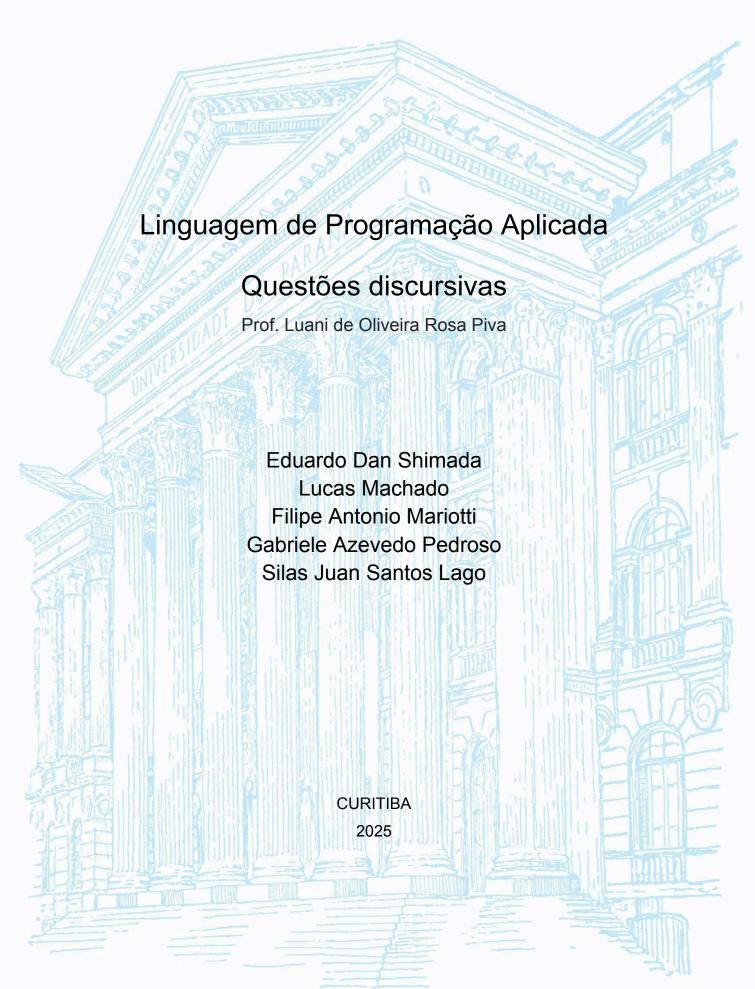
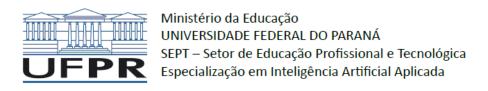
UNIVERSIDADE FEDERAL DO PARANÁ







Respostas Discursivas

1.f. Dê um breve explicação (máximo de quatro linhas) sobre os principais resultados encontrados na Análise Exploratória dos dados.\

A base de dados importado do Kaggle estava em sua maior parte já tratada, havendo apenas 3 linhas contendo valores duplicados e nenhuma célula com valores em branco, não havendo assim a necessidade de realizar tratamento de preenchimento de dados nulos. O dataset é composto por 4 colunas do tipo numérica e 7 colunas do tipo categórica, sendo as colunas 'year_of_reference', 'engine_size', 'year_model' e 'avg_proce_brl' do tipo numérica; as colunas `mont_of_reference', 'fipe_code', 'authentication', 'brand', 'model', 'fuel' e 'gear' do tipo categórica.

2.e. Dê uma breve explicação (máximo de quatro linhas) sobre os resultados gerados no item d

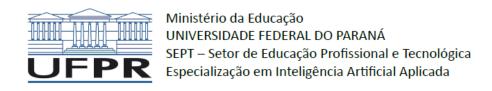
Os carros das marcas Fiat, Renault e VW - Volkswagen apresentam os menores valores de mercado em todas as categorias de engrenagem, além de não possuírem modelos com mais de seis marchas. Em contraste, Ford e GM - Chevrolet estão presentes em todas as categorias e demonstram uma variação significativa de preços. Já a Nissan se destaca por sua forte presença em carros com quatro marchas e um alto valor médio de mercado.

2.g. Dê uma breve explicação (máximo de quatro linhas) sobre os resultados gerados no item f

Fiat, Ford, GM - Chevrolet e VW - Volkswagen possuem modelos movidos a três tipos de combustível. A Volkswagen lidera em valor de mercado para carros a diesel, enquanto a Ford apresenta os modelos mais baratos nessa categoria. No álcool, a Ford tem os menores preços, enquanto a Chevrolet possui os mais elevados. No caso da gasolina, a Nissan tem os modelos mais caros, e a Renault, os mais acessíveis. Nissan e Renault não possuem veículos a álcool no conjunto de dados analisado.

3.f. Dê uma breve explicação (máximo de quatro linhas) sobre os resultados encontrados na análise de importância de variáveis

A análise utilizando Random Forest revelou que *engine_size* é a variável mais relevante, com um peso de 42% na previsão dos preços. No entanto, ao empregar XGBoost para redistribuir a importância entre outras variáveis, *engine_size* caiu para a terceira posição, enquanto *fuel_numerico* aumentou sua relevância no modelo.





4.h. Dê uma breve explicação (máximo de quatro linhas) sobre qual modelo gerou o melhor resultado e a métrica de avaliação utilizada

Utilizando o altoritmo Random Forest sem o uso limitado de folhas notou-se uma maior taxa de acertos por não haver um limite em sua profundidade de aprendizagem. Já com a sua limitação houve uma queda de acurácia de 3%, o que não é uma diferença tão alta mas que fica abaixo da taxa de acerto do XGBoost ao melhorar algumas de suas variáveis, tendo uma taxa de acerto de 95%