



## PREVISÃO DE PREÇO DE CARROS USADOS: um estudo utilizando regressão linear para análise de dados

**Leonardo H. J. da SILVA<sup>1</sup>; Leonardo E. da SILVA<sup>2</sup>; Diego SAQUI<sup>3</sup>;**

### RESUMO

O presente artigo consiste em realizar um estudo sobre o preço de automóveis usados, utilizando um algoritmo para relacionar a variação no preço dos veículos com as características e informações obtidas sobre eles. Para desenvolver este estudo foi aplicado um algoritmo de regressão linear, capaz de prever um possível preço para os carros e classificar quais características são mais relevantes para essa oscilação do preço. Como resultado foi possível prever os valores dos veículos de forma próxima aos valores reais, utilizando como base o ano de fabricação, a quilometragem, quantidade de lugares e tamanho.

**Palavras-chave:** Inteligência Artificial; Precificação; Veículos Usados; Algoritmo.

### 1. INTRODUÇÃO

O mercado automobilístico brasileiro é um componente significativo da economia nacional (SANTOS, 2015). De acordo com a Associação Nacional dos Fabricantes de Veículos Automotores (ANFAVEA), o Brasil se destaca como o sétimo maior mercado consumidor de automóveis no mundo. A venda de veículos usados continua a crescer no Brasil, com a Federação dos Revendedores de Veículos Usados (FENAUTO) relatando um aumento de 10,7% nas vendas em abril de 2023 em comparação com o mesmo mês em 2022 (PAIXÃO, 2023). A ANFAVEA revelou que a relação entre carros novos e usados é de um novo para cinco usados, ressaltando a necessidade de examinar as variações de preço nesse mercado e suas tendências. Essas variações, frequentemente, são influenciadas pelas características dos veículos e pelas informações fornecidas pelos vendedores (MEDEIROS e VIDIGAL, 2018).

Identificar quais as características mais significantes para a variação do preço no momento da compra é um desafio recorrente (FONSECA, 1996). Resende e Scarpel (2009) destacaram a necessidade de que os modelos de precificação de veículos considerem especialmente o conjunto de atributos presentes nos carros.

Por meio do *Machine Learning*, método usado para ensinar computadores a lidar com informações da melhor maneira utilizando diferentes fórmulas para resolver problemas (MAHESH,

<sup>1</sup>Docente do Bacharelado em Ciência da Computação, IFSULDEMINAS – *Campus* Muzambinho. E-mail: 12201001331@muz.ifsuldeminas.edu.br

<sup>2</sup>Discente do Bacharelado em Ciência da Computação, IFSULDEMINAS – *Campus* Muzambinho. E-mail: 12191000383@muz.ifsuldeminas.edu.br.

<sup>3</sup>Docente do Bacharelado em Ciência da Computação, IFSULDEMINAS – *Campus* Muzambinho. E-mail: diego.saqui@muz.ifsuldeminas.edu.br.

2020), é possível analisar os dados de forma a encontrar uma relação entre as características dos veículos e a variação nos seus preços.

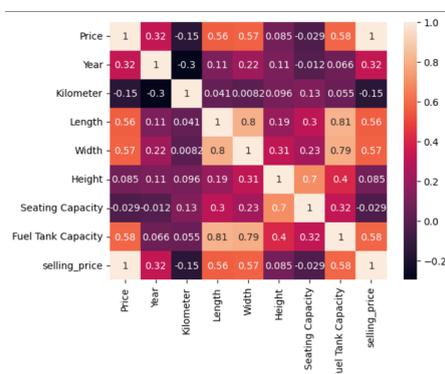
Nesse contexto, este estudo visa compreender as variações nos preços dos carros usados, utilizando um modelo de regressão linear, uma técnica capaz de prever valores desconhecidos com base em valores relacionados conhecidos. Neste projeto, um algoritmo foi empregado para prever possíveis preços de veículos usados, levando em consideração o grau de relevância de cada característica para a variação de preço.

### 3. MATERIAL E MÉTODO

Este estudo foi realizado com base em um *dataset* de precificação de carros usados na qual possuem 2.090 registros, com 20 instâncias<sup>4</sup>, que conta com várias características - como ano de fabricação, quilômetros rodados, tipo de combustível, entre outros - a fim de encontrar relações entre essas características e a variação no preço dos veículos. Foi desenvolvido um código na linguagem de programação Python no ambiente de desenvolvimento Google Colab<sup>5</sup> para aplicar o algoritmo de Regressão Linear.

Inicialmente, coletamos um conjunto de dados contendo informações sobre diversos veículos, como marca, modelo, ano de fabricação, quilometragem, tipo de transmissão, entre outros. Com base nesses dados, buscamos estabelecer uma correlação entre as variáveis conforme mostrado na Figura 1. Após isto, foi realizada a análise de desempenho do modelo, sendo 41% dos dados para treino e 37% para teste na qual revelou um coeficiente de determinação ( $R^2$ ) de 0,46, indicando que aproximadamente 46% da variabilidade nos preços dos carros pode ser explicada pelas variáveis independentes consideradas no modelo.

Figura 1 - Matriz de correlação



<sup>4</sup> Disponível em :

<<https://www.kaggle.com/datasets/nehalbirla/vehicle-dataset-from-cardekho/discussion?select=CAR+DETAILS+FROM+CAR+DEKHO.csv>>. Último acesso em 23/06/2023

<sup>5</sup> Disponível em :

<<https://colab.research.google.com/drive/1Zc-HY7qfZQXY8j011XlyInbcESpYhqsG?usp=sharing>>. Último acesso em 01/07/2023

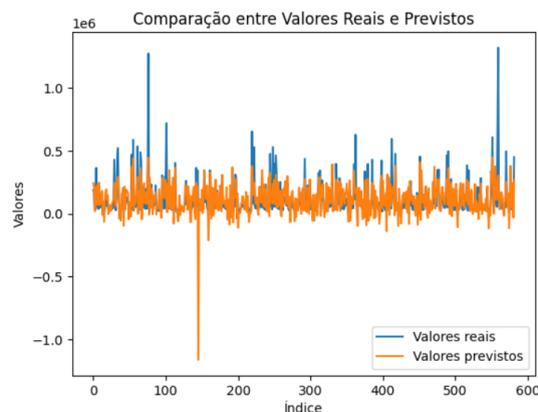
#### 4. RESULTADOS E DISCUSSÃO

Neste estudo, aplicamos uma análise de regressão linear para estimar o preço de venda de carros com base em variáveis independentes relevantes. Os resultados obtidos revelaram informações importantes sobre a relação entre essas variáveis e os preços dos veículos.

Ao explorar os dados, identificamos uma correlação significativa entre as variáveis independentes consideradas (como ano de fabricação, quilometragem, assentos, tamanho, entre outras) e o preço de venda dos carros. No entanto, o coeficiente de determinação ( $R^2$ ) obtido foi de 46%, indicando que apenas 46% da variação nos preços dos veículos pode ser explicada pelas variáveis independentes incluídas no modelo de regressão linear. Isso significa que há uma quantidade considerável de variação nos preços que não é capturada pelas variáveis utilizadas neste estudo.

Apesar do  $R^2$  relativamente baixo, os coeficientes estimados pelo modelo de regressão linear fornecem informações relevantes sobre a direção e a magnitude da influência das variáveis independentes nos preços dos veículos, conforme mostrado na Figura 2. Esses resultados podem ser úteis para compradores e vendedores de carros usados, fornecendo uma referência inicial para a precificação dos veículos.

**Figura 2** - Comparação entre valores Reais e previstos



Um exemplo de resultado obtido através deste algoritmo, utilizando como parâmetro as variáveis ano, lugares do carro, e quilometragem, pode ser visto a seguir na Figura 3, onde foi possível estimar o valor do carro com as informações coletadas.

Figura 3 - Exemplo de resultado

```
▶ Year = 2020
Length = 5
Kilometer = 0
Width = 5
entrada2=[[Year,Length,width, Kilometer]]

print('Valor do carro é: R$ {}'.format(abs(float(modelo.predict(entrada2)[0].round(2))*0.066)))

Valor do carro é: R$ 60731.07942
```

## 5. CONCLUSÃO

Neste estudo, utilizamos regressão linear para estimar os valores de carros. Os resultados revelaram que o ano de fabricação, quilometragem, capacidade de passageiros e tamanho do automóvel tiveram um impacto significativo nos preços dos veículos. Carros mais recentes, menos quilometragem e com capacidade de passageiros maior tendem a ter preços mais altos. O modelo ajustado permitiu quantificar a influência dessas variáveis nos preços dos carros. O coeficiente de determinação ( $R^2$ ) de 0,46 indicou que aproximadamente 46% da variação nos preços pode ser explicada pelas variáveis consideradas, com *outliers* no *dataset* foi o motivo na qual ( $R^2$ ) tenha ficado um pouco inferior do esperado. mas a comparação entre os valores reais e os previstos estão próximos. Portanto, concluímos que a regressão linear foi uma abordagem adequada e os resultados podem ser úteis para compradores, vendedores e profissionais do setor automotivo.

## REFERÊNCIAS

FONSECA, Renato. Quality change in brazilian automobiles. 1996.

MAHESH, Batta. Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet], v. 9, n. 1, p. 381-386, 2020.

MEDEIROS, Rodrigo de Vasconcellos Viana; VIDIGAL, Naiara Costa. Modelagem econométrica dos preços de carros populares do tipo hatch utilizando a regressão linear múltipla. Cadernos de Ciências Sociais Aplicadas, p. 22-22, 2018.

PAIXÃO, André. Venda de carros usados cresce, mas brasileiro está comprando modelos mais velhos. [S. l.], 9 maio 2023. Disponível em: <https://autoesporte.globo.com/carros/usados-e-seminovos/noticia/2023/05/venda-de-carros-usados-cresce-mas-brasileiro-esta-comprando-modelos-mais-velhos.ghtml>. Acesso em: 29 jun. 2023.

RESENDE, Camilo Brandão de; SCARPEL, Rodrigo Arnaldo. Importância das características na precificação de veículos nacionais. Production, v. 19, p. 345-358, 2009.

SANTOS, Alexandre Magno Helou. Testando a validade da lei do preço único: análise do mercado de carros usados no Brasil. 2015.