# Entrega 2 – Aplicação de um modelo de Machine Learning

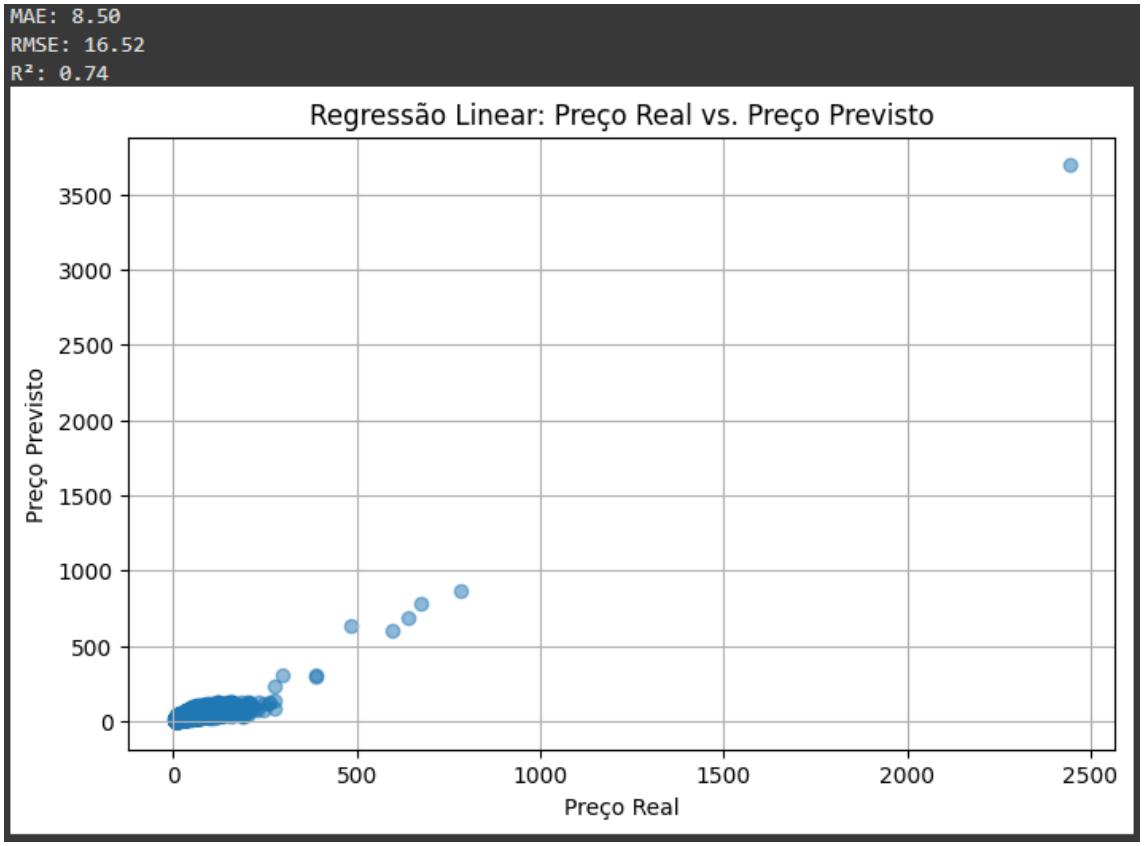
Abordado no documento anterior, o modelo inicial escolhido foi o de Regressão Linear, por decorrência de ser um modelo muito utilizado no mercado e com a capacidade de lidar bem com previsões de preços.

As bibliotecas incluem: sklearn, matplotlib e geaopy.

O modelo foi empregado usando a base de dados derivada na matéria de “Ciência de Dados e Big Data”, as variáveis categóricas foram: 'ano', 'mes', 'dia_semana', 'productid' enquanto as variáveis preditoras e “y” foi a ‘price’.

Foi utilizado o “geodesic” para cálculo de distância a partir da latitude e longitude de origem e destino, não era a abordagem que desejávamos inicialmente (por conta da precisão), porém até o momento não encontramos uma alternativa.

Para treinamento do modelo carregamos uma base com 100 mil linhas (uma parte da base total para termos uma amostragem) os resultados foram o da imagem abaixo:




Logo temos um erro médio de aproximadamente de R$8,50 nas previsões para mais ou para menos (MAE), o RMSE se encontra relativamente alto (não tanto) o

que pode trazer alguns erros consideráveis e por último o $R^2$ que aponta 0.74 que explica 74% das variações de preços que apresenta um resultado muito bom para um 1° treinamento.

## Ações futuras:

Derivar mais dados da base (separar data e hora) datas de feriados, tempo médio da distância percorrida;

Tratar os outliers;

Passar mais dados para treinar o modelo;

Gerar um modelo para cada categoria;

Trocar de modelo se for preciso (ainda essa semana se for o caso).

**Conclusão**

As métricas não se apresentaram ruins para o início do modelo (inclusive criamos um bloco de código e uma requisição teste que apresentou poucas variações nos valores), e foi uma etapa útil para sanar algumas dúvidas acerca de ML e criação do motor de cotação de modo geral.