

Desafio cientista de dados

Descreva graficamente os dados disponíveis, apresentando as principais estatísticas descritivas. Comente o porquê da escolha dessas estatísticas.

Os dados fornecidos de máquinas contendo seus modelos e referências apresentando o processo de manufatura, no qual é identificado tipos de falhas!

A Cardinalidade foi fator determinante para a escolha em que a coluna `failure_type` e `type` apresentaram uma maior frequência para análise e remoção de colunas com um menor de generalização. Foi analisado as colunas que podiam gerar uma análise que fornecia um score maior pelo pacote árvore de decisão

Explique como você faria a previsão do tipo de falha a partir dos dados. Avaliando os dados e identificando padrões para gerar hipóteses para a busca informações mais generalizado para treinar os dados onde o machine learning apresentará uma eficiência maior.

Qual tipo de problema estamos resolvendo (regressão, classificação)?

Classificação, no qual determina alguma categoria para os dados para a previsão da análise.