**Entregáveis IoT e IA**

**Global Impact**

1 – Escolha do pré-processamento de dados

O pré-processamento foi feito a partir na normalização dos dados para números entre 0 e 1. A normalização é importante para o processamento mais rápido e maior precisão dos pesos de cada atributo (coluna). Para tal, foi utilizada a biblioteca MinMaxScaler do kit sklearn.preprocessing.

2 – Atributos categóricos e não categóricos

Como o dataset disponibilizado não possui dados não categóricos, como informações de gênero que são escritas por extenso (macho ou fêmea), por exemplo, não foi necessário transformar esses dados em números, já que só assim poderia ser possível fazer o cálculo da distância entre pontos.

3 – Escolha do algoritmo

Por se tratar de uma quantidade de amostras pequena, apenas 132, foi utilizado o KNN (K Nearest Neighbors), que basicamente verifica a distância da amostra de teste, em relação às amostras com menor distância dos atributos analisados no treinamento. Caso a amostra tenha mais “vizinhos” da classe x ela será classificada como tal, caso tenha mais “vizinhos” da classe y ela será classificada como y e assim sucessivamente dependendo da quantidade possível de classes.

4 – Métricas para avaliar performance do modelo

Para avaliar a performance do modelo, foram separados dois conjuntos de amostras: teste e treinamento. O conjunto de treinamento foi usado para a máquina aprender quais atributos podem influenciar na classificação de uma amostra. Após realizar este aprendizado, defini o k (número de vizinhos a ser considerado) e passei o conjunto de teste para validar sua eficácia, vendo quantas amostras o modelo conseguiu acertar. Depois de ter o número de acertos definido, apenas dividi pelo total de amostras de testes e multipliquei por cem para transformar em porcentagem.