# **5ª Lista de Exercícios - Árvores de Decisão e SVM**

**Curso:** Ciência da Computação  
**Instituição:** Instituto Federal do Ceará, Campus Maracanaú  
**Professor:** Hericson Araújo (hericson.araujo@ifce.edu.br)  
**Disciplina:** Reconhecimento de Padrões

## **Instruções**

1. Salvo nos casos explicitamente indicados, os algoritmos e modelos devem ser implementados do início em qualquer linguagem de programação, como Python, Swift ou Octave.
2. É permitido o uso de pacotes auxiliares (ex.: sklearn, matplotlib) somente para manipulação de dados e criação de gráficos.
3. A entrega da solução deve ser feita via arquivo PDF ou Jupyter Notebook na plataforma Classroom. O código deve ser enviado junto.

## **Exercícios**

### **1. Conjunto de Dados: bostonbin.csv**

**Descrição:**O conjunto de dados possui 18 colunas:

* As 17 primeiras colunas são atributos para prever preços de casas em Boston na década de 1970.
* A última coluna é a saída, sendo **binarizada** em:
  + **0:** abaixo da média.
  + **1:** acima da média.

**Mais informações sobre os dados:** [OpenML Dataset 825](https://www.openml.org/d/825).

### **Tarefa (a): Avaliação de Modelos de Classificação Binária**

#### **Divisão dos Dados:**

* 70% para treinamento.
* 30% para teste.

#### **Abordagens:**

1. **SVM (Support Vector Machine):**
   * Kernel: **RBF**.
   * Ajuste de hiperparâmetros usando **Grid Search**:
     + Parâmetro C: valores 2^{-5},2^{−3},2^{−1},…,2^{11}, 2^{13}, 2^{15}.
     + Parâmetro gamma valores 2^{-15}, 2^{-13}, 2^{-11} … 2^{1}, 2^{3}
2. **Árvore de Decisão:**
   * Ajuste de hiperparâmetros usando **Grid Search**:
     + Altura máxima da árvore: valores de 1 a 10.
     + Número mínimo de amostras em um nó folha: valores de 1 a 5.
   * Critério de divisão: **Índice de Gini**.

### **Tarefa (b): Análise e Relatórios**

1. **Para cada modelo:**
   * Reportar os **hiperparâmetros selecionados**.
   * Calcular as seguintes métricas nos dados de teste:
     + **Acurácia.**
     + **Revocação.**
     + **Precisão.**
     + **F1-Score.**
2. **Visualizações:**
   * Plotar a curva **ROC**.
   * Plotar a curva **Precision-Recall**.

## **Observações Finais**

1. Use validação cruzada com **10 folds** dentro do Grid Search.
2. É permitido utilizar as seguintes implementações existentes:
   * **SVM.**
   * **Curva Precision-Recall.**
   * **KFold.**
   * **StandardScaler.**
   * **MinMaxScaler.**
   * **Curva ROC**