



TRABALHO 3 STUDENT PERFORMANCE DATASET

1 Descrição do Dataset

Neste trabalho, você irá classificar o rendimento escolar de alunos do ensino fundamental de dois colégios (*Gabriel Pereira* e *Mousinho da Silveira*) em um ano letivo. Para isso, iremos utilizar um dataset com 30 anotações (algumas numéricas e outras categóricas), a respeito do aluno e de sua família. Alguns exemplos destas anotações são:

- | | |
|---------------------------|--|
| – Sexo | – Profissão da mãe |
| – Idade | – Profissão do pai |
| – Endereço | – Duração do trajeto casa-escola |
| – Tamanho da família | – Tempo de estudo semanal |
| – Se os pais moram juntos | – Número de reprovações passadas |
| – Escolaridade da mãe | – Se faz reforço escolar |
| – Escolaridade do pai | – Se pratica atividades extra-curriculares |

O objetivo será classificar, a partir destas informações, se o aluno foi aprovado naquele ano letivo (atributo “*approved*” do dataset). Para uma explicação de cada feature e os valores que elas podem assumir, veja o arquivo “*dataset_info.txt*”.

2 Tarefas

Pedimos que você:

1. Carregue os dados e os inspecione.
2. Como baseline, treine uma árvore de decisão para o problema.
3. Treine também florestas aleatórias variando o número de árvores geradas.
4. Calcule a matrix de confusão e a acurácia normalizada para o conjunto de validação/teste e compare os modelos treinados.

3 Arquivos

Os arquivos disponíveis no Moodle são:

- *student_performance_train.data*: conjunto de dados para treinamento;
- *student_performance_val.data*: conjunto de dados para validação;
- *student_performance_test.data* (**apenas durante a avaliação**): conjunto de dados retido pelo professor a ser utilizado na avaliação;
- *dataset_info.txt*: descrição de cada feature e os valores que elas podem assumir;

4 Avaliação

O dataset foi previamente dividido aleatoriamente em três conjuntos — treino, validação e teste — e apenas os dois primeiros serão disponibilizados para os alunos implementarem suas soluções.

No dia da avaliação, nos 40 minutos finais da aula, o professor irá disponibilizar o conjunto de teste e validar a solução de cada aluno (ou dupla de alunos). A avaliação consiste da análise dos resultados reportados sobre o conjunto de teste e possíveis perguntas a respeito das decisões tomadas pelo aluno durante a implementação de sua solução.

Observações sobre a avaliação:

- **NÃO** haverá submissão do código ou relatório do trabalho no Moodle;
- O trabalho poderá ser feito em duplas, podendo haver repetição das duplas a cada trabalho;
- Pelo menos um membro da dupla deverá estar presente no momento da avaliação;
- As notas do trabalho serão divulgadas em até uma semana após a avaliação;