

AVALIAÇÃO DE MODELOS DE MACHINE LEARNING NA PREVISÃO DE EXPORTAÇÕES BRASILEIRAS DE AMENDOIM (*ARACHIS HYPOGEA L.*)

Rafael Zanette ¹
Mateus Bonacina Zanguettin ¹
Deise Deolindo Silva ²
João Ricardo Favan ²

Resumo

O amendoim (*arachis hypogea L.*) é uma oleaginosa que vem ganhando força nas exportações brasileiras ao mesmo tempo em 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 às exportações brasileiras de amendoim, ajustados em séries temporais e aplicados aos modelos Linear Regression, Random Forest e Multilayer Perceptron para avaliar o coeficiente de determinação (R²), 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 ocasionadas a partir do péssimo desempenho dos modelos. Dessa forma, houve êxito na avaliação dos modelos, no entanto foram ineficazes os resultados para realizar as previsões.

Palavras-chave: Tecnologia da Informação. Séries temporais. Comércio Exterior.

Abstract. Evaluation of machine learning models in forecasting Brazilian peanut (*arachis hypogea L.*) exports.

Peanut (*arachis hypogea L.*) is an oilseed that has been gaining strength in Brazilian exports while Information Technology (IT) has increasingly helped in the decision-making process in various segments. Therefore, studies on export forecasts using IT are necessary. Therefore, the present work aims to evaluate different machine learning models, which is a subarea of IT, in forecasting Brazilian peanut exports. To carry out the study, data relating to Brazilian peanut exports were collected, adjusted in time series, and applied to the Linear Regression, Random Forest, and Multilayer Perceptron models to evaluate the coefficient of determination (R²), mean squared error (MSE) and the root mean square error (RMSE). During the application of the models, variations in the trend and distribution of data on Brazilian peanut exports were observed, resulting in poor performance of the models. Thus, the evaluation of the models was successful, however the results were ineffective in making predictions.

Keywords: Information Technology. Time series. Foreign Trade.

¹ Discentes do Curso Superior de Tecnologia em Big Data no Agronegócio da FATEC Pompéia (E-mails: rafael.zanette91@hotmail.com e mateus.bonacina.b@gmail.com, respectivamente).

² Professores de Ensino Superior do Curso Superior de Tecnologia em Big Data no Agronegócio da FATEC Pompéia (E-mails: deise.deolindo@fatec.sp.gov.br e joao.favan@fatecpompeia.edu.br, respectivamente).

1 Introdução

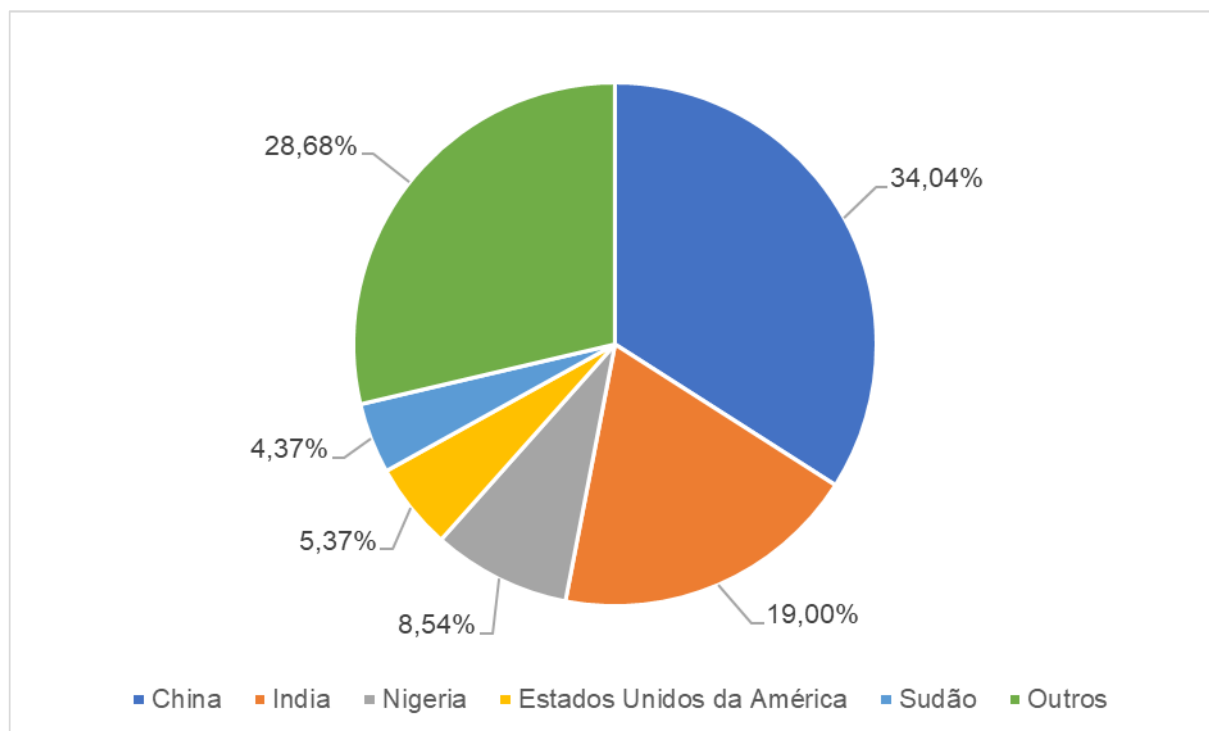
O amendoim (*arachis hypogea* L.) é uma leguminosa muito utilizada em vários países do mundo. Sua importância econômica está conectada ao sabor agradável dos grãos e por serem ricos em óleo (cerca de 50%) e proteína (22% a 30%), além de conter carboidratos, sais minerais e vitamina E e complexo B, sendo 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, sendo introduzido na Europa no século XVIII e no século XIX houve uma difusão do Brasil para a África, do Peru para as Filipinas, China, Japão e Índia (SUASSUNA, 2014).

Essa oleaginosa é bastante consumida no mundo. Segundo dados disponibilizados pela FAO (2023) a somatória da produção mundial foi cerca de 53,9 milhões de toneladas no ano de 2021, sendo a China a maior produtora com cerca de 34,04% de toda a produção, 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 negociado, visto que foram exportados em todo mundo no ano de 2021, mais de US\$ 4,2 bilhões 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) (em US\$ mil).

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 US\$ 2.465.132 em 2018 para US\$ 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 & 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 Alcool (Proálcool), 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 à mudança na dinâmica do cenário econômico mundial, houve incentivos às culturas voltadas à exportação (*tradeables*) que resultaram no declínio da produção de amendoim que foi dando lugar à soja (LOURENZANI & LOURENZANI, 2009). A diminuição da produção dessa cultura foi marcada pela baixa rentabilidade, fatores tecnológicos e qualidade (Id., 2010).

Essa situação continuou até a década de 1990, no entanto no início dos anos 2000 a produção do amendoim começou a mudar devido a 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 RDC172 de 2003 que estabelece o regulamento técnico de boas práticas de fabricação para estabelecimentos que industrializam amendoins processados e derivados

e à publicação da Instrução Normativa nº 3 em 2009 pelo MAPA (Ministério da Agricultura, Pecuária e Abastecimento) que dispõe de critérios e de 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. 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 aos padrões do mercado europeu, conseguindo bons resultados e mostrando comprometimento com o meio ambiente e, a partir das inovações tecnológicas, proporcionando uma melhor comercialização no mercado (ASCANIO, MALAGOLI & FREIRE, 2018).

Atualmente o mercado de amendoim passa por um período de crescimento. De acordo com a CONAB (2022) houve um aumento de produção de mais de 100% em 8 anos no país, 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 - cerca de 100% quando os períodos são comparados. Cerca de 70% dessa produção é exportada. A exportação é o principal fator que tem estimulado o aumento na produção de amendoim e este fato se justifica por duas grandes tendências globais: a primeira é o mercado de carbono, levando-se em consideração que a pegada de carbono do amendoim nacional é 59% inferior à média mundial e a segunda está relacionada às novas gerações e às mudanças nos seus hábitos alimentares com um crescente mercado de alimentos à base de ingredientes de origem vegetal (*plant-based*) (ROSSATO JÚNIOR, 2022).

Portanto, torna-se cada vez mais necessário realizar previsões de exportações do amendoim brasileiro. 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 cinco políticas para o setor por se tornar uma ferramenta eficiente para a tomada de decisões (SOUZA, 2021).

Uma série temporal (ST) é definida 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 (MIGUEL, 2021). Um dos benefícios das previsões de séries temporais está na possibilidade de realizar um bom planejamento e evitar gastos desnecessários (PAULI, KLEINA & BONAT, 2019). Além disso, observa-se que é de extrema importância a previsão de séries temporais macroeconômicas para os decisores das políticas monetária e orçamentária avaliarem o desenvolvimento econômico. 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 (TI) para realizar previsões com séries temporais, uma delas é com o uso de Machine Learning (ML) que é uma subárea da inteligência artificial oferece suporte para previsão de séries temporais (CAMPOS, 2020).

Ao contrário dos métodos estatísticos tradicionais, os de ML 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 *apud* Id., 2020).

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 (BREIMAN, 2001). Alguns estudiosos

utilizam nos estudos de previsão as séries temporais (CARVALHO *et al.*, 2021) os modelos de Linear Regression e Random Forest e outros os de Linear Regression e Multilayer Perceptron (PAULI, KLEINA & BONAT, 2019).

Considerando a importância do amendoim para as exportações brasileiras e a capacidade dos modelos de ML em realizar previsões com o uso de séries temporais é necessário o uso deles para exportações da oleaginosa com o intuito de fornecer informações importantes para os agentes envolvidos na cadeia produtiva desta cultura.

Assim, o presente trabalho tem o objetivo de 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) indicadores de avaliação.

Todas essas etapas estão descritas a seguir.

Coleta e tratamento de dados

Foram utilizados dados históricos relacionados à cultura do amendoim como exportação, produção, área colhida e as cotações das moedas Euro e Dólar dos Estados Unidos da América 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 do Brasil e dos dois maiores exportadores mundiais de amendoim (Argentina e Índia), visto que esses podem interferir nas exportações (LIU *et al. apud* YANG, 2021).

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

As informações relacionadas às áreas colhidas e à produção brasileira de amendoim e produção dos três principais produtores mundiais dessa cultura (China, Índia e Nigéria) foram coletadas do *site* da FAO (2023).

A produção de cana-de-açúcar no Brasil está relacionada à do amendoim, pois é usada na entressafra da cana-de-açúcar para recuperação do solo (ASCANIO, MALAGOLLI & FREIRE, 2021), assim, considerou-se a produção dessa gramínea a partir de dados também coletados na FAOSTAT (FAO, 2023).

Foi realizado, a partir de toda a coleta de dados, o tratamento que constituiu nas etapas de criação de *data frames* para os arquivos, substituição dos valores faltantes com o uso do método de imputação pela média como sugerido por Silva Júnior (2020) e centralização em um único *data frame*. 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, MILLMAN & VAN DER WALT, 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 líquidos) 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) e *brazil_export_groundnut_last_month* (exportações brasileira de amendoim do mês anterior em quilogramas líquido).

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).

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 com a validação sempre sendo feita com o *fold* do ano posterior aos usados para o treinamento, conforme demonstrado 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: próprios autores (2023).

Configurações dos modelos

Foram utilizados três modelos de Machine Learning supervisionado de regressão contidos na biblioteca SkLearn, sendo eles os de 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. Ajustou-se o de Random Forest Regressor com um número de 1.000 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.

Indicadores de avaliação

Os modelos ajustados foram avaliados utilizando os seguintes indicadores de desempenho 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) que tem como função melhorar a interoperabilidade da métrica (AZANK, 2020).

3 Resultado e discussão

Os resultados e discussões foram divididos em três fases, sendo:

- a) seleção das features, onde foram selecionadas as que têm 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;
- c) considerações adicionais.

Seleção das features

Para a seleção das features mais importantes foi aplicada 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>	<i>Relação com o 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 (2023).

Para a aplicação dos modelos de Machine Learning (ML) foram selecionadas 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*.

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 (2023).

Tabela 5. Classificação de desempenho.

Classes	R^2	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, Sentelhas & Ferraz *apud* Carvalho *et al.* (2021).

Comparando os resultados da Tabela 4 com os da Tabela 5, que classifica o desempenho dos modelos considerando o R^2 , pode-se observar que o desempenho de todos os modelos é considerado péssimo de acordo com a média do R^2 . O MSE e o RMSE também indicam que os modelos não tiveram bom desempenho, pois segundo Azank (2020) quanto mais próximo de 0 os valores 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 *folds* em R².

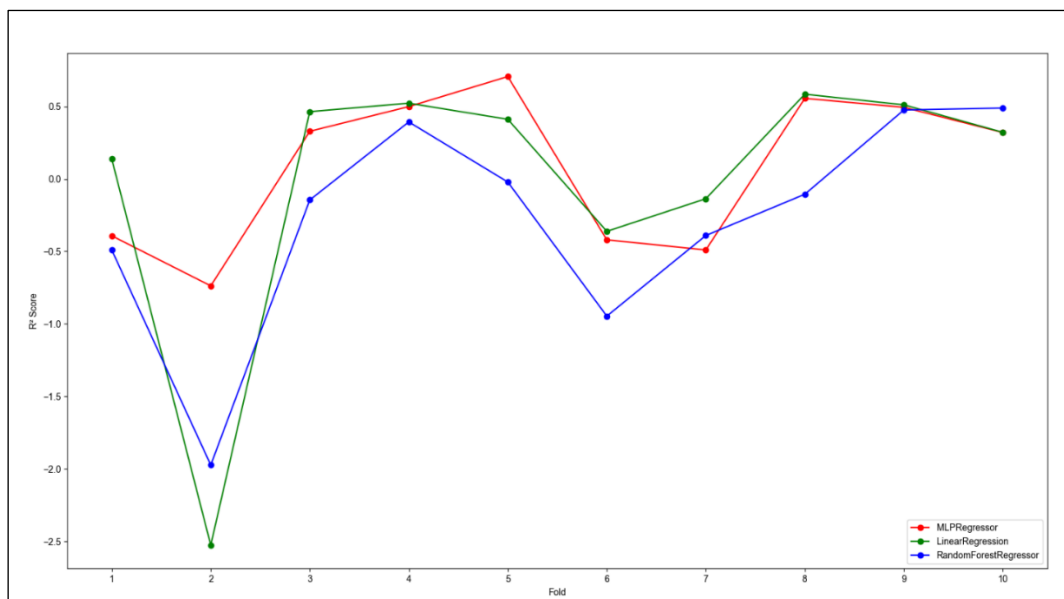
<i>Folds</i>	Linear Regressor	Random Forest Regressor	MLP Regressor
1	0,1373	-0,4899	-0,3923
2	-2,5248	-1,9727	-0,7372
3	0,4637	-0,1451	0,3281
4	0,5218	0,3926	0,4984
5	0,4109	-0,0214	0,7055
6	-0,3614	-0,9458	-0,4199
7	-0,1365	-0,3901	-0,4897
8	0,584	-0,1056	0,5551
9	0,5106	0,4751	0,4934
10	0,3209	0,4885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

Fonte: autores (2023).

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 *folds* positivos dos dez avaliados. De todos *folds* avaliados, o Linear Regression obteve o R² 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* 2 e 6 que foi determinante para a média de desempenho ser considerada péssimo, visto que foram valores negativos e proporcionalmente altos. A Figura 2 mostra o desempenho dos modelos com a indicador R² em cada *fold*.

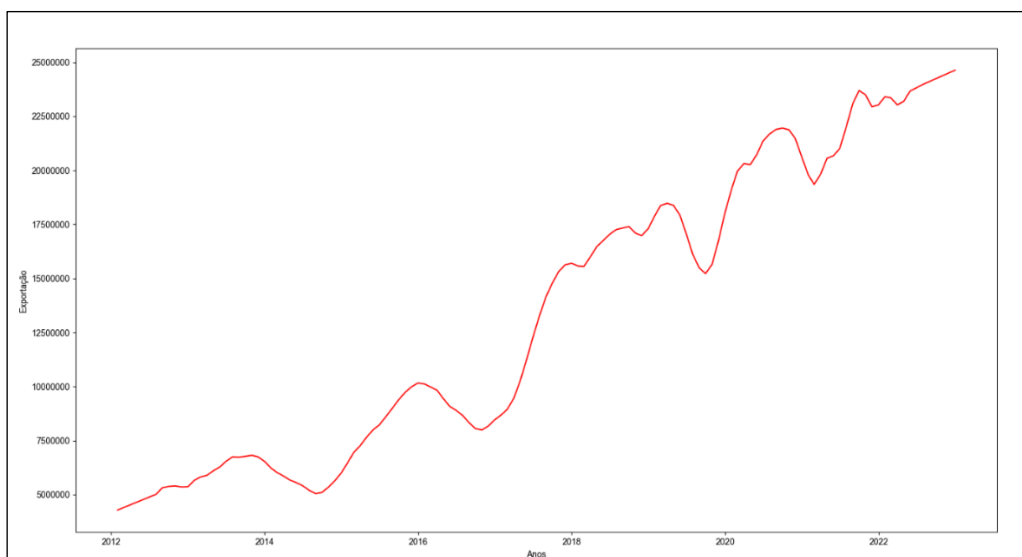
Figura 2. Desempenho dos modelos em R^2 em decorrência dos *fold*s.



Fonte: autores (2023).

O péssimo desempenho dos modelos no *fold* 2 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 (2023).

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

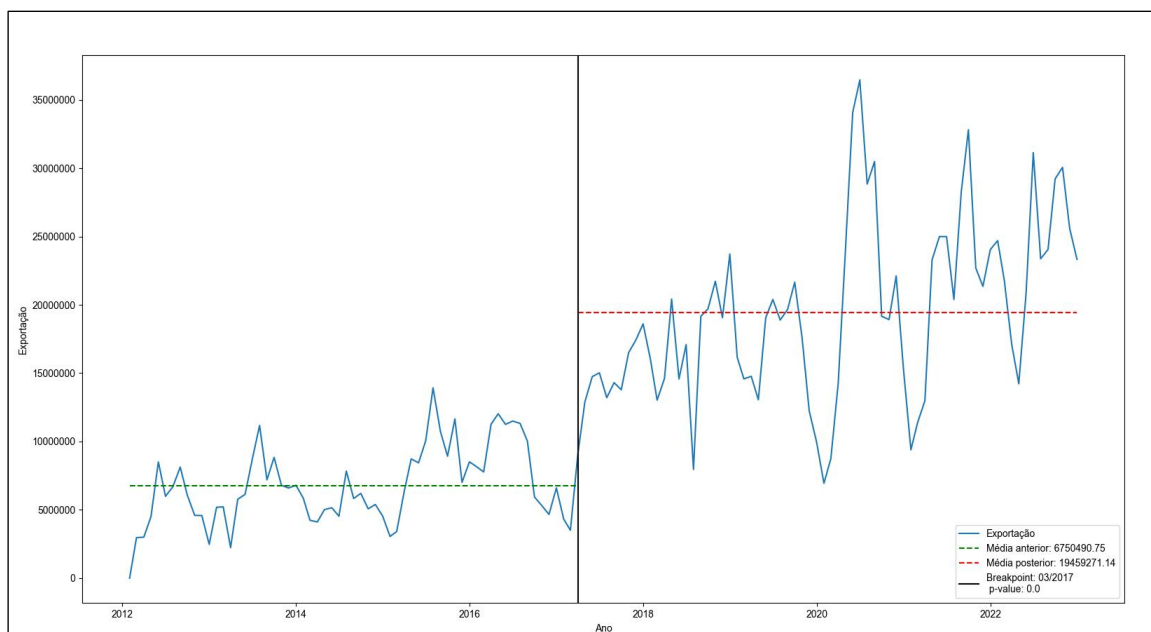
A mudança de tendência ocorrida com os dados de validação do *fold 2* (ano de 2014) pode estar relacionado com a queda de exportações brasileiras ocorridas nesse ano, pois houve diminuições nos preços das *commodities* como relatado por Amato (2015) e Máximo (2014).

Deve se notar que para os próximos *folds* (3, 4 e 5) os modelos Linear Regressor e MLP Regressor tiveram melhora no desempenho se comparado com o *fold 2*.

No *fold 6* houve uma queda brusca no desempenho causada por um *breakpoint* (ponto de mudança) na série temporal, que segundo Hussain, Mahmud & 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 teste de homogeneidade, que é um teste estatístico que detecta em séries temporais um ou mais *breakpoints* (Id., 2023). O teste de homogeneidade utilizado foi o de Pettitt que é usado para identificar o ponto de início de impacto (ASSIS, 2014). O resultado desse teste pode ser analisado na Figura 4.

Figura 4. Teste de Pettitt.

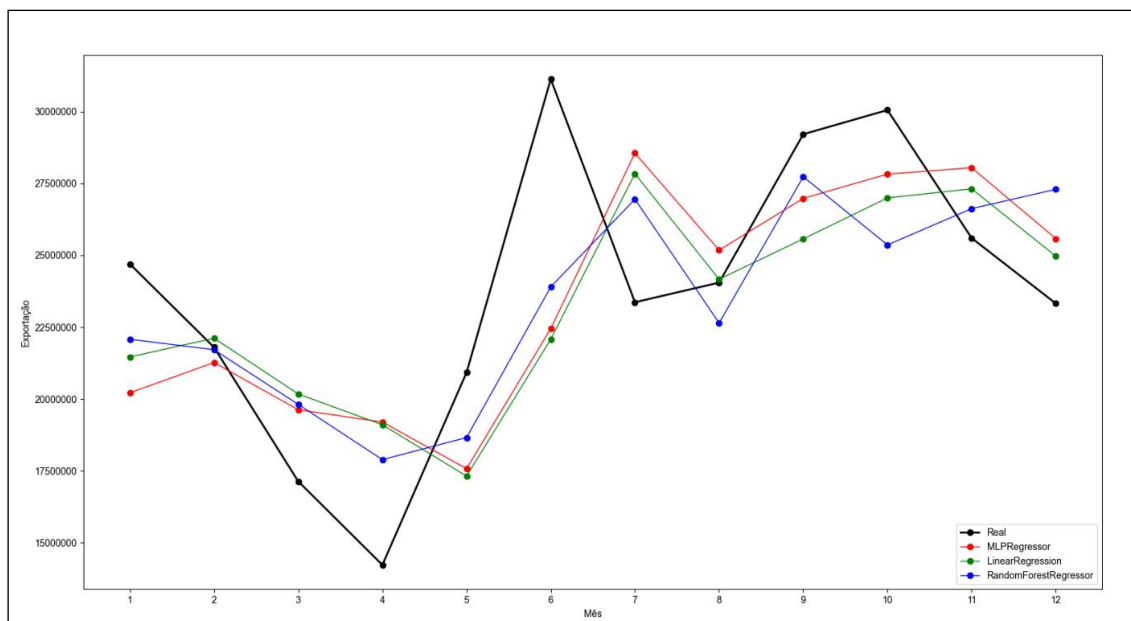


Fonte: autores (2023)

Esse *breakpoint* prejudicou o desempenho dos modelos nos *folds 6* e *7*, pois os modelos validaram com dados as diferentes distribuições utilizadas para treinamento.

Nos *fold*s seguintes (8, 9 e 10) 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.



Fonte: autores (2023).

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 os reais é grande.

Considerações adicionais

Embora, o desempenho dos modelos de modo geral foi classificado como péssimo, outros trabalhos, conduzidos por de Souza (2021) e Pauli, Kleina & Bonat (2019), conseguiram os 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 (ML).

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

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

4 Considerações finais

Os modelos de Machine Learning (ML) 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.

Referências

- AMATO, F. **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 mai. 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 2018. 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 mai. 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%A3oc2c8d73dab96>>. Acesso em: 2 mai. 2023
- BREIMAN, L. **Randon Forest**. *Machine Learning* **45**, 5-32, oct. 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, nov. 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

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

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.

HUSSAIN, M., MAHMUD, I., BARI, S. H. pyHomogeneity: A Python Package for Homogeneity Test of Time Series Data. **Journal of Open Research Software**. 14 feb. 2023. Disponível em: <<https://openresearchsoftware.metajnl.com/articles/10.5334/jors.427>>. Acesso em: 5 mai. 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>. 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 mai. 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 mai. 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 mai. 2023.

ONU – ORGANIZAÇÃO DAS NAÇÕES UNIDAS. **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**. 2019. IX Congresso Brasileiro de Engenharia de Produção. Ponta Grossa – PR. Disponível em: <https://aprepro.org.br/conbrepro/2019/anais/arquivos/08092019_090853_5d4d6ca5e4964.pdf>. Acesso em: 8 mai. 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.

ROSSATO JÚNIOR, 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 JÚNIOR, 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=f3f28db226e2c7a68e556b6fdac17e66>>. 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. **Análise e previsão de séries temporais da exportação de produtos florestais brasileiros utilizando a metodologia boxjenkins e redes neurais artificiais**. Viçosa (MG). 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_sistemasdeproducao>

lf6_1galceportlet&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_topicoId=3445>. Acesso em: 1 mai. 2023.

VAN ROSSUM, G., DRAKE JR., F. L. **Python reference manual**. Centrum voor Wiskunde en Informatica. Amsterdam. 1 jan. 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 jan. 2023.