



08, 09, 10 e 11 de novembro de 2022  
ISSN 2177-3866

## **DESENVOLVIMENTO DE MODELO DE PREDIÇÃO DE PREÇO MÉDIO PARA PET RECICLADO DE GRAU ALIMENTÍCIO**

**JOÃO PAULO FUMES BENETON**

FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DA UNIVERSIDADE DE SÃO PAULO - FEA

**ADRIANA MAROTTI DE MELLO**

FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DA UNIVERSIDADE DE SÃO PAULO - FEA

# DESENVOLVIMENTO DE MODELO DE PREDIÇÃO DE PREÇO MÉDIO PARA PET RECICLADO DE GRAU ALIMENTÍCIO

## 1. Introdução

A Economia Circular vem ganhando muita força na agenda das empresas e das políticas públicas em muitas partes do mundo. Esse conceito vem direcionando pesquisas e influenciando políticas visando um mundo mais sustentável. (Genovese et al., 2017). Dentro desse tema, a gestão de cadeias de suprimentos fechadas – conceito que consiste em reaproveitar o material descartável – na forma de matérias-primas – para a produção de outros bens de consumo, vem ganhando considerável destaque na agenda de pesquisas em operações (Braz e Mello, 2022). Entretanto, um dos empecilhos para a implementação em larga escala dessa proposta é o preço do insumo gerado através dos materiais descartados, uma vez que, há forte incerteza em relação ao comportamento desse preço, se comparado aos insumos primários, o que prejudica o planejamento e a gestão das empresas que trabalham com esse tipo de material. Além disso, há muito poucos estudos na literatura que tratem desse tema – boa parte deles são dedicados a materiais metálicos, e não plástico – material amplamente consumido e uma das maiores fontes de poluição ambiental.

Diante deste cenário, este artigo pretende, utilizando-se do plástico reciclado (PET) como material de estudo e com o auxílio do aprendizado de máquina (*machine learning*), propor uma forma de previsão do preço médio – predição – do material secundário advindo dos plásticos descartados e expor os fatores que afetam diretamente o preço destes materiais, para que, com isso, haja uma maior conciliação entre a demanda e a oferta do material em voga. Sob essa prerrogativa, acredita-se que o presente estudo poderá auxiliar na gestão de cadeias de suprimentos de ciclo fechado, uma vez que a possibilidade de previsão do preço do insumo reciclado poderá impactar na conciliação entre a demanda e a oferta deste, o que afetará, posteriormente, a lucratividade daqueles que fazem parte dessa cadeia.

De modo semelhante, o modelo de predição poderá auxiliar na diminuição da incerteza sobre a volatilidade do preço dos materiais reciclados, no caso, os derivados do plástico. Sendo que essa volatilidade expõe muitas organizações a riscos financeiros e regulatórios devido às incertezas geradas. Portanto, nossa contribuição será em duas vias: para a literatura, através da discussão dos fatores e a aplicação de uma metodologia inovadora na análise do problema: para os gestores de empresas e formuladores de políticas públicas, a ferramenta de previsão de preço auxilia no processo de tomada de decisão sobre o uso de materiais reciclados.

Esse artigo está estruturado da seguinte forma: na seção 2, é apresentada a fundamentação teórica do texto

## 2. Fundamentação Teórica

### 2.1 Fatores que influenciam o preço dos materiais reciclados

Desde o século passado há um enorme interesse em estabelecer uma relação de causa e efeito envolvendo os preços dos materiais primários e secundários – materiais reciclados. Os estudos, a priori, apontaram que as mudanças nos preços dos insumos primários eram determinantes para explicar as mudanças no preço do material reciclado, os quais, muitas vezes, eram derivados de metais, uma vez que, havia maior interesse para entender o comportamento do preço de insumos metálicos, como cobre e alumínio. Fisher e Owen (1981) apontavam tal relação entre o alumínio primário e seu respectivo material reciclado.

Entretanto, estudos apontavam que outros fatores influenciavam o preço do material secundário para além do preço primário. De acordo com Taylor (1979), que estudou o mercado da sucata derivada do cobre, o processo de refinamento do metal juntamente da demanda

envolvendo o mercado primário poderiam afetar o preço final da sucata. Com isso, havia a necessidade de contabilizar mais variáveis para explicar o preço do material reciclado.

Neste aspecto, Ackerman e Gallagher (2002) apresentaram um artigo envolvendo a alta dos preços no mercado de papel reciclado em 1994 e 1995 nos Estados Unidos e, posteriormente, em 1996, o retorno do preço deste material aos “níveis tradicionais”, ocasionando custos econômicos e políticos substanciais. A especulação foi apontada, segundo os autores, como a responsável pelo aumento súbito do preço, que criou um cenário enganoso para o investimento no setor.

Diante dessa prerrogativa, Xiarchos e Fletcher (2009) buscaram apontar uma correlação de curto e longo prazo entre os preços do cobre, chumbo e zinco e a respectiva sucata derivada desses insumos. Os autores, neste aspecto, buscaram averiguar se, tanto no curto quanto no longo prazo, os fatores que afetam a variação do preço/volatilidade dos materiais reciclados eram os mesmos. Para tal averiguação foram utilizados métodos envolvendo séries temporais dentro do período de 1984 até 2001. Foi constatado pelos autores que, no longo prazo, há uma forte correlação entre os preços dos produtos primários e secundários para os metais analisados, todavia, no curto prazo, as flutuações arbitrárias do mercado também influenciavam o preço final da sucata, tornando o ato de prever, neste prazo, os preços dos materiais reciclados algo pouco aplicável até então.

Posteriormente e em relação ao Brasil, segundo entrevistas realizadas por Santos (2020), há algumas variáveis que possuem uma maior influência no preço do material reciclado. Segundo o estudo citado, tais variáveis são:

- Oferta e demanda de materiais recicláveis;
- Preço do material primário;
- Preço do dólar;
- Preço do combustível;
- Energia/custo logístico.

Parte destes preços elencados envolve práticas vinculadas à Cadeias de Suprimentos Fechadas, as quais, segundo Guide e Van Wassenhove (2009), são sistemas complexos que englobam o fluxo reverso de produtos, adquiridos e consumidos pelos indivíduos, para o início da cadeia produtiva, ou seja, o produto consumido retorna como insumo para o fabricante e esse processo envolve custos, os quais são incorporados ao preço dos insumos secundários.

## **2.2 Modelo de regressão linear simples envolvendo previsões de materiais reciclados**

Segundo Govindan e Soleimani (2015), há poucos modelos de previsão em geral no que tange cadeias de abastecimento circulares – Cadeias de Suprimentos Fechadas. Diante disso, apresenta-se o modelo de previsão de preço de materiais reciclados proposto por Angus, Casado e Fitzsimons (2011) dentro do mercado do Reino Unido. A análise em voga buscou estabelecer uma melhor compreensão sobre a volatilidade do preço do material reciclável – plástico – por meio de um modelo de regressão linear simples utilizando-se, como base, o preço do insumo primário.

O modelo de regressão proposto por Angus, Casado e Fitzsimons (2011) iniciava-se com a estimação da correlação entre os preços dos insumos primários e secundários. Nessa primeira etapa registrou-se uma forte correlação entre o preço do petróleo bruto, eteno e nafta e o preço dos plásticos reciclados.

Posteriormente, uma vez averiguado que as variáveis envolvendo o plástico possuem correlação significativa – e utilizando-se de métricas, como o Teste F, para corroborar com tal análise de correlação –, Angus, Casado e Fitzsimons (2011) apresentaram o valor da constante e o valor do coeficiente da variável primária – valores presentes na fórmula da regressão linear.

Nesta prerrogativa aponta-se que, Angus, Casado e Fitzsimons (2011) propuseram uma equação linear que apresentava um valor previsto para o preço dos materiais secundários com base nos preços primários. Entretanto, diante do exposto, Angus, Casado e Fitzsimons (2011), também afirmaram que, cálculos mais sofisticados – regressões múltiplas – seriam mais adequados para um modelo de previsão de preço, uma vez que, o próprio artigo informou que a hipótese de haver apenas uma variável para determinar o preço do material reciclado é pouco realista.

### 3. Metodologia

#### 3.1 Machine Learning

Diante da complexidade de se alcançar um modelo de predição de preço factível para os materiais reciclados utilizando-se apenas de métodos relacionados a regressão linear simples, Angus, Casado e Fitzsimons (2011) propõem a aplicação de técnicas mais robustas para alcançar melhores resultados.

Nessa perspectiva, para alcançar maior robustez, o presente estudo utilizará do aprendizado de máquina (*machine learning*) – junto às bibliotecas existentes no software Python. Sobre esse tema, Wade (2020, p. 55) afirma que *machine learning* “é a capacidade dos computadores de aprender a partir de dados”, ou seja, é uma tecnologia na qual a máquina tem a possibilidade de aprender e criar associações com os dados fornecidos. E dentro deste tema, afirma-se que a subárea de machine learning destacada neste estudo é a regressão, a qual busca estimar um valor numérico para os alvos do estudos envolvendo o aprendizado de máquinas. Neste contexto, a implementação de machine learning consiste na utilização de algoritmos que “aprendem” com o conjunto de dados estabelecido e, posteriormente, realizam previsões. (WADE, 2020).

De acordo com Wade (2020, p. 393), “a correlação é uma medida estatística entre -1 e 1 que indica a força da relação linear entre dois pontos. Uma correlação de 1 é uma linha perfeitamente reta, enquanto uma correlação de 0 indica uma relação não linear”.

Diante da definição apresentada acima é desejável averiguar se as variáveis que irão compor o modelo de regressão possuem uma alta correlação entre si – aproximando-se de 1 (forte correlação positiva) ou -1 (forte correlação negativa). Ou seja, como apontado por Angus, Casado e Fitzsimons (2011), é necessário que haja um alto coeficiente de correlação entre as variáveis para que o modelo de regressão proposto seja verossímil. Ademais, estudos gráficos podem corroborar na “visualização” das análises de correlação.

#### - Algoritmos

Os algoritmos de *machine learning* presentes neste estudo consistem em: regressão linear (linear regression), árvore de decisão (decision tree) e florestas aleatórias (random forest). Abaixo encontram-se os conceitos vinculados a esses algoritmos.

Sobre o algoritmo de regressão linear é dito que esse algoritmo “ajusta um modelo linear com coeficientes  $w = (w_1, \dots, w_p)$  para minimizar a soma residual dos quadrados entre os alvos observados no conjunto de dados e os alvos previstos pela aproximação linear.” *scikit-learn user guide* (2017, p. 197)

Sobre o algoritmo de árvore de decisão afirma-se que seu objetivo “é criar um modelo que preveja o valor de uma variável através do aprendizado de regras de decisão simples inferidas a partir dos recursos de dados.” *scikit-learn user guide* (2017, p. 216)

Em sequência, sobre florestas aleatórias apresenta-se que esse algoritmo “é um meta estimador que se ajusta a um número de árvores de decisão de classificação em várias subamostras de o conjunto de dados e usa a média para melhorar a precisão preditiva e controlar o ajuste excessivo.” *scikit-learn user guide* (2017, p. 508)

Por fim, expõe-se ainda que, para uma maior comparação dos resultados vinculados aos algoritmos de *machine learning*, foi introduzido o *Dummy Regressor*, o qual, de acordo com o

guia oficial, é um “regressor que faz previsões usando regras simples”. (SCIKIT LEARN, 2022c, [sem página]). Esse regressor é útil como uma linha de base para comparação com outros regressores reais (SCIKIT-LEARN, 2017).

Neste estudo em questão, utilizou-se das médias e das medianas para compor o *Dummy Regressor*, ou seja, esse algoritmo em questão irá prever a média e a mediana do conjunto de treino, valores estes que irão compor a previsão desse algoritmo.

#### - **Treino, teste e validação**

Segundo Matt Harrison (2020), o conjunto de treino é responsável pela criação do modelo – envolvendo os algoritmos – e o conjunto de validação é destinado à implementação das métricas – responsáveis por averiguar o desempenho dos modelos criados, a partir das linhas de treino. Por fim, as linhas de teste são utilizadas para constatar se o modelo proposto é capaz de apontar valores satisfatórios para dados não antes expostos ao modelo criado.

#### - **Métricas**

As métricas a serem utilizadas neste estudo são:  $R^2$ , o erro quadrático médio (MSE) e a raiz do erro quadrático médio (RMSE).

O  $R^2$ , também conhecido como coeficiente de determinação, representa a proporção da variância (de  $y$ ) que foi explicada pelas variáveis independentes no modelo.” *scikit-learn user guide* (2017, p. 566)

Neste contexto, afirma-se ainda que esta métrica “fornece uma indicação da qualidade do ajuste e, portanto, uma medida de quão bem as amostras não vistas provavelmente serão previstas pelo modelo, por meio da proporção da variância explicada.” *scikit-learn user guide* (2017, p. 566)

O erro quadrático médio (MSE) é “uma métrica de risco correspondente ao valor esperado do erro ou perda quadrática.” *scikit-learn user guide* (2017, p. 564)

Sobre essa métrica, Azank afirma que um de seus pontos negativos é “a falta de interpretabilidade direta” (AZANK, 2022, [sem página]). Nessa perspectiva, o autor pontua que a raiz do erro quadrático médio, a qual é “derivada” da métrica acima, atua como uma forma de melhorar/auxiliar a interpretação do erro em questão (MEDIUM, 2020).

#### - **Dados Utilizados**

As variáveis independentes ou previsoras utilizadas neste presente estudo foram:

- Preço médio do material-virgem na região sudeste;
- Preço médio de outros materiais secundários na região sudeste/sul;
- Preço médio do combustível na região sudeste;
- Preço do dólar.

Os materiais secundários – elencados acima – a serem utilizados são:

- PET (polietileno tereftalato);
- PVC (policloreto de vinila);
- PS (poliestireno);
- PP (polipropileno);
- PEBD (polietileno de baixa densidade).

Diante do exposto, apresenta-se que o preço médio do material-virgem na região sudeste, assim como o preço médio de outros materiais secundários, será obtido através da assinatura de uma base privada de dados, que se encontra dentro do ramo de avaliação de negócios na indústria química com foco em análise de mercados e competitividade. Ademais, o preço do combustível na região sudeste será obtido através da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP), uma agência do governo brasileiro cuja função é fiscalizar, regular e controlar as atividades do setor petrolífero brasileiro. Por fim, sobre a cotação do dólar, será utilizado as informações retiradas da Associação Comercial, Industrial e de Serviços de Novo Hamburgo (ACI).

## 4. Resultados

### 4.1.Criação do Dataframe

O preço médio do plástico secundário selecionado como variável dependente quantitativa – variável em estudo – é o preço médio do PET reciclado de Grau Alimentício – cujos dados, assim como os demais preços médios dos materiais secundários, constam na base de dados da empresa utilizafa

Ademais, visto que um dos objetivos deste estudo é prever o preço futuro do PET reciclado de Grau Alimentício, modificou-se a coluna desta variável – também denominada PET Condensado –, sendo que, esta modificação consiste em alterar o valor do preço médio no mês presente para o valor do preço médio no mês subsequente – por exemplo, de acordo com a Tabela exposta a seguir, o preço médio existente no *dataframe* para o mês de janeiro de 2015 envolvendo a variável em questão é de quatro reais. Este preço diz respeito ao valor médio observado no mês de fevereiro de 2015, do mesmo modo que o valor de três reais e noventa centavos exposto na segunda linha (2015-02-01), consiste no valor médio para variável no mês de março de 2015 e assim sucessivamente. Enquanto os valores para as demais variáveis representam o preço médio para o mês em questão, ou seja, o valor presente, um real e quarenta e cinco centavos, na primeira linha (2015-01-01) na coluna sucata representa o preço desta variável no mês de janeiro de 2015. O mesmo ocorre com a variável combustível, assim como as demais quinze variáveis preditoras, as quais, devido ao espaço delimitado, não foram contabilizadas nesta imagem em questão.

Nesta prerrogativa, nota-se que este estudo analisou dezessete variáveis preditoras – como observado na figura abaixo – para apontar o preço médio do PET reciclado de Grau Alimentício (PET Condensado). Ademais, a série temporal analisada inicia-se em janeiro de 2015 e estende-se até dezembro de 2021, sendo essa série delimitada pelos dados fornecidos pela empresa assinada.

Para a criação desta tabela a seguir – juntamente das informações dos dados – utilizou-se da biblioteca *Pandas* (2022c), a qual foi desenvolvida por Wes McKinney em 2008.

Tabela - Análise dos dados

	sucata	combustivel	pet_condensado	DatetimeIndex: 84 entries, 2015-01-01 to 2021-12-01	
				Data columns (total 18 columns):	
data				#	Column
2015-01-01	1.45	2.529108	4.0	0	sucata
2015-02-01	1.45	2.680011	3.9	1	pet_cristalizado
2015-03-01	1.40	2.679371	3.9	2	pet_verde
2015-04-01	1.40	2.665130	3.9	3	pet_cristal
2015-05-01	1.40	2.634260	3.9	4	pet_verde_sul
...	...	...	...	5	pvc_flexível
2021-08-01	3.85	5.135897	8.1	6	pvc_rígido
2021-09-01	4.00	5.225937	8.5	7	ps_natural
2021-10-01	4.20	5.436353	8.7	8	ps_colorido
2021-11-01	4.10	5.669871	8.6	9	pp_natural
2021-12-01	4.05	5.583845	8.7	10	pp_branco
				11	pp_colorido
				12	pebd_natural
				13	pebd_branco
				14	pebd_colorido
				15	dolar
				16	combustivel
				17	pet_condensado

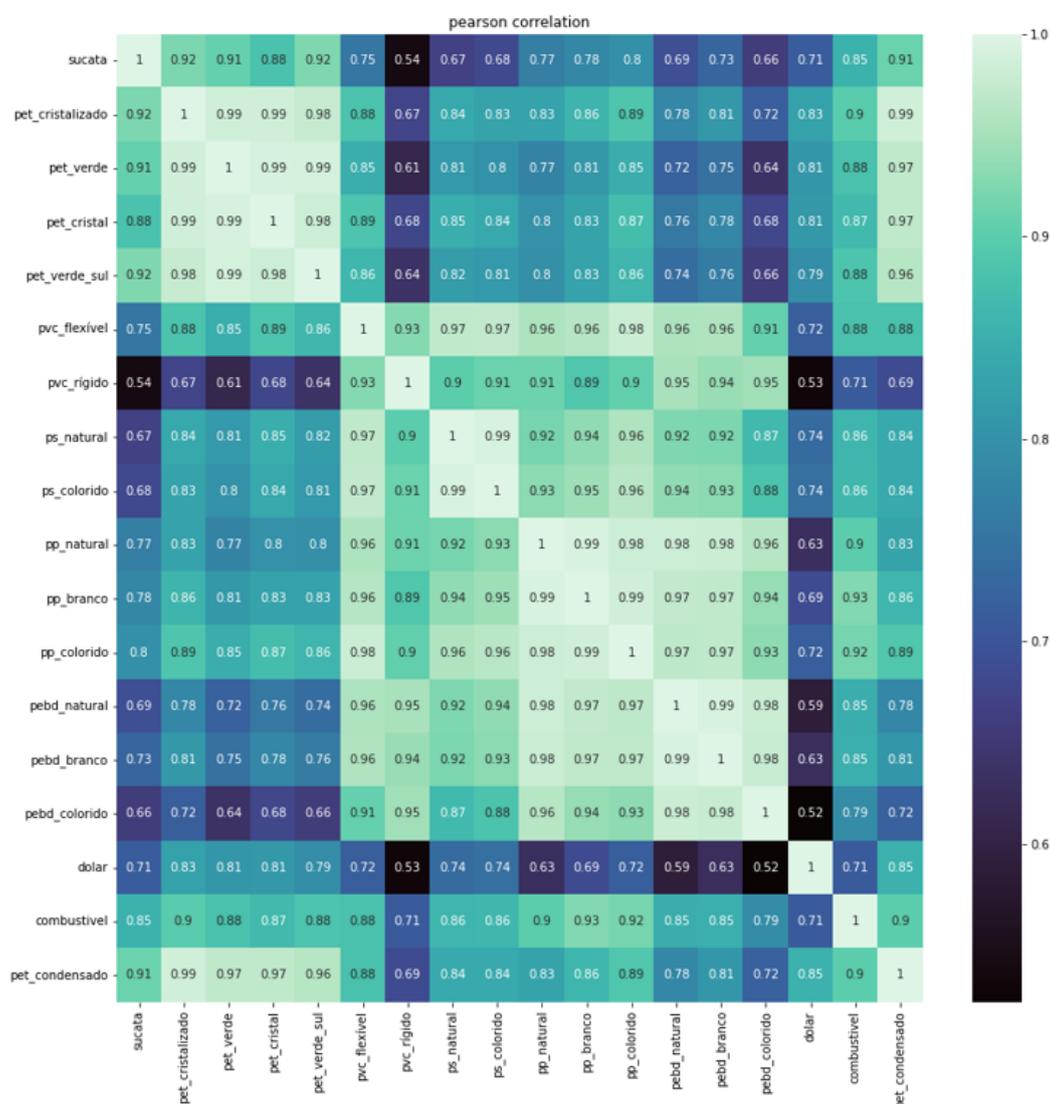
Fonte: produção do autor, 2022.

Para a análise das correlações e para a criação do modelo de regressão, posteriormente, utilizou-se, como inputs, os valores do dataframe criado e exposto parcialmente acima.

## 4.2. Correlação

A imagem com as correlações das variáveis se encontra na figura a seguir – para a criação deste conjunto de correlações utilizou-se da biblioteca *Seaborn* (2022), a qual foi desenvolvida por Michael Waskom e da biblioteca *Pandas* (2022b), utilizada anteriormente. Analisando a figura abaixo, percebe-se que há uma alta correlação – próximas ou superiores a 0.70, observada na última linha da figura – entre as variáveis e a variável dependente quantitativa. Diante disso, julgou-se válido manter todas variáveis predictoras para a criação do modelo, visto que estas variáveis conseguem, em maior ou menor grau, explicar o comportamento da variável dependente quantitativa.

Figura - Correlação

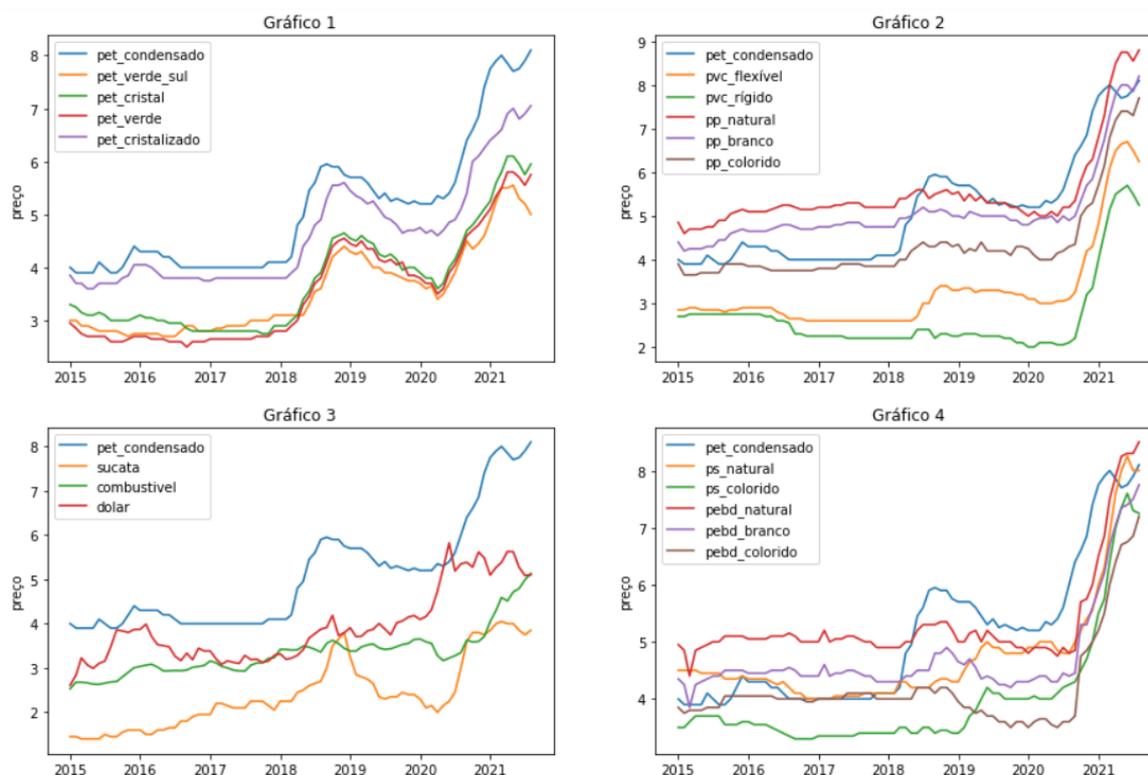


Fonte: produção do autor, 2022.

Para complementar este estudo das correlações criou-se o gráfico a seguir que demonstra o comportamento do preço médio das variáveis ao longo do tempo. Este gráfico – produzido através da biblioteca *Matplotlib*, criada por John D. Hunter (MATPLOTLIB, 2022) – corrobora com a hipótese de que todas as variáveis consideradas como *input* podem auxiliar

na previsão do *PET Condensado*, uma vez que o comportamento das linhas possui certa semelhança.

Figura 5 - Análise gráfica



Fonte: produção do autor, 2022.

Sobre os gráficos acima afirma-se que, para a maior compreensão da relação das variáveis predictoras com o *PET Condensado*, buscou-se separar essas variáveis em quatro diferentes gráficos e adicionou-se a variável dependente quantitativa – *PET Condensado*

Nota-se no primeiro gráfico, no qual todos os materiais oriundos do *PET* foram separados, que o *PET Condensado*, dentro da linha temporal analisada, apresenta o maior valor de compra e, nitidamente, observa-se que há uma semelhança entre o comportamento das linhas, o que corrobora para alta correlação. No segundo gráfico contendo o *PET Condensado* juntamente com os derivados do *PVC* e do *PP*, nota-se, novamente, certa semelhança entre as linhas do gráfico, tal semelhança também pode ser observada no quarto gráfico. Sobre o terceiro gráfico – contendo o *PET Condensado*, a sucata (insumo primário), o preço do dólar e do combustível – percebe-se que a relação do *PET Condensado* com as demais variáveis contidas no gráfico, possui maiores inconstâncias no curto prazo – principalmente envolvendo o preço do dólar e o preço médio do combustível –, apesar de haver certa paridade entre o *PET Condensado* e o preço médio da sucata.

### 4.3 Treino, teste e validação

Diante do cenário – análise exploratória feita acima – em que as variáveis elencadas como predictoras conseguem explicar o preço médio variável *PET Condensado* e tendo conhecimento de que os dados utilizados na pesquisa são mensais, buscou-se separar as oitenta e quatro linhas do *dataframe* em linhas para treino, linhas para validação e linhas para teste.

Para tal formulação, as primeiras oitentas linhas do *dataframe* – criado anteriormente – serão destinadas ao treino e validação e as últimas quatro linhas serão destinadas ao conjunto teste. Utiliza-se, para divisão entre o conjunto treino e validação, do comando *random states*.

Sobre esse parâmetro, Wade (2020) afirma que esse comando é utilizado para que os mesmos conjuntos de treinamento e validação permaneçam durante as diferentes execuções, ou seja, o comando em questão permite que haja a delimitação dos dados que irão compor o conjunto de treino e os dados que irão compor o conjunto de testes. A função, na qual o comando em questão foi introduzido, pertence à biblioteca de código aberto *scikit-learn* (SCIKIT LEARN, 2022g) desenvolvida por David Cournapeau.

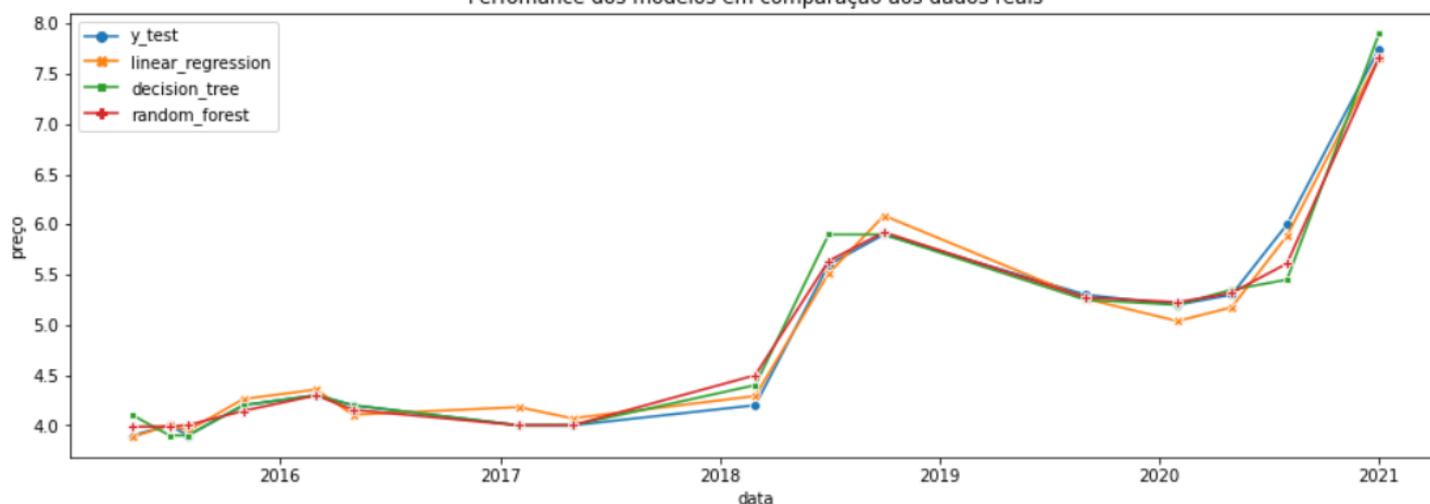
#### 4.4 Criação dos Modelos

Mediante à separação dos dados que irão compor o conjunto de treino e o conjunto de validação, aplica-se comandos da biblioteca *scikit-learn* (SCIKIT LEARN, 2022d, 2022e, 2022f) no conjunto de treino para a concepção dos modelos, utilizando os algoritmos de regressão linear (*linear regression*), árvore de decisão (*decision tree*) e florestas aleatórias (*random forest*). Esses modelos criados serão utilizados, posteriormente, para prever o preço futuro da variável *PET Condensado* (*target*).

Criados os modelos de regressão – mediante aos comandos desenvolvidos pela biblioteca *scikit-learn* –, expõe-se o gráfico abaixo, o qual possibilita visualizar o desempenho/performance dos algoritmos selecionados. Neste contexto, a linha azul do gráfico consiste nos preços médios reais da variável dependente quantitativa, e um “modelo perfeito” iria sobrepor-se a essa linha. A linha de cor laranja, verde e vermelha representam, respectivamente, as previsões dos algoritmos regressão linear, árvore de decisão (*decision tree*) e florestas aleatórias (*random forest*).

Nesta perspectiva, percebe-se que o desempenho desses três modelos, graficamente, é semelhante. Os três algoritmos, em um panorama geral, seguem as tendências dos preços reais – representados pela linha “*y\_test*”.

Figura 6 - Desempenho dos modelos  
Performance dos modelos em comparação aos dados reais



Fonte: produção do autor, 2022.

#### 4.5 Métricas e Comparações

Uma vez que, graficamente, seria pouco factível de analisar quais dos três modelos criados detêm um melhor desempenho, utilizou-se das métricas – utilizando-se da biblioteca *scikit-learn* (SCIKIT LEARN, 2022a, 2022b) – para averiguar qual modelo detêm o maior poder explicativo frente aos dados reais.

A comparação entre as três métricas elencadas na metodologia e a adesão dos algoritmos aleatórios (*Dummy Regressor*) – cujo intuito é expor a utilidade de usar os algoritmos de machine learning frente à aleatoriedade – encontram-se na tabela abaixo.

Tabela - Métricas

	MSE	RMSE	R-quadrado
<b>dummy_regressor_mean</b>	1.174	1.084	-0.076
<b>dummy_regressor_median</b>	1.138	1.067	-0.043
<b>linear_regression</b>	0.011	0.104	0.990
<b>decision_tree</b>	0.032	0.179	0.971
<b>random_forest</b>	0.017	0.131	0.984

Fonte: produção do autor, 2022.

No que tange à métrica “erro quadrático médio” (MSE), observa-se que o modelo de regressão linear múltipla apresenta o melhor resultado (0.011) – menor valor, erro, dentre os modelos. Com isso, pode-se afirmar que o modelo de regressão linear possui uma menor variância se comparado com os demais modelos. Sobre a raiz do erro quadrático médio (RMSE), o modelo de regressão linear, também, apresenta o menor valor (0.104).

Ademais, analisando o  $R^2$  – e sabendo que, o melhor *score* possível, para esta métrica, é 1.0 – percebe-se, novamente, que o algoritmo de regressão linear, apresenta o valor mais próximo a um, e isso demonstra que há poucos resíduos para cada observação da amostra – indicando que a variabilidade da variável  $Y$  é, quase que totalmente, explicada pelo vetor das variáveis que compõem as variáveis  $X$  (FÁVERO; BELFIORE, 2017).

Observa-se, também, que o desempenho do *Dummy Regressor* é o pior frente aos demais algoritmos, ou seja, de certo modo, os modelos criados possuem um melhor desempenho se comparado com a aleatoriedade proposta pelo *Dummy Regressor*.

#### 4.6 Conjunto Teste

De acordo com as métricas expostas acima, o algoritmo de regressão linear, até então, se mostrou o mais aplicável. Para averiguar se esse modelo, assim como os demais, conseguiria prever os preços médios futuros da variável  $Y$  – PET reciclada de Grau Alimentício –, utilizou-se dos dados de teste (separados anteriormente) para mensurar os preços médios em dados não antes aplicados, ou seja, utilizou-se de dados recentes para analisar se os valores apresentados pelos algoritmos – em especial, o de regressão linear – eram iguais (ou próximos) aos preços médios reais. Os resultados encontram-se na tabela abaixo.

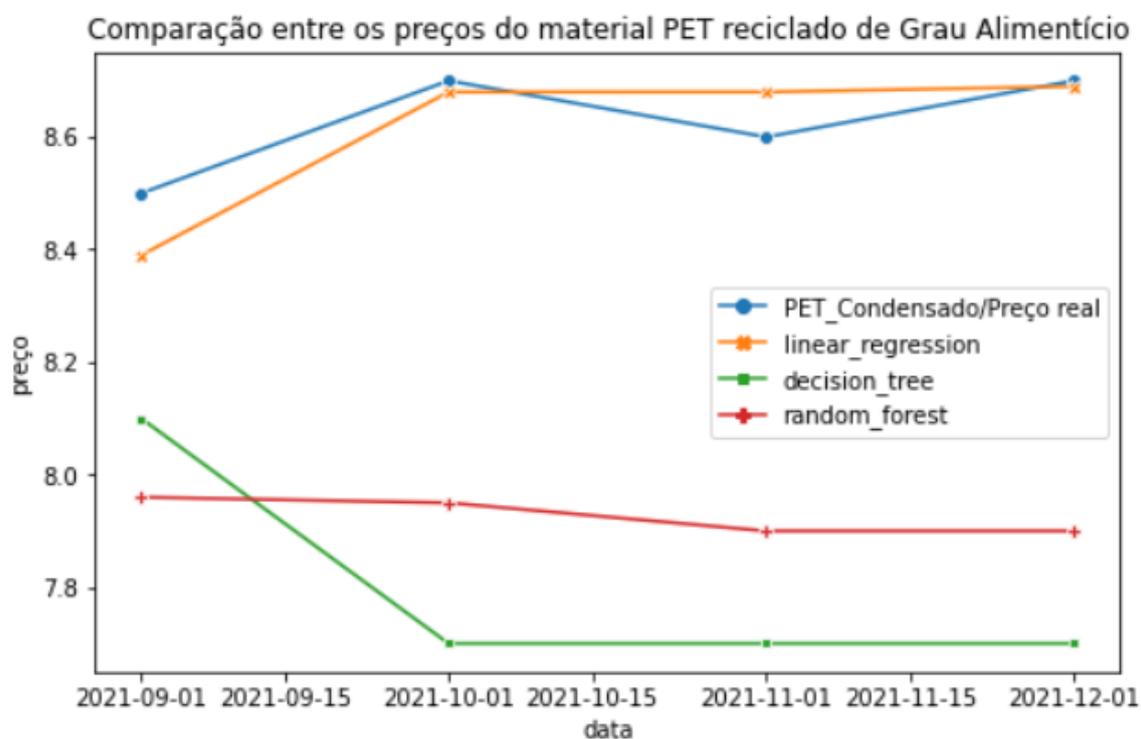
Tabela - Previsão do preço médio

	<b>pet_condensado</b>	<b>linear_regression</b>	<b>decision_tree</b>	<b>random_forest</b>
<b>data</b>				
<b>2021-09-01</b>	8.5	8.39	8.1	7.96
<b>2021-10-01</b>	8.7	8.68	7.7	7.95
<b>2021-11-01</b>	8.6	8.68	7.7	7.90
<b>2021-12-01</b>	8.7	8.69	7.7	7.90

Fonte: produção do autor, 2022.

A tabela em questão elucida a escolha do modelo de regressão linear para prever o preço médio do PET Pellet Grau Alimentício. Analisando a primeira linha (01/09/2021), observa-se que o valor previsto para o mês de outubro corresponde a R\$ 8,50 (valor arredondado) – coluna *PET Condensado*. O algoritmo de regressão linear previu um valor, arredondado, de R\$ 8,40, enquanto os demais algoritmos – *decision tree* e *random forest* – previam, respectivamente, R\$ 8,10 e R\$ 7,96. A linha subsequente indica que o modelo de regressão linear apresenta um valor arredondado idêntico ao preço real, ou seja, ele seria capaz de prever o preço para o mês de novembro – o mesmo ocorre na última linha da tabela. Constatou-se, também, que o algoritmo de regressão linear melhor compreendeu a tendência temporal dos dados reais, ou seja, o algoritmo atribuiu um valor maior para o mês de outubro se comparado ao mês anterior, enquanto os outros algoritmos atribuíram um valor menor.

Gráfico - Comparação entre os preços médios



Fonte: produção do autor, 2022.

A imagem acima é uma representação gráfica da tabela. Nesta imagem é possível visualizar o desempenho dos três modelos criados frente aos dados reais. Observa-se, nesta perspectiva, que o algoritmo de regressão linear (linear regression) apresenta valores, ao preço médio do PET reciclado de Grau Alimentício, mais verossímeis frente aos demais algoritmos – como pode-se observar na tabela apresentada anteriormente.

## 5. Conclusão

O presente estudo teve como proposta a ampliação da discussão que circunda a cadeia de suprimentos fechada envolvendo o material plástico, e em especial o PET. Para tanto, buscou-se apontar os fatores que influenciam a precificação desse material. Apontou-se, com base no estudo de correlações e análises gráficas, que o preço médio do material virgem, da sucata plástica, o preço de outros materiais secundários (PVC, PP, PS, PEBD e PET), o preço médio do combustível e o preço do dólar afetam o preço final do PET reciclado de Grau Alimentício (plástico reciclado estudado).

Uma vez comprovada certa correlação entre as variáveis preditoras e a variável *PET Condensado* (variável dependente quantitativa), o presente estudo buscou estabelecer um modelo de regressão, utilizando de artifícios vinculados à *machine learning*. Utilizou-se de três algoritmos diferentes, sendo eles: regressão linear (*linear regression*), árvore de decisão (*decision tree*) e florestas aleatórias (*random forest*). Averiguou-se que, com base em métricas estabelecidas –  $R^2$ , o erro quadrático médio (MSE) e a raiz do erro quadrático médio (RMSE) – e com as previsões feitas pelos algoritmos com base nos dados de teste, a regressão linear foi mais assertiva em comparação aos outros dois modelos – o algoritmo em voga apresentou valores previstos muito próximos (ou idênticos) aos valores reais da amostra destinada ao teste, além de apresentar  $R^2$  de 0.99 e RMSE de 0.104.

Como contribuição teórica, espera-se que este estudo possa agregar discussões futuras envolvendo temas relacionados a economia circular e projetos de previsão de preço do material reciclado – em especial, o plástico –, uma vez que, às implicações resultantes deste presente estudo tendem a proporcionar maior entendimento sobre a volatilidade do preço médio do PET reciclado. Como contribuição para a prática, espera-se que os gestores privados e públicos possam se beneficiar desse modelo para elaborar planejamentos mais robustos de sua cadeia, reduzindo incertezas quanto a preços praticados, melhorando sua posição competitiva e poder de barganha, auxiliando assim, no desenvolvimento de projetos mais robustos que envolvam o uso de materiais reciclados.

No entanto, esse estudo apresenta algumas limitações – só foi estudado um tipo de material – PET grau alimentício (ainda que esse material seja um dos mais utilizados em embalagens de uso único). Sugere-se para futuros estudos, a aplicação desse modelo para outros tipos de materiais plásticos, como forma de avaliar sua robustez.

## Bibliografia

ACKERMAN, F.; GALLAGHER K. Mixed signals: market incentives, recycling, and the price spike of 1995. **Resour Conserv Recy.** [S. l.], p. 275-295, 2002.

ANGUS, A.; CASADO, R.; FITZSIMONS, D. Exploring the usefulness of a simple linear regression model for understanding price movements of selected recycled materials in the UK. **Resour Conserv Recy.** [S. l.], p. 10-19, 2012.

AZANK, F. Como avaliar seu modelo de regressão. **Medium.** [S. l.], 2020. Disponível em: <https://cutt.ly/YK2PC4Z>. Acesso em: 23 nov. 2022.

Braz, A. C., & Mello, A. M. (2022). Circular economy supply network management: A complex adaptive system. *International Journal of Production Economics*, 243. <https://doi.org/10.1016/j.ijpe.2021.108317>

FISHER, A.; OWEN, A. An economic model of the US aluminium market. **Resources Policy.** [S. l.], p. 150-160, 1981.

GOVINDAN, K.; SOLEIMANI, H. A review of reverse logistics and closed-loop supply chains. **Journal of Cleaner Production.** [S. l.], p. 371-384, 2017.

GOVINDAN, K.; SOLEIMANI, H. Reverse logistics and closed-loop supply chain: a comprehensive review explore the future. **European Journal of Operational Research.** [S. l.], v. 240, n. 3, p. 603-626, 2015.

GUIDE, D.; WASSENHOVE, L. OR FORUM: the evolution of closed-loop supply chain research. **Operations Research**. [S. l.], p. 10-18, 2009.

HARRISON, M. **Machine Learning**. O'Reilly Media, 2020.

MATPLOTLIB. Matplotlib.pyplot.hist. **Matplotlib**. [S. l.], 2022. Disponível em: <https://cutt.ly/3K269DI>. Acesso em: 22 out. 2022.

PANDAS. Pandas.data.frame. **Pandas**. [S. l.], 2022a. Disponível em: <https://cutt.ly/NK9qSwo>. Acesso em: 03 ago. 2021.

PANDAS. Pandas.data.frame. **Pandas**. [S. l.], 2022b. Disponível em: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.corr.htm>. Acesso em: 22 maio 2021.

PANDAS. Pandas.data.frame.info. **Pandas**. [S. l.], 2022c. Disponível em: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.corr.htm>. Acesso em: 22 maio 2021.

SANTOS, M. **Desafios para o fechamento de cadeias de suprimento de produtos de baixo valor agregado**: o caso das embalagens plásticas e de vidro. 2020, 75 f. Dissertação (Mestrado em Administração) – Departamento de Administração da Universidade de São Paulo, São Paulo, 2020.

SCIKIT LEARN. 1.10. Decision Trees. **Scikit Learn**. [S. l.], 2022e. Disponível em: <https://cutt.ly/oK27iY4>. Acesso em: 16 set. 2022.

SCIKIT LEARN. Metrics and scoring: quantifying the quality of predictions. **Scikit Learn**. [S. l.], 2022c. Disponível em: <https://cutt.ly/oH27iz4>. Acesso em: 16 set. 2022.

SCIKIT LEARN. Scikit learns user guide. **Scikit Learn**. [S. l.], 2017. Disponível em: [https://scikit-learn.org/stable/user\\_guide.html](https://scikit-learn.org/stable/user_guide.html). Acesso em: 03 jul. 2021.

SCIKIT LEARN. Sklearn.linear\_model.LinearRegression. **Scikit Learn**. [S. l.], 2022d. Disponível em: <https://cutt.ly/oK27iz4>. Acesso em: 16 set. 2022.

SCIKIT LEARN. Sklearn.metrics.accuracy\_score. **Scikit Learn**. [S. l.], 2022a. Disponível em: <https://cutt.ly/OK2417T>. Acesso em: 10 set. 2022.

SCIKIT LEARN. Sklearn.metrics.mean\_squared\_error. **Scikit Learn**. [S. l.], 2022b. Disponível em: <https://cutt.ly/oK27iz4>. Acesso em: 21 set. 2022.

SCIKIT LEARN. Sklearn.model\_selection.train\_test\_split. **Scikit Learn**. [S. l.], 2022g. Disponível em: <https://cutt.ly/oKr7iz4>. Acesso em: 25 set. 2022.

SCIKIT LEARN. Slearn.ensemble.RandomForestClassifier. **Scikit Learn**. [S. l.], 2022f. Disponível em: <https://cutt.ly/oK2tiz4>. Acesso em: 28 set. 2022.

SEABORN. Seaborn.heatmap. **Seaborn**. [S. l.], 2022. Disponível em: <https://cutt.ly/nK9w1r6>. Acesso em: 06 jun. 2021.

TAYLOR, C. A Quarterly domestic copper industry model. **The Review of Economics and Statistics**. [S. l.], p. 410-422, 1979.

WADE, C. **Hands-On Gradient Boosting with XGBoost and scikit-learn**. Packt Publishing Ltd, 2020.

XIARCHOS, I.; FLETCHER, J. Price and volatility transmission between primary and scrap metal markets. **Resour Conserv Recy**. [S. l.], p. 664–673, 2009.