



## OBJETIVO

Este trabalho tem como objetivo aplicar os conceitos de *Machine Learning* (ML) vistos em sala de aula, utilizando uma base de dados real para desenvolver um projeto completo de ML. Os grupos deverão realizar todas as etapas do processo, desde a análise dos dados até a avaliação final do modelo. O trabalho será desenvolvido em grupos de até cinco pessoas. Além disso, os grupos deverão entregar um relatório técnico explicando as decisões de design e a abordagem utilizada.

## INSTRUÇÕES

### 1. Base de Dados:

A base de dados é o conjunto *Heart Failure Prediction Dataset*. Disponível na plataforma Kaggle, no seguinte link:

- <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction/data>

Esse conjunto de dados contém informações clínicas de pacientes, com o objetivo de prever a ocorrência de insuficiência cardíaca. As variáveis incluem idade, sexo, presença de doenças como diabetes e hipertensão, nível de creatinina, e outras métricas relevantes obtidas durante o acompanhamento hospitalar.

### 2. Etapas do Projeto:

- Análise exploratória dos dados (EDA): Identificação de padrões, correlações, outliers, distribuição das variáveis etc.
- Pré-processamento: Limpeza dos dados, tratamento de valores ausentes, codificação de variáveis categóricas, normalização/padronização se necessário.
- Divisão dos dados: Separação em conjunto usando o *k fold cross validation*
- Escolha e implementação de algoritmo de Machine Learning: O grupo é livre para escolher o algoritmo
- Treinamento e ajuste do modelo: Pode-se utilizar validação cruzada, tuning de hiperparâmetros, etc.
- Avaliação do modelo: Utilização de métricas, tais como acurácia, precisão, recall, F1-score, matriz de confusão etc.

- Interpretação dos resultados e conclusão: Discussão sobre o desempenho do modelo, limitações e possíveis melhorias.

## ENTREGA

1. Código-fonte do projeto (Colab ou Github).
2. Relatório técnico, contendo:
  - a. Descrição da base de dados
  - b. Etapas de pré-processamento
  - c. Escolha do algoritmo e justificativas
  - d. Resultados e métricas de avaliação
  - e. Conclusões e considerações finais
3. Apresentação oral:
  - a. Cada grupo terá até 15min para apresentar
  - b. Uso de slides é recomendado
  - c. Todos os membros devem participar

## AVALIAÇÃO

O trabalho será avaliado com base em três componentes principais: o desempenho do modelo no conjunto de dados (40%), a qualidade do relatório técnico (20%) e a apresentação do trabalho (40%). Caso o discente não compareça no dia da apresentação, as notas correspondentes as etapas anteriores serão zeradas.

É válido destacar que apesar do trabalho ser desenvolvido em Grupo, a nota será individual!

## PRAZOS

- Data de entrega: 15/06/2025 – até às 18h:00min
- Deverá ser entregue para o e-mail [geraldo.rocha@uesb.edu.br](mailto:geraldo.rocha@uesb.edu.br) com o assunto “Aprendizado de Máquina”

**Obs:** Caso não sejam entregues até as 18h00min, haverá um desconto de 1 ponto a cada duas horas de atraso.