

Para previsão do preço a partir dos dados, foi necessário realizar o problema de regressão, uma vez que estamos tentando prever um valor numérico contínuo (o preço).

Como a escolha das variáveis a serem utilizadas na previsão do preço depende do contexto do problema e dados disponíveis. Geralmente, são consideradas variáveis como marca, modelo, ano de fabricação, quilometragem, tipo de câmbio, número de portas, entre outras, que têm relação direta com o preço de um veículo.

Além disso, podemos realizar transformações nas variáveis, como a criação de variáveis dummy para as variáveis categóricas, normalização das variáveis numéricas, e a inclusão de interações ou polinômios para capturar possíveis relações não lineares entre as variáveis.

Quanto à escolha do modelo, acredito que poderíamos usar coeficiente de determinação  $R^2$  indicar a proporção da variância e ou utilizar métrica erro médio absoluto (MAE).