











# **Data Mining com Clementine**

### <u>Suporte passivo</u>

- o utilizador formula a hipótese inicial ou pesquisa de padrões
  - Permite que a <u>experiência</u> desempenhe um <u>papel</u> <u>fundamental</u> na análise
  - Interacção com o Clementine permite:
    - Manipular, visualizar, explorar os dados através de tabelas e gráficos
    - Efectuar estatísticas e teste de hipóteses para avaliar relacionamentos de que já se suspeita



#### Análise Inteligente de Dados



Instituto Superior Politécnico de VISEU Escola Superior de Tecnologia



5

# **Data Mining com Clementine**

# <u>Suporte activo</u>

- o utilizador <u>tem a iniciativa</u> no processo de Data Mining, mas o <u>computador desempenha um papel muito mais relevante</u> na análise
  - O computador identifica por si regras e relacionamentos nos dados
  - Utiliza técnicas de modelação e "machine learning":
    - Redes neuronais
    - Árvores de decisão
    - Regras de associação
    - Regressão linear e logística
    - Clustering





## **Data Mining com Clementine**

#### Aprendizagem Supervisionada

- Compreende: redes neuronais, árvores de decisão, regressão linear e logística e outras
- Envolvem a modelação de uma variável de saída baseada em várias variáveis de entrada
- Os modelos criados podem ser utilizados para predizer casos futuros, onde o alvo é desconhecido

### <u>Aprendizagem Não Supervisionada</u>

- Compreende técnicas de clustering (redes Kohonen, K.-means e Two Steps)
  - São métodos que permitem agrupar registos semelhantes
- Não é dado ao algoritmo um dado campo a predizer, mas apenas explorar os relacionamentos nos dados para descobrir a sua estrutura geral



Análise Inteligente de Dados



Instituto Superior Politécnico de VISEU Escola Superior de Tecnologia



7

# **Data Mining com Clementine**

#### <u>Validade da abordagem Data Mining</u>

- Antes de escolher a técnica apropriada para um dado caso:
  - Há que conhecer o problema do negócio a resolver
  - Avaliar os dados necessários

#### • Aspectos a considerar:

- Os dados estão disponíveis?
- Os dados cobrem os factores relevantes?
- Os dados têm muito "ruído"?
- Há dados suficientes?
- Está disponível o perito relativo a esses dados?







# **Data Mining com Clementine**

#### • Há dados suficientes?

- A resposta depende do problema em questão
- Muitas vezes não é o tamanho dos dados que dificulta o processo, mas a sua natureza representativa e cobertura de possíveis resultados
- Quanto maior a complexidade dos padrões ou relacionamentos a procurar, maior número de registos serão necessários
- Mas:
  - se os dados proporcionarem boa cobertura dos diversos resultados e não tiverem ruído,
  - resultados razoáveis serão de esperar com poucos milhares ou mesmo centenas de registos

































































#### Analisar a Distribuição dos Dados (Numéricos)





















# Para modificar valores de dados ou criar novos campos em função de outros

Permite criar um de seis tipos de novos campos:

| Any         | O novo campo é o resultado uma qualquer expressão CLEM   |
|-------------|--|
| Flag        | O campo resultante terá uma resposta True ou False (flag), reflectindo uma dada expressão                                  |
| Set         | O novo campo terá valores assignados a partir de membros de um dado conjunto   |
| Count*      | O novo campo é baseado no número de vezes que uma condição especificada for verdadeira                                     |
| State*      | O novo valor do campo representa um de dois estados. A comutação entre esses estados é ditada por condições especificadas. |
| Conditional | O novo campo é o resultado de uma de duas expressões, dependendo do valor de uma condição.                                 |
|             | * não tratados aqui  |

























