Atividade prática: Algoritmo k-Nearest Neighbors (KNN)

Objetivo da atividade:

- Realizar uma implementação própria do algoritmo KNN
- Avaliar o uso do algoritmo KNN com diferentes valores de k
- Verificar o efeito da normalização de dados sobre os resultados

1. Faça uma implementação própria do algoritmo KNN na sua linguagem de programação de preferência. Utilize como base o pseudocódigo fornecido no Slide 24 da aula sobre KNN.
    - Sua implementação deve permitir informar/variar o valor do hiperparâmetro k para cada execução do algoritmo
    - Utilize como medida de distância padrão a *Distância Euclidiana*. Se desejar, pode permitir que esta medida seja configurada, por exemplo, que seja usada a *Distância de Manhattan*.
    - Assuma que sua implementação se destina à análise de dados quantitativos, isto é, não é necessário se preocupar em implementar suporte ao tratamento de dados qualitativos.

2. Utilize os conjuntos de dados de treinamento e teste fornecidos no diretório *Dados_Normalizados* para avaliar a aplicação do KNN com diferentes valores de k. Como o nome do arquivo indica, estes dados foram previamente normalizados, tal que todos os atributos variam no mesmo intervalo ([0,1]). Os dados se referem à classificação de tumores de mama em maligno (1) ou benigno (0), de acordo com a coluna *target*. Os atributos (29) descrevem características dos núcleos celulares presentes em uma imagem digitalizada do material coletado na biópsia pelo método *fine needle aspirate* (FNA).
    - A partir dos dados de treinamento disponíveis, classifique os dados de teste usando **k=1, k=3, k=5, e k=7** (se desejar, avalie valores adicionais para k).
    - Avalie o desempenho do modelo usando a métrica de *acurácia* (taxa de acerto), reportando para cada valor de k a porcentagem de instâncias de teste classificadas corretamente.

3. Refaça o item 2, agora para o conjunto de dados no arquivo *Dados_NaoNormalizados*.
    - Observe o intervalo em que varia cada atributo no dado de treinamento: há uma grande diferença entre os valores máximo e mínimo de cada atributo?
    - Analise e comente se houveram ou não diferenças em relação à taxa de acerto do algoritmo treinado e testado com dados normalizados. A normalização impactou? De que forma: melhorando ou piorando o desempenho? Esta tendência foi observada para todos os valores de *k*?

Entregáveis:

- Código com a implementação do algoritmo. Pode ser em formato "notebook", se o aluno preferir.
- Relatório (em **pdf**) devidamente identificado, com a apresentação dos resultados para os itens 2 e 3 acima. A apresentação pode ser feita por meio de gráficos ou tabelas. O aluno deve interpretar e comentar os resultados, apontando os principais achados em relação a cada experimento.

Atenção: para esta atividade **não serão aceitas** soluções que aplicam implementações prontas do KNN de bibliotecas como sklearn (Python), caret (R), ou ferramentas como Weka, dentre outros.

O prazo final de entrega deste exercício é dia **07 de fevereiro às 23:59h**.