

## PROVA 2 - INTELIGENCIA ARTIFICIAL

Elabore um único relatório com todos os procedimentos (Metodologia) e resultados obtidos (com discussão) de cada questão, associando o método aos resultados encontrados e sugerindo melhorias.

### Classificação

1. Utilizando a base disponível em: <https://www.kaggle.com/srinivas365/multilabel-classification-emotions> elabore uma solução utilizando dois algoritmos de aprendizagem de máquina do seu conhecimento para classificar a emoção da música (escolha como você representará a saída, uma vez que algumas músicas podem pertencer a duas categorias). Justifique a escolha dos dois algoritmos de aprendizagem de máquina utilizados e discuta os resultados obtidos em ambos. Divida (aleatoriamente) o conjunto em treinamento e teste: 80% e 20%. Os resultados dessa questão deverão ser descritos detalhadamente no relatório através de três pontos principais:

- Análise da base de dados: identificar instancias com atributos incompletos, gerar matriz de correlação, identificar a presença de *outliers* e verificar se as classes estão balanceadas. O balanceamento devera ser ilustrado por meio de gráficos (e.g. histograma);
- Análise dos resultados considerando matriz de confusão, especificidade, sensibilidade, medida f1 e acurácia. Descreva detalhadamente os resultados obtidos por cada métrica, justificando a diferença entre eles.

**Bônus:** Altere dois parâmetros de cada algoritmo de aprendizagem de máquina utilizados na questão e discuta os resultados obtidos. Exemplo: alterar a quantidade de k vizinhos e a função de distancia utilizada, alterar a função kernel do SVM, alterar a arquitetura da rede neural (e.x. camadas e função de ativação, alterar o otimizador e a taxa de aprendizado).

### Regressão

2. Utilizando as bases de treinamento e testes disponível no link:

<https://archive.ics.uci.edu/ml/machine-learning-databases/00492/>

Teste dois métodos de aprendizagem de máquina para identificar o que apresenta melhores resultados na previsão do volume de tráfego. Avaliemos o impacto de usar validação cruzada (com 10 grupos) e split (70% para treinamento e 30% para teste). Para esse último, execute o experimento duas vezes selecionando as instâncias aleatoriamente para o treinamento e teste e avalie o impacto nos resultados. Outras informações como dia da semana podem ser incluídas.

### Clusterização

3. Utilizando a base:

<https://archive.ics.uci.edu/ml/datasets/Facebook+Live+Sellers+in+Thailand>

- Execute os algoritmos de agrupamento K-means e Hierárquico com diferentes valores de clusters e escolha o melhor resultado.

Para o hierárquico, varie entre o single e complete linkage e analise os resultados obtidos.

Para o k-means, inicialize os centroides aleatoriamente e avalie o impacto da inicialização nos resultados encontrados.

**Bônus:** normalizar os dados e executar novamente os algoritmos para analisar os resultados gerados. Fazer visualização dos clusters encontrados com PCA ou t-SNE.