

**CENTRO UNIVERSITÁRIO UNINORTE**  
**CURSO DE PÓS-GRADUAÇÃO EM:** Pós  
Graduação em Gerência de Banco de Dados.  
**DISCIPLINA: Mineração de Dados**

---



# Introdução

Prof.º: Manoel Limeira  
juniorlimeiras@gmail.com

# Agenda

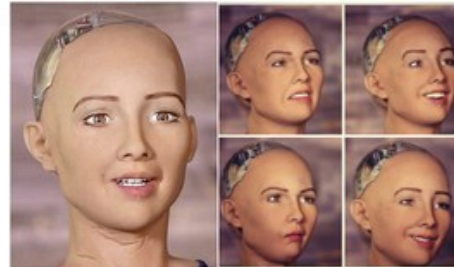
---

- Mineração de dados, o que é isso?
- Exemplos Práticos
- Padrões
- Machine Learning
- Visões sobre *Data Mining* (Mineração de Dados)
- Exemplos de Representações
- Tipos de Aprendizado
- *Data Mining* e a Ética

# O que é isso? Por que estudar?



**NETFLIX**



# Quais imagens são pessoas reais?



# Exemplos

---

- Fertilização *in vitro* (Inglaterra)
  - Coleta de ovários e espermatozoides para produção de embriões
  - Como selecionar os melhores?
  - 60 características
- Produção de Leite (Nova Zelândia)
  - Quais animais manter no rebanho e quais vender para um matadouro?
  - 700 características
  - Histórico de reprodução e produção de leite (8 anos)
  - Problemas de saúde, parto difícil, temperamento

# Por que estudar?

- A quantidade de dados no mundo parece cada vez maior e não há fim à vista.
- Discos baratos e armazenamento on-line
- Nossas escolhas são registradas na internet
- Lacuna entre geração e utilidade dos dados
- A busca (semi) automatizada por padrões nos dados
- A impressionante quantidade de oportunidades
  - O caso da fidelidade (ou não) dos clientes



# *Machine Learning*

---

- O que é **aprender**, afinal?
  - Obter conhecimento por estudo, experiência ou aprendizado; tornar-se consciente pela informação; memorizar; ser informado, verificar; receber instruções.
- Aprender implica **pensamento** e propósito
  - É uma grande discussão
- E nas máquinas/computadores...
  - Uma máquina aprende ou se torna consciente?
  - Questão filosófica
- Melhorar o **desempenho** em novas situações
  - Também chamado de '**treinamento**'

# *Data Mining*

---

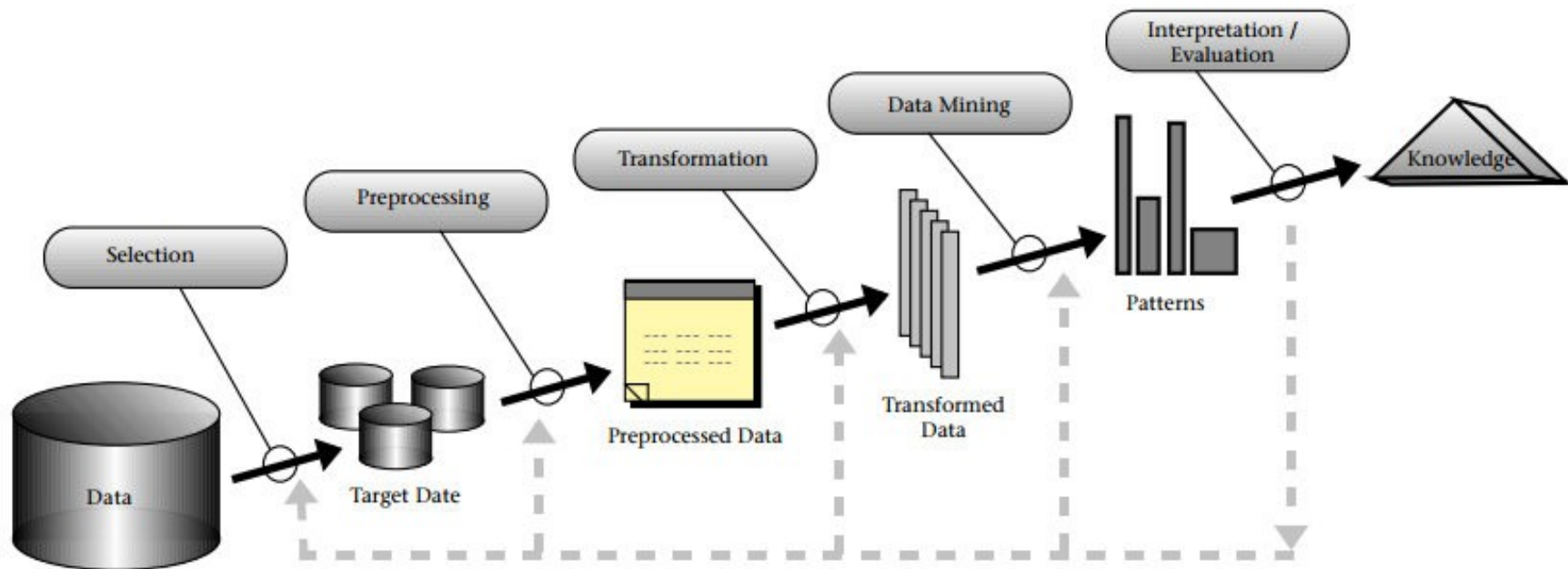
- Descobrir padrões, automaticamente, em grandes quantidades de dados e os padrões devem ser úteis
- Envolve aprendizado em um sentido prático
- Técnicas que permitam uma tomada de decisão
  - Utilizam estruturas de representação para descrever o aprendizado
  - Regras, árvores, probabilidades, equações
  - Obtenção de conhecimento
  - Avaliação do desempenho



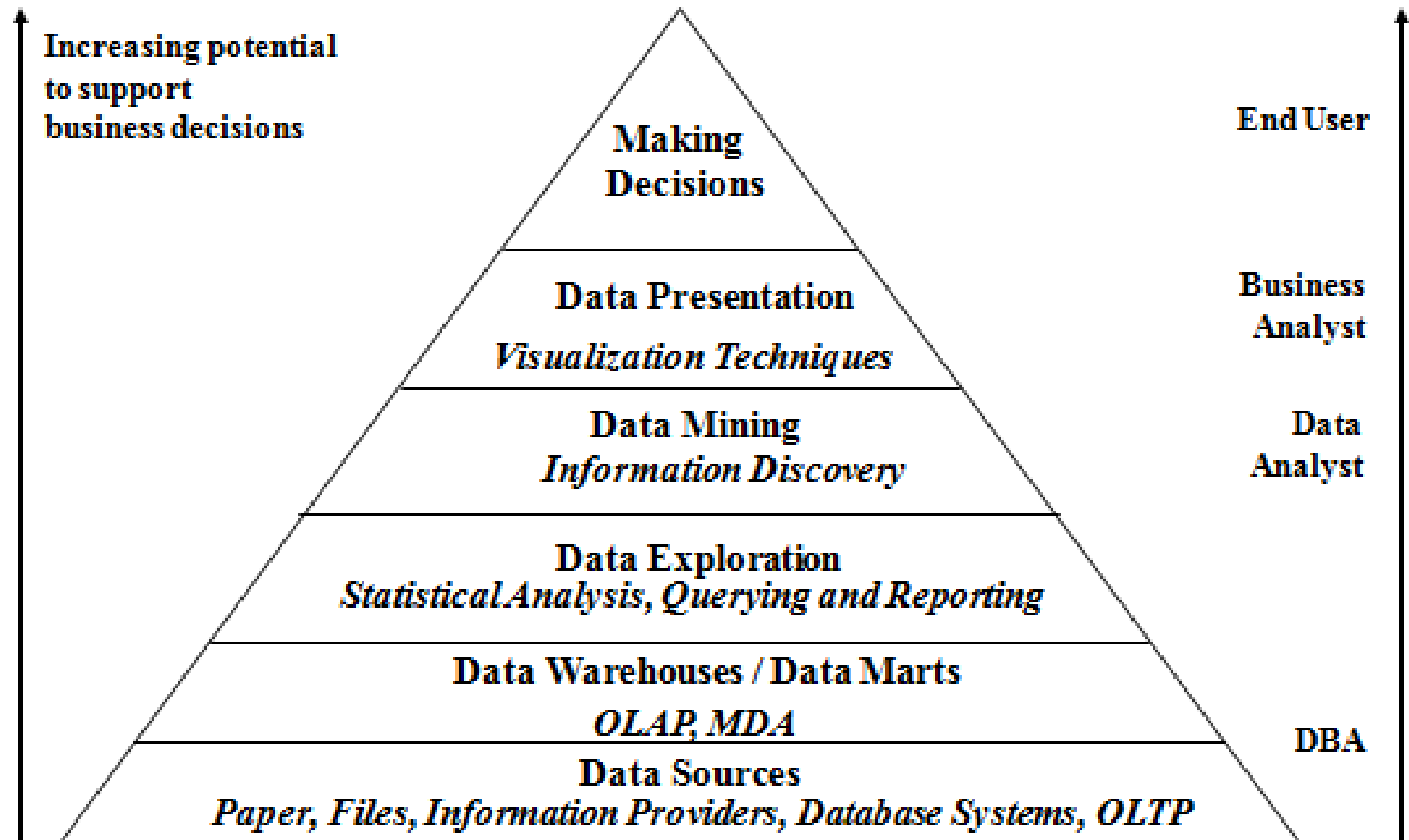
# Machine Learning e Estatística

- Diferença histórica (excessivamente simplificada)
  - **Estatística:** testar hipóteses
  - **Machine Learning:** encontrar a hipótese certa
- Enorme sobreposição
  - Árvores de decisão (C4.5 e CART)
  - Métodos vizinhos próximos
- Hoje as perspectivas convergiram
  - A maioria dos algoritmos de aprendizado de máquina emprega técnicas estatísticas

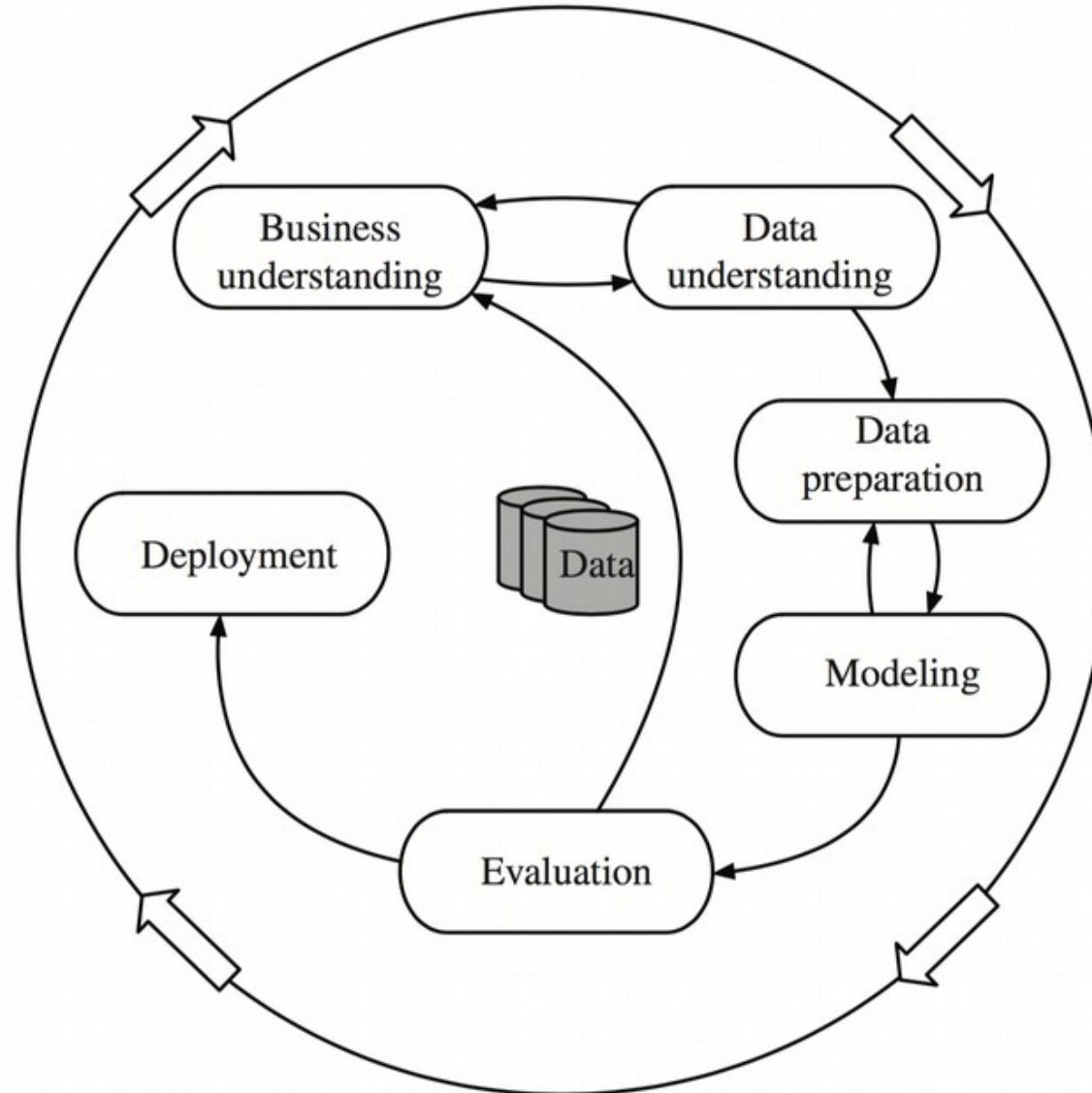
# Na comunidade de Banco de dados



# Na comunidade de *Business Intelligence*



# Na comunidade de *Machine Learning*



# Exemplos de estruturas de representação

**Table 1.2** The Weather Data

Outlook	Temperature	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

If outlook = **sunny**  
and humidity = **high**  
then play = **no**

If outlook = **rainy**  
and windy = **true**  
then play = **no**

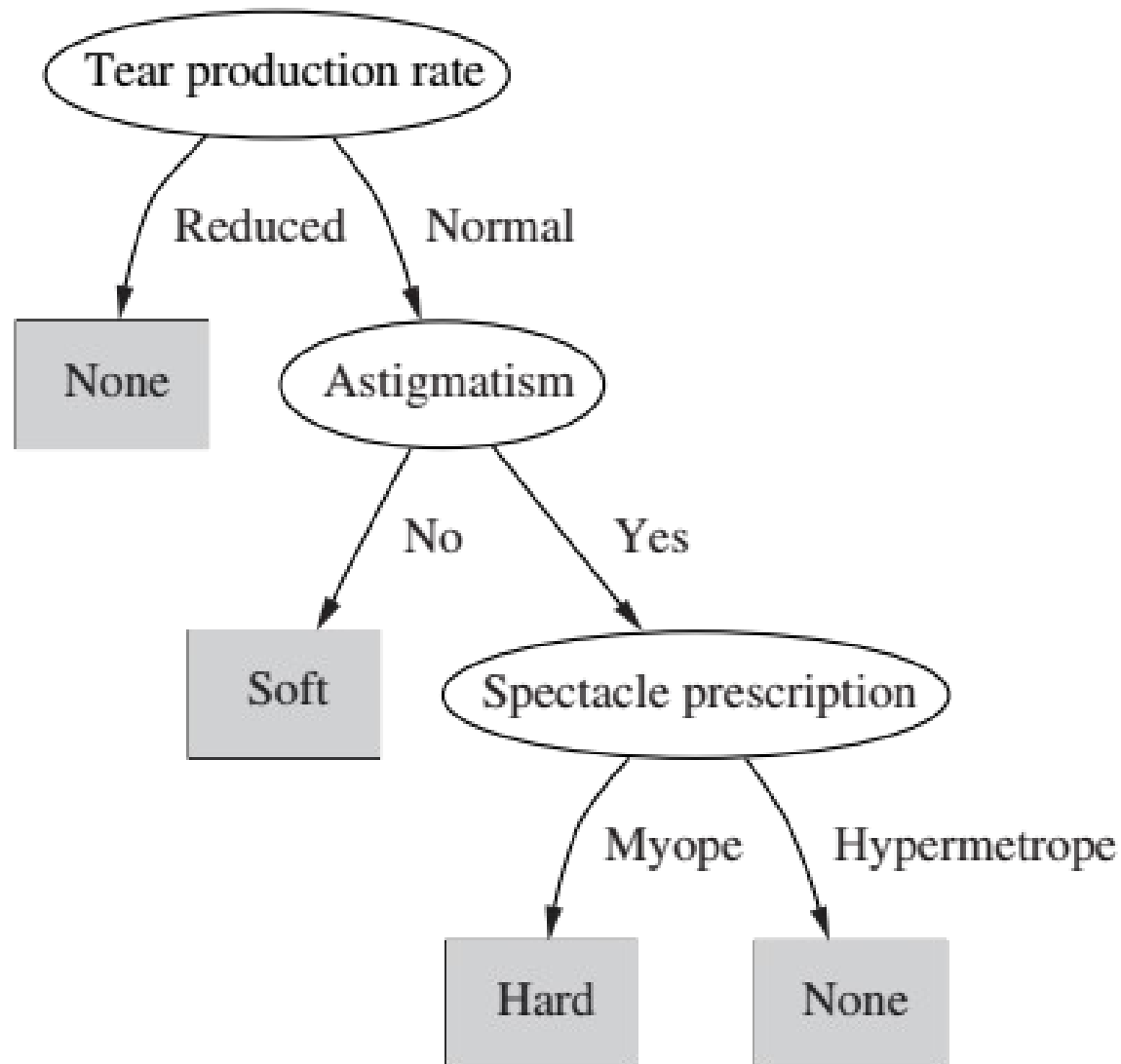
If outlook =  
**overcast**  
then play = **yes**

If humidity =  
**normal**  
then play = **yes**

**Table 1.1** The Contact Lens Data

Age	Spectacle Prescription	Astigmatism	Tear Production Rate	Recommended Lenses
Young	Myope	No	Reduced	None
Young	Myope	No	Normal	Soft
Young	Myope	Yes	Reduced	None
Young	Myope	Yes	Normal	Hard
Young	Hypermetrope	No	Reduced	None
Young	Hypermetrope	No	Normal	Soft
Young	Hypermetrope	Yes	Reduced	None
Young	Hypermetrope	Yes	Normal	Hard
Prepresbyopic	Myope	No	Reduced	None
Prepresbyopic	Myope	No	Normal	Soft
Prepresbyopic	Myope	Yes	Reduced	None
Prepresbyopic	Myope	Yes	Normal	Hard
Prepresbyopic	Hypermetrope	No	Reduced	None
Prepresbyopic	Hypermetrope	No	Normal	Soft
Prepresbyopic	Hypermetrope	Yes	Reduced	None
Prepresbyopic	Hypermetrope	Yes	Normal	None
Presbyopic	Myope	No	Reduced	None
Presbyopic	Myope	No	Normal	None
Presbyopic	Myope	Yes	Reduced	None
Presbyopic	Myope	Yes	Normal	Hard
Presbyopic	Hypermetrope	No	Reduced	None
Presbyopic	Hypermetrope	No	Normal	Soft
Presbyopic	Hypermetrope	Yes	Reduced	None
Presbyopic	Hypermetrope	Yes	Normal	None

# Exemplos de estruturas de representação





# Exemplos de estruturas de representação

**Table 1.4** The Iris Data

	Sepal Length	Sepal Width	Petal Length	Petal Width	Type
1	5.1	3.5	1.4	0.2	<i>Iris setosa</i>
2	4.9	3.0	1.4	0.2	<i>I. setosa</i>
3	4.7	3.2	1.3	0.2	<i>I. setosa</i>
4	4.6	3.1	1.5	0.2	<i>I. setosa</i>
5	5.0	3.6	1.4	0.2	<i>I. setosa</i>
...					
51	7.0	3.2	4.7	1.4	<i>Iris versicolor</i>
52	6.4	3.2	4.5	1.5	<i>I. versicolor</i>
53	6.9	3.1	4.9	1.5	<i>I. versicolor</i>
54	5.5	2.3	4.0	1.3	<i>I. versicolor</i>
55	6.5	2.8	4.6	1.5	<i>I. versicolor</i>
...					
101	6.3	3.3	6.0	2.5	<i>Iris virginica</i>
102	5.8	2.7	5.1	1.9	<i>I. virginica</i>
103	7.1	3.0	5.9	2.1	<i>I. virginica</i>
104	6.3	2.9	5.6	1.8	<i>I. virginica</i>
105	6.5	3.0	5.8	2.2	<i>I. virginica</i>
...					

# Exemplos de estruturas de representação

```
If petal-length < 2.45 then Iris-setosa
If sepal-width < 2.10 then Iris-versicolor
If sepal-width < 2.45 and petal-length < 4.55 then Iris-versicolor
If sepal-width < 2.95 and petal-width < 1.35 then Iris-versicolor
If petal-length  $\geq$  2.45 and petal-length < 4.45 then Iris-versicolor
If sepal-length  $\geq$  5.85 and petal-length < 4.75 then Iris-versicolor
If sepal-width < 2.55 and petal-length < 4.95 and petal-width < 1.55 then
    Iris-versicolor
If petal-length  $\geq$  2.45 and petal-length < 4.95 and petal-width < 1.55 then
    Iris-versicolor
If sepal-length  $\geq$  6.55 and petal-length < 5.05 then Iris-versicolor
If sepal-width < 2.75 and petal-width < 1.65 and sepal-length < 6.05
    then Iris-versicolor
If sepal-length  $\geq$  5.85 and sepal-length < 5.95 and petal-length < 4.85
    then Iris-versicolor
If petal-length  $\geq$  5.15 then Iris-virginica
If petal-width  $\geq$  1.85 then Iris-virginica
If petal-width  $\geq$  1.75 and sepal-width < 3.05 then Iris-virginica
If petal-length  $\geq$  4.95 and petal-width < 1.55 then Iris-virginica
```

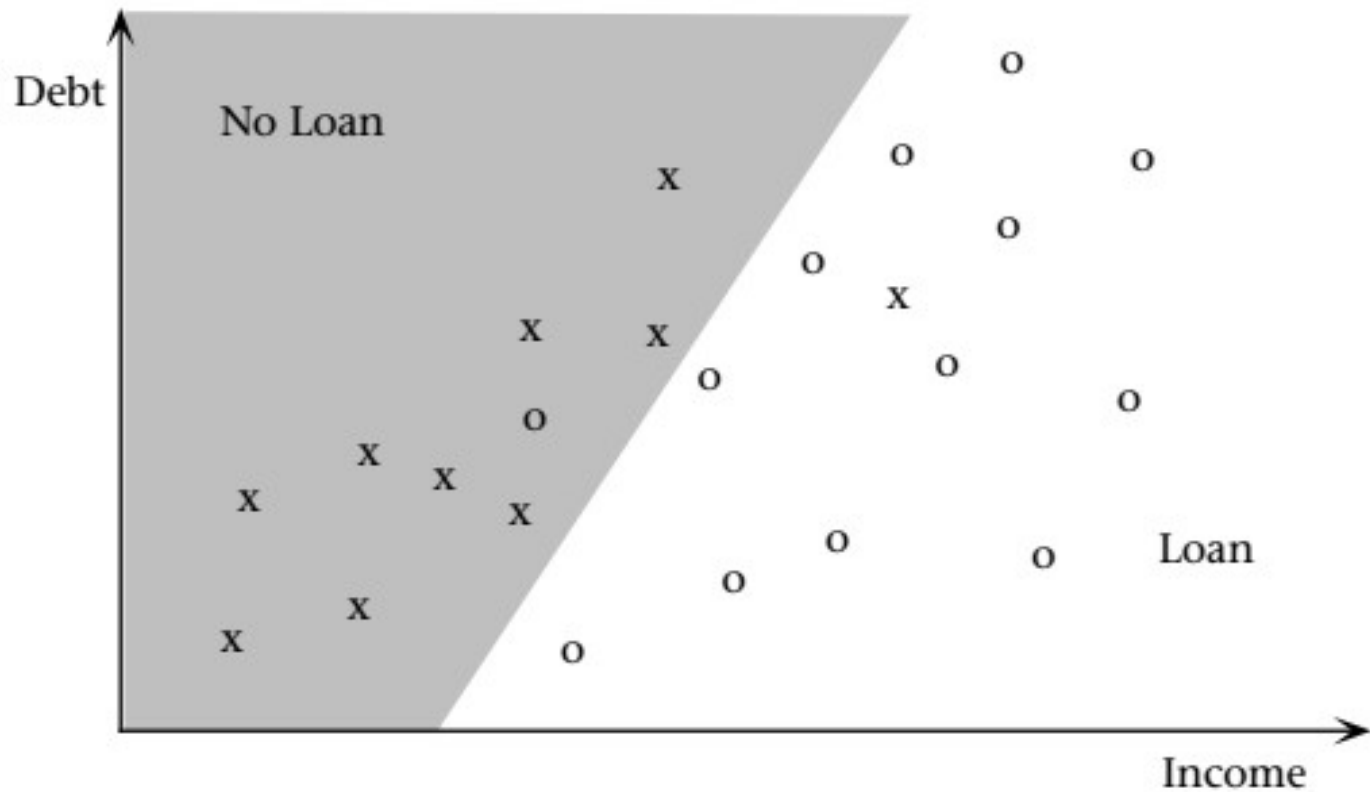
# Exemplos de estruturas de representação

**Table 1.5** The CPU Performance Data

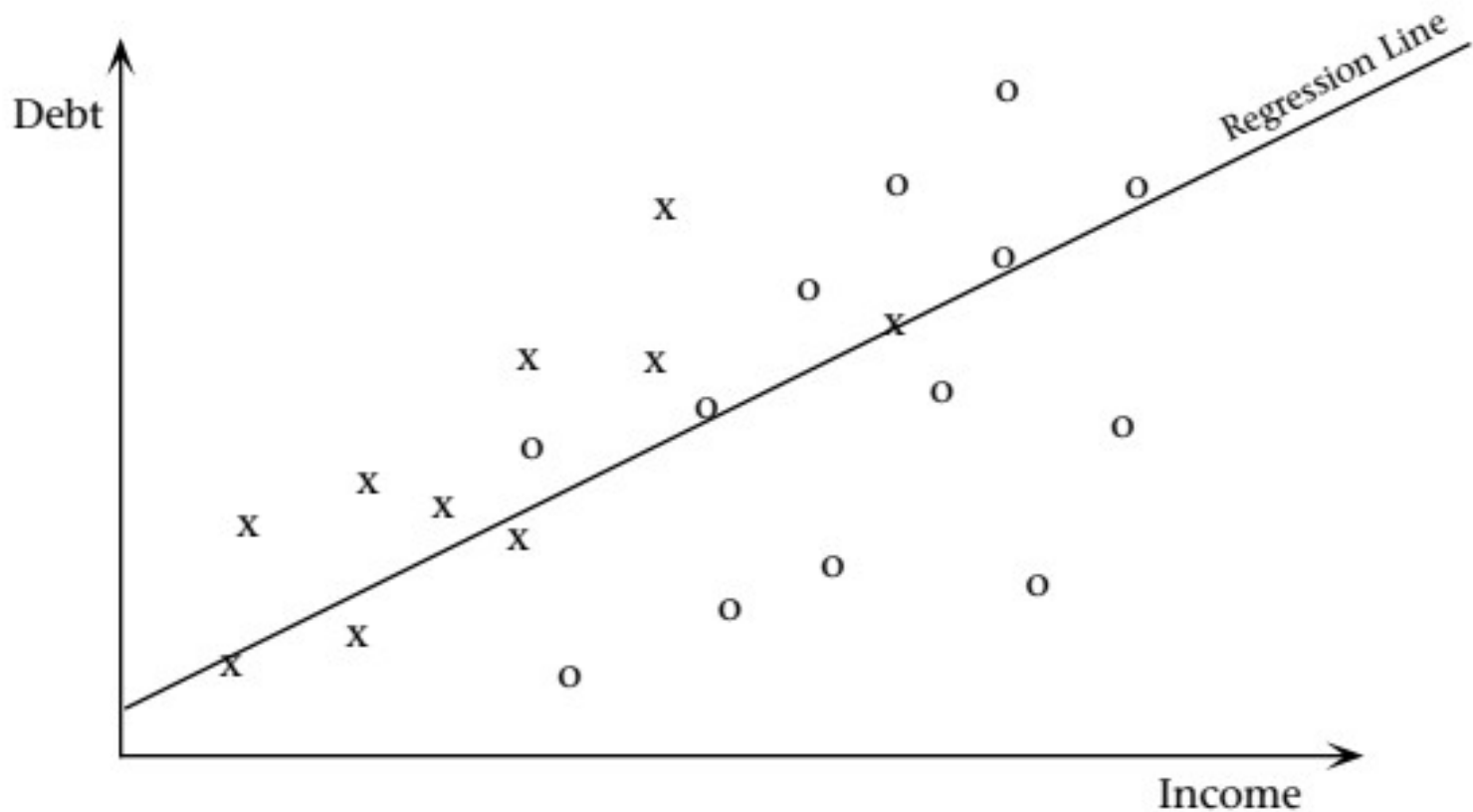
	Cycle Time (ns)	Main Memory (Kb)		Cache (KB)	Channels		Performance
		Min	Max		Min	Max	
	MYCT	MMIN	MMAX	CACH	CHMIN	CHMAX	PRP
1	125	256	6000	256	16	128	198
2	29	8000	32,000	32	8	32	269
3	29	8000	32,000	32	8	32	220
4	29	8000	32,000	32	8	32	172
5	29	8000	16,000	32	8	16	132
...							
207	125	2000	8000	0	2	14	52
208	480	512	8000	32	0	0	67
209	480	1000	4000	0	0	0	45

$$\text{PRP} = -55.9 + 0.0489 \text{ MYCT} + 0.0153 \text{ MMIN} + 0.0056 \text{ MMAX} \\ + 0.6410 \text{ CACH} - 0.2700 \text{ CHMIN} + 1.480 \text{ CHMAX}$$

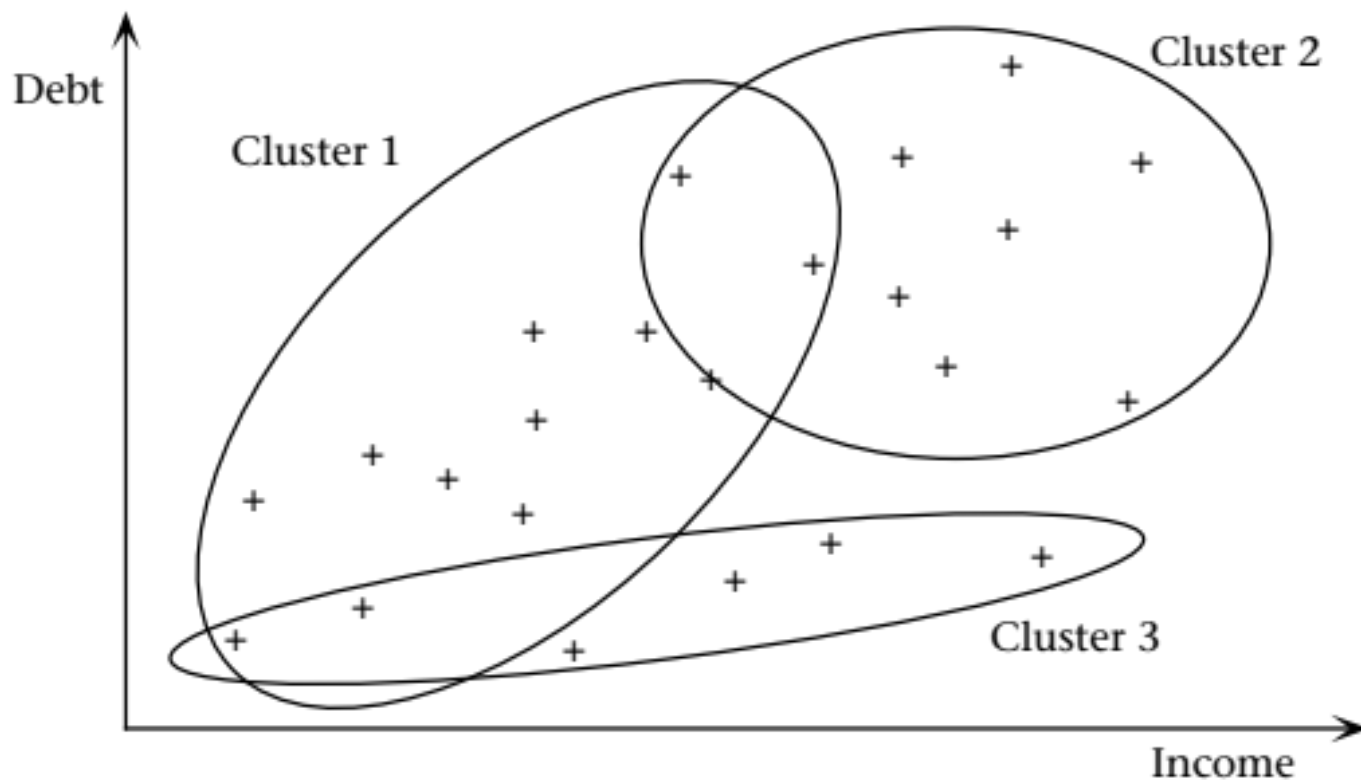
# Tipos de Aprendizado - Classificação



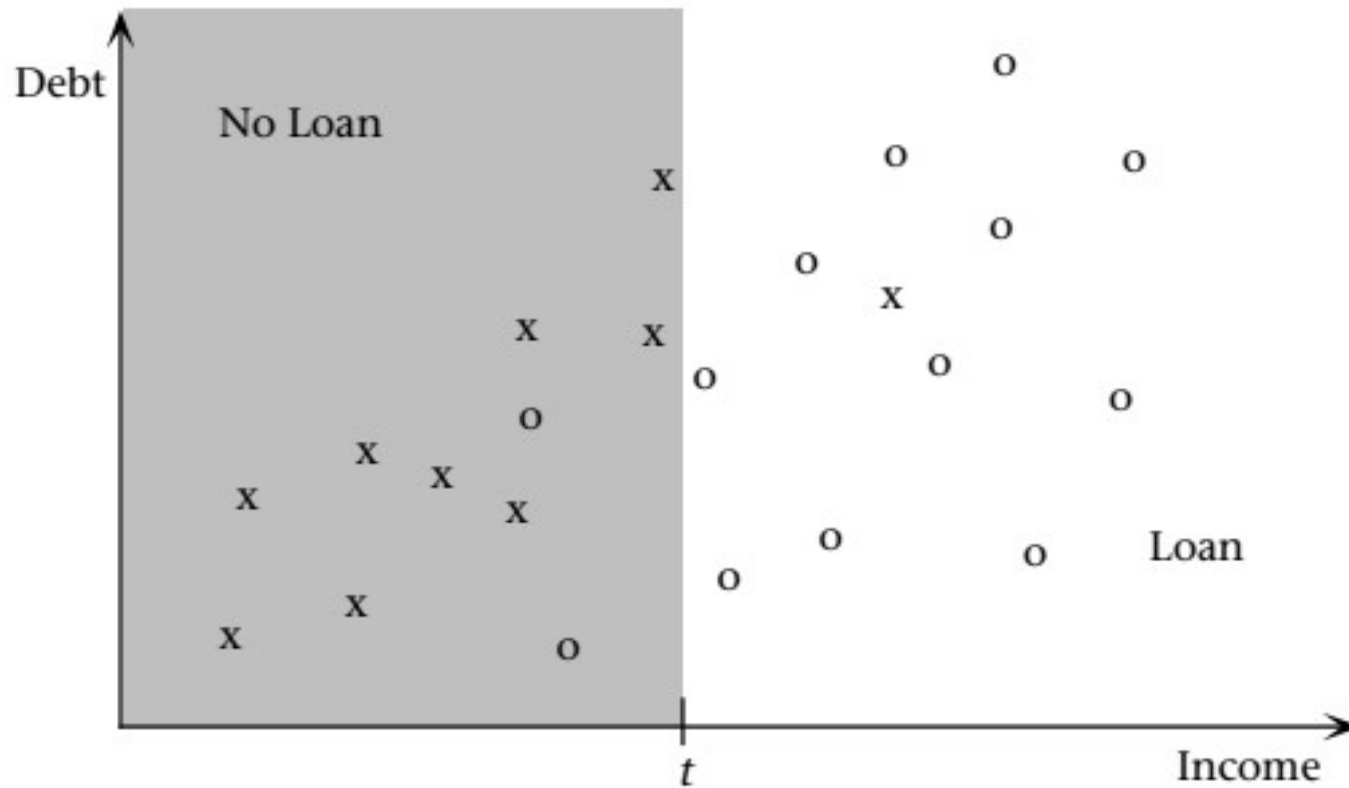
# Tipos de Aprendizado - Regressão



# Tipos de Aprendizado - Clusterização



# Tipos de Aprendizado - Regras





# Data Mining e Ética

- Questões éticas surgem em aplicações práticas
- *Data mining* não pode ser usado para discriminar
  - Empréstimo: usar algumas informações (por exemplo, sexo, religião, raça) é antiético
- Situação ética depende da aplicação
  - Por exemplo, esta mesma informação pode ser usada em uma aplicação médica
- Atributos podem conter informações problemáticas
  - Por exemplo, o código de área (CEP) pode se correlacionar com a raça

# *Data Mining e Ética*

---

- Outras Questões importantes:
  - Quem tem permissão para acessar os dados?
  - Para que finalidade os dados foram coletados?
  - Que tipo de conclusões podem ser legitimamente tiradas disso?
  - Advertências devem ser anexadas aos resultados
  - Argumentos puramente estatísticos nunca são suficientes!

**CENTRO UNIVERSITÁRIO UNINORTE**  
**CURSO DE PÓS-GRADUAÇÃO EM:** Pós  
Graduação em Gerência de Banco de Dados.  
**DISCIPLINA: Mineração de Dados**

---



# Introdução

Prof.º: Manoel Limeira  
juniorlimeiras@gmail.com