



CLASSIFICAÇÃO

1

Daricélio Moreira Soares

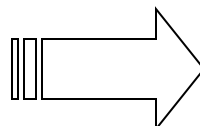
Classificação

Classificação: identificação da classe a qual um elemento pertence a partir de suas características. O conjunto de possíveis classes é discreto e predefinido.

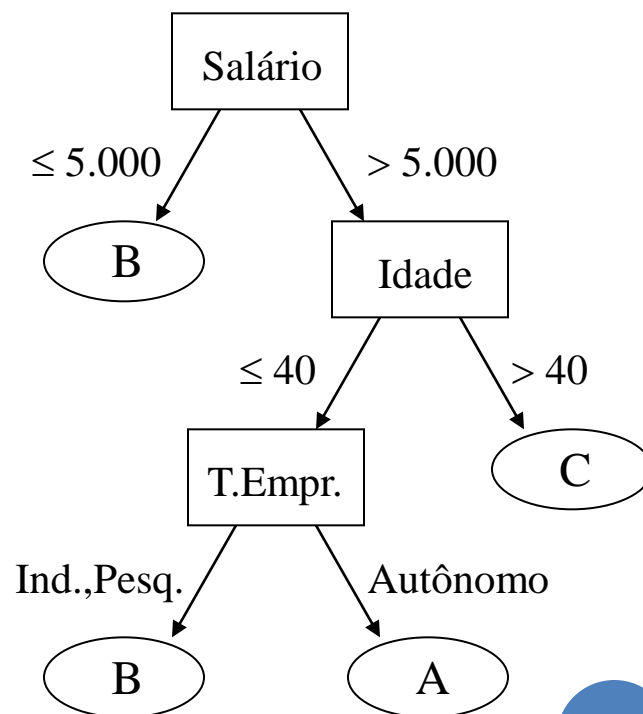
Exemplo: a partir das características de um candidato a um empréstimo bancário, prever se este candidato será um bom ou mau pagador.

Mineração de um Modelo de Classificação

Atributos Independentes				Atributo Dependente
ID	Salário	Idade	Tipo Emprego	Classe
1	3.000	30	Autônomo	B
2	4.000	35	Indústria	B
3	7.000	50	Pesquisa	C
4	6.000	45	Autônomo	C
5	7.000	30	Pesquisa	B
6	6.000	35	Indústria	B
7	6.000	35	Autônomo	A
8	7.000	30	Autônomo	A
9	4.000	45	Indústria	B

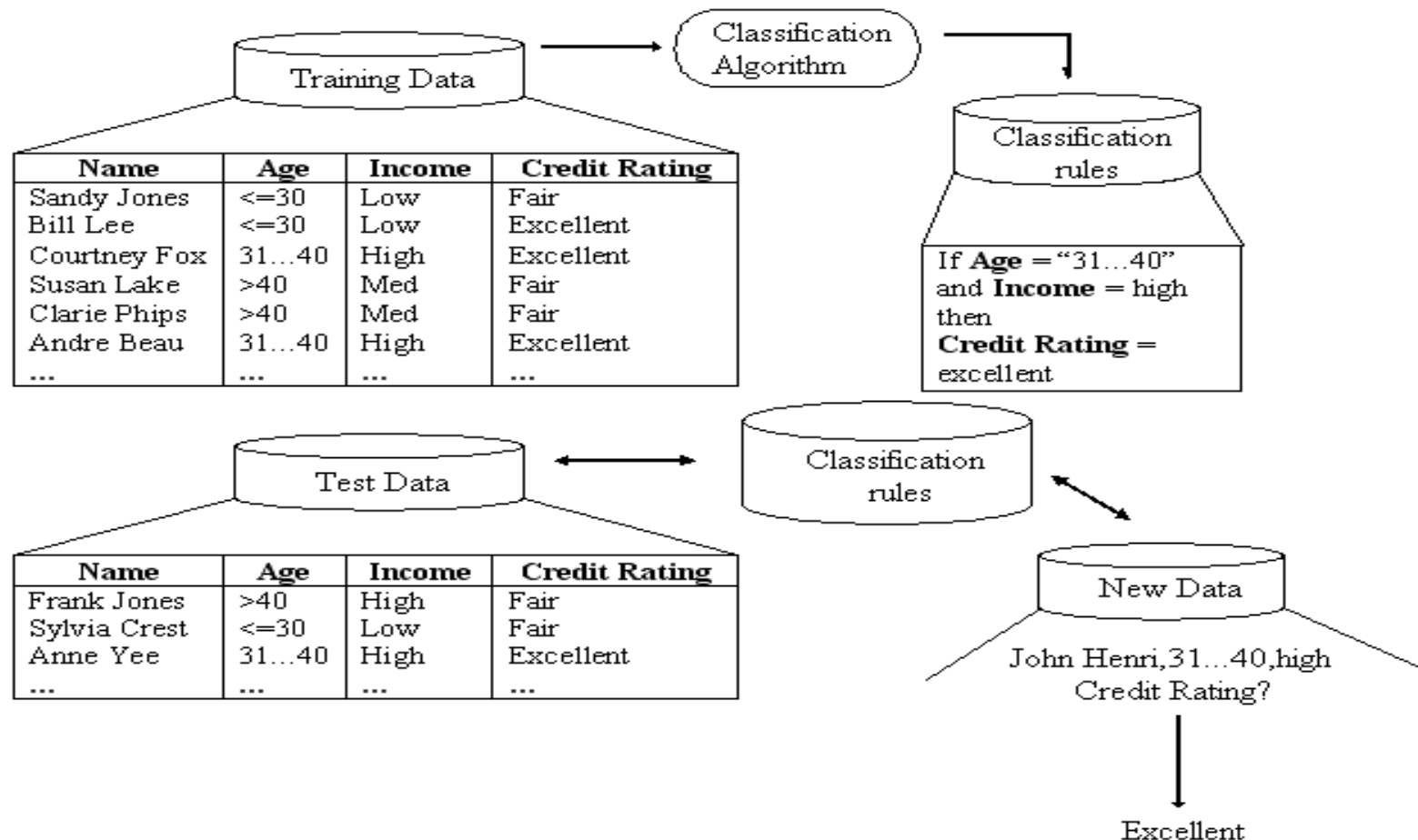


Árvore de Decisão ou
Árvore de Classificação



A partir de uma base de treinamento,
extraí-se o modelo de classificação
(árvore de decisão, p.e.).

Mineração, Avaliação e Utilização do Modelo

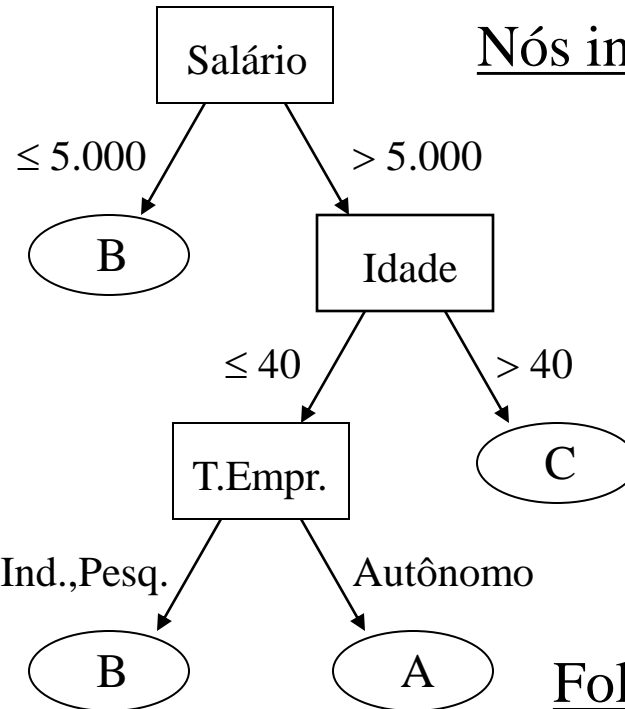


Árvores de Decisão

- ✓ A sua representação é intuitiva e torna o modelo de classificação resultante fácil de se utilizar e de ser entendido.
- ✓ A precisão de suas previsões, em geral, possui taxas de acertos competitivas em relação a de outros modelos.
- ✓ Algoritmos rápidos e escaláveis podem ser implementados para a construção de árvores de decisão, considerando-se grandes bases de treinamento.
- ✓ Outras técnicas: métodos Bayesianos, k-vizinhos mais próximos, classificação associativa, entre outros.

Árvores de Decisão

Nós internos: atributos



Arestas: predicados

Folhas: valores de classes

Nó interno + Aresta = condição

Regras de Classificação

Forma Geral:

$(\text{var}_1 \text{ satisfaz cond}_1) \text{ and } (\text{var}_2 \text{ satisfaz cond}_2) \text{ and } \dots$

$\dots \text{ and } (\text{var}_n \text{ satisfaz cond}_n) \Rightarrow \text{objeto pertence à classe } C_j$

Exemplo:

$(\text{Salário} > 5000) \wedge (\text{Idade} > 40) \Rightarrow \text{Classe} = C$

Árvores de Decisão \Rightarrow Regras de Classificação

- Cada caminho da raiz até a folha representa uma regra, definida como a conjunção das condições percorridas, implicando no valor da classe encontrada na folha em questão
- A árvore deve ser definida de forma que, para um mesmo registro, haja um e apenas um caminho da raiz até a folha.

Regras de classificação obtidas a partir da árvore de decisão:

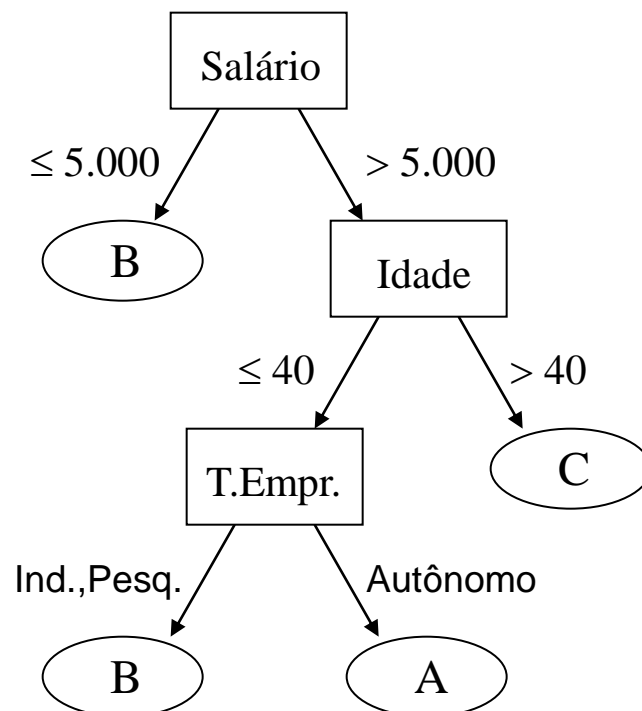
$(\text{Sal} \leq 5k) \Rightarrow \text{Classe} = B$

$(\text{Sal} > 5k) \wedge (\text{Idade} > 40) \Rightarrow \text{Classe} = C$

$(\text{Sal} > 5k) \wedge (\text{Idade} \leq 40) \wedge (\text{TEmpr} = \text{Autônomo}) \Rightarrow \text{Classe} = A$

$(\text{Sal} > 5k) \wedge (\text{Idade} \leq 40) \wedge ((\text{TEmpr} = \text{Indústria}) \vee (\text{TEmpr} = \text{Pesquisa})) \Rightarrow \text{Classe} = B$

Árvore de Decisão ou
Árvore de Classificação



Mineração de Árvores de Decisão

Algoritmo ID3

⇒ utilizado para construir árvores de decisão.

Entrada:

- ✓ base de treinamento que contém os registros,
- ✓ lista dos atributos independentes,
- ✓ definição do atributo dependente (alvo ou objetivo).

Saída:

- ✓ Árvore de decisão que permite classificar (definir o valor do atributo dependente) um novo registro a partir de seus atributos independentes.

Referência:

- Seção 6.3 do livro *Data Mining: Concepts and Techniques*
J.Han e M.Kamber, Morgan Kaufmann, segunda edição, 2006.