**Projeto A3 para a UC de Inteligência Artificial.**

**Variáveis/Colunas**

* Price - preço dos diamante | float
* Carat - peso em quilates dos diamantes | float
* cut - qualidade do corte (Regular, Bom, Muito Bom, Premium, Ideal) | String
* color - cor do diamante, de J (pior) a D (melhor) | String
* clarity - clareza do diamante uma medida de quão claro é o diamante (I1 (pior), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (melhor)) | String
* x - comprimento em mm | float
* y - largura em mm | float
* z - profundidade em mm | float
* depth - profundidade total em porcentagem | float
* table - do topo do diamante em relação ao ponto mais largo | float

Escolhemos o preço do diamante como o principal foco de estudo porque entender o preço nos ajuda a descobrir o que torna um diamante caro ou barato. Isso é útil para pessoas que fazem ou vendem joias, para aquelas que querem investir em diamantes e até mesmo para quem gosta de estudar essas pedras preciosas. Através disso, podemos aprender o que as pessoas valorizam em um diamante e como isso afeta os negócios e os preços no mercado.

**Explicação variáveis**

As variáveis que precisarão ser transformadas são "carat" (quilates), devido às diferentes escalas de valores, exigindo padronização para garantir que todas tenham o mesmo peso ao serem inseridas na etapa de machine learning. Além disso, também devemos transformar as variáveis "cut", “color” e “clarity”, por serem Strings, considerando que a maioria dos algoritmos apenas aceita valores numéricos para as colunas. Para isso, é necessário aplicar a codificação adequada para garantir sua utilização nos algoritmos de machine learning. Nesse caso, nossa variável-alvo é o preço do diamante, que é o indicador principal para avaliar um diamante.

Escolhemos esta base de dados devido à sua importância econômica e comercial. Compreendendo os fatores que influenciam o preço do diamante nos segmentos da indústria, consumo e investimentos, podemos abordar a falta de especificações no comércio de diamantes, que não atende a uma ampla gama de público, incluindo a classe média e a classe média baixa. Os dados nos permitirão explorar essa curiosidade em relação a esse nicho de mercado.

**Explicação variáveis transformadas**

**Cut / Color / Clarity**

Texto

Descrição gerada automaticamente

Tela de computador com fundo preto

Descrição gerada automaticamente

Texto

Descrição gerada automaticamente

Nesses três casos se tornou necessário transformar as variáveis pois as mesma eram Strings que precisavam ser transformadas em números para que a IA pudesse seguir com seu processamento.

Nosso Target é prever o valor dos diamantes com os dados passados.

**Primeiro Método**

**Método de rede neural**

**com 1493 iterações**

**Resultados:**

**MSE: 3323.494797002967**

**MAE: 25.268863099271645**

**RMSE: 57.64975973066121**

**MAPE: 1.8696014488082793 %**

**Média de valores: 3906.0357804968485**

**Exemplo de previsão**

**index: 10005**

**Previsão: 2925.872249280787**

**Valor correto: 2913**

**com 10493 iterações (diamond-relu-10k.pkl)**

**Resultados:**

**MSE: 2576.665712026238**

**MAE: 17.603135639596992**

**RMSE: 50.76086792034035**

**MAPE: 1,187004661096045 %**

**Média de valores: 3906.0357804968485**

**Exemplo de previsão**

**index: 7793**

**Previsão: 756.1300119369904**

**Valor correto: 764**

**Segundo Método**

Árvore de decisão

Resultados:

MSE: 536837.5160131628

MAE: 351.1618001483129

RMSE: 732.6919652986259

Média de valores: 3906.0357804968485

Exemplo de previsão

index: 249

Previsão: 11899.0

Valor correto: 11792

**Terceiro Método**