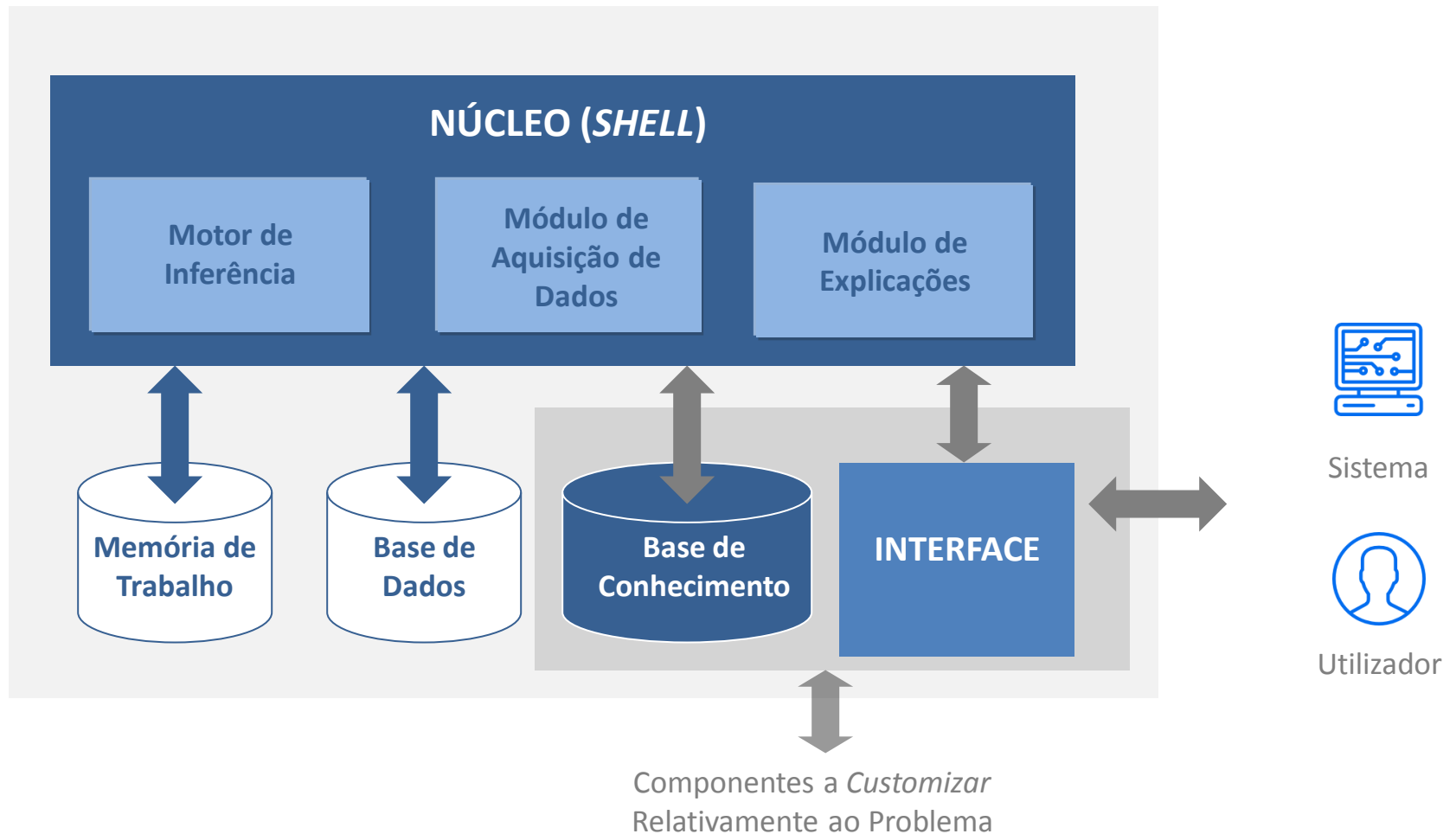


Aplicação de Técnicas de Aquisição de Conhecimento Baseadas em Aprendizagem Automática

Manuel Filipe Santos



Sistemas Baseados em Conhecimento



Aquisição e Extração do Conhecimento

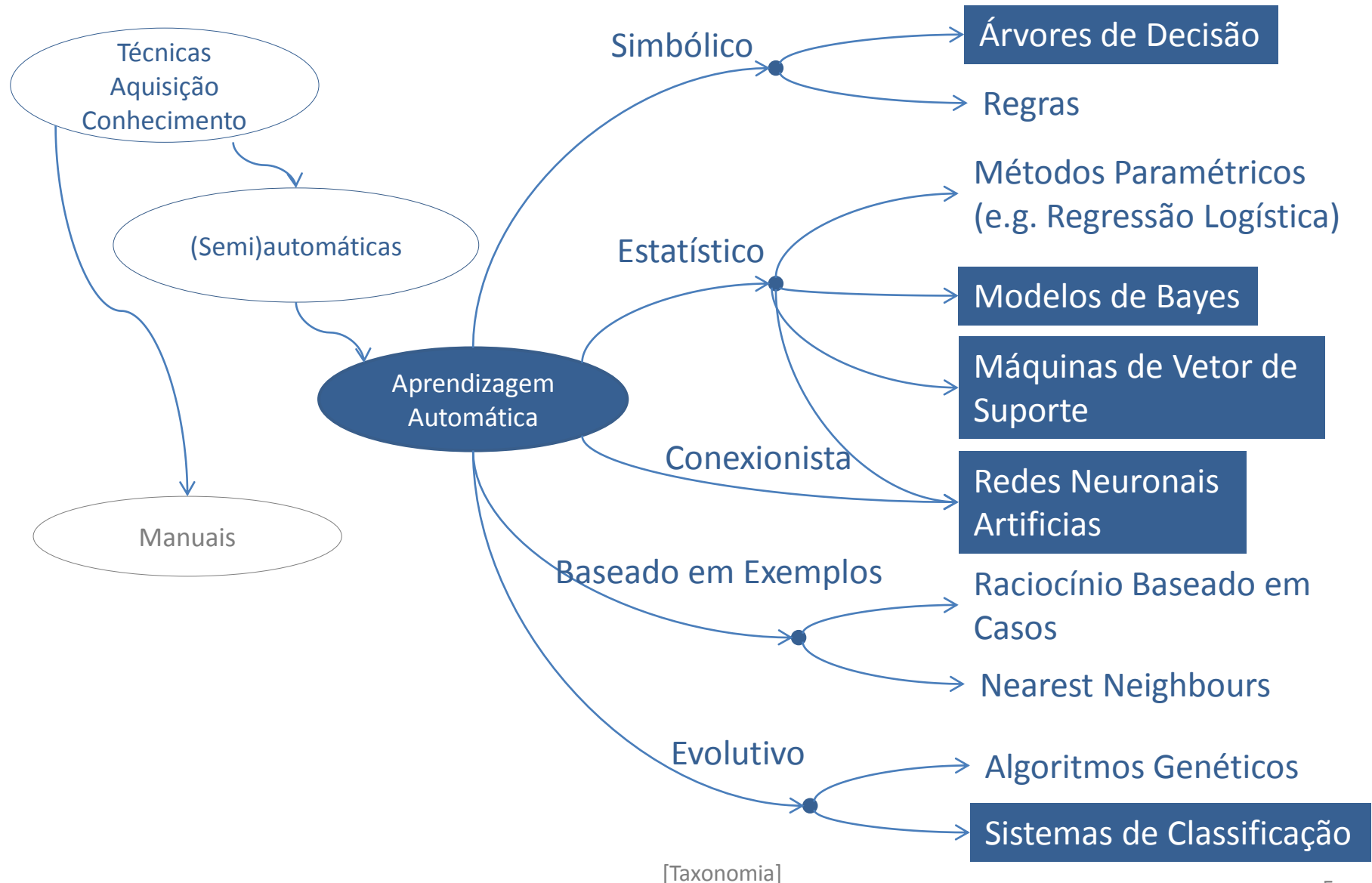
A problemática da aquisição ou extração do conhecimento assume uma importância vital na construção de qualquer SBC

Os métodos e aquisição de conhecimento (semi)automáticos são cada vez mais aplicados em detrimento dos manuais pois permitem, fazendo uso de técnicas de Aprendizagem Automática (AA) e com grande acuidade, descobrir conhecimento a partir de conjuntos de dados (e.g. documentos, bases de dados, data warehouses) que de outra forma seria praticamente impossível

Objetivos

- Técnicas de AA supervisionada aplicadas a problemas de classificação considerando vários tipos de dados
- Estudo comparativo do seu desempenho
- Discutir as vantagens e desvantagens
- Interpretar o conhecimento adquirido/descoberto

Aquisição e Extração do Conhecimento



Aquisição e Extração do Conhecimento



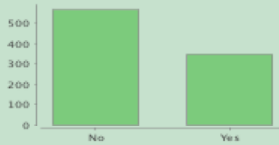
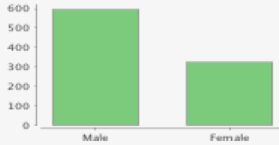
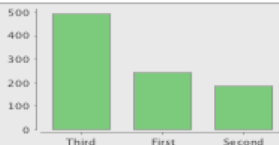
Aquisição e Extração do Conhecimento

ExampleSet (916 examples, 1 special attribute, 6 regular attributes)

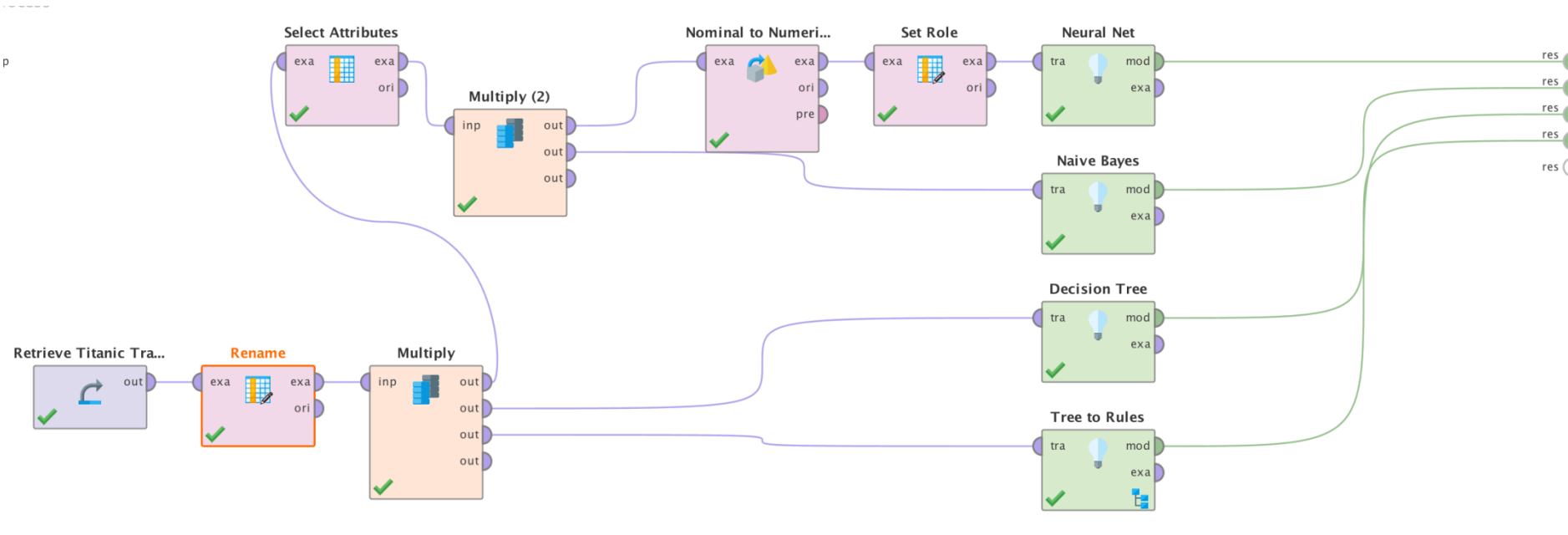
| Row No. | Sobreviveu | Idade | Classe | Sexo | Nº Irmãos/Cônjuges | Nº Pais/Filhos | Tarifa ↓ |
|---------|------------|-------|--------|--------|--------------------|----------------|----------|
| 39 | Yes | 36 | First | Male | 0 | 1 | 512.329 |
| 225 | Yes | 35 | First | Female | 0 | 0 | 512.329 |
| 85 | Yes | 24 | First | Female | 3 | 2 | 263 |
| 86 | Yes | 23 | First | Female | 3 | 2 | 263 |
| 87 | No | 19 | First | Male | 3 | 2 | 263 |
| 88 | No | 64 | First | Male | 1 | 4 | 263 |
| 89 | Yes | 60 | First | Female | 1 | 4 | 263 |
| 51 | Yes | 36 | First | Female | 0 | 0 | 262.375 |
| 188 | Yes | 13 | First | Male | 2 | 2 | 262.375 |
| 189 | Yes | 18 | First | Female | 2 | 2 | 262.375 |
| 190 | No | 61 | First | Male | 1 | 3 | 262.375 |
| 191 | Yes | 48 | First | Female | 1 | 3 | 262.375 |
| 12 | No | 24 | First | Male | 0 | 1 | 247.521 |
| 13 | Yes | 50 | First | Female | 0 | 1 | 247.521 |
| 8 | Yes | 18 | First | Female | 1 | 0 | 227.525 |
| 18 | Yes | 42 | First | Female | 0 | 0 | 227.525 |

[Conjunto de Dados de Treino]

Aquisição e Extração do Conhecimento

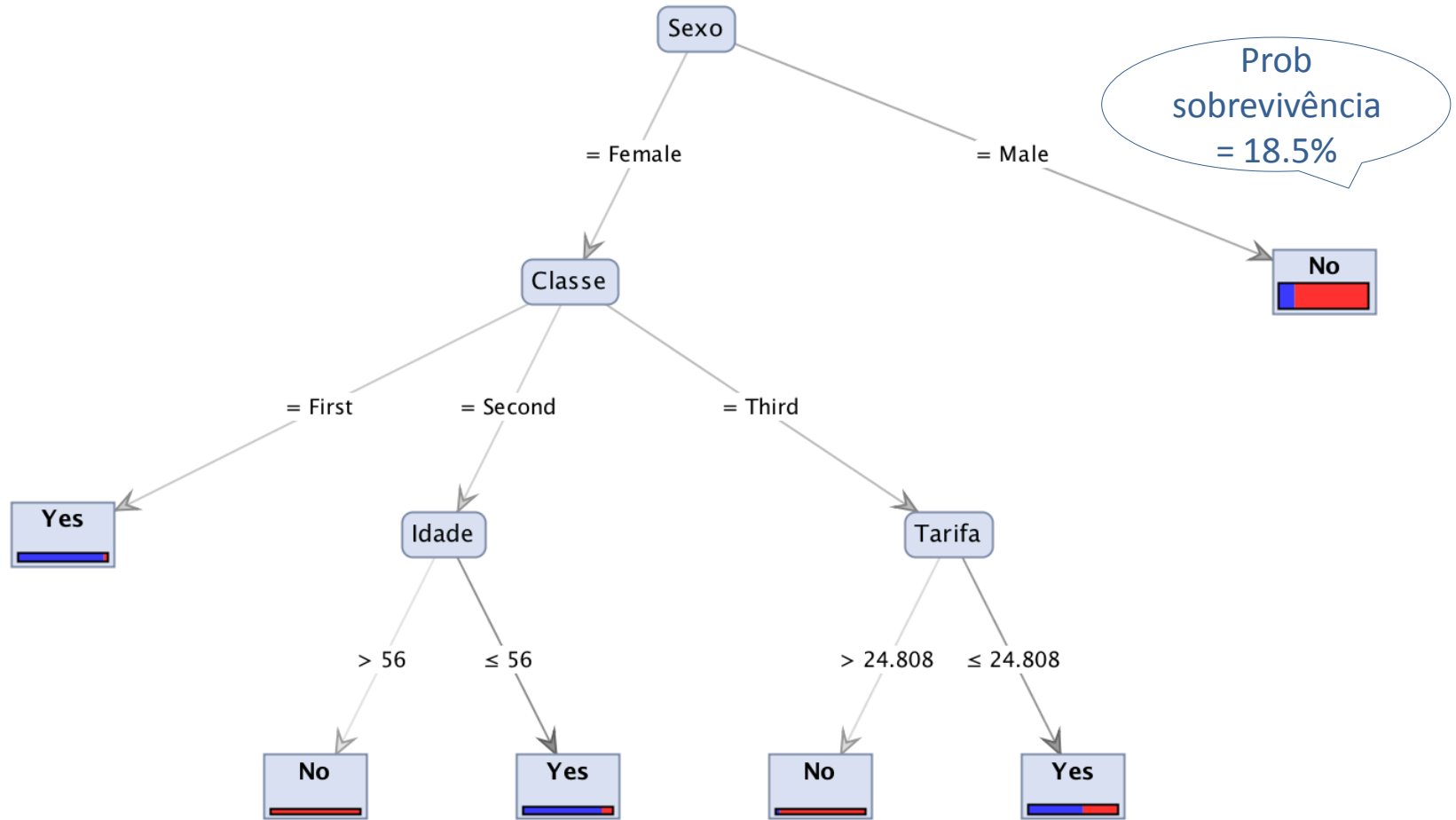
| Name | Type | Missing | Statistics | Filter (7 / 7 attributes): <input type="text" value="Search for Attribute."/> |
|--|------------|---------|--|---|
| <div>Label</div> <div> <div>▲</div> Sobreviveu </div> | Binominal | 0 | <div>  <p>Least: Yes (349) Most: No (567)</p> <p>Values: No (567), Yes (349) Details...</p> <p>Open chart</p> </div> | |
| <div>▲</div> Sexo | Binominal | 0 | <div>  <p>Least: Female (322) Most: Male (594)</p> <p>Values: Male (594), Female (322) Details...</p> <p>Open chart</p> </div> | |
| <div>▼</div> Nº Irmãos/C... | Integer | 0 | <div> <div>Min: 0 Max: 8 Average: 0.514</div> </div> | |
| <div>▼</div> Nº Pais/Filh... | Integer | 0 | <div> <div>Min: 0 Max: 9 Average: 0.386</div> </div> | |
| <div>▼</div> Tarifa | Numeric | 0 | <div> <div>Min: 0 Max: 512.329 Average: 33.564</div> </div> | |
| <div>▲</div> Classe | Polynomial | 0 | <div>  <p>Least: Second (184) Most: Third (491)</p> <p>Values: Third (491), First (241), Second (184) Details...</p> <p>Open chart</p> </div> | |
| <div>▼</div> Idade | Real | 0 | <div> <div>Min: 0.167 Max: 80 Average: 29.932</div> </div> | |

Aquisição e Extração do Conhecimento



Aquisição e Extração do Conhecimento

Titanic (modelo de conhecimento simbólico)



[Árvore de Decisão C4.5, Gini]

Aquisição e Extração do Conhecimento

Titanic (modelo de conhecimento simbólico)

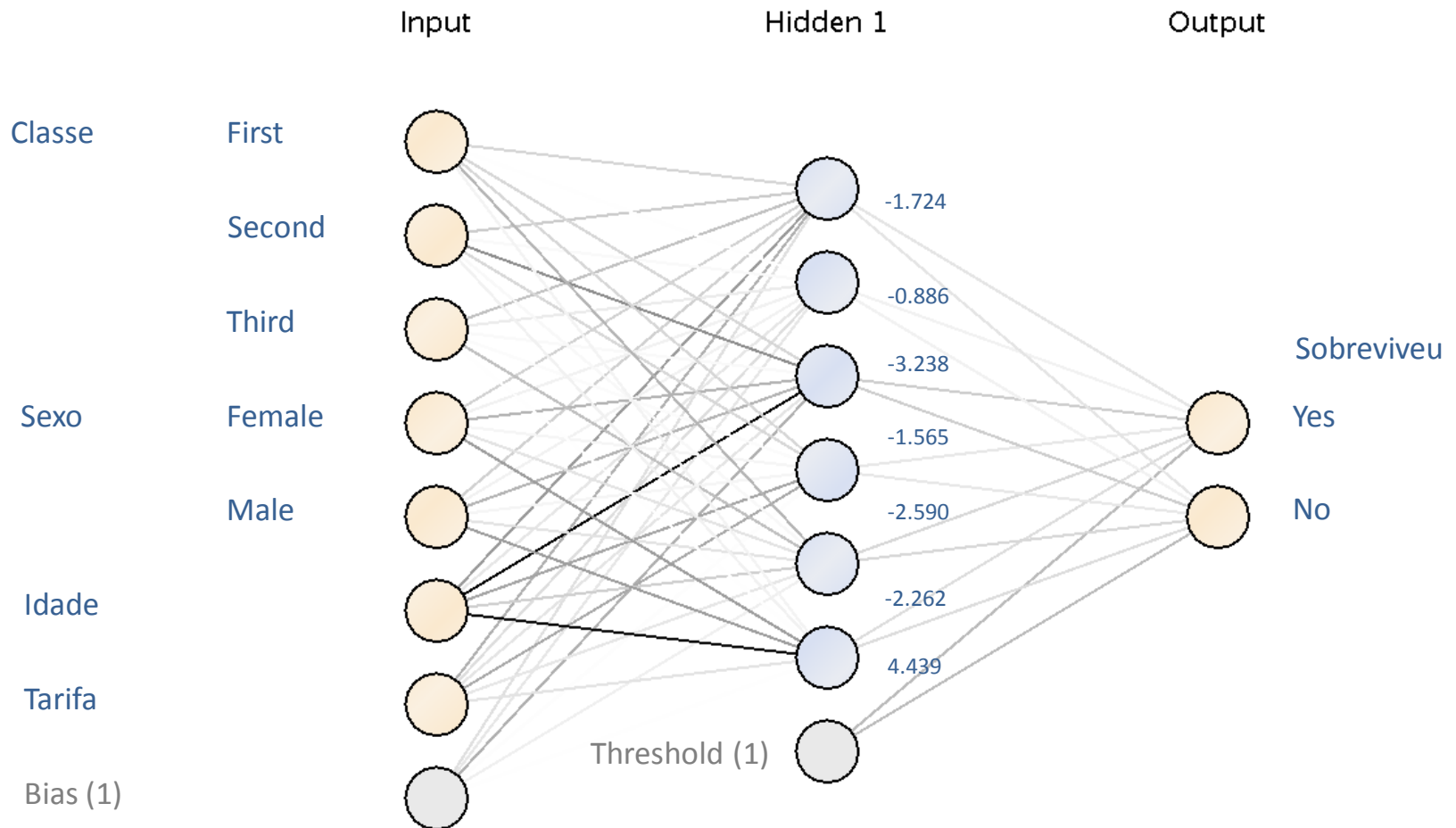
if Sexo = Female and Classe = First then Yes (97 / 4)
if Sexo = Female and Classe = Second and Idade > 56 then No (0 / 2)
if Sexo = Female and Classe = Second and Idade ≤ 56 then Yes (59 / 8)
if Sexo = Female and Classe = Third and Tarifa > 24.808 then No (1 / 18)
if Sexo = Female and Classe = Third and Tarifa ≤ 24.808 then Yes (82 / 51)
if Sexo = Male then No (110 / 484)

Quem sobreviveu (perfil)?

if Sexo = Female and
 (Classe = First
 or
 Classe = Second and Idade ≤ 56
 or
 Classe = Third and Tarifa ≤ 24.808)

Aquisição e Extração do Conhecimento

Titanic (modelo de conhecimento conexionista/estatístico)



Aquisição e Extração do Conhecimento

Titanic (modelo de conhecimento estatístico)

| Attribute | Parameter | Yes | No |
|-----------|--------------------|--------|--------|
| Idade | mean | 29.006 | 30.502 |
| Idade | standard deviation | 14.219 | 12.621 |
| Classe | value=First | 0.413 | 0.171 |
| Classe | value=Second | 0.206 | 0.198 |
| Classe | value=Third | 0.381 | 0.631 |
| Classe | value=unknown | 0.000 | 0.000 |
| Sexo | value=Female | 0.685 | 0.146 |
| Sexo | value=Male | 0.315 | 0.854 |
| Sexo | value=unknown | 0.000 | 0.000 |
| Tarifa | mean | 48.517 | 24.360 |
| Tarifa | standard deviation | 64.355 | 36.209 |