Licap

Formação Cientista de Dados

PUC Minas

Licap

Formação do Cientista de Dados

Exploração do espaço problema e do espaço solução – Módulo Básico

Luis Enrique Zárate

Conteúdo do Curso

Licap

Exploração do espaço problema

1. Identificação de potenciais problema para o negócio
2. Matriz *pairwise* para identificação e priorização de problemas

Exploração do espaço solução

1. Revisão ilustrativa de técnicas de aprendizado de máquina
2. Revisão ilustrativa das principais abordagens para representação do conhecimento extraído (visualização).
3. Exploração do espaço solução que definirá as expectativas globais sobre o resultado e as saídas esperadas.

Exploração do Espaço Problema

Licap

- Os **especialistas de domínio** participam na identificação e listagem dos potenciais problemas de interesse que podem ser alvos para projetos em Ciência de dados na empresa.
- O **Espaço Problema** é composto por todos os **domínios de problema** listados pelos **especialistas de domínio**.
- O conhecimento acerca dos **domínios de problema**, bem como as regras de negócio envolvidas, são transmitidos ao Cientista de dados pelo **especialista de domínio**.

Exploração do Espaço Problema

Licap

- Os especialistas de domínio e diretores devem auxiliar na identificação de aspectos relevantes para a Empresa. Estes aspectos podem ser: Relevância estratégica, Relevância Administrativa, Retorno financeiro, etc.

- Os especialistas de domínio devem auxiliar na atribuição de pesos para os diferentes aspectos, visto que a resposta do processo KDD para certo problema pode gerar maior valor para os negócios do que outros problemas.

- Uma matriz de problemas, **pairwise**, deve ser utilizada para identificar o problema de maior relevância a ser tratado.

Matriz problema - Pairwise

Licap

Preenchendo a coluna RELEVÂNCIA ESTRATÉGICA por meio de perguntas comparativas:

(1,2)	(2,3)	(3,4)	...	(N-1,N)
(1,3)	(2,4)	...	...	
(1,4)	...	(3,N)		
...	(2,N)			
(1,N)				

Problema	Relevância Estratégica	Relevância Administrativa	Retorno Financeiro
PROB 1	✔ ✔ ✔		
PROB 2	✔ ✔		
...			
PROB N	✔		

Matriz problema - Pairwise

Licap

Completar as colunas RELEVÂNCIA ESTRATÉGICA E RETORNO FINANCEIRO:

Ponderar Aspectos

IMP	DIF	RET
0,5	0,25	0,25

Problema	Relevância Estratégica	Relevância Administrativa	Retorno Financeiro	Resultado
PROB 1	5	3	2	3,75
PROB 2	2	1	4	2,25
PROB 3	1	2	6	2,25
PROB 4	6	6	3	5,25
PROB 5	3	4	1	2,75
PROB 6	4	5	5	4,5

Licap

Prática 1 – Identificar 3 potenciais problema e construir a Matriz de Pairwise para identificação do domínio de problema relevante.

PUC Minas

Exploração do Espaço solução

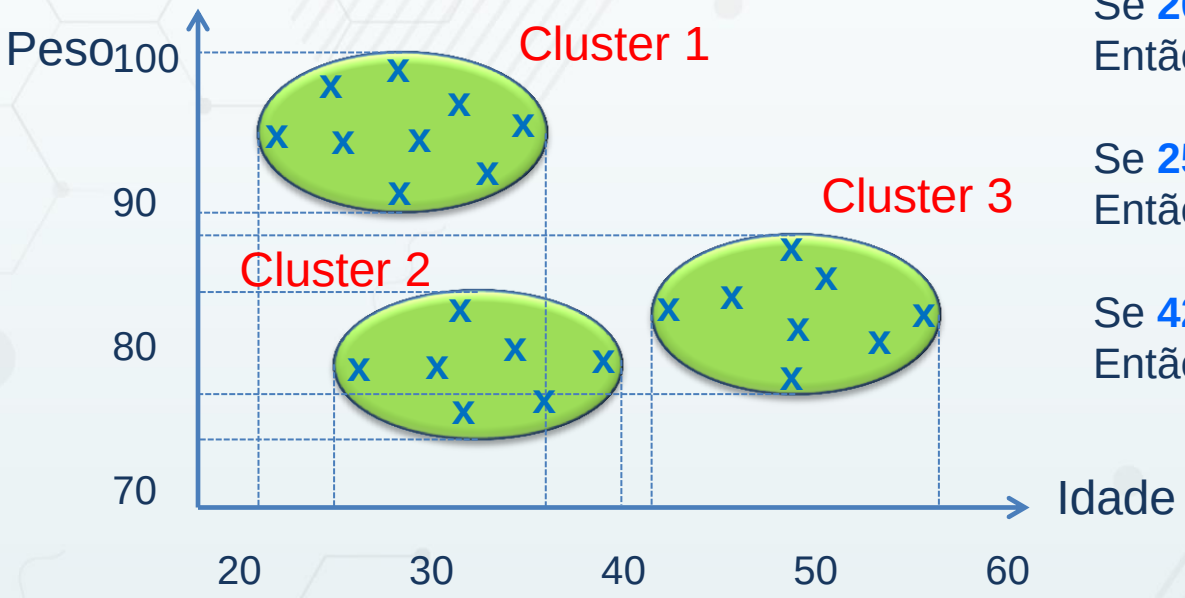
Licap

- Após a identificação do problema de maior relevância, o Cientista de dados, com o auxílio do Especialista de Domínio, definirá as **expectativas** sobre o resultado e as **saídas esperadas**.

- **Definir as técnicas de mineração de dados (Agrupamento, Classificação, Associação, etc.)** e de **visualização** que correspondam às expectativas do Especialista de Domínio.

Típicas técnicas de Data Mining

Técnicas de Agrupamento



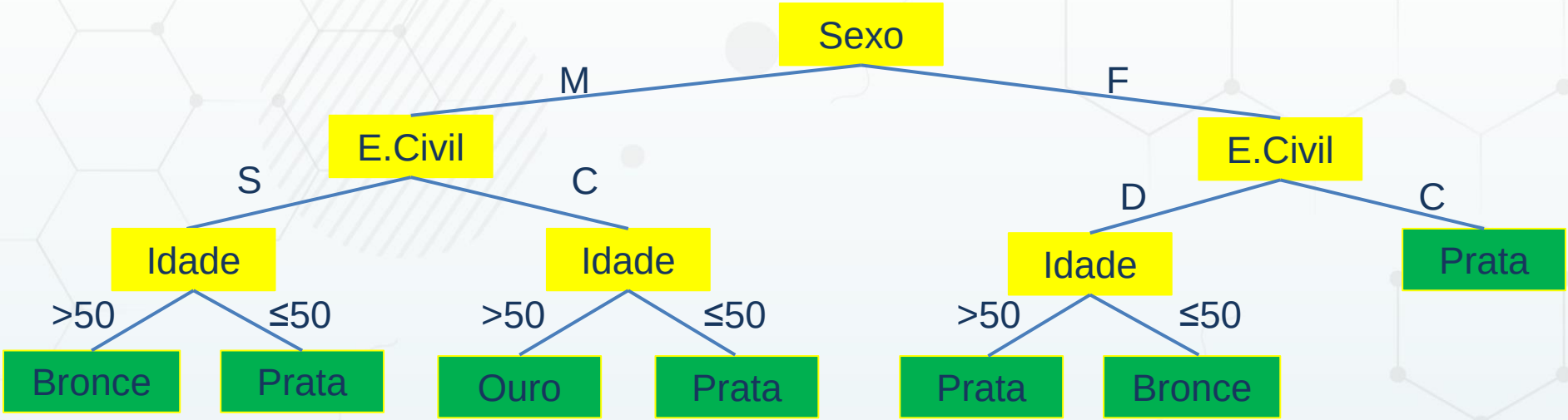
Descrição dos Agrupamentos:

Se **20<Idade<35 & 90<peso<100**
Então registro pertence ao **Cluster 1**

Se **25<Idade<40 & 74<peso<84**
Então registro pertence ao **Cluster 2**

Se **42<Idade<57 & 78<peso<89**
Então registro pertence ao **Cluster 3**

Técnicas de Classificação – Árvore de decisão



Se Sexo=M **&** Estado Civil = C **&** Idade > 50 **Então** cliente = Ouro

Se Sexo=F **&** Estado Civil = C **Então** cliente = Prata

Será que é suficiente ser do sexo = F e estado civil = Casada para ser um cliente PRATA**???**.

Isto mostra a importância do conjunto de dados. O Modelo pode não ser global**!!!**

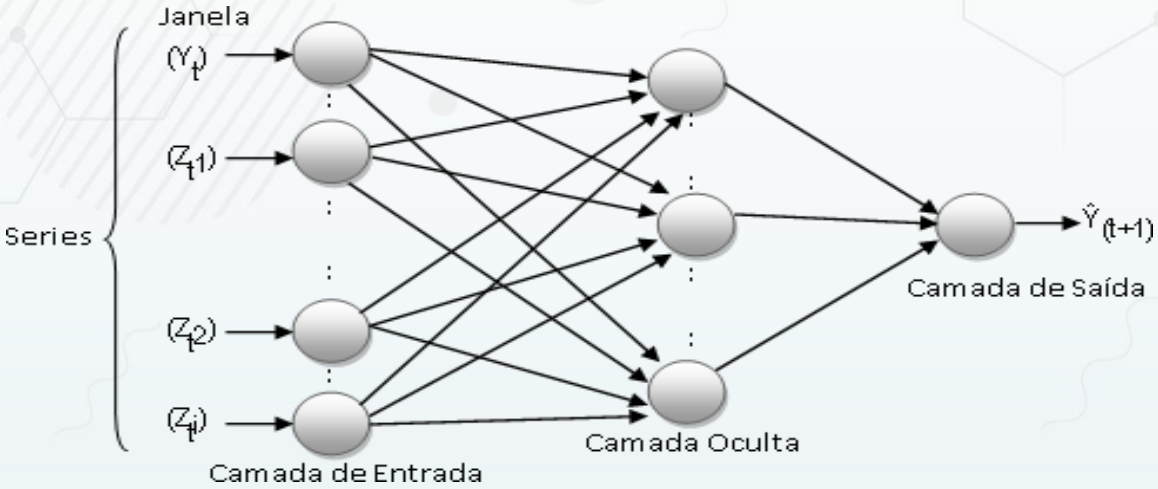
Técnica de Associação

Transc	Ítem 1	Ítem 2	Ítem 3	Ítem 4	...	...	...	Ítem 100
#1	X			X	X		X	
#2	X		X		X	X	X	X
#3	X	X	X	X	X			X
...	X	X	X		X		X	
#n		X				X		X

Se *compra* ITEM 1 e ITEM 3 então *compra* ITEM 6
Suporte = 0,80, Confiança = 0,75

Suporte = 80%: Os ítens 1, 3 e 6 foram comprados juntos em 80% das compras realizadas.
Confiança = 75%: O ítem 6 foi comprado em 75% das transações que compraram os ítens 1 e 3 juntos.

Técnica de Redes Neurais Artificiais



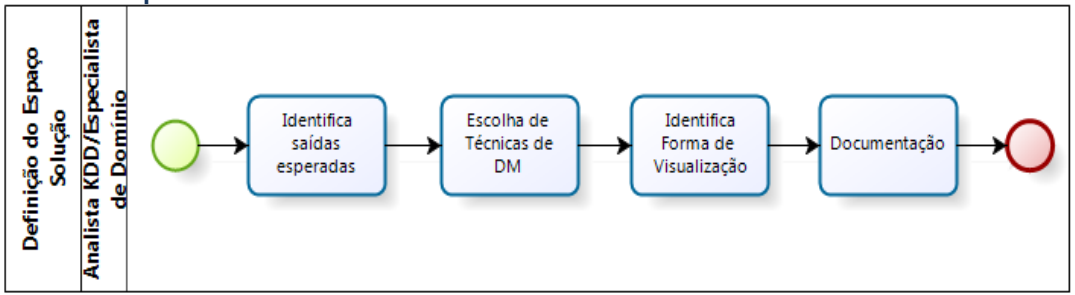
Dado uma Entrada = (Yt, Z1,....Zi) => Saída = Yt+1 = f(Yt, Z1,....Zi)

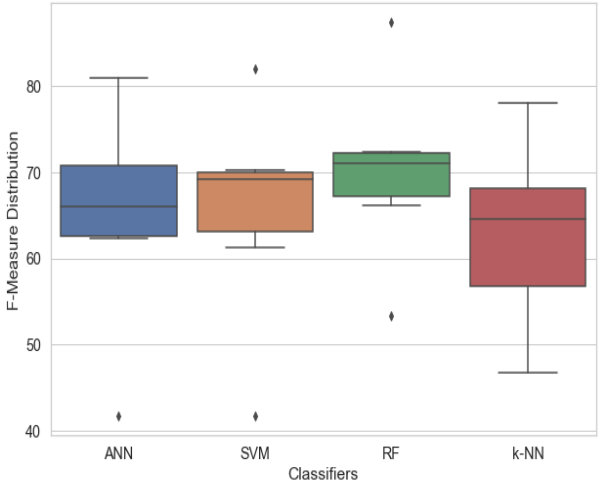
Os modelos baseados em Redes Neurais são considerados aproximadores universais. O desafio é que não conseguem explicar seu reciocínio e por isso são considerados BLACK BOX.

Definição do Espaço solução

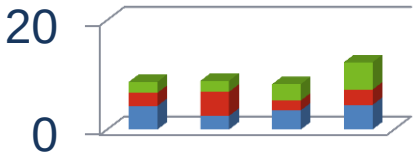
Licap

- Para cada problema e sua específica tarefa de Mineração de Dados, há um tipo de visualização que deve ser definida.






```
void main( )
{
...
}
```



Favourable conditions and actions for longevity		
A1 >6	The individual practices regular physical activities, combining light, moderate and vigorous exercises. A good frequency of light exercises (more than once a week) combined with smaller frequencies of moderate and vigorous exercises (one to three times a month) is enough to greatly increase chances of longevity.	Physical exercise practices must promote flexibility, strenght and aerobic condition. Being physically active helps prevent blood pressure diseases, reduce cholesterol and weight control, as well as increasing well-being and life satisfaction.
A2 >2	The individual does not have a serious or chronic health condition (reporting either psychiatric problems and angina was not enough to classify the individual as short-lived)	These conditions are usually associated with a genetic predisposition, but there are ways to reduce the risk of chronic diseases manifesting through general healthcare.
A4 >4	The individual did not receive cancer treatment in the last 2 years. The individual did not present all 3 other health issues at the same time (joint replacement, hip replacement, cataracts surgery).	The mobility-related health issues can be mitigated by creating safe environments for elderly people, reducing the risks of accidents.
A8 >2.62	The individual never smoked. If former smoker, the individual does not drink frequently (twice a week or less).	There must be public policies in place to discourage smoking and heavy drinking, especially at an older age. Services to help people quit smoking and drinking also play an important role in increasing a society's longevity and well-being.

Licap

Prática 2 – Para o problema identificado na prática anterior, indique a técnica de mineração de dados mais apropriada e o tipo de visualização mais adequada.

PUC Minas

Licap

Formação Cientista de Dados

Obrigado!

PUC Minas