

## Trabalho Aprendizado de Máquina (em Duplas)

O objetivo deste trabalho é aplicar técnicas de aprendizado de máquina para construir e avaliar modelos de classificação de doenças renais crônicas (CKD - Chronic Kidney Disease). O trabalho envolve investigar e construir modelos de classificação distintos para os seguintes rótulos de um dataset de CKD: CKD\_Stage (problema de classificação multiclasse) e CKD\_Progression (problema de classificação binária). O conjunto de dados contendo registros clínicos e laboratoriais de pacientes de CKD está sendo enviado em anexo. O dataset possui as seguintes features:

"Sex": "Gênero do paciente (masculino ou feminino).", "Age": "Idade do paciente em anos.", "Systolic\_Pressure": "Pressão arterial sistólica (mm Hg).", "BMI": "Índice de Massa Corporal (IMC).", "CKD\_Cause": "Causa da DRC.", "Hemoglobin": "Hemoglobina no sangue (g/dL).", "Albumin": "Albumina sérica (g/dL).", "Creatinine": "Creatinina sérica (mg/dL).", "eGFR": "Taxa de filtração glomerular (mL/min/1,73m²).", "CKD\_Stage": "Estágio atual da DRC.", "CKD\_Risk": "Categoria de risco da DRC.", "Dipstick\_Proteinuria": "Resultado para proteinúria.", "Proteinuria": "Proteinúria (sim ou não).", "Occult\_Blood\_in\_Urine": "Sangue oculto na urina.", "Protein\_Creatinine\_Ratio": "Relação proteína/creatinina.", "UPCR\_Severity": "Gravidade da UPCR.", "Hypertension": "Hipertensão (sim ou não).", "Previous\_CVD": "Histórico de CVD (sim ou não).", "Diabetes": "Diabetes (sim ou não).", "RAAS\_Inhibitor": "Uso de inibidores da RAA.", "Calcium\_Channel\_Blocker": "Uso de bloqueadores de cálcio.", "Diuretics": "Uso de diuréticos.", "CKD\_Progression": "Progressão da DRC (sim ou não)."

O dataset está previamente tratado, mas vocês deverão realizar as etapas de pré-processamento necessárias à modelagem.

Tarefas a serem realizadas:

- **Análise Exploratória de Dados:** Realizar análise estatística dos atributos (médias, desvios, valores faltantes, etc.). Gerar visualizações (gráficos de dispersão, histogramas, correlações) que ajudem na compreensão dos dados.
- **Pré-processamento:** Tratar valores faltantes. Normalizar ou padronizar atributos quando necessário. Separar o conjunto de dados (sugestão: 70% treino, 15% teste, 15% validação).
- **Implementar três modelos de classificação:** Árvores de Decisão, Redes Neurais Artificiais (MLP), e outra técnica de aprendizado de máquina escolhida pelo grupo do trabalho (entre as técnicas estudadas na disciplina).
- **Avaliação dos Modelos:** Avaliar e comparar o desempenho dos modelos com as seguintes métricas: Acurácia, Precisão, Revocação (Recall), F1-Score, Matriz de Confusão e Área sob a curva ROC (AUC), se aplicável. Comentar sobre os casos de falso positivo e falso negativo, e suas implicações na área da saúde.
- **Discussão e Conclusões:** Comparar os resultados obtidos entre os modelos. Discutir qual técnica apresentou melhor desempenho e por quê. Apontar limitações do estudo e possíveis melhorias.

O trabalho deve ser entregue por email para o prof. em formato de relatório (PDF), contendo: descrição das etapas realizadas, tabelas e gráficos com os resultados, e discussão dos resultados e conclusão. Em anexo, enviar o código-fonte comentado. Identificar todas as submissões e anexos com o nome dos alunos.

Data entrega: 07/07/2025, até 23.59

Luís Alvaro