

PACD1 – Projeto Aplicado para Ciência de Dados I

Diana Mendes (MTCD)
Sérgio Moro (MTCD)

E-mails:

diana.mendes@iscte-iul.pt

smcmo@iscte-iul.pt

1

Tema do Projeto

Tema: Previsão do n.º de sets para conclusão de um jogo de ténis profissional (ranking ATP)

- Conjunto de dados sobre o ranking ATP (Ténis) já usado nas UCs de FGBD e ABD
- A fonte de dados para extração foi o site do ATP (<https://www.atptour.com/>)
- Os docentes atribuem a cada grupo de alunos um conjunto de países onde se realizaram jogos registados do ATP

Todos os grupos têm de utilizar pelo menos as seguintes variáveis para todos os jogos realizados nos países atribuídos:

Dados sobre ambos os jogadores (player & opponent): (1) altura, (2) país de naturalidade, (3) se é destro ou esquerdino, (4) data de nascimento (portanto, uma variável por cada um dos 2 jogadores – total de 8 variáveis)

Dados sobre o jogo: (5) qual o torneio, (5) data, (6) tipo de piso, (7) prémio monetário, (8) round, e (9) quantos sets para a tomada de decisão da partida

Os alunos têm de desenvolver um modelo que permita estimar o n.º de sets que decorreram no jogo.

2

Entregáveis

- Relatório (em formato pdf)
- Código
 - tem de poder ser executado AS-IS, diretamente (atenção a paths, nomes ficheiros, etc.)
 - o ZIP fornecido tem de ter tudo o necessário (código) e ser de execução direta;
 - o output final tem de ficar visível após a execução (e.g., gráficos, métricas, etc.)
- Entrega por email para os 2 docentes, com Cc para todos os elementos do grupo

3

Relatório

- 30 a 40 pg. (indicativo)
- Deve descrever todas as fases do **CRISP-DM**:
 - Compreensão do problema
 - Compreensão dos dados (análise de todas as variáveis e das instâncias)
 - Preparação dos dados (detalhar toda a preparação de dados)
 - Modelação (justificar a escolha do modelo)
 - Avaliação (avaliação dos resultados)
 - Implementação (proposta de implementação)
- É necessário que sejam detalhadas as iterações desenvolvidas dentro do ciclo do CRISP-DM ao longo do projeto

4

Apresentação do Trabalho

Apresentação final e Discussão:

- Decorrerá em datas a propor pelos docentes, após o fim das aulas, durante a 1ª época
- **Os docentes é que decidirão que grupo é que faz em que dia, não havendo lugar a negociação**
- Duração máxima: 1h00 hora
- O docente colocará perguntas individualmente e que serão decisivas para a nota atribuída **a cada aluno**
- Todos os elementos do grupo têm de participar na apresentação e responder às questões colocadas

Outras Regras

- **Código:**
 - Deverá estar devidamente comentado
 - Deverá incluir todas as experiências feitas, incluindo a análise e preparação de dados, modelação, e avaliação do modelo
 - Deverá ter identificadas as fases do CRISP-DM percorridas
- **Deverá ser claramente identificado benefício para o negócio decorrente do projeto**

Nota: outras regras poderão vir a ser identificadas posteriormente