

## CARROS USADOS

Desenvolva um modelo de Machine Learning, suportado no algoritmo Random Forest, capaz de prever os preços de carros usados, com base nas suas características. O objetivo é dispormos de uma ferramenta que nos auxilie na tomada de decisão quando andarmos à procura de um carro usado. O conjunto de dados fornecido, para indução do modelo, contém informações sobre vários carros usados, incluindo características como o ano de registo, quilometragem, o tipo de viatura, marca, potência do motor, tipo de combustível, entre outras. Tente ainda perceber quais as características que mais influenciam o preço das viaturas.

### Tarefas a realizar

- Implementar num documento Jupyter Notebook a solução para o problema enunciado, desenvolvendo, com os dados disponibilizados no ficheiro `dataset.csv`, um modelo de ML de elevado desempenho, que cumpra integralmente as especificações descritas. [gerar solucao.ipynb]
- Com o modelo desenvolvido, estimar os preços do conjunto de viaturas com as características indicadas no ficheiro `just_features.csv`, para as quais não é dado a conhecer o preço real. [gerar precos\_estimados.csv]
- Elaborar uma breve apresentação do trabalho desenvolvido. [gerar apresentacao.pdf]

### Considerações a ter em conta na implementação

- Para o desenvolvimento do modelo de ML solicitado deve ser usado o Scikit-Learn e outros *packages* de suporte do Python.
- Os ficheiros `dataset.csv` e `just_features.csv`, contendo, respetivamente, o *dataset* para indução do modelo (com 25.000 instâncias – cada instância contém as características duma viatura usada, incluindo o preço pelo qual foi vendida) e o *dataset* que o irá pôr à prova (com 22.222 instâncias sem a indicação do preço), podem ser descarregados da área da unidade curricular da plataforma `ipb.virtual`, em Recursos/avaliacao/trabPratico.
- A apresentação, depois de elaborada no MS PowerPoint ou noutra aplicação adequada, à escolha, deve ser convertida para pdf, conter entre 3 e 7 diapositivos (sem contar com o 1º), e o tamanho da fonte do corpo do texto situar-se entre os 16 e 20 pt. Não esquecer de incluir no 1º diapositivo o nome e o número de cada elemento que integra o grupo de trabalho, e evitar incluir na apresentação código de implementação.
- O ficheiro a submeter `precos_estimados.csv` deve conter apenas as colunas `id` e `preco`, devendo, por isso, apresentar o seguinte aspeto:  

```
id, preco
23346, 1520.2
12344, 10345.3
12112, 5034.5
...
```
- Na avaliação do trabalho vai ser tido em conta, nomeadamente, o acerto, de acordo com a métrica  $R^2$ , das estimativas contidas no ficheiro `precos_estimados.csv`, geradas pelo modelo para as viaturas com preços não divulgados.

### Considerações gerais

- Este trabalho prático deverá ser realizado por grupos de 2 ou 3 alunos e tem um carácter obrigatório para aprovação à unidade curricular (trabalhos individuais terão uma penalização de 1 valor).  
(Sugestão de cooperação nas tarefas a realizar: cada um dos elementos do grupo pode começar por desenvolver o seu próprio modelo de regressão; depois, em conjunto, preparam um único modelo para submissão, que aproveite as melhores ideias e opções consideradas nas duas propostas.)

- É expressamente proibida a cópia integral ou parcial de código de outras fontes que não a documentação disponibilizada pelos docentes da unidade curricular.
- O trabalho deverá ser entregue apenas por um dos elementos do grupo, dentro do prazo estabelecido, obrigatoriamente no portal de e-learning (em <http://virtual.ipb.pt/>, escolher <Trabalho Pratico> no separador <Atividades>, dentro da área de IA), e em nenhuma situação poderá ser remetido por e-mail.
- Deverão ser submetidos, em anexos separados e não compactados, os 3 ficheiros solicitados (solucao.ipynb, precos\_estimados.csv e apresentacao.pdf). O não cumprimento desta regra implicará uma penalização de 1 valor.
- O trabalho apenas poderá ser submetido com um atraso máximo de 5 dias, sendo subtraído, ainda assim, um valor à sua nota por cada dia de atraso.
- Não serão permitidas resubmissões (quando submeter, certifique-se de que se trata da versão final).
- Os alunos poderão ter de vir a defender os seus trabalhos.