



Aprendizagem Automática

Mestrado em Engenharia Informática

Trabalho Prático 2024/2025

-- Classificação de florestas --

Objetivo

Prever o tipo de floresta a partir de variáveis cartográficas.

Descrição do trabalho

Utilizando um desafio na plataforma kaggle (www.kaggle.com), submeta modelos que classificam florestas, com base em informação cartográfica. O desafio está disponível em <https://www.kaggle.com/t/440cc105d3214fdf8e73961179081527>. Lá encontra informação sobre o problema, o processo de submissão, a medida de desempenho 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:

- Caracterização do conjunto de dados;
- Experiências realizadas para pesquisa dos melhores modelos (atributos utilizados e suas transformações, algoritmos e parâmetros testados, estratégia de avaliação, etc);
- Escolha dos modelos submetidos na plataforma kaggle (subconjunto dos modelos criados);
- Caracterização dos dois 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);
- Discussão e conclusões

Condições gerais

O trabalho deverá ser efetuado em grupos de 2 alunos e será apresentado em dia e horário a combinar. O notebook (em ambos os formatos .ipnb e .pdf) deve ser submetido no moodle através de um ficheiro .zip.

O notebook deve ser submetido até dia **2 dias após a conclusão do desafio**.