1. **Problema e objetivo do projeto**

O mercado automotivo nos dias atuais, com uma crescente gama de veículos novos e seminovos, apresenta ao comprador um desafio conhecido como o paradoxo da escolha. Essa abundância de opções, somada à multiplicidade de especificações técnicas (potência, consumo, itens de segurança) e categorias de mercado (SUV, Sedan, Esportivo), frequentemente resulta em uma sobrecarga de informação. Diante disso, o consumidor, seja ele um comprador experiente ou alguém em busca do seu primeiro carro, enfrenta dificuldades para comparar as alternativas de forma eficiente, alinhar suas necessidades e preferências pessoais ao seu orçamento, e, por fim, tomar uma decisão de compra confiante e bem-informada.

1. **Descrição do dataset e justificativa para sua escolha**

O dataset utilizado é um dataset público, disponível no kaggle, onde o mesmo inclui as características principais de um veículo, como marca, modelo,ano e categoria, trazendo também seu preço inicial(MSRP) e seus números de economia.

Decidi por escolher esse dataset, já que no primeiro trabalho(machine learning) eu utilizei um dataset parecido, mas gerado por IA, algo que me deu diversos problemas, com isso, decidi pegar um público, onde eu poderia adaptar melhor as ideias e fazer um trabalho com um resultado mais real.

1. **Análise exploratória e pré-processamento dos dados.**

A análise exploratória inicial teve como missão obter os primeiros insights sobre a estrutura e distribuição do dataset. Através de funções da biblioteca Pandas, foi possível verificar os tipos de dados de cada coluna e a presença de valores ausentes.

Para garantir a qualidade e consistência dos dados para a modelagem, um processo de limpeza foi executado. Inicialmente, foi selecionado um conjunto das 14 colunas mais relevantes para o problema. Em seguida, foram removidas todas as linhas que continham valores ausentes e registros duplicados.

Dado que o objetivo era criar um sistema de recomendação, e o dataset não possuía uma variável alvo explícita que definisse a "qualidade" ou "desejabilidade" de um carro, foi necessário criar uma através de engenharia de feature. Foi desenvolvida uma métrica sintética, a coluna score, para quantificar a atratividade de um veículo em uma escala de 0 a 1.

Com o dataset limpo e a variável alvo criada, a etapa de pré-processamento preparou os dados para serem consumidos pelos algoritmos. Todas as variáveis de entrada categóricas, como Make, Model, e Vehicle Style, foram transformadas em formato numérico através da técnica de **One-Hot Encoding(**algo que a IA me apresentou quando precisei resolver os problemas numéricos).

**4.Descrição do modelo escolhido e processo de treinamento.**

Durante os testes, vários modelos foram testados, mas o que se saiu melhor foi o MLP, construído com as API’s do Keras, foram criadas as camadas, sendo a de entrada com 256 Neurônios e utilizando ReLu, depois, foram criadas 2 camadas ocultas, com 128 e 64 Neurônios, também com ReLu. Tentando evitar um overfitting, foi usado dropout e early stopping, com isso, foi criada a camada de saída com 1 neurônios

O treinamento foi executado através da função .fit() por um máximo de 100 épocas (epochs), com um tamanho de lote (batch\_size) de 32. Uma parte dos dados de treino (20%) foi reservada para validação a cada época. Crucialmente, foi empregado um *callback* de **Early Stopping**, configurado para monitorar a perda de validação (val\_loss). Este callback interrompeu o treinamento automaticamente caso o desempenho no conjunto de validação não melhorasse por 10 épocas consecutivas, e restaurou os pesos do modelo da melhor época encontrada.

**5.Resultados e análise da performance do modelo.**

Com todos os testes realizados, eu cheguei a um resultado bem satisfatório com o deep learning, porém nos meus testes a performance poderia ter sido melhor, algo mais próximo à árvore de decisão, um modelo que teve uma performance melhor.