

Nome: João Victor Maia Cordeiro

Email: jvmc@cesar.school

Dataset Utilizado: [Vehicle Dataset](#)

workflow do modelo de treinamento (Regressão Linear)

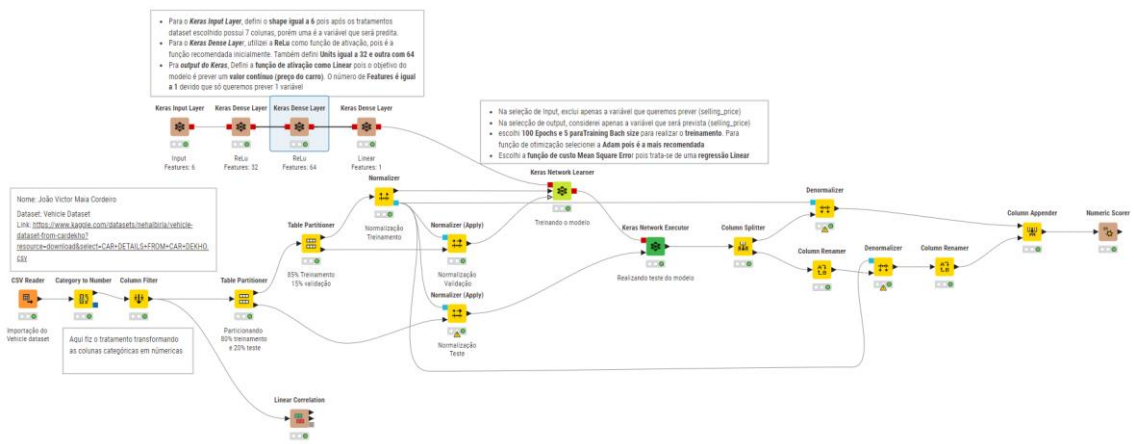


Figura 1- Workflow completo

Pré Processamento dos dados

Como etapa inicial, realizei o tratamento do dataset transformando as variáveis categóricas em numéricas. Para isso, utilizei o nó “**Category to Number**”, que converte os valores das colunas categóricas em números inteiros.

Após o tratamento, foi possível gerar a **matriz de correlação**. Avalio que não há correlações muito fortes entre as variáveis, mas existe uma correlação mediana (0,414) entre o **ano do carro** e o **preço de venda**, o que indica que veículos mais novos tendem a apresentar preços mais elevados.

RowID	year Number (Float)	selling_price Number (Float)	km_driven Number (Float)	fuel (to number) Number (Float)	seller_type (to number) Number (Float)	transmission (to number) Number (Float)	owner (to number) Number (Float)
year	1	0.414	-0.42	0.09	0.182	0.144	-0.393
selling_	0.414	1	-0.192	0.227	0.241	0.53	-0.19
km_dri	-0.42	-0.192	1	0.275	-0.187	-0.12	0.276
fuel (to	0.09	0.227	0.275	1	0.011	0.028	0.015
seller_1	0.182	0.241	-0.187	0.011	1	0.2	-0.209
transm	0.144	0.53	-0.12	0.028	0.2	1	-0.075
owner	-0.393	-0.19	0.276	0.015	-0.209	-0.075	1

Configuração do modelo em Keras

- **Input Layer:** defini o *shape* igual a 6, já que o dataset, após o pré-processamento, possui 7 colunas, sendo uma delas a variável alvo (preço de venda).
- **Dense Layers:** utilizei a função de ativação **ReLU**, recomendada como padrão inicial, configurando duas camadas densas, uma com **32 units** e outra com **64 units**.
- **Output Layer:** a função de ativação escolhida foi **Linear**, adequada para problemas de regressão, já que o objetivo do modelo é prever um valor contínuo (preço do carro). O número de *features* de saída foi definido como 1, correspondendo à variável que será prevista.

Treinamento do modelo

- **Input:** foram utilizadas todas as variáveis preditoras, exceto a variável alvo (*selling_price*).
- **Output:** apenas a variável *selling_price* foi considerada.
- **Configurações de treinamento:** defini **100 epochs** e um **batch size de 5**.
- **Função de otimização:** utilizei o **Adam**, por ser recomendado.
- **Função de custo:** Defini o **Mean Squared Error (MSE)** por ser apropriado para modelos de regressão linear.

Resultados

Após realizar algumas mudanças de parâmetros, cheguei ao resultado final do modelo com as variáveis descritas acima. Ao avaliar a predição e as métricas de avaliação, percebo que o modelo não possuiu uma boa performance de previsão.

o Valor de R^2 está razoável, porém ao avaliar MSE e MAE nota-se que o modelo está com grande desvio no valor previsto e o valor real de venda. Mesmo com mudanças nos parâmetros não consegui melhorar os resultados

RowID	predition_selling_price <i>Number (Float)</i>
R^2	0.627
mean absolute error	185,391.47
mean squared error	125,333,386,814.272
root mean squared error	354,024.557
mean signed difference	7,382.046
mean absolute percentage error	0.478
adjusted R^2	0.627