



MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL DO PIAUÍ  
CAMPUS SENADOR HELVÍDIO NUNES DE BARROS – CSHNB  
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO  
DISCIPLINA DE TÓPICOS EM COMPUTAÇÃO

**Atividade Prática IX**

A base *Pima Indians Diabetes* é um conjunto de dados médicos amplamente utilizado em estudos de aprendizado de máquina, voltado à predição da ocorrência de diabetes tipo II em mulheres indígenas Pima com idade igual ou superior a 21 anos. O conjunto contém 768 registros e 8 variáveis clínicas — como número de gestações, glicose, pressão arterial, espessura da pele, insulina, índice de massa corporal (IMC), histórico familiar de diabetes e idade — além do rótulo alvo que indica a presença ou ausência da doença. Todos os atributos são numéricos, embora alguns apresentem valores zero que representam dados ausentes, exigindo tratamento prévio. A base apresenta leve desbalanceamento, com aproximadamente 65% de amostras não diabéticas e 35% diabéticas. Nesse contexto, desenvolva modelos de classificação supervisionada utilizando os algoritmos Multilayer Perceptron (MLP) e Support Vector Machine (SVM), com a otimização de hiperparâmetros por meio das abordagens *Grid Search*, *Random Search*, Otimização *Bayesiana* e Algoritmos Genéticos. A seguir, são apresentados os passos para a execução.

### **1. Preparação dos Dados**

- Divisão: 70% treino / 30% teste.
- Use o StandardScaler.
- Tratamento de valores inválidos por substituição da mediana.

### **2. Modelos**

- SVM (SVC com probability=True)
- MLP (MLPClassifier)

### **3. Estratégias de Otimização (KFold cv=5)**

- *Grid Search* – GridSearchCV
- *Random Search* – RandomizedSearchCV (n\_iter=20)
- *Bayesiana* – Optuna (TPE) ou skopt (BayesSearchCV)
- Genética – TPOT (AutoML genético) ou sklearn-genetic (GeneticSearchCV)



MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL DO PIAUÍ  
CAMPUS SENADOR HELVÍDIO NUNES DE BARROS – CSHNB  
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO  
DISCIPLINA DE TÓPICOS EM COMPUTAÇÃO

#### **4. Espaços de Hiperparâmetros:**

##### **SVM**

- C: [1e-3, 1e3] (log-uniforme)
- kernel: ['linear', 'rbf']
- gamma: [1e-4, 1] (log-uniforme, se rbf)
- class\_weight: [None, 'balanced']

##### **MLP**

- hidden\_layer\_sizes: [(64,), (128,), (64,32), (128,64), (256,128)]
- activation: ['relu', 'tanh']
- solver: ['adam', 'lbfgs']
- alpha: [1e-6, 1e-2] (log-uniforme)
- learning\_rate\_init: [1e-4, 1e-2]
- max\_iter: [300, 600]
- early\_stopping: [True]

#### **5. Métricas de Avaliação**

- Acurácia, Precisão, Recall, F1, AUC ROC e Kappa.
- Relatar média  $\pm$  desvio-padrão no CV e desempenho final no conjunto de teste.

#### **6. Experimentos**

- Para cada modelo (SVM e MLP), otimize com as quatro estratégias.
- Registre o melhor score, combinação de hiperparâmetros e tempo de execução.

#### **7. Análise Comparativa**

- Qual estratégia obteve melhor equilíbrio entre desempenho e custo?
- Qual modelo se beneficiou mais da otimização?
- Há *overfitting* na busca?



MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL DO PIAUÍ  
CAMPUS SENADOR HELVÍDIO NUNES DE BARROS – CSHNB  
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO  
DISCIPLINA DE TÓPICOS EM COMPUTAÇÃO

- Discuta a estabilidade (desvio-padrão entre folds) e impacto de parâmetros como class\_weight (SVM) e early\_stopping (MLP).

#### **8. Escrita do Relatório**

- Elabore um relatório (utilizando o template da SBC) contendo:
  - a) Introdução, desenvolvimento e conclusão.
  - b) Descrição detalhada das etapas e métodos aplicados e achados.
- Submeta, via SIGAA, o relatório e o código-fonte com todo o processo descrito.