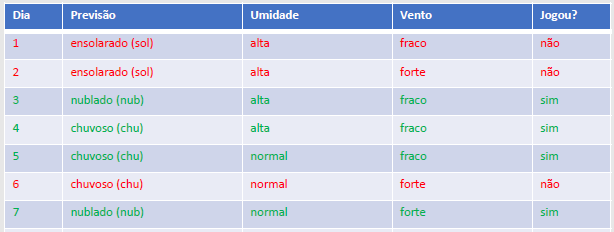
**Machine Learning & Modelling**

# 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*”?**R: 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:O aprendizado supervisionado é aquele que o programador fornece labels para a machina, um exemplo seria de classificacao já o não supervisionado não tem labels, e um exemplo seria de Clusterização
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?**Overfitting é quando o modelo esta muito ajustada ao treino,ao ponto que quando é apresentada aos dados de teset, ela não tem o mesmo desempenho por se “Acostumar” com os dados ao invez de entender a solução, é como um estudante que decora um gabarito de uma prova ao invez de estudar para a mesma, ele ira tirar um 10 naquela prova, porem, não conseguira fazer outras provas da mesma matéria.  
   **e *underfitting*** seria o contrario disso, seria um modelo que esta extremamente mal ajustado ao treino, resultando em um desempenho ainda pior nos testes, algo como um aluno que nem estuda para a prova e vai fazer ela.
5. **Neste exemplo de dataset visto em sala de aula, conceitue:** 
   1. **Quais são as variáveis?**Previsao, Umidade, Vento
   2. **Qual é o rótulo (ou *label*)?**Se jogou ou nao
   3. **Quantos exemplos nós temos?**7
   4. **Se fossemos criar uma modelagem para prever se o jogador “jogou ou não”, faríamos um modelo de classificação ou regressão? Justifique.**Classificacao, pois temos apenas 2 labels.

****

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 serve para o modelo aprender a resolver o problema, pode ser representado por 80% dos dados disponives, e já o teste, seria para poder avaliar a eficiência do podelo, nesse caso representando o restante dos dados, 20%.
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?** |
| **Precisao** | **classificacao** | **dos que foram chamados de proibido, quantos eram de fato proibido** | TP/(TP+FP) |
| **Recall** | **classificacao** | **de todos os proibidos, quantos nao passaram batido** | TP/(TP+FN) |
| **F1-Score** | **classificacao** | **relacao de precisao com recall** | **2\*((Precisao\*Recall)**  **/**  **(Precisao +Recall))** |
| **Acuracia** | **classificacao** | **resultados certos em relacao ao total** | **(TP+TN)/(TP+TN+FP+FN)** |

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?**