

Análise estatística de dados de motores

Inicialmente foi realizada uma análise gráfica dos dados de treinamento, para identificar possíveis correlações existentes entre os parâmetros captados pelos sensores e as possíveis falhas (Anexo 1).

Após realizado a plotagem combinado em par de duas variáveis, foi possível identificar que alguns grupos tem se encontrado em algumas regiões específicas, independente da combinação que fosse. Mas antes dessas análises, foi visto que a quantidade de dados de classificação *No Failure* era muito superior à das falhas, um total de 6435 dados, e isso atrapalhou a visualização dos dados de falhas, por isso para a plotagem desses dados foram retirados via filtro.

As falhas de *Power Failure*, exceto quando aplicado as características de temperatura de processo x temperatura ambiente, desgastes de ferramentas x temperatura de processo e desgaste e temperatura ambiente, tende a se dividir em duas regiões dos gráficos distintas, próximas aos limites inferiores do gráfico.

As falhas de *Heat Dissipation Failure*, em todas as análises tendem a se agrupar próximas, facilitando sua segregação das outras falhas. Assim como a *Tool Wear Failure*, apesar de acabarem se misturando com outras falhas, conseguimos definir bem uma região para sua alocação.

Já as falhas de *Overstrain Failure*, acabam sempre se misturando outros grupos nas combinações de Desgaste x Temperatura de processo e Desgaste x Temperatura do ambiente, nas outras é possível a determinação de uma região de operação.

E são essas regiões distintas que devem ser focos de separação via rede neural, por isso se trata de um problema de classificação. Pois precisamos focar na classificação de cada região, para assim obtermos um modelo que consiga nos dizer em qual região se encontra um motor para prever sua falha.

Além disso, foi realizada uma análise para saber quais dados eram mais importantes e relevantes para o estudo. A coluna UID e product ID foram inicialmente retiradas por serem dados de identificação, não teriam valor que ajudasse na classificação. Já a coluna *type*, foi retirada por não demonstrar grandes diferenciações entre as falhas (Anexo 2), com visto no gráfico *Failure_type* x *Count*, sempre temos a predominância de um tipo de peça em relação a outra, e isso é um fator em comum e não uma característica que as distingue-as.

Em sequência, antes de realizar o treinamento da rede, foi feita a separação de alguns dados da classe *No Failure*, devido a grande quantidade de dados ocasionarem um “vício” na rede, foi separado apenas 214 dados desse grupo para participarem do treinamento, e também ocorreu a separação da coluna de *failure_type* (a classificação) das outras colunas, por fim essas colunas restantes

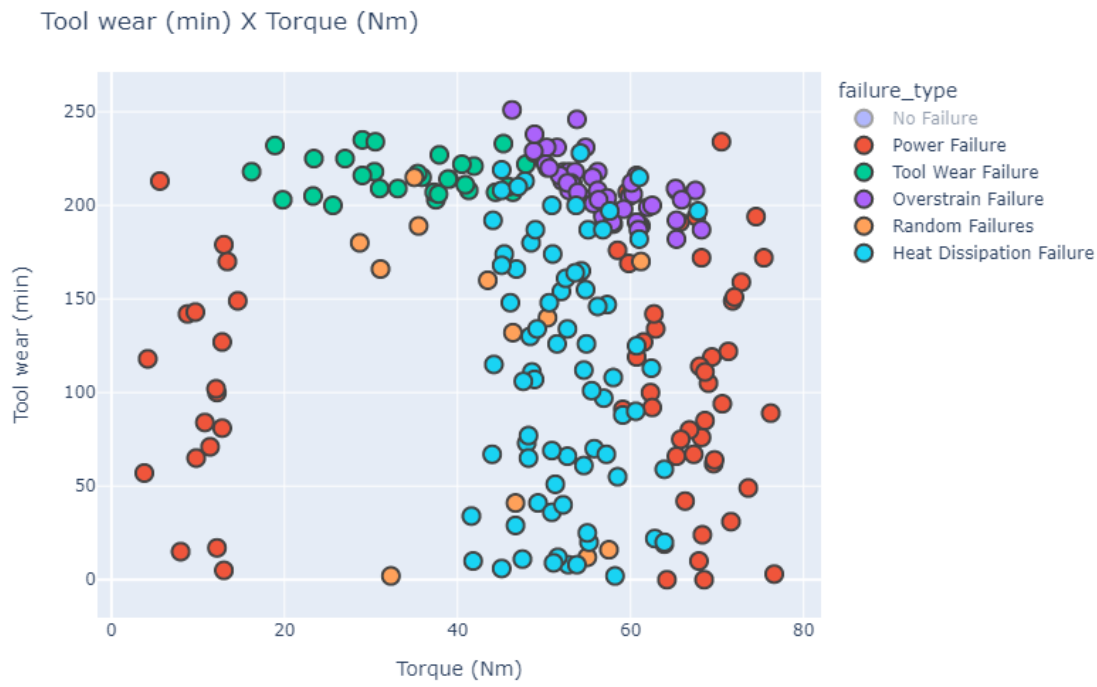
sofreram uma normalização dos seus dados, dessa forma conseguimos mais performance computacional e agilizamos o treinamento da rede.

Por último, realizamos a transformações de cada classe em vetores *dummy*, que nos facilitam a classificação pela rede neural, atribuindo a cada falha um vetor de saída específico da rede, indicando a possibilidade daqueles dados serem classificados em cada classe.

A rede neural utilizada foi uma rede neural profunda, pois elas são mais eficientes em encontrar relações mais complexas como a desse estudo. Já que temos muitos dados sobrepostos, ela seria a melhor escolha, apesar de ter um gasto maior de processamento e exponencialmente mais variáveis que uma rede mais comum, e outro problema foi a quantidade total de dados ser muito pequena para explorar melhor essas correlações, pois não seria prudente colocar tantos dados de uma mesma classe para um treinamento.

Após alguns testes realizados, quanto maior o número de camadas, maior foi a precisão categórica e menor foi o valor da função de perda categórica da rede nos dados de teste. A precisão obtida nos dados de teste foi de 90% e o custo da função de perda foi em torno de 15%, o que seria um resultado excelente caso houvesse mais dados para o estudo, essas duas métricas são eficientes para esse tipo de estudo no qual queremos saber se de fato a classificação tem uma precisão boa e ambas servem para essa conclusão.

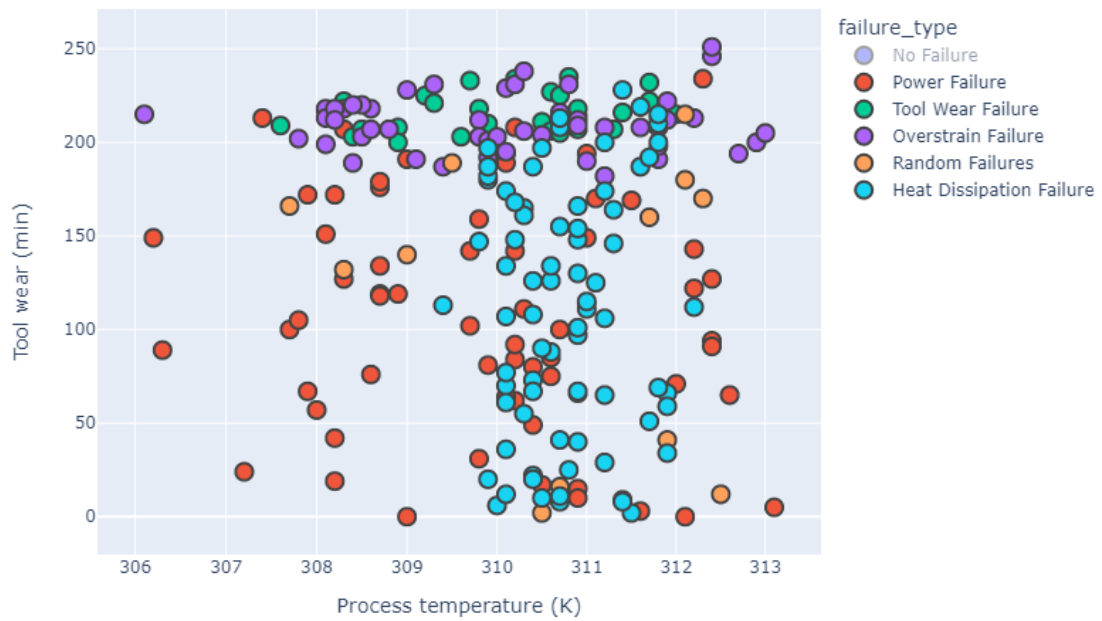
Anexo 1



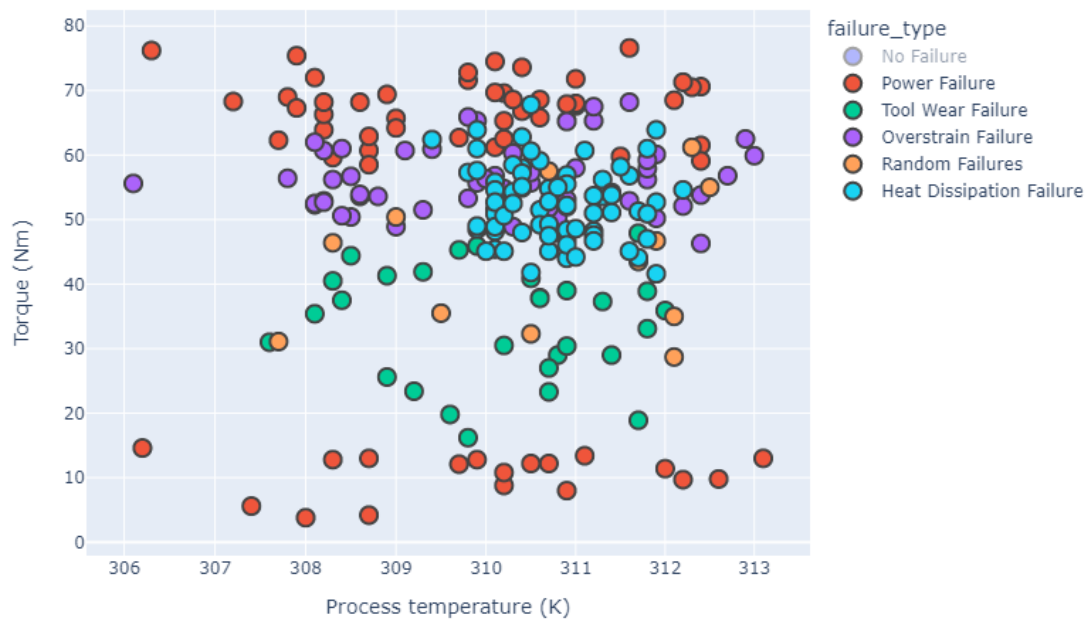
Torque (Nm) X Rotational speed (rpm)



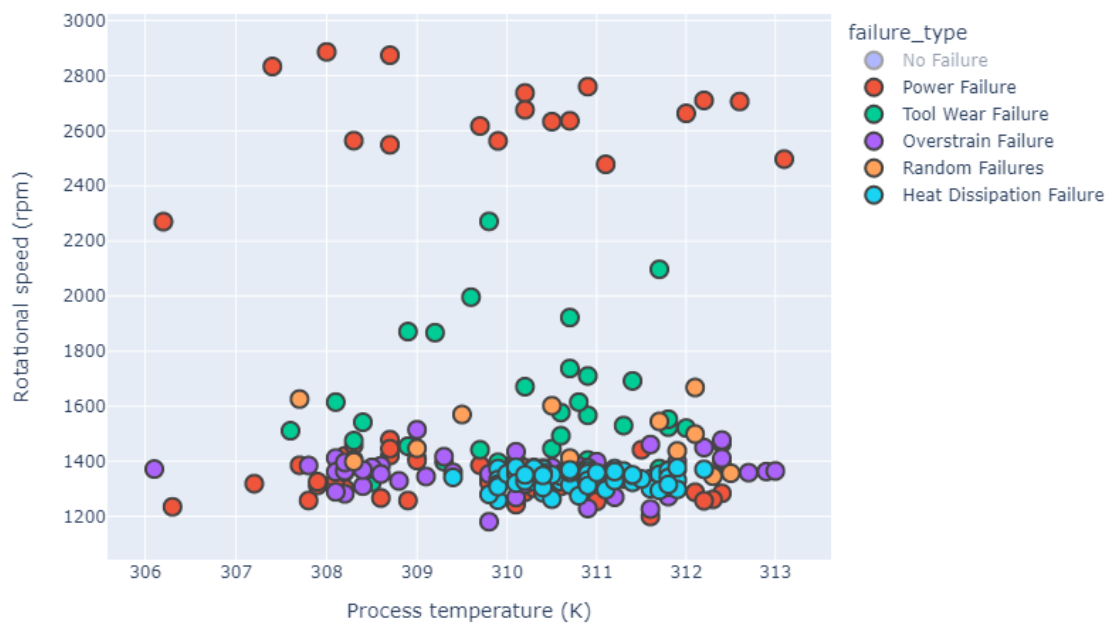
Tool wear (min) X Process temperature (K)



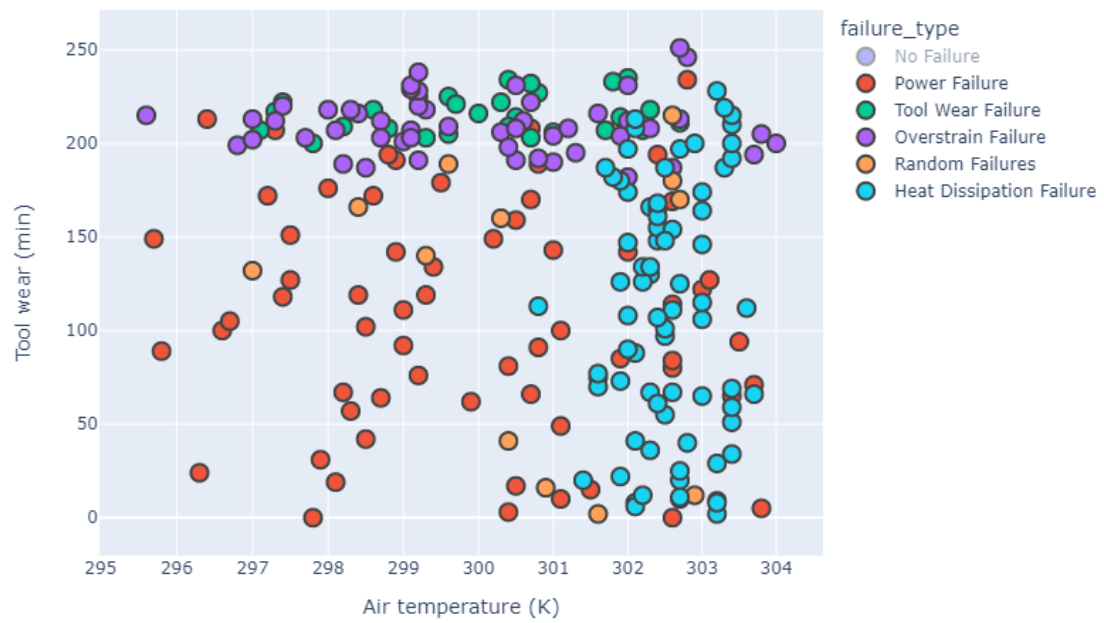
Torque (Nm) X Process temperature (K)



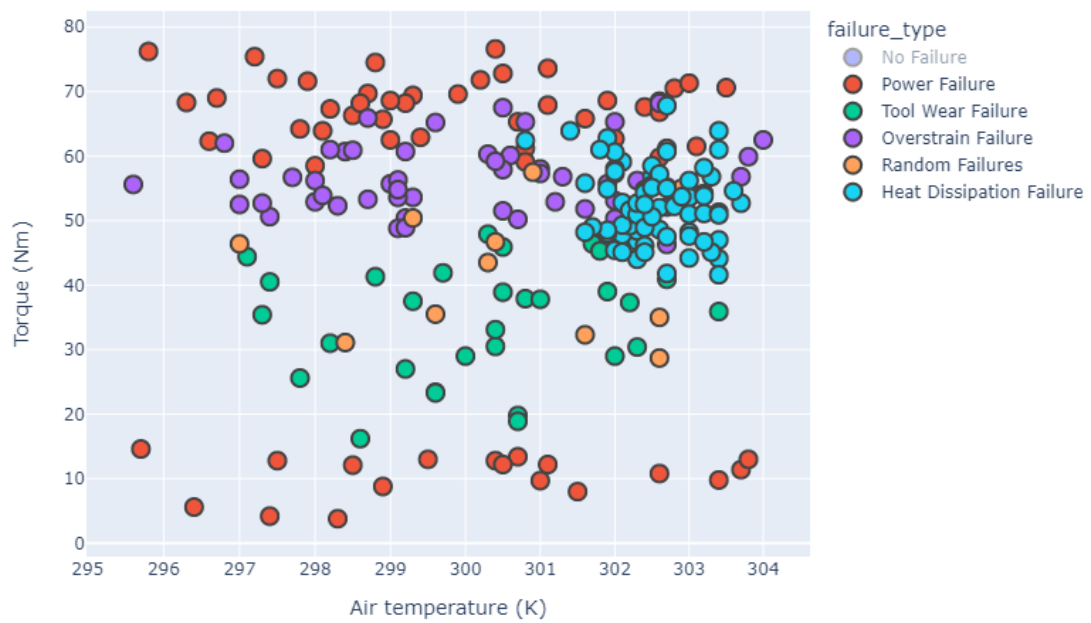
Rotational speed(rpm) X Process temperature (K)



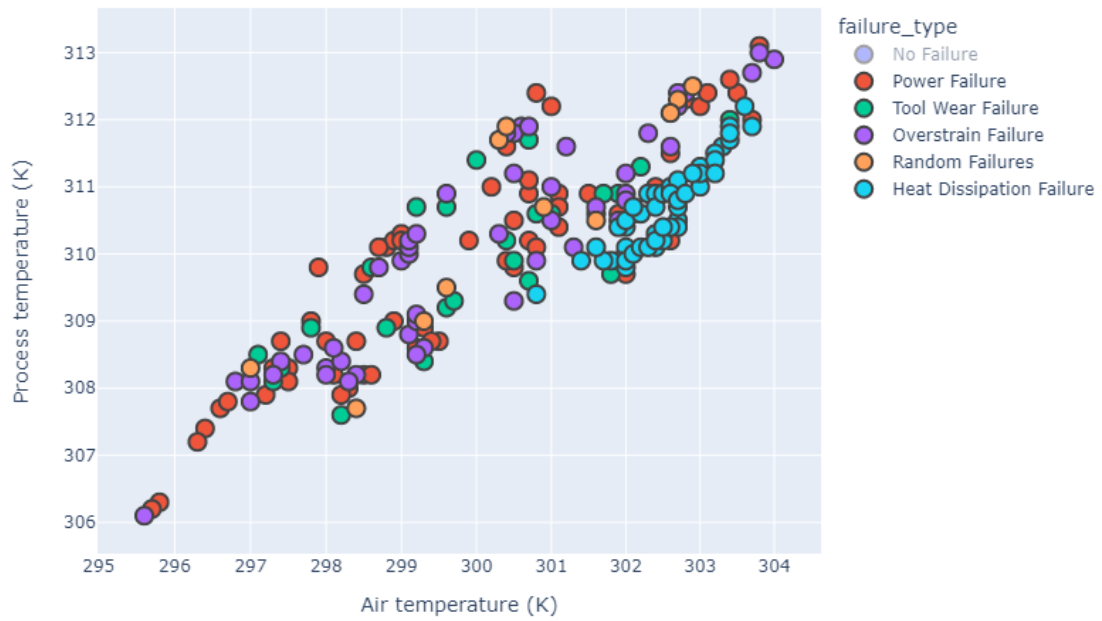
Tool wear (min) X Air temperature (K)



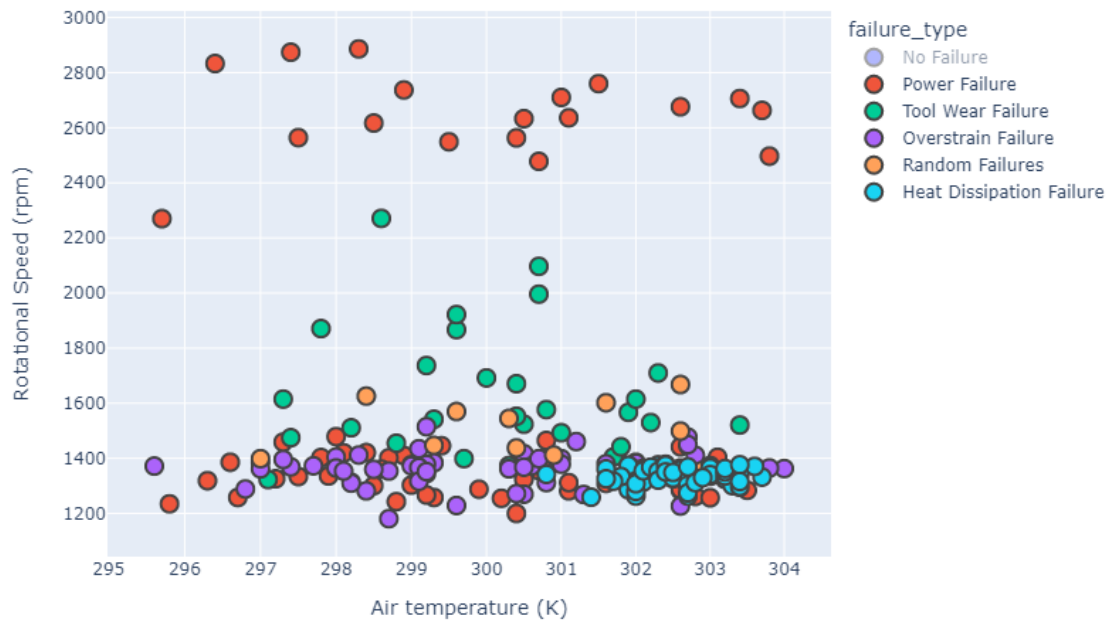
Torque (Nm) X Air temperature (K)



Process temperature (K) X Air temperature (K)

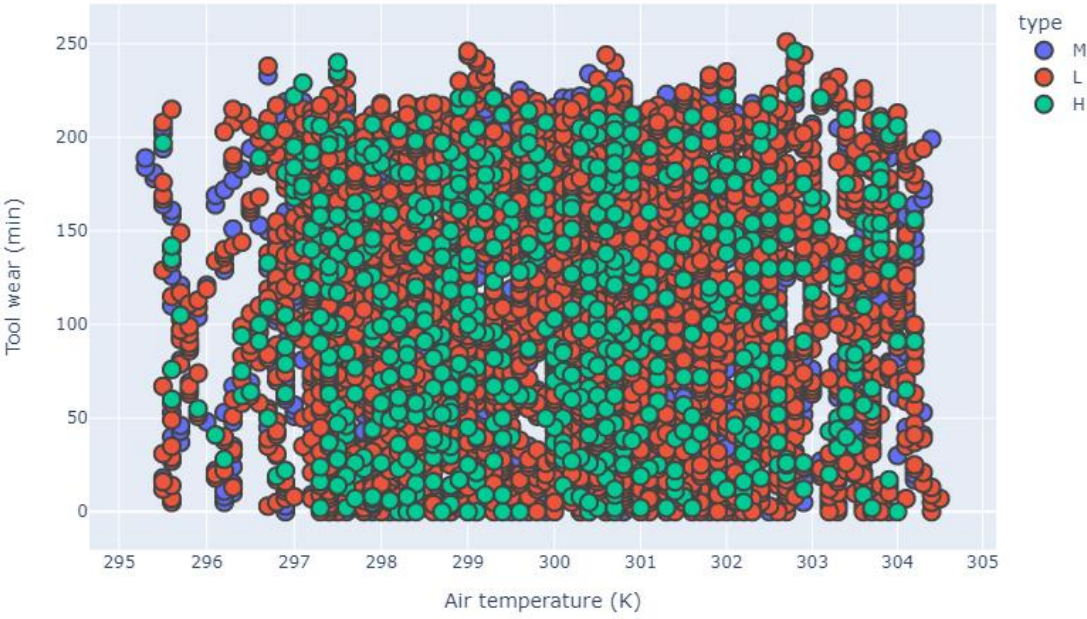


Rotational speed (rpm) X Air temperature (K)

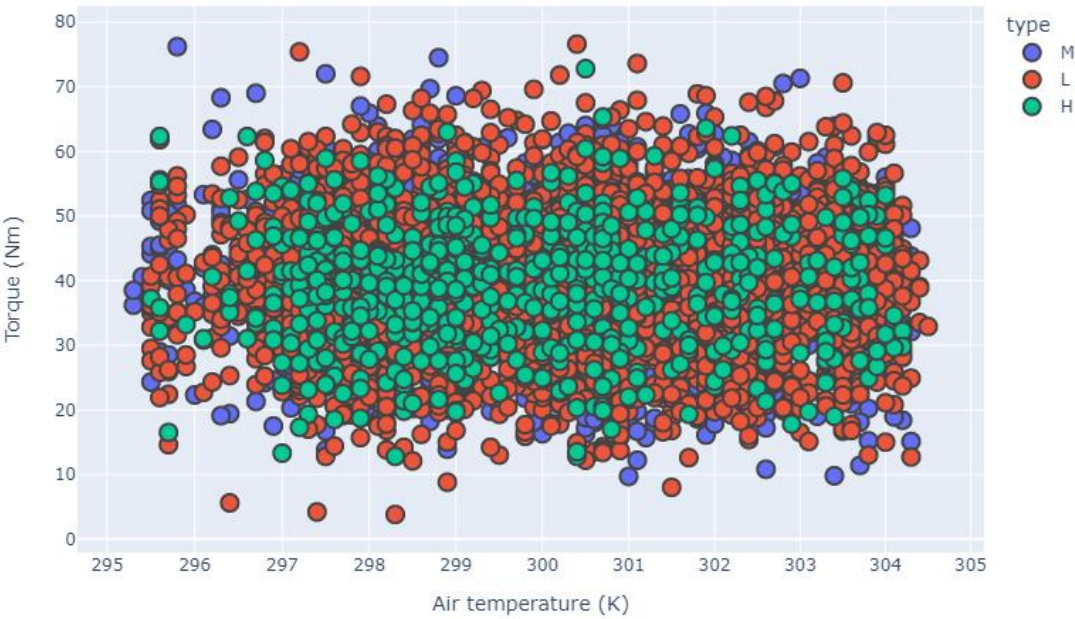


Anexo 2

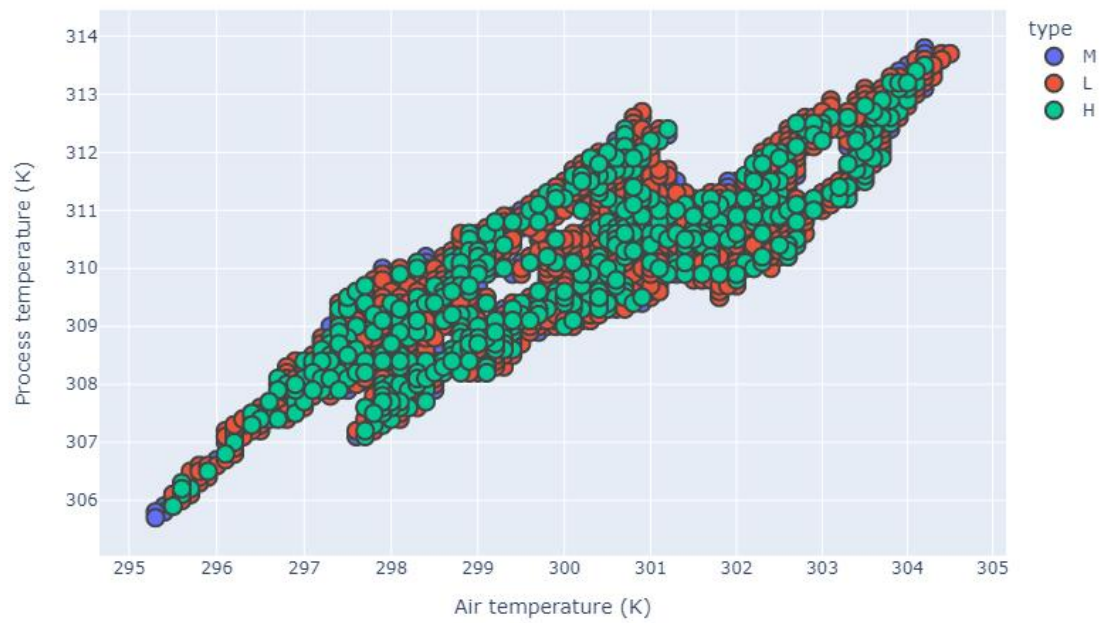
Tool wear (min) X Air temperature (K)



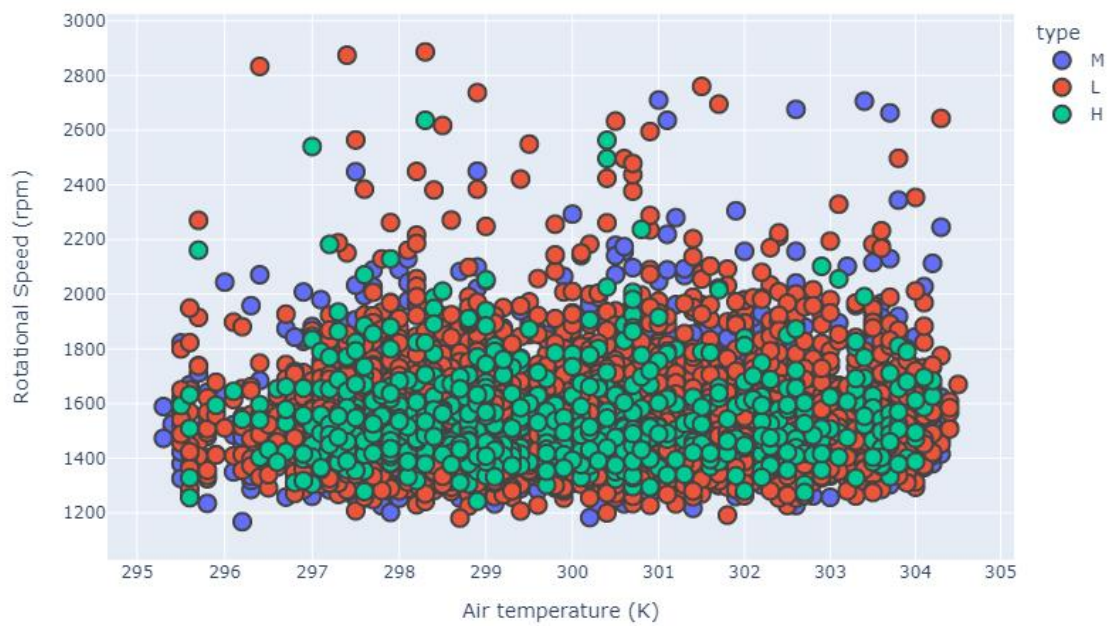
Torque (Nm) X Air temperature (K)



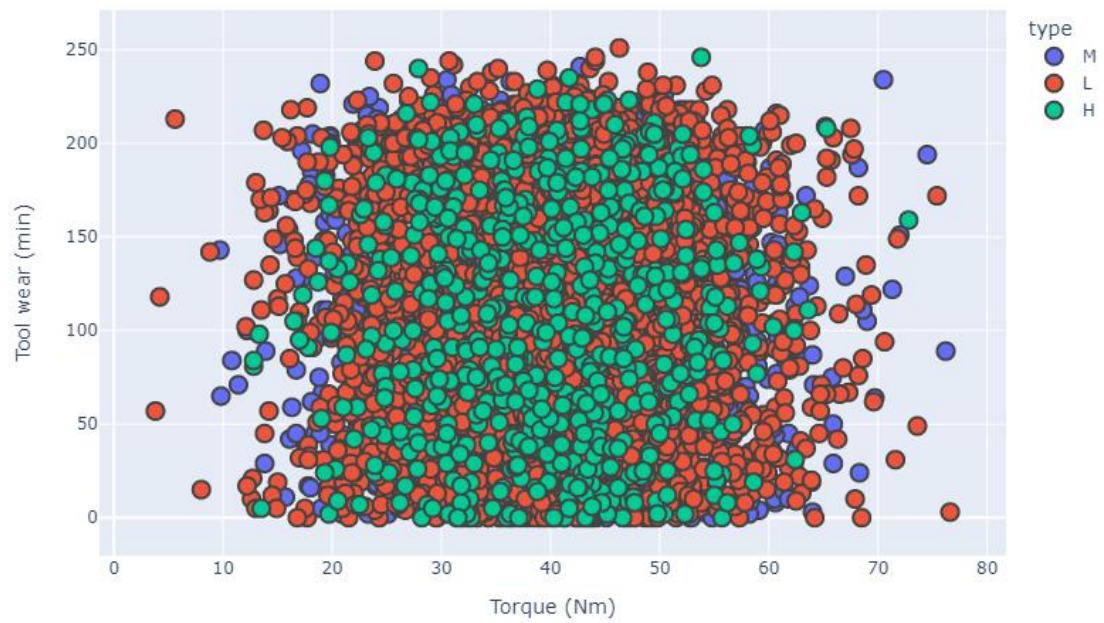
Process temperature (K) X Air temperature (K)



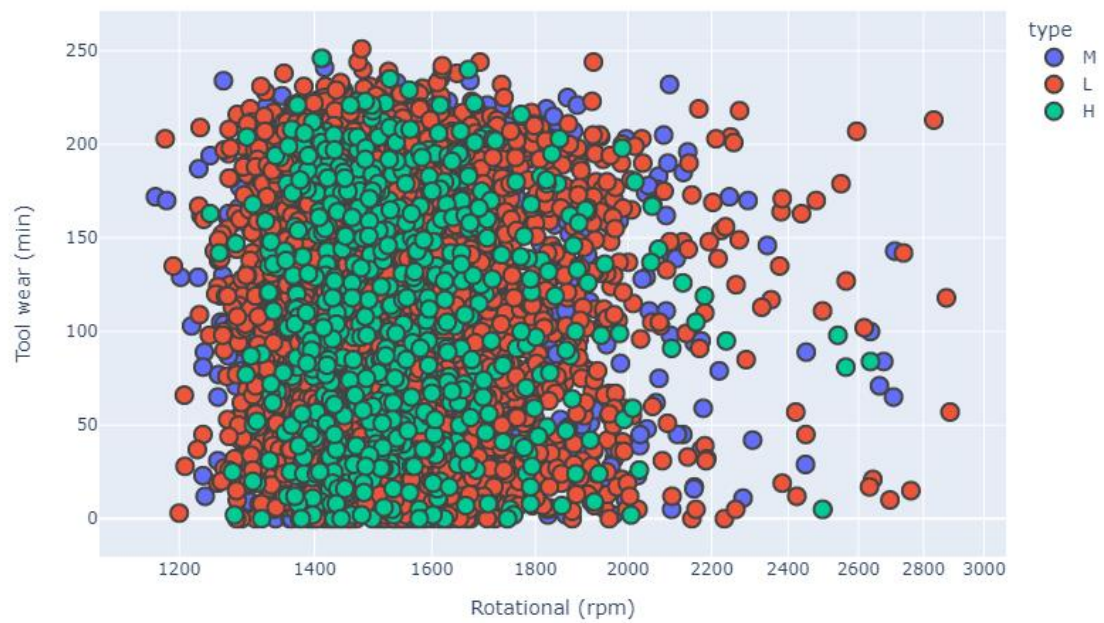
Rotational speed (rpm) X Air temperature (K)



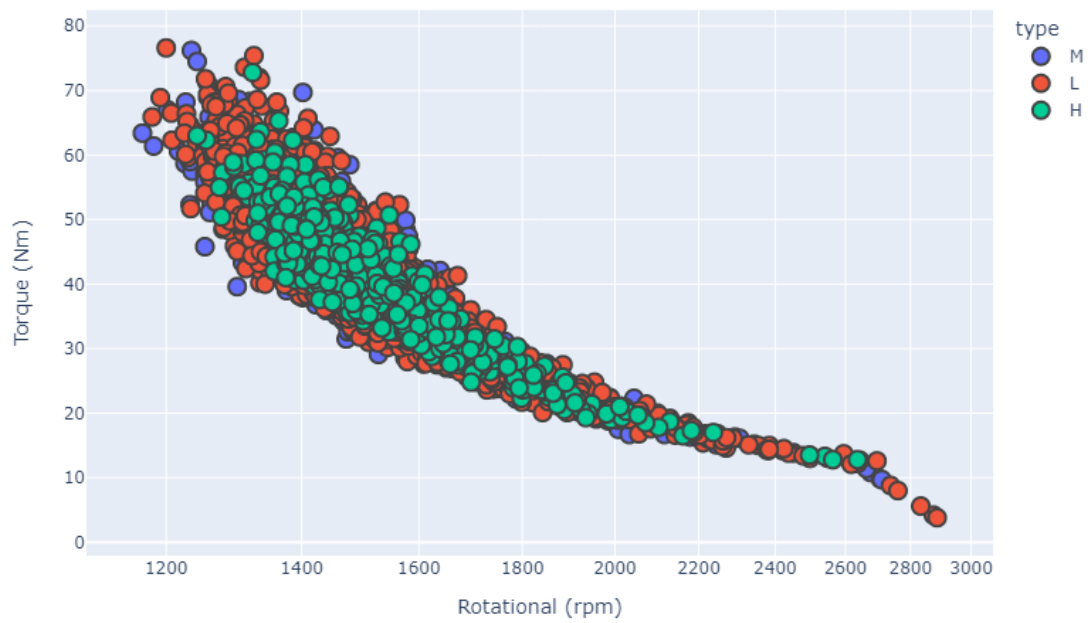
Tool wear (min) X Torque (Nm)



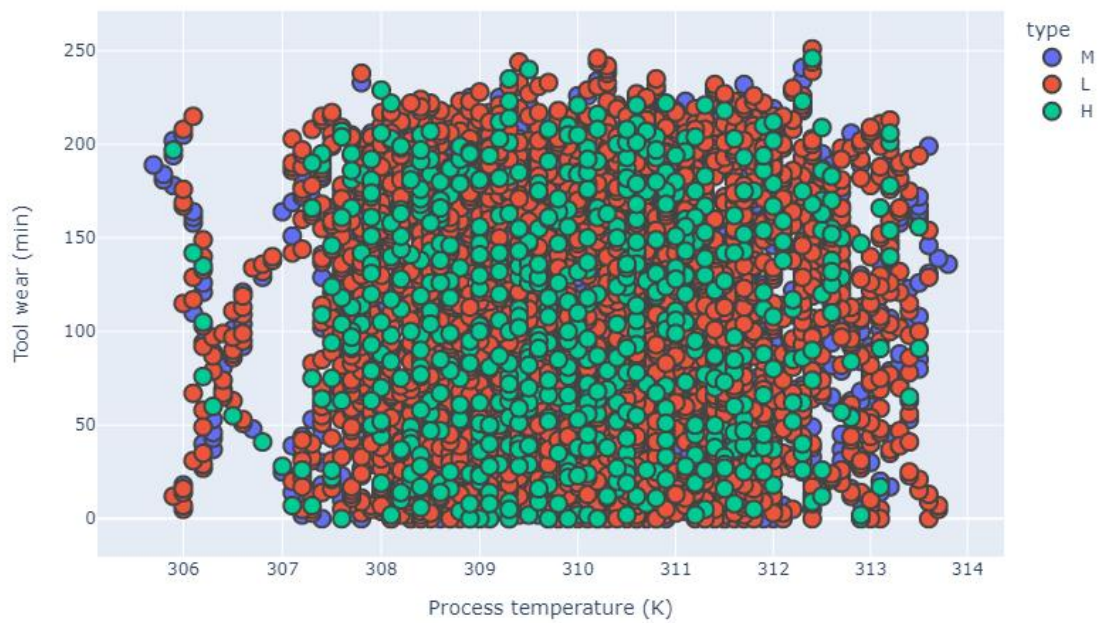
Tool wear (min) X Rotational speed (rpm)



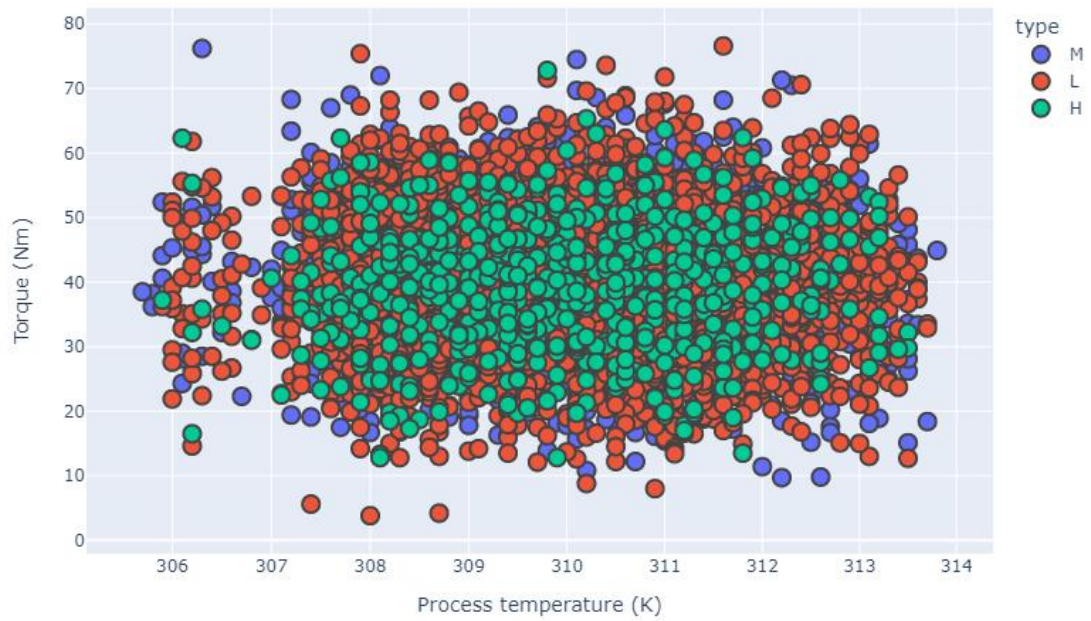
Torque (Nm) X Rotational speed (rpm)



Tool wear (min) X Process temperature (K)



Torque (Nm) X Process temperature (K)



Rotational speed(rpm) X Process temperature (K)

