

Atlântico Avanti

Nome: Allan Vitor Marques dos Santos Almeida

Atividade 01 - Machine Learning

1 - É a identificação de padrões que auxilie em previsões e/ou decisões de forma automatizada. Ao invés de realizar um algoritmo para realizar essas atividades a máquina faz isso de forma independente, pois há uma grande massividade de dados e podendo identificar esses padrões.

2 - Conjunto de treinamento: É usado para ajustes de parâmetros internos onde o modelo aprende padrões, relações e características dos dados e sua porcentagem típica é de 60 a 80%.

Conjunto de validação: Ajusta o hiperparâmetros e avalia o desempenho do modelo durante o processo e avalia os dados para evitar o overfitting.

Conjunto de teste: É usado para avaliar o desempenho final após o treinamento e correção do hiperparâmetro, tendo uma porcentagem típica de 10 a 20% dos dados.

3 - Se houver poucos dados ausentes é mais indicado sua remoção. Caso seja uma coluna que faça sentido para minha regra de negócio é indicado imputação de dados com média/mediana/moda e se for uma coluna irrelevante com muitos nulos é mais indicado a remoção da coluna.

4 - A Matriz de confusão é uma tabela que faz comparação dos valores reais com os valores de previsão pelo modelo. Ela é importante quando é necessário a classificação binária, mas que pode ser expandida para múltiplas classes.

A partir da matriz pode se calcular:

Acurácia: para ver a proporção de acertos em relação ao total.

Precisão: entre o que o modelo previu, quantos são positivos e quantos realmente são positivos?

Recall: Entre os que realmente são positivos, quantos o modelo detectou?

F1 Score: Média harmônica entre a precisão e o recall.

5 - Acredito que não há uma área específica que os algoritmos de machine learning se aplica melhor, mas sim em quantidades de padrões que há. Por exemplo a construção civil se a pilastra não tiver com as quantidades certas de cimento, pedras e entre outras coisas vai ser um problema, pois não sustentará o peso, mas na área veterinária é onde pode ser interessante a aplicação, pois para auxiliar no tratamento de animais domésticos e um diagnóstico mais preciso e mais rápido.