

## Lighthouse - Indicium - Matheus Ventura

### Problema de Negócio

O objetivo da previsão é determinar o preço de cada veículo, sendo este um valor contínuo, o problema em questão se enquadra no uso de modelos de regressão (não estamos interessados, por exemplo, se o carro será ou não vendido na próxima semana).

Utilizei transformações de MinMaxScaler para reescalar as features numéricas, e encoders de Frequencia, Target (Alvo) e Ordinal (Rótulo) para transformar as features categóricas. Aplicando algoritmo Boruta e Importância de features com Random Forest, cheguei na seguinte lista de features destacadas para teste dos modelos:

- *Identificação*: combinação de Marca e Modelo do carro
- *Hodometro*
- *Ano\_modelo*
- *Cilindradas*: extraída da versão do carro
- *Endereco*: combinação de cidade e estado do vendedor
- *Tipo*
- *Combustivel*: extraída da versão do carro
- *Cor*
- *Cambio*

Após a validação cruzada dos dados, o modelo com melhor desempenho foi CatBoost, suas vantagens são capacidade de lidar bem com relações não-lineares, rápida velocidade de previsão, devido ao seu balanceamento simétrico das árvores. Sua desvantagem, como todo modelo baseado em árvores de decisão, é a limitação na extrapolação, não sendo capaz de prever valores maiores ou menores dos presentes nos dados de treino.

As métricas escolhidas para avaliação dos modelos foram RMSE (Root Mean Squared Error), usada para comparar a performance entre os modelos pois dá maior peso a erros maiores, e para avaliar os resultados de negócio foram escolhidas MAE (Mean Absolute Error) e MAPE (Mean Absolute Percentage Error) que permitem uma análise não técnica mais fácil.