOSÉ ZANETTE <sup>1</sup>; JOÃO RICARDO FAVAN <sup>2</sup>; DEISE DEOLINDO SILVA <sup>2</sup>.

<sup>1</sup> Discente no curso Big Data no Agronegócio, FATEC Pompéia, Pompéia-SP.

<sup>2</sup> Docente do curso Big Data no Agronegócio, FATEC Pompéia, Pompéia-SP.

## RESUMO

O amendoim (*arachis hypogea L.*) é uma oleaginosa que vem ganhando força nas exportações brasileiras, ao mesmo tempo que, a Tecnologia da Informação (TI) vem auxiliando cada vez mais no processo de tomada de decisão em diversos segmentos. Portanto, são necessários estudos sobre previsões de exportações utilizando TI. Assim, o presente trabalho tem como objetivo avaliar diferentes modelos de *machine learning*, que é uma subárea da TI, na previsão de exportações brasileiras de amendoim. Para a realização do estudo, foram coletados dados referentes as exportações brasileiras de amendoim, ajustados em séries temporais e aplicados os modelos *Linear Regression*, *Random Forest* e *Multilayer Perceptron*, sendo avaliados pelo coeficiente de determinação ( $R^2$ ), erro quadrático médio (MSE) e a raiz do erro quadrático médio (RMSE). Durante a aplicação dos modelos constatou-se variações de tendência e distribuição dos dados de exportações brasileiras de amendoim ocasionando em um péssimo desempenho dos modelos. Dessa forma, houve êxito na avaliação dos modelos, no entanto foram ineficazes para realizar as previsões.

**Palavras-chave:** Tecnologia da informação. Series temporais. Comércio exterior.

## 1. INTRODUÇÃO

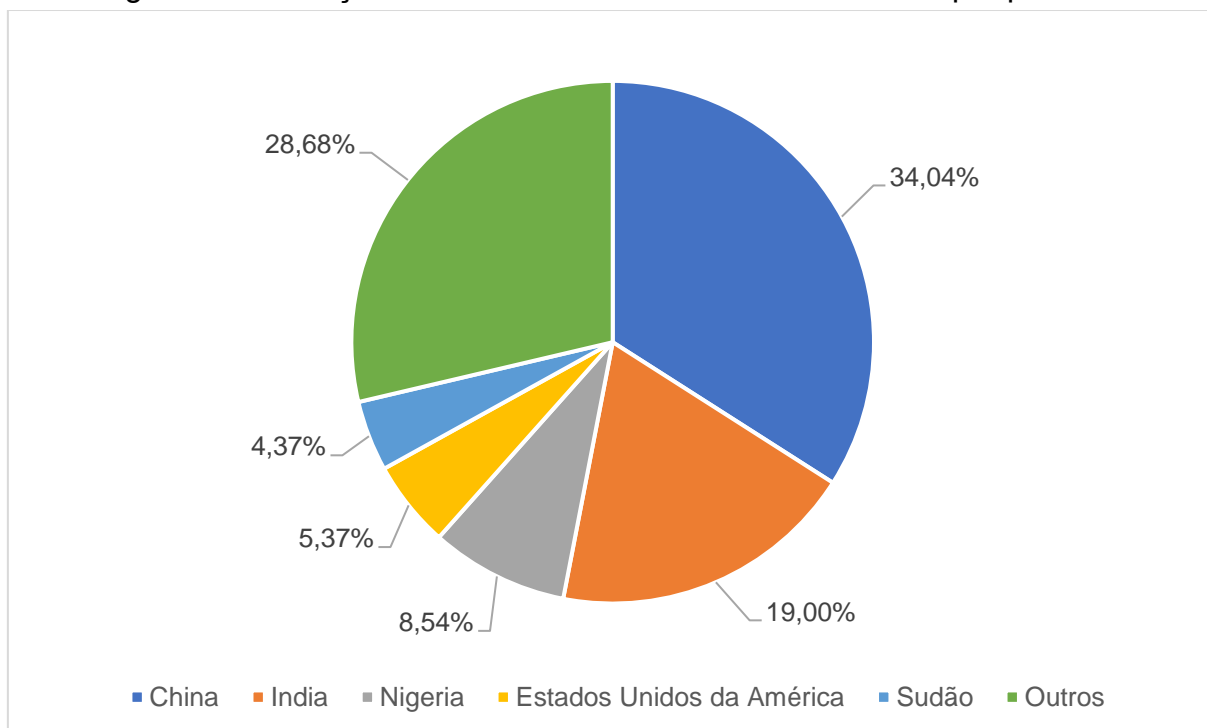
O amendoim (*arachis hypogea L.*) é uma leguminosa muito utilizada em vários países do mundo. Sua importância econômica está conectada ao sabor agradável dos grãos e por estes serem ricos em óleo (cerca de 50%) e proteína (22% a 30%), além de conter carboidratos, sais minerais e vitaminas E e do complexo B, sendo um alimento bastante energético (585 calorias/100g) (BULGARELLI, 2008).

Consumido *in natura* ou processado, esse produto é usado para extração de óleo, como matéria prima no ramo de confeitaria, na indústria alimentícia para a produção de balas, doces, bombons e ainda a pasta de amendoim que em muitos países substitui manteigas e margarinas convencionais no preparo de sanduíches (LOURENZANI; LOURENZANI, 2009).

A provável origem dessa planta são os vales dos rios Paraná e Paraguai, com estudos que indicam que os nativos das Américas já o consumiam regularmente desde o século XVI, e desde o seu descobrimento pelos colonizadores a cultura foi se espalhando pelo mundo, no século XVIII foi introduzido na Europa; no século XIX houve uma difusão do Brasil para a África, do Peru para as Filipinas, China, Japão e Índia (SUASSUNA, 2014).

No mundo essa oleaginosa é bastante consumida. Segundo dados disponibilizados pela FAO (2023) a somatória da produção mundial foi cerca de 53,9 milhões no ano de 2021, sendo a China a maior produtora com cerca de 34,04% de toda a produção mundial, seguida pela Índia com 19%, como observado na Figura 1.

Figura 1 – Produção mundial de amendoim no ano de 2021 por país



Fonte: Adaptado pelos autores a partir de FAO (2023)

No comércio internacional esse produto também é bastante negociando, visto que, foram exportados em todo mundo no ano de 2021, mais de 4,2 milhões de dólares dessa oleaginosa, sendo os principais exportadores Argentina, Índia, Estados Unidos da América, Sudão e Brasil, de acordo com dados extraídos da plataforma Trade Map (ITC, 2023) utilizando a classificação do Sistema Harmonizado (SH) 1202 - Amendoins não torrados nem de outro modo cozidos, mesmo descascados ou triturados. O Sistema harmonizado é definido por Siscomex (2022) como um sistema

único e de âmbito mundial para a designação e codificação de mercadorias. A Tabela 1 apresenta os maiores exportadores mundiais de amendoim no ano de 2021.

Tabela 1 - Exportação de amendoim em dólar (código SH 1202)

Exportador	2018	2019	2020	2021
Mundial	2.465.132	3.164.257	4.127.154	4.265.980
Argentina	151.201	426.953	855.391	818.781
Índia	457.559	618.957	750.371	688.352
Estados Unidos da América	454.090	475.763	577.747	496.431
Sudão	58.629	123.363	306.285	487.887
Brasil	230.778	228.722	318.904	330.532

Fonte: Adaptado pelos autores a partir de ITC (2023).

As exportações de amendoim aumentaram consideravelmente nos últimos anos, passando de USD 2.465.132 em 2018 para USD 4.265.980 em 2021 (aumento aproximado de 73%), em destaque para a Argentina, Sudão e Brasil que aumentaram as exportações em 441,51%, 732,15% e 43,22% respectivamente.

No Brasil o amendoim tem um histórico de altos e baixos. Martins e Vicente (2010) afirmam que durante a década de 1960 o país (em especial os estados de São Paulo e Paraná) ocupava destaque mundial na produção de óleo e na exportação de farelo.

Na década de 1970 a produção nacional do grão ganhou força com o Programa Nacional Do Álcool (Proalcool), pois começou a ser usado na entressafra da cana-de-açúcar, para que as terras não ficassem ociosas nesse período, eliminando a sazonalidade econômica e mantendo o mercado aquecido o ano todo (ASCANIO; MALAGOLI; FREIRE, 2018).

Nas décadas seguintes, devido a mudança na dinâmica do cenário econômico mundial houve incentivos a culturas voltadas à exportação (*tradeables*) resultando no declínio da produção de amendoim que foi dando lugar a soja (LOURENZANI; LOURENZANI, 2009). Martins e Vicente (2010) ainda acrescentam que a diminuição da produção dessa cultura foi marcada pela baixa rentabilidade, fatores tecnológicos e qualidade.

Essa situação continuou até a década de 1990, no entanto no início dos anos 2000 a produção do amendoim começou mudar devido vários fatores, dos quais se destacam: o lançamento em 2001 do Pró-amendoim feito pela ABICAB (Associação Brasileira de Chocolate, Cacau, Amendoim, Balas e Derivados); a Resolução RDC-172 (em 2003) que estabelece o regulamento técnico de boas práticas de fabricação para estabelecimentos que industrializam amendoins processados e derivados; a publicação da instrução normativa n.3 em 2009 pelo MAPA (Ministério da Agricultura, Pecuária e Abastecimento) que dispõe de critérios e procedimentos de controle higiênico-sanitário na cadeia de produção dessa cultura (MARTINS; VICENTE, 2010).

Essa recuperação foi muito importante para a que a oleaginosa brasileira conseguisse ampliar seus mercados. Para Ascanio, Malagolli e Freire (2018) a produção nacional de amendoim se tornou um exemplo de como modificações nas técnicas de produção e avanços tecnológicos resultaram em melhores produtividade e qualidade, satisfazendo assim as métricas do mercado Europeu, conseguindo bons resultados e mostrando comprometimento com o meio ambiente, inovações tecnológicas proporcionando uma melhor comercialização no mercado.

Atualmente o amendoim passa por um período de crescimento, de acordo com Conab (2022) no país houve um aumento de produção de mais de 100% em 8 anos, saltando de 346,8 mil toneladas na safra 2014/15 para 746,7 mil toneladas na safra 2021/22 (aumento de 115%), impulsionado principalmente pelo aumento da área cultivada, visto que, há um aumento de cerca de 100% quando os períodos são comparados, de modo que cerca de 70% dessa produção é exportada.

Para Rossato Junior (2022) a exportação é o principal fator que tem estimulado o aumento na produção de amendoim, e o autor destaca ainda duas grandes tendências globais com oportunidades para a inserção estratégica do grão no mercado mundial: a primeira é o mercado de carbono, levando em consideração que a pegada de carbono do amendoim nacional é 59% inferior à média mundial; e a segunda, está relacionada as novas gerações e as mudanças nos seus hábitos alimentares com um crescente mercado de alimentos à base de ingredientes de origem vegetal (*plant-based*).

Portanto, torna-se cada vez mais necessário realizar previsões de exportações do amendoim brasileiro. Souza (2021) destaca que esse tipo de previsão, através de análises de séries temporais, pode subsidiar o planejamento dos agentes (produtores, vendedores, compradores, exportadores, *players* de mercado e investidores) de

políticas para o setor, por se tornar uma ferramenta eficiente para a tomada de decisões.

Uma série temporal (ST) é definida por Miguel (2021) como uma série de pontos de dados ordenados no tempo, de modo que, o tempo costuma ser a variável independente e geralmente a meta é fazer uma previsão do futuro. Para Pauli, Kleina e Bonat (2019) um dos benefícios das previsões de séries temporais está na possibilidade de realizar um bom planejamento, e evitar gastos desnecessários.

Observa-se também, que é de extrema importância a previsão de séries temporais macroeconômicas para os decisores da política monetária e orçamental avaliarem o desenvolvimento econômico, e atualmente os avanços recentes nesses tipos de estudo têm sido amparados com o uso de grandes dimensões de dados disponíveis e que são manuseadas com o uso da tecnologia da informação (DIAS; LOURENÇO; RUA, 2018).

Dentre as diversas maneiras proporcionadas pela tecnologia da informação para realizar previsões com séries temporais uma delas é com o uso de *machine learning* (ML) que segundo Campos (2020) é uma subárea da inteligência artificial que tem suporte para previsão de séries temporais e vem sendo muito utilizada para esse fim.

Ao contrário dos métodos estatísticos tradicionais, os métodos de *machine learning* procuram descrever as propriedades dos dados sem conhecimento prévio da distribuição dos mesmos e por essa razão são mais simples de serem ajustados, e em muitos casos proporcionam considerável desempenho mesmo na aplicação de séries temporais complexas e não lineares (PARMEZAN, 2016 apud CAMPOS, 2020 p. 23).

Vários modelos de *machine learning* podem ser utilizados para a previsão de exportações como *Linear Regression* (LEGGETTER; WOODLAND, 1994), *MLP Regressor (Multilayer Perceptron)* (FAUSSET, 1994) e *Random Forest Regressor (Random Forest)* (BREIMAN, 2001). Nos estudos de previsão utilizando séries temporais Carvalho *et al.* (2021) utilizaram *Linear Regression* e *Random Forest* e Pauli, Kleina e Bonat (2019) usaram *Linear Regression* e *Multilayer Perceptron*.

Considerando a importância do amendoim para as exportações brasileiras e a capacidade dos modelos de *machine learning* em realizar previsões com o uso de séries temporais é necessário o uso desses modelos na previsão das exportações

dessa oleaginosa com o intuito de fornecer informações importantes para os agentes envolvidos na cadeia produtiva dessa cultura.

Assim, o presente trabalho teve como objetivo avaliar os diferentes modelos de *machine learning* (*Linear Regression*, *Random Forest Regressor* e *MLP Regressor*) na estimativa de previsões de exportação do amendoim no Brasil.

## **2. METODOLOGIA**

A metodologia utilizada no trabalho engloba cinco fases, sendo elas: a) coleta e tratamento de dados, b) análise das *features* mais importantes, c) validação cruzada, d) configurações dos modelos e e) indicadores de avaliação. Todas essas etapas estão descritas detalhadamente nesse capítulo.

### **2.1. Coleta e tratamento de dados**

Foram utilizados dados históricos relacionados a cultura do amendoim como exportação, produção, área colhida e as cotações das moedas euro e dólar referentes ao período de 01/01/2012 a 31/12/2022.

Os dados de exportação foram extraídos em formato CSV em períodos mensais na plataforma Comtrade (ONU, 2023) utilizando o código do Sistema Harmonizado 1202, de modo, que foram coletados os dados de exportação do Brasil e dos dois maiores exportadores mundiais de amendoim (Argentina e Índia), visto que, esses fatores podem interferir nas exportações, como explicado por Liu *et al.* (2019 apud YANG, 2021, p. 4).

As cotações do último dia de cada mês das moedas euro e dólar dos Estados Unidos foram extraídas, em formato CSV, da plataforma Investing.com (INVESTING, 2023). As taxas de câmbio foram consideradas no estudo adaptando-se ao método utilizado por Anefalos e Margarido (2006).

As informações relacionadas área colhida e produção brasileira de amendoim e produção dos três principais produtores mundiais dessa cultura (China, Índia e Nigéria) foram coletadas do site da FAOSTAT (FAO, 2023).

A produção de cana-de-açúcar no Brasil está relacionada a do amendoim, pois segundo Ascanio, Malagolli e Freire (2021) é usada na entressafra da cana-de-açúcar para recuperação do solo, assim, considerou-se a produção dessa gramínea no presente estudo, cujos dados também foram coletados FAOSTAT (FAO, 2023).

Com todos os dados coletados foi realizado o tratamento que constituiu nas etapas de: criação de *dataframes* para os arquivos, substituição dos valores faltantes com o uso do método de imputação pela média como sugerido por Silva Junior (2020) e centralização em um único *dataframe*. Para realizar essas etapas, foi utilizado a linguagem de programação Python, na versão 3.11.2 (VAN ROSSUM; DRAKE JR, 1995), com o auxílio das bibliotecas Pandas, versão 1.5.3 (MCKINNEY, 2023), e Numpy, versão 1.24.2 (HARRIS, 2020), na IDE (*integrated development environment*) PyCharm 2022.3.2 (Community Edition) (JETBRAINS, 2023).

Após o tratamento dos dados foi definido como *target* a coluna `brazil_export` (exportações brasileiras de amendoim em quilogramas líquido), e estabelecidas as seguintes *features*: `brazil_production_groundnuts` (produção brasileira de amendoim em quilogramas), `brazil_production_sugar_cane` (produção brasileira de cana-de-açúcar em quilogramas), `china_production_groundnuts` (produção chinesa de amendoim em quilogramas), `india_production_groundnuts` (produção indiana de amendoim em quilogramas), `nigeria_production_groundnuts` (produção nigeriana de amendoim em quilogramas), `brazil_area_groundnuts` (área brasileira colhida de amendoim em hectares), `quotation_dolar` (cotação do dólar), `quotation_euro` (cotação do euro), `argentina_export` (exportação argentina de amendoim em quilogramas líquido), `india_export` (exportação indiana de amendoim em quilogramas líquido), `brazil_export_groundnut_last_year` (exportações brasileira de amendoim do ano anterior em quilogramas líquido), `brazil_export_groundnut_last_month` (exportações brasileira de amendoim do mês anterior em quilogramas líquido).

## **2.2. Análise das *features* mais importantes**

Objetivando a seleção dos atributos mais relevantes para a predição de exportação do amendoim, foi utilizada a função *f-regression* da biblioteca *Sklearn*, versão 0.0. post1, (PEDREGOSA *et al.*, 2011). Dessa forma, foram descartados os atributos que obtiveram pontuação inferior ou igual à pontuação média de todos os atributos adaptando a metodologia utilizada por Demsar *et al.* (2013) e Carvalho *et al.* (2021).

## **2.3. Validação cruzada**

Foi utilizado o método de validação cruzada em séries temporais sugerido por Sher (2020) para a aplicação dos modelos, que consiste em separar o conjunto de

dados em *folds* (subconjuntos) com dados de cada ano utilizando-os para treinamento e teste dos modelos, de modo que foram realizadas várias aplicações dos modelos com a validação sempre sendo feita com o *fold* do ano posterior aos usados para o treinamento, como pode ser mostrado na Tabela 2.

Tabela 2 – Separação dos dados em *folds* anuais para a validação cruzada

<i>Fold</i>	Treinamento	Validação
1	[2012]	[2013]
2	[2012 a 2013]	[2014]
3	[2012 a 2014]	[2015]
4	[2012 a 2015]	[2016]
5	[2012 a 2016]	[2017]
6	[2012 a 2017]	[2018]
7	[2012 a 2018]	[2019]
8	[2012 a 2019]	[2020]
9	[2012 a 2020]	[2021]
10	[2012 a 2021]	[2022]

Fonte: Autores

## 2.4. Configurações dos modelos

Foram utilizados três modelos de *machine learning* supervisionado de regressão contidos na biblioteca *SkLearn*, sendo eles: *Linear Regression*, *Random Forest Regressor* e *MLP Regressor*.

O modelo *Linear Regression* foi definido para calcular a interceptação e forçar os coeficientes a serem positivos.

Foi ajustado o modelo *Random Forest Regressor* com um número de 1000 estimadores (árvores), o critério de *absolute\_error* para medir a qualidade de uma divisão e o valor de *random\_state* igual a 0 para controlar a aleatoriedade na construção das árvores.

O *MLP Regressor* contou com 5 neurônios na camada intermediária (oculta), cada um utilizando a função de ativação identidade, com o treinamento do modelo

limitado à cem mil épocas e o valor de *random\_state* igual a 0 para controlar a geração de números aleatórios para pesos e inicialização de tendência.

## **2.5. Indicadores de avaliação**

Os modelos ajustados foram avaliados utilizando os seguintes indicadores de desempenho sugeridos por Azank (2020) para modelos de regressão: coeficiente de determinação ( $R^2$ ) que é medida que calcula a porcentagem de variância que o modelo consegue prever, erro quadrático médio (MSE) sendo a média do erro das previsões elevado ao quadrado e a raiz do erro quadrático médio (RMSE) de modo que essa última tem como função melhorar a interoperabilidade da métrica.

## **3. RESULTADO E DISCUSSÃO**

Os resultados e discussões foram divididos em três fases, sendo: a) seleção das *features*, onde foram selecionadas as *features* que tem maior relação com o *target*, b) análise das métricas dos modelos e dos dados, que tratou de analisar os resultados obtidos pelos modelos buscando justificativas através do comportamento dos dados e da literatura vigente e c) considerações adicionais.

### **3.1. Seleção das *features***

Para a seleção das *features* mais importantes, foi aplicado a função *f-regression* que pontuou a relação de cada atributo com o *target*. A Tabela 3 apresenta as pontuações.

Tabela 3 - Relação das *features* com o *target*

<i>Features</i>	Relação com o <i>target</i>
brazil_production_sugar_cane	0,4994
india_export	6,0064
india_production_groundnuts	37,393
nigeria_production_groundnuts	61,7408
argentina_export	63,8014
china_production_groundnuts	81,0486
brazil_area_groundnuts	82,7043
brazil_production_groundnuts	94,2613
brazil_export_groundnut_last_year	187,5814
quotation_euro	210,3729
quotation_dolar	234,7464
brazil_export_groundnut_last_month	480,27808
Média	128,3694983

Fonte: Autores

Para a aplicação dos modelos de *machine learning* foram selecionados as *features* que obtiveram uma pontuação superior a 128, 36 (média), sendo elas *brazil\_export\_groundnut\_last\_year*, *quotation\_euro*, *quotation\_dolar*, *brazil\_export\_groundnut\_last\_month*.

### 3.2. Análise das métricas dos modelos e dos dados

Foram utilizadas as médias das métricas  $R^2$ , MSE e RMSE de todos os *folds* para avaliação dos modelos. Os resultados estão expostos na Tabela 4.

Tabela 4 - Resultados das médias de todos os *folds*

Indicador	Regressão		
	Linear	Random Forest Regressor	MLP Regressor
$R^2$	-0,0073	-0,2714	0,0861
MSE	$13,68 \times 10^{12}$	$22,2 \times 10^{12}$	$14,1 \times 10^{12}$
RMSE	$3,43 \times 10^6$	$4,19 \times 10^6$	$3,45 \times 10^6$

Fonte: Autores.

Tabela 5 - Classificação de desempenho

Classes	R <sup>2</sup>	Desempenho
1	>0,85	Ótimo
2	0,76 a 0,85	Muito Bom
3	0,66 a 0,75	Bom
4	0,61 a 0,65	Regular
5	0,51 a 0,60	Fraco
6	0,41 a 0,50	Muito Fraco
7	<0,41	Péssimo

Fonte: Adaptado de Camargo e Sentelhas (1997) e Ferraz (2013) apud Carvalho et al. (2021)

Comparando os resultados da Tabela 4 com a Tabela 5 que classifica o desempenho dos modelos considerando o R<sup>2</sup>, pode-se observar que o desempenho de todos os modelos é considerado péssimo de acordo com a média do R<sup>2</sup>. O MSE e o RMSE também indicam que os modelos não tiveram bom desempenho, pois segundo Azank (2020) quanto mais próximo de 0 o valor desses indicadores melhor é o desempenho.

Para entender melhor o comportamento dos modelos, deve-se analisar os desempenhos individuais dos modelos em cada *fold* obtidos durante a validação cruzada, cujos resultados estão expostos na Tabela 6.

Tabela 6 - Desempenhos dos *fold*s em R2

<i>Folds</i>	Regressão Linear	Random Forest Regressor	MLP Regressor
1	0,1373	-0,4899	-0,3923
2	-2,5248	-1,9727	-0,7372
3	0,4637	-0,1451	0,3281
4	0,5218	0,3926	0,4984
5	0,4109	-0,0214	0,7055
6	-0,3614	-0,9458	-0,4199
7	-0,1365	-0,3901	-0,4897
8	0,584	-0,1056	0,5551
9	0,5106	0,4751	0,4934
10	0,3209	0,4885	0,3198

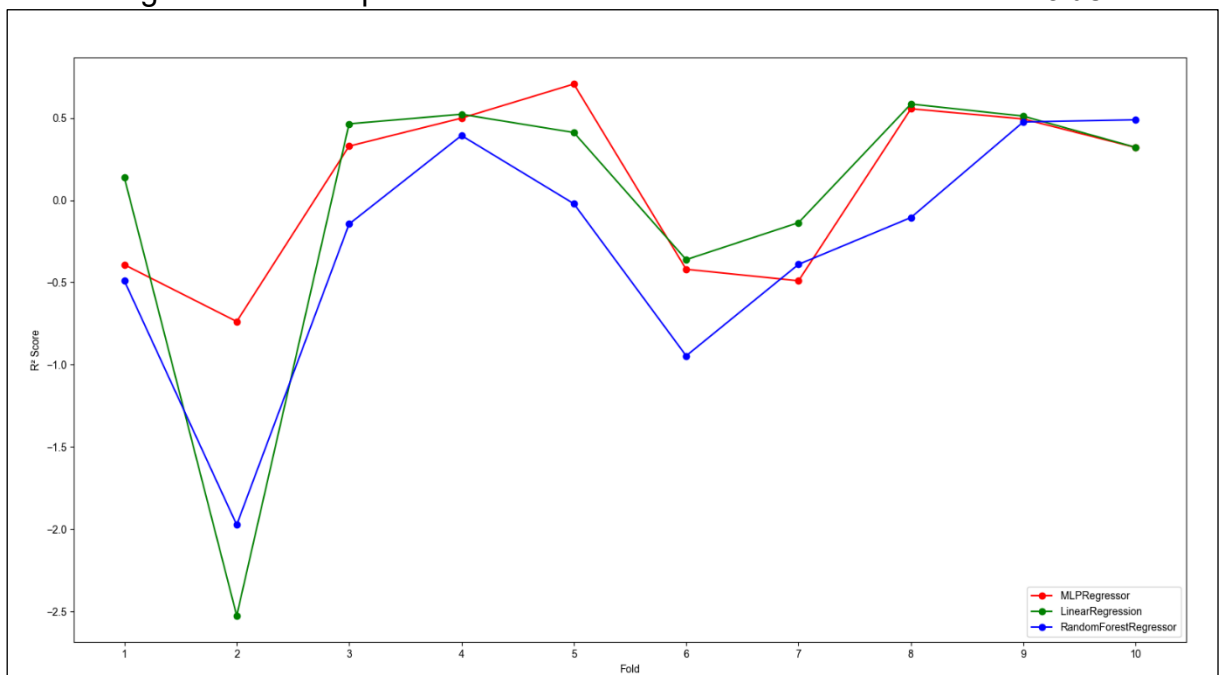
Fonte: Autores

Comparando os resultados da Tabela 6 com a classificação da Tabela 5, observa-se que o modelo com pior desempenho foi o *Random Forest Regressor* com apenas três *fold*s positivos dos dez avaliados, divergindo do resultado obtido por

Carvalho *et al.* (2021) que obtiveram três *folds* com o desempenho considerado ótimo de quatro avaliados. Dos dez *folds* avaliados, o *Linear Regression* obteve o  $R^2$  negativo em três e os demais resultados foram classificados como fraco, muito fraco ou péssimo. O *MLP Regressor* foi o modelo com melhor desempenho dentre os três, sendo o único que alcançou um *fold* com um desempenho classificado como bom (5° *fold*).

Todos os modelos tiveram uma queda acentuada de desempenho nos *folds* dois e seis que foi determinante para a média de desempenho ser considerado péssimo, visto que, foram valores negativos e proporcionalmente altos. A Figura 2 mostra o desempenho dos modelos com a indicador  $R^2$  em cada *fold*.

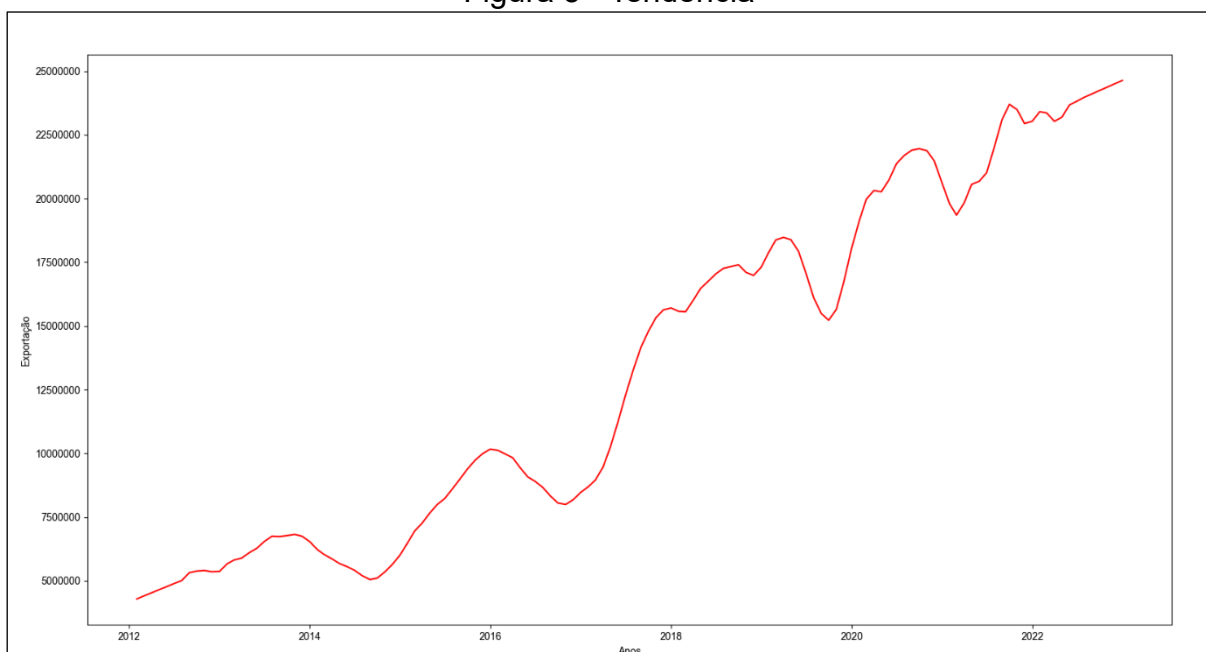
Figura 2 - Desempenho dos modelos em  $R^2$  em decorrência dos *folds*



Fonte: Autores

O péssimo desempenho dos modelos no *fold* dois é explicado pela variação na tendência, visto que, nesse *fold* os modelos foram treinados com dados dos anos de 2012 e 2013 com tendência de alta, no entanto a validação ocorreu com dados do ano de 2014 que possui uma tendência de queda. A Figura 3 retrata a tendência do *target* em relação aos anos.

Figura 3 - Tendência



Fonte: Autores

Uma situação semelhante ocorreu no trabalho de Avaliação de algoritmos de Machine Learning na cotação de preço no contrato futuro de milho, apresentado por Carvalho *et al.* (2021) em que três modelos tiveram isoladamente desempenho péssimo em um determinado *fold*, cujo treinamento foi feito com tendência lateralizada e a validação com tendência de queda.

A mudança de tendência ocorrida com os dados de validação do *fold* dois (ano de 2014) pode estar relacionado com a queda de exportações brasileiras ocorridas nesse ano, onde houve diminuições nos preços das *commodities*, como relatado por Amato (2015) e Maximo (2014).

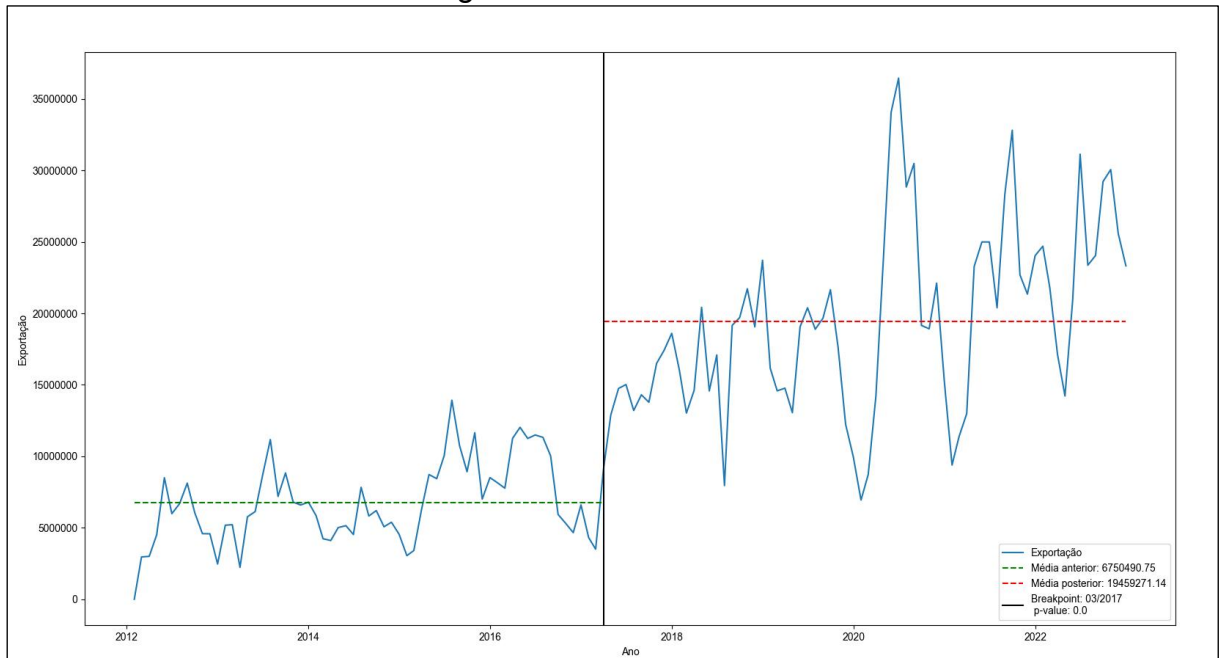
Deve se notar, que para os próximos *folds* (três, quatro e cinco) os modelos *Regressão Linear* e *MLP Regressor* tiveram melhora no desempenho se comparado com o *fold* 2.

No *fold* seis houve uma queda brusca no desempenho causada por um *breakpoint* (ponto de mudança) na série temporal, que segundo Hussain, Mahmud e Bari (2023) ocorre quando o conjunto de dados altera a sua distribuição.

Para identificar o *breakpoint* e a homogeneidade dos dados de exportação brasileira de amendoim foi necessário o uso de um teste de homogeneidade, que segundo Hussain, Mahmud e Bari (2023) é um teste estatístico que detecta em séries temporais um ou mais *breakpoints*. O teste de homogeneidade utilizado foi o de Pettitt,

que de acordo com Assis (2014) é usado para identificar o ponto de início de impacto. O resultado desse teste pode ser analisado na Figura 4.

Figura 4 - Teste de Pettitt

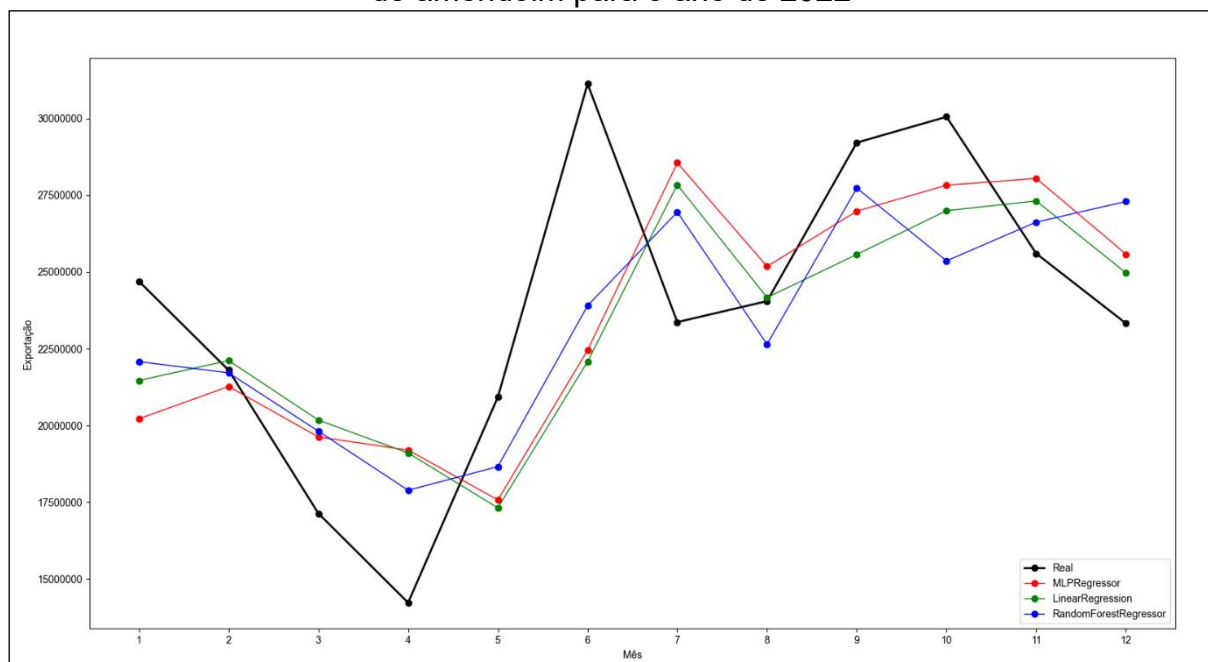


Fonte: Autores

Esse *breakpoint*, prejudicou o desempenho dos modelos nos *folds* seis e sete, visto que, os modelos validaram com dados contendo diferentes distribuição daqueles que foram utilizados para treinamento.

Nos *folds* seguintes (oito, nove e dez) os modelos tiveram recuperação, e em alguns casos, obtiveram o  $R^2$  classificado como fraco. A Figura 5, compara os valores preditos dos modelos com os valores reais para o ano de 2022 (*fold* 10).

Figura 5 - Valores reais e valores preditos pelos modelos de exportações brasileira de amendoim para o ano de 2022



Fonte: Autores

No *fold* 10, o modelo *Random Forest Regressor* conseguiu uma pontuação melhor se comparado com os demais, entretanto a diferença entre os valores preditos e o reais é grande.

### 3.3. Considerações adicionais

Embora, o desempenho dos modelos de modo geral foi classificado como péssimo, outros trabalhos como o de Souza (2021) e Pauli, Kleina e Bonat (2019) conseguiram melhores resultados na previsão de exportações de outros produtos, mostrando que pode ser possível obter melhores desempenhos na previsão de exportação brasileira de amendoim utilizando *machine learning*.

É importante destacar que durante o trabalho verificou-se a dificuldade de encontrar materiais em língua portuguesa referente a estudos de previsão de exportação, previsão de exportação com *machine learning* e previsão de exportação de amendoim.

Não foram encontrados dados relacionados à produção e área colhida em períodos mensais, dessa forma, seguiu-se uma adaptação da metodologia utilizada por Carvalho *et al.* (2021) replicando-se dados anuais para os períodos mensais.

#### **4. CONCLUSÃO**

Os modelos de *machine learning* foram avaliados com êxito, entretanto não tiveram eficácia na previsão das exportações brasileiras de amendoim.

Indica-se a necessidade de outros estudos na área da previsão de exportações brasileiras de amendoim.

#### **AGRADECIMENTOS**

Primeiramente a Deus e Nossa Senhora Aparecida por ter nos sustentado durante o desenvolvimento deste estudo. Agradecemos também aos docentes doutor João Ricardo Favan e doutora Deise Deolindo Silva, orientador e coorientadora respectivamente deste estudo, pelos conhecimentos e ensinamentos a nós fornecidos. Obrigado a Antonia Bonacina Zanguettin e Marino Zanguettin (país do discente Mateus Bonacina Zanguettin) e Josefa Angelina Cassani Zanette e Celso Donisete Zanette (país do discente Rafael José Zanette) pelo apoio incondicional.

#### **REFERÊNCIAS:**

AMATO, Fabio. Balança comercial registra em 2014 primeiro déficit desde 2000. G1, Brasília, 5 jan. 2015. Disponível em: <https://www.google.com/search?q=como+fazer+referencia+de+noticias&oq=como+fazer+referencia+de+noticias&aqs=chrome..69i57.7963j0j7&sourceid=chrome&ie=UTF-8> . Acesso em 10 maio 2023.

ANEFALOS, L. C.; MARGARIDO, M.A. Modelos de séries temporais aplicados ao setor de exportação brasileira de flores de corte. XLIV Congresso do Saber. Fortaleza, 23 a 27 de julho de 2006. Disponível em: <http://www.iea.sp.gov.br/ftp/iea/congressos/sober-790a.pdf>. Acesso em 5 jan. 2023.

ASCANIO, E.; MALAGOLI, G. A.; FREIRE, J. E. Um panorama da exportação do amendoim no interior paulista. Desenvolvimento Rural Sustentável: pesquisas emergentes no contexto da Agricultura e Agroindústria, p.76-89, 01 de janeiro de

2021. Disponível em: <https://www.editoracientifica.com.br/artigos/um-panorama-da-exportacao-do-amendoim-no-interior-paulista>. Acesso em 5 jan. 2023.

ASSIS, W. T. Avaliação do impacto do uso consuntivo de água para irrigação na Bacia Hidrográfica do Alto rio das Mortes. Programa de pós-graduação em recursos hídricos – UFTM. Cuiabá, 2014. Disponível em: <https://ufmt.br/ppgrh/dissertacao/avaliacao-do-impacto-do-uso-consuntivo-de-agua-para-irrigacao-na-bacia-hidrografica-do-alto-rio-das-mortes/>. Acesso em 1 maio 2023.

AZANK, F. Como avaliar seu modelo de regressão: As principais métricas para avaliar seus modelos de regressão. Turing Talks, 03 de agosto 2020. Disponível em: <https://medium.com/turing-talks/como-avaliar-seu-modelo-de-regress%C3%A3o-c2c8d73dab96> . Acesso em: 2 maio 2023.

BREIMAN, I. Random Forests. *Machine Learning* 45, 5-32, outubro de 2001. Disponível em: <https://link.springer.com/article/10.1023/a:1010933404324> . Acesso em 7 fev. 2023.

BULGARELLI, E. M. B. Caracterização de variedades de amendoim cultivadas em diferentes populações. Jaboticabal, novembro 2008. Disponível em: <https://repositorio.unesp.br/handle/11449/92696>. Acesso em: 29 out. 2022.

CAMPOS, B. A. R. M. Análise comparativa de técnicas para a previsão de séries temporais no contexto de mercados financeiros. Florianópolis, 2020. Disponível em <https://repositorio.ufsc.br/handle/123456789/223843>. Acesso em 24 mar. 2023.

CARVALHO, R. et al. Avaliação de algoritmos de *Machine Learning* na cotação do Preço do Contrato Futuro de Milho. *Revista eF@tec*, Garça, v.11, a.1, outubro 2021. Disponível em: <https://pesquisafatec.com.br/ojs/index.php/efatec/article/view/249>. Acesso em 24 ago. 2022.

CONAB – COMPANHIA NACIONAL DE ABASTECIMENTO. Produção de amendoim cresce mais de 100% nos últimos 8 anos, 28 de setembro de 2022. Disponível em: <https://www.conab.gov.br/ultimas-noticias/4768-producao-de-amendoim-cresce-mais-de-100-nos-ultimos-8-anos>. Acesso em 10 jan. 2023.

PEDREGOSA ET AL. Scikit-learn: Machine Learning in Python, *JMLR* 12, pp. 2825-2830, 2011. Disponível em: <https://scikit-learn.org/stable/about.html#citing-scikit-learn>. Acesso em 5 jan. 2023.

DEMSAR, J. et al. Orange: data mining toolbox in python. *The Journal of Machine Learning Research*, v.14, n. 1, p.2349-2353, 1 janeiro 2013. Disponível em <https://jmlr.org/papers/volume14/demsar13a/demsar13a.pdf>. Acesso em 1 maio 2023.

DIAS, F; LOURENÇO N.; Rua, A. Previsão de exportações com seleção de previsores. Portugal, 2018. Disponível em <https://www.bportugal.pt/paper/previsao-das-exportacoes-com-selecao-de-previsores>. Acesso em 9 jan. 2023.

FAO - FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. Production. Disponível em: <https://www.fao.org/faostat/en/#data/QI>. Acesso em: 7 jan. 2023.

FAUSSET, L. Fundamentals of Neural Network: Architecture, Algorithm, and Application. New Jersey: Prentice Hall, 1994.

HUSSAIN, M; MAHMUD, I.; BARI, S. H. pyHomogeneity: A Python Package for Homogeneity Test of Time Series Data. Journal of open research software. 14 de fevereiro de 2023. Disponível em: <https://openresearchsoftware.metajnl.com/articles/10.5334/jors.427>. Acesso em 5 maio 2023.

INVESTING. Currencies. Disponível em: <https://br.investing.com/currencies/eur-brl-historical-data>. Acesso em: 7 jan. 2023.

ITC - INTERNATIONAL TRADE CENTRE. Export. Disponível em: <https://www.trademap.org/index.aspx>. Acesso em: 7 jan. 2023.

JETBRAINS. Pycharm - Community Edition. Disponível em: <https://www.jetbrains.com/pt-br/pycharm/>. Acesso em 12 fev. 2023.

LEGGETTER, C.J; WOODLAND, P.C. Speaker adaptation of HMMS using linear regression. Technical Report. CUED/F-INFENG/TR.181, Cambridge University Engineering Department, 1994. Disponível em: [http://mi.eng.cam.ac.uk/reports/svr-ftp/auto-pdf/leggetter\\_tr181.pdf](http://mi.eng.cam.ac.uk/reports/svr-ftp/auto-pdf/leggetter_tr181.pdf). Acesso em 10 fev. 2023.

LOURENZANI, W.L.; LOURENZANI, A.E.B.S. Perspectivas do agronegócio brasileiro de amendoim. Informações Econômicas, São Paulo, v.39, n.2, fevereiro, 2009. Disponível em <http://www.iea.sp.gov.br/ftpiea/publicacoes/IE/2009/tec6-0209.pdf>. Acesso em 10 jan. 2023.

MARTINS, R.; VICENTE, J.R. Demandas por inovação no amendoim paulista. Informações Econômicas, São Paulo, v.40, n.5, maio., 2010. Disponível em <http://www.iea.sp.gov.br/ftpiea/publicacoes/IE/2010/tec5-0510.pdf>. Acesso em 7 maio 2023.

MÁXIMO, W. Queda de preço nas commodities reduz valor das exportações agrícolas do país. Agência Brasil. Brasília, 08 de novembro 2014. Disponível em: <https://agenciabrasil.ebc.com.br/economia/noticia/2014-11/queda-nos-precos-e-nos-embarques-fazem-exportacoes-do-agronegocio>. Acesso em 10 maio 2023.

MCKINNEY, W. Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference, Volume 445, 2010. Disponível em: <https://pandas.pydata.org/>. Acesso em: 30 jan. 2023.

MIGUEL, L. O que é série temporal? ICMC Junior, 04 de outubro de 2021. Disponível em: <https://icmcjunior.com.br/o-que-e-serie-temporal/#:~:text=Uma%20s%C3%A9rie%20temporal%20%C3%A9%20simplesmente,uma%20previs%C3%A3o%20para%20o%20futuro>. Acesso em 5 maio 2023.

HARRIS, C.R.; MILLMAN, K.J.; VAN DER WALT, S.J. et al. Array programming with NumPy. Nature 585, 357–362 (2020). DOI: 10.1038/s41586-020-2649-2. Disponível em: <https://numpy.org/>. Acesso em: 30 jan. 2023.

ONU – UNITED NATIONS. Trade Data. Disponível em: <https://comtradeplus.un.org/TradeFlow>. Acesso em: 7 jan. 2023.

PAULI, S. T. Z.; KLEINA, M.; BONAT, W. H. Técnicas de Redes Neurais Artificiais e Regressão Linear Múltipla aplicadas na previsão de valores de importação e exportação no setor de bebidas no Brasil. IX Congresso Brasileiro de Engenharia de Produção. Ponta Grossa – PR. Disponível em [https://aprepro.org.br/conbrepro/2019/anais/arquivos/08092019\\_090853\\_5d4d6ca5e4964.pdf](https://aprepro.org.br/conbrepro/2019/anais/arquivos/08092019_090853_5d4d6ca5e4964.pdf). Acesso em 8 maio 2023.

ROSSATO JUNIOR, J. A. Amendoim brasileiro: um gigante adormecido. Revista Agroanalysis. p.20-21, junho 2022. Disponível em: <https://bibliotecadigital.fgv.br/ojs/index.php/agroanalysis/article/download/88065/82808>. Acesso em 9 jan. 2023.

SHER, V. Times Series Modeling using Scikit, Pandas and Numpy. Towards Data Science, 2020. Disponível em: <https://towardsdatascience.com/time-series-modeling-using-scikit-pandas-and-numpy-682e3b8db8d1>. Acesso em: 2 abr. 2023.

SILVA JUNIOR, G.A. Método de Imputação Incremental, Simples e Eficiente para Dados Ausentes em sistemas Evolutivos. Centro Federal de Educação Tecnológica de Minas Gerais, 2020. Disponível em: [https://sig.cefetmg.br/sigaa/verArquivo?idArquivo=3010142&key=f3f28db226e2c7a68e5\\_56b6fdac17e66](https://sig.cefetmg.br/sigaa/verArquivo?idArquivo=3010142&key=f3f28db226e2c7a68e5_56b6fdac17e66). Acesso em: 2 abr. 2023.

SISCOMEX - SISTEMA INTEGRADO DE COMÉRCIO EXTERIOR. Sistema Harmonizado. Março 2022. Disponível em <https://www.gov.br/siscomex/pt-br/servicos/aprendendo-a-exportar/planejando-a-exportacao-1/sistema-harmonizado>. Acesso em 10 abr. 2023.

SOUZA, J. R. M de. Análise e previsão de séries temporais da exportação de produtos florestais brasileiros utilizando a metodologia boxjenkins e redes neurais artificiais. Viçosa, setembro de 2021. Disponível em: <https://www.locus.ufv.br/handle/123456789/29739>. Acesso em 5 jan. 2023.

SUASSUNA, T. M. F. Sistema de Produção de Amendoim. Embrapa Algodão, fevereiro de 2014. Disponível em: [https://www.spo.cnptia.embrapa.br/conteudo?p\\_p\\_id=conteudoportlet\\_WAR\\_sistema\\_sdeproducao1f6\\_1qa1ceportlet&p\\_p\\_lifecycle=0&p\\_p\\_state=normal&p\\_p\\_mode=view&p\\_p\\_col\\_id=column-1&p\\_p\\_col\\_count=1&p\\_r\\_p\\_-76293187\\_sistemaProducaoId=3803&p\\_r\\_p\\_-996514994\\_topicId=3445](https://www.spo.cnptia.embrapa.br/conteudo?p_p_id=conteudoportlet_WAR_sistema_sdeproducao1f6_1qa1ceportlet&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-1&p_p_col_count=1&p_r_p_-76293187_sistemaProducaoId=3803&p_r_p_-996514994_topicId=3445). Acesso em 1 maio 2023.

VAN ROSSUM, G; DRAKE JR, F. L. Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam. 1 de janeiro de 1995. Disponível em: <https://ir.cwi.nl/pub/5008>. Acesso em 25 jan. 2023.

YANG L. Foreign Trade Export Forecast Based on Fuzzy Neural Network. Complexity, vol. 2021. Disponível em: <https://www.hindawi.com/journals/complexity/2021/5523222/>. Acesso em 5 de jan. 2023.

## **APÊNDICE A – Repositório contendo o código desenvolvido durante o estudo**

O código desenvolvido durante este estudo está hospedado no GitHub, disponível em: <https://github.com/MateusBonacinaZ/avaliacao-de-modelos-ml-exportoes-amendoim>.