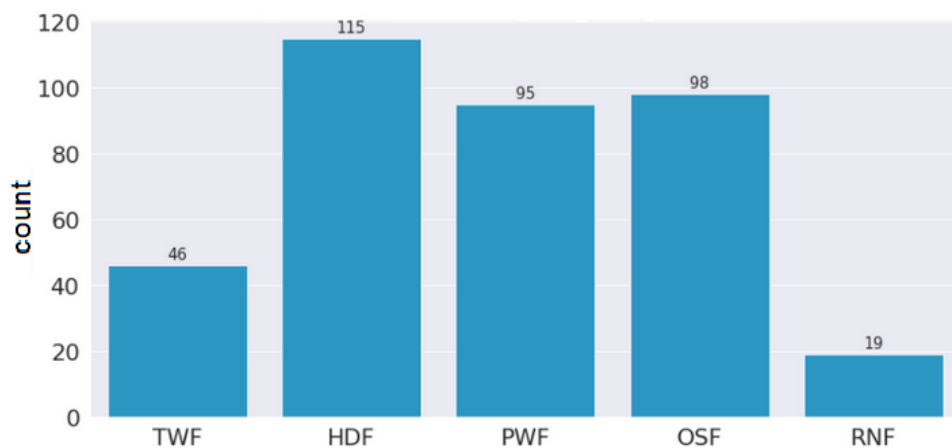
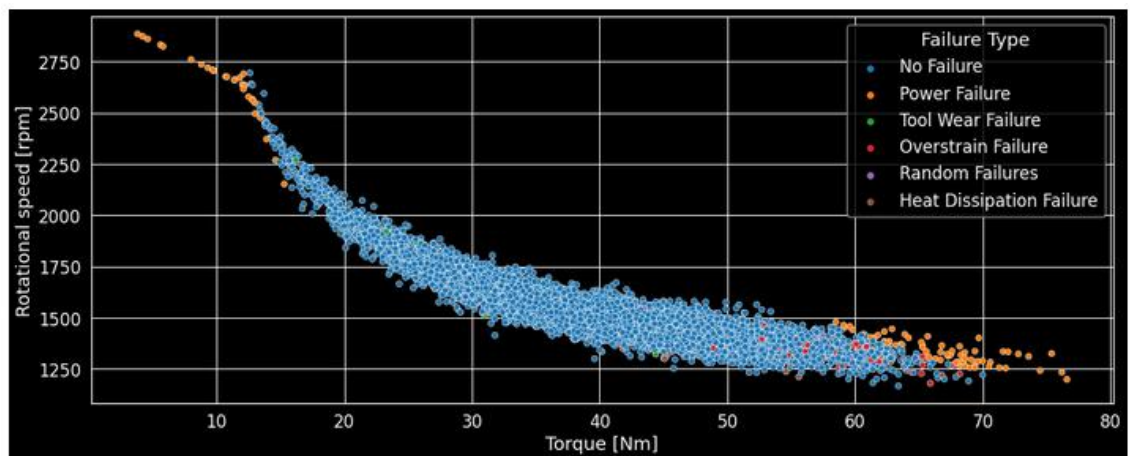


Relatório Lighthouse 2023 - Desafio Cientista de Dados

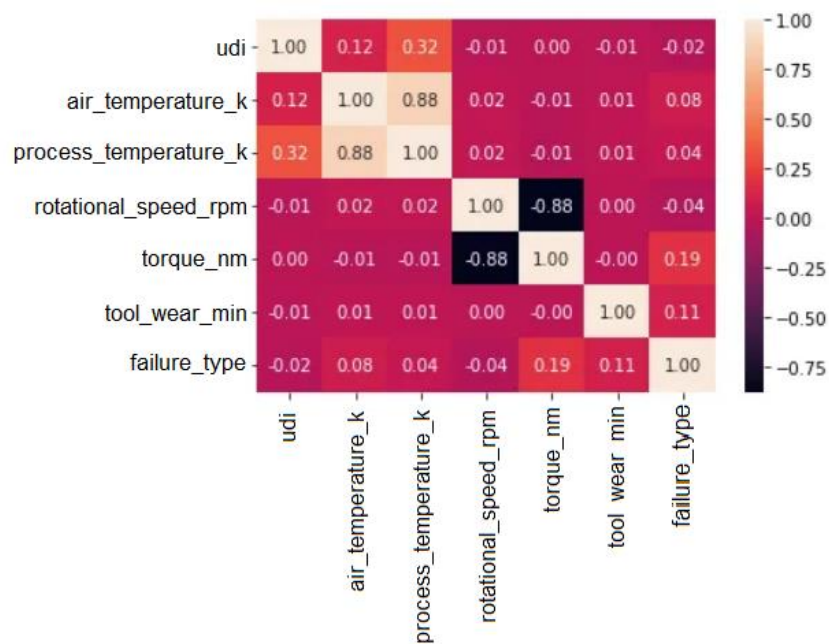
- 1- As estatísticas descritivas nos ajudam a interpretar melhor o problema e diferentes formas de chegar em sua conclusão, aqui apresento algumas que melhor representam o conjunto de dados disponibilizado.



Nesse gráfico podemos perceber o desbalanceamento entre os tipos de falhas que precisamos levar em consideração no nosso modelo, aqui podemos perceber que alguns tipos são mais frequentes.



Nesse gráfico conseguimos ver a distribuição dos tipos de falhas, é possível observar que um tipo de falha (power failure) é mais frequente onde há pouco torque e muitas rotações. Assim como em baixas rotações e muito torque.



Aqui apresentamos uma correlação entre todas as variáveis numéricas, essa correlação nos dá uma noção de quais interações são mais prováveis de estar relacionadas a falhas e quais interações entre as variáveis são mais importantes.

2- Para esse tipo de problema, utilizei duas abordagens, optei, primeiramente, pela simplicidade dos modelos e eficiência para produzir as respostas, assim como a confiabilidade. Como se trata de um problema de classificação com várias classes (multiclassificação) utilizei um modelo de árvore de decisões (Decision Tree), disponível na figura abaixo. e um modelo de regressão logística (logit). O modelo de maior confiabilidade foi a árvore de decisões, com aproximadamente 99% de acurácia. Para avaliar a performance do modelo foi realizada uma curva-ROC, pela simplicidade e eficiência.

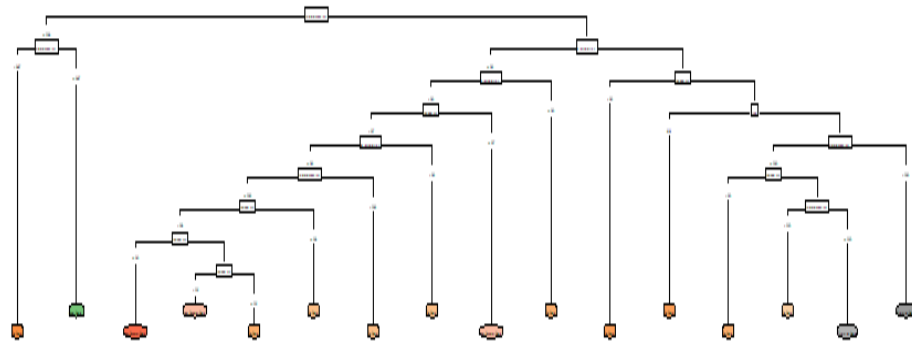


FIGURA. Árvore de decisões gerada pelo modelo.