1. Machine Learning é o processo em qual a máquina adquire conhecimento sobre um tópico determinado. Esse processo é composto por treinamento, onde você popula o modelo com seus dados, que podem ser rotulados ou não. Em sequência, faz-se o teste do modelo e verifica se ele conseguiu aprender com os dados fornecidos, utilizando métricas como acurácia, precisão, recall etc.
2. **Conjunto de treino:** Parte do conjunto de dados disponíveis que o modelo terá acesso para realizar seu aprendizado, ajustar seus hiperparâmetros e aprender com os dados.

**Conjunto de validação:** Dados utilizados para verificar a capacidade de generalização dos modelos dos dados de testes para novos dados. Geralmente escolhe-se o modelo com melhor capacidade de generalização no conjunto de validação para ser testado.

**Conjunto de teste:** Responsável por avaliar o modelo escolhido no treinamento e validação. É um conjunto de dados que o modelo não teve acesso nas fases de treinamento e validação e avalia a capacidade do modelo de generalizar e classificar corretamente novos dados.

1. Para tratar valores ausentes em um conjunto de treinamentos, existem várias técnicas possíveis, como apagar os registros que estão em branco, o que pode diminuir consideravelmente sua base de treino e prejudicar o aprendizado do modelo. Outra solução é preencher os valores com base em dados estatísticos retirados da base, podendo usar por exemplo a média ou mediana de todos os valores presentes para substituir valores ausentes.
2. A matriz de confusão é uma ferramenta muito util para verificar as predições em relação aos valores reais, mostrando os valores verdadeiros positivos (modelo previu positivo e acertou), falsos positivos (modelo previu positivo e era negativo), verdadeiros negativos (previsão negativa e acerto) e falsos negativos (modelo previu negativo e era positivo). Sua utilização pode facilitar a ter uma noção mais clara das métricas de desempenho, como acurácia, precisão e recall.
3. Acho que a aplicação de algoritmos de aprendizado de máquina na área da saúde é o mais interessante no momento atual, principalmente em países em desenvolvimento, que apresentam super lotação em hospitais e cargas de trabalho prejudiciais aos seus médicos e demais funcionários. A aplicação pode ajudar a diminuir a sobrecarga médica, diminuir a quantidade de erros em diagnósticos e detectar o surgimento de novas doenças em estágios iniciais com um tratamento menos prejudicial aos pacientes.