**Machine Learning & Modelling**

# Checkpoint 01

|  |
| --- |
| Respondam as questões abaixo **individualmente** relativas às redes neurais e enviem as respostas em Word (formato editável) no Teams.  **Data limite**: 04/maio/2022 |

1. **Como você define “*machine learning*”?**O machine learning, ou aprendizado de máquina, explora o estudo e construção de algoritmos que podem aprender de seus erros e fazer previsões sobre dados.
2. **Quais são as principais diferenças entre aprendizado supervisionado e não-supervisionado? Dê exemplos de problemas modelados por cada uma destas abordagens.**R: a maior diferença, é que o supervisionado, você fornece labels para a máquina, já o não supervisionado, a máquina não tem acesso a labels já existentes na hora de classificar seus dados.  
   um exemplo de supervisionado, seria uma maquina que quer classificar dados em 2 tipos, como por exemplo a lista de sobreviventes do Titanic, que precisava classificar em vivo ou morto.  
   um exemplo de não supervisionado seria de clusterizacao, onde uma empresa precisa classificar grupos de clientes baseado em diversas características.
3. **Quais são as etapas do processo de Ciência de Dados? Qual a importância de cada uma delas?**

Entendendo o problema > sem entender o problema corretamente, é impossível conseguir solucionar o mesmo

Coleta de dados > extrair e coletar dados, para ter recursos suficientes para solucionar o problema

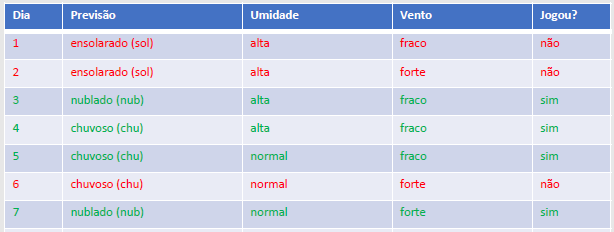
Processamento /Análise de dados > limpar os dados de registros duplicados, faltantes, formatados de forma não-convencional etc.

Exploração de dados > com base na observação dos dados, pensar em ideias e hipóteses a serem validadas

Comunicação de resultados > Comunicação de resultados

Feedback > retorna se o resultado do projeto foi suficiente ou não, se resolveu o problema ou nao

1. **Um dos principais desafios do aprendizado de máquina é lidar com problemas de *overfitting* e *underfitting* dos modelos. Qual é a diferença destes desafios e por que eles são problemáticos à aprendizagem?**R: Overfitting, é quando uma maquina de certa forma se acostuma com o treino, cumpre o seu papel de forma exemplar, porem, quando apresentada outras situações e problemas minimamente diferente que os usados em seu treino, a maquina não consegue resolver, como um aluno que decora exatamente o gabarito da prova ao invés de estudar, e tira 10, porem se fizer outra prova com o mesmo conteúdo da primeira, não irá conseguir resolver.  
     
   underfitting acontece quando um modelo de aprendizado de máquina não é complexo o suficiente para capturar com precisão as relações entre os recursos de um conjunto de dados e uma variável de destino. Um modelo insuficiente resulta em resultados problemáticos ou errôneos em novos dados, ou dados nos quais não foi treinado, e muitas vezes tem um desempenho ruim, mesmo em dados de treinamento.
2. **Neste exemplo de dataset visto em sala de aula, conceitue:** 
   1. **Quais são as variáveis?**R: Previsao, umindade, Vento
   2. **Qual é o rótulo (ou *label*)?**R: Se jogou ou não jogou
   3. **Quantos exemplos nós temos?**R: 7 exemplos
   4. **Se fossemos criar uma modelagem para prever se o jogador “jogou ou não”, faríamos um modelo de classificação ou regressão? Justifique.**R: Classificação, pois a saída tem 2 opções, se Jogou ou se não jogou.

****

1. **Explique a diferença entre conjuntos de *treino*, *validação* e *teste*. Diga de onde eles provêm, como são construídos, a importância e utilização de cada um deles.**

R: O Treino e Teste, é uma divisão proporcional dos dados e dos labels para diferenciar, onde a maquina vai ser trabalhada, e onde ela vai ser testada, para evitar problemas de Overfitting entre outros.

1. **Liste 4 métricas diferentes, que vimos em sala de aula ou que pesquisou na Internet, preenchendo o quadro abaixo:**

|  |  |  |  |
| --- | --- | --- | --- |
| **NOME DA MÉTRICA** | **USADA PARA CLASSIFICAÇÃO OU REGRESSÃO?** | **O QUE SIGNIFICA?** | **COMO É INTERPRETADA?** |
| **Acurácia** | **Classificação** | **É basicamente o número de acertos (positivos) divido pelo número total de exemplos** |  |
| **F1 Score** | **Classificação** | **O F1 Score é uma média harmônica entre precisão e recall** |  |
| **Precisão** | **Classificação** | **úmero de exemplos classificados como pertencentes a uma classe, que realmente são daquela classe (positivos verdadeiros), dividido pela soma entre este número, e o número de exemplos classificados nesta classe, mas que pertencem a outras (falsos positivos).** |  |
| **Recall** | **Classificação** |  |  |

1. **Utilizando o código do notebook “Aula05\_ArvoresDecisao ---AULA2022.ipynb”, altere os parâmetros da árvore: *profundidade*, *grau de pureza mínimo aceito para as folhas*, *número de árvores na floresta*. O que você observou nos resultados com cada alteração? O que você conclui com isso?**