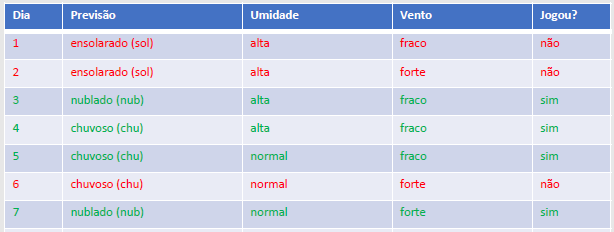
**Machine Learning & Modelling**

# 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*”?
2. Quais são as principais diferenças entre **aprendizado supervisionado** e **não-supervisionado**? Dê exemplos de problemas modelados por cada uma destas abordagens.
3. Quais são as etapas do **processo de Ciência de Dados**? Qual a importância de cada uma delas?
4. 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?
5. Neste exemplo de **dataset** visto em sala de aula, conceitue:
   1. Quais são as variáveis?
   2. Qual é o rótulo (ou *label*)?
   3. Quantos exemplos nós temos?
   4. Se fossemos criar uma modelagem para prever se o jogador “jogou ou não”, faríamos um modelo de classificação ou regressão? Justifique.



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.
2. 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?** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

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?