



Sistemas de Informação II - Trabalhos Práticos

Cotação Total: 10 valores

-
1. Os grupos de trabalho são de 2 alunos
 2. A implementação é em SQL Server BIDS (ETL, OLAP e Data-Mining)
 3. O trabalho é apresentado numa curta sessão Zoom a decorrer antes do exame da 1ª chamada em data a agendar
-

Trabalho 1 - Data-warehouses e OLAP (5 Valores)

O ficheiro Chinook.bak contém um backup de uma base de dados OLTP de uma empresa fictícia que comercializa música pela Internet.

- a) Implemente uma datawarehouse destinada a suportar uma análise de vendas:

Métricas: número de pistas, sua duração e valor de venda
Dimensões: semana, mês, ano, cidade, país, artista e álbum

- b) Crie um projeto ETL e carregue a datawarehouse

- c) Crie um projeto Analysis Services e o respectivo servidor OLAP (cubo)

- d) Defina um KPI para vendas por mês

- e) Defina algumas pivot tables e gráficos num ficheiro Excel com ligação ao serviço OLAP

- f) No PowerBIDesktop implemente um dashboard que inclua pelo menos um gráfico de vendas com geolocalização (mapa mostrando o total de vendas em cada cidade); uma pivot table; um filtro; um gráfico de barras; um gráfico em queijo

- g) Crie um projeto Report e desenhe um relatório de vendas em quantidade e valor, por cidade, com sub-totais por país.
-

Trabalho 2 - Data-Mining (5 Valores)

Os ficheiros flag.data e flag.names contêm respetivamente dados e a sua descrição relativos às bandeiras dos países do mundo.

1. Usando os classificadores disponíveis no Visual Studio, implemente 4 tarefas de classificação em que, a partir das cores e características das bandeiras (atributos 9 a 30) os alvos são:

- a) O continente a que o país pertence (atributo 2)
- b) O quadrante geográfico (atributo 3)
- c) A língua (atributo 6)
- d) A religião (atributo 7)

2. Utilize técnicas de clustering e interprete os resultados obtidos

3. Usando regras de associação, determine quais as cores que aparecem mais frequentemente em conjunto

4. Elabore um relatório do trabalho onde apresente clara e resumidamente as conclusões de cada estudo, em particular: eficácia (accuracy), medida F, lift charts, influência dos fatores discriminantes do alvo, interpretação dos clusters, itens que surgem em conjunto e importância das regras obtidas.

NOTA IMPORTANTE: para o modelo Decision Tree, de modo a conseguir a criação de ramos na árvore (split) vai ter de alterar os parâmetros do algoritmo: Minimum Support (inferior a 10, por exemplo 2 ou 3 ou mais) e Score Method = 1
