



Aprendizagem Automática

Licenciatura em Engenharia Informática

Trabalho Prático 2025/2026

-- Diagnóstico de tumores cerebrais --

Objetivo

Prever o diagnóstico de tumores cerebrais a partir das características demográficas do paciente e de textura das imagens ADC (Apparent Diffusion Coefficient). ADC é uma técnica de imagem por ressonância magnética (MRI).

Descrição do trabalho

Utilizando um desafio da plataforma kaggle (<https://www.kaggle.com>), submeta modelos que prevêem o diagnóstico relativo a um tumor cerebral de um paciente. O desafio está disponível em <https://www.kaggle.com/t/8cc7a8457fc24640a1e34293c44d155a>. Lá encontra informação sobre o problema, o processo de submissão, a medida de desempenho utilizada e o prazo para a avaliação final.

Implementação

Sugere-se a utilização da linguagem Python com o scikit-learn e bibliotecas consideradas adequadas para manipulação (pandas, numpy, ...) e visualização de dados (pyplot, matplotlib, seaborn, ...). O trabalho deve ser desenvolvido como um **notebook Jupyter** (<https://jupyter.org/>) que será submetido no moodle.

Para além de outra informação que achar relevante, o notebook deve incluir a seguinte informação:

1. Caracterização e análise do conjunto de dados;
2. Experiências realizadas para pesquisa dos melhores modelos (atributos utilizados e suas transformações, algoritmos e parâmetros testados, estratégia de avaliação, etc);
3. Descrição dos modelos submetidos na plataforma kaggle (subconjunto dos modelos criados);
4. Caracterização dos 3 modelos submetidos com melhor desempenho na (parte pública do conjunto de teste); desempenho de cada um dos modelos na parte pública e privada do conj. de teste);
5. Discussão e conclusões

O notebook é **muito mais que o código** desenvolvido para construção e submissão dos modelos. O **notebook funciona como relatório**: isto significa que toda a informação que permite perceber o trabalho realizado deve estar lá presente (desde a análise dos atributos, seleção e transformações dos mesmos, estratégia para pesquisa e escolha dos modelos, setup dos modelos usados para as previsões submetidas na plataforma e a discussão dos resultados intermédios e finais, entre outros).

Condições gerais

O trabalho deverá ser efetuado em grupos de 3 alunos e será apresentado em dia e horário a combinar. O notebook (em ambos os formatos .ipynb e respetivo .pdf com inclusão de resultados da execução do notebook) deve ser submetido no moodle através de um ficheiro .tar.gz ou .zip.

Cada aluno deve criar uma conta no kaggle e formar a equipa na plataforma. O nome da equipa deve conter entre parêntesis os números dos elementos do grupo.

O notebook deve ser submetido **no dia seguinte ao fecho do desafio**.