

Trabalho Final de Inteligência Artificial 2025.2

A atividade prática será realizada em grupos de 4 alunos, cada grupo deve escolher um algoritmo de aprendizado de máquina para criar um modelo preditivo de doença cardíaca usando o *Heart Disease Health Indicators Dataset*. Os grupos realizarão um seminário sobre o algoritmo em 08/01/2026 e entregará um artigo de até 6 páginas no formato da SBC relatando os resultados da implementação.

*Cada grupo deve indicar qual algoritmo irá trabalhar na planilha a seguir:
[Divisão Grupos](#)

Organização geral da atividade

- Grupos de 4 alunos selecionam um classificador (ex.: Regressão Logística, SVM, Random Forest, Gradient Boosting, Redes Neurais).
- Utilizar o dataset para treinar o modelo e predizer as classes de 1000 exemplos na planilha de validação fornecida.
- Seminário foca no algoritmo; artigo com ênfase nos resultados.

Seminário (08/01/2026)

Apresentação de 15-20 minutos cobrindo:

- História: Origem, criadores e evolução do algoritmo.
- Funcionamento: Mecânica de aprendizado.
- Parâmetros principais: Hiperparâmetros chave e suas influências.
- Exemplos práticos: Aplicações reais onde o algoritmo é utilizado.
- Curiosidades: Fatos, limitações e variações do algoritmo.

Artigo no formato SBC (até 6 páginas) (15/01/2026)

Seguindo o template SBC (disponível em sbc.org.br), com seções padrão para artigos curtos de conferências:

- Resumo e Palavras-chave.
- Introdução: Problema de doença cardíaca, dataset (22 features + target), objetivos e algoritmo escolhido.
- Metodologia:
 - Pré-processamento (codificação, normalização, divisão treino/teste 70/30 ou k-fold).
 - Treinamento e parametrização (grid search, métricas: acurácia, F1, AUC-ROC).

- Aplicação na planilha de validação (previsões para 1000 instâncias).
- Resultados: Matriz de confusão, métricas no teste, importância de features, curva ROC. Incluir tabela comparativa se aplicável.
- Discussão e Conclusão: Análise, comparação com literatura e sugestões futuras.
- Referências

Dicionário de Dados:

Variável alvo

- HeartDisease (HeartDiseaseorAttack): indica se o respondente já foi informado de que teve doença coronariana (CHD) ou infarto do miocárdio.

Indicadores clínicos e de doenças

- BMI: índice de massa corporal (contínuo ou categorizado).
- Stroke: se já foi informado de que teve AVC.
- Diabetic: se já foi informado de que tem diabetes.
- Asthma: se já foi informado de que tem asma.
- KidneyDisease: se já foi informado de doença renal (exclui pedra, infecção de bexiga ou incontinência).
- SkinCancer: se já foi informado de câncer de pele.

Indicadores de estilo de vida e saúde geral

- Smoking: já fumou pelo menos 100 cigarros na vida (sim/não).
- AlcoholDrinking: bebedor pesado (homem >14 doses/semana, mulher >7).
- PhysicalActivity: reportou atividade física fora do trabalho nos últimos 30 dias.
- GenHealth: autoavaliação da saúde (Excelente, Muito boa, Boa, Regular ou Ruim).
- PhysicalHealth: dias de saúde física ruim nos últimos 30 dias.
- MentalHealth: dias de saúde mental ruim nos últimos 30 dias.
- DiffWalking: dificuldade séria para caminhar ou subir escadas.
- SleepTime: horas médias de sono em 24 horas.

Variáveis demográficas

- Sex: sexo do respondente (masculino/feminino).
- AgeCategory: categoria de idade (14 faixas, por exemplo 18–24, 25–29, ..., 80+).
- Race: raça/etnia imputada.