

# Sistemas de Apoio a Decisão

(Inteligência nos Negócios - *Business Inteligente*)



## Sistemas de Informação

Aran Bey Tcholakian Morales, Dr. Eng.

Apostila 7: Análises de Dados: Modelos Preditivos - Arvore de Decisão

# Análises de Dados: Modelos Descritivos e Preditivos

Modelos de Análises Descritiva/Exploratória

**Modelos de Análises Preditivas**

Modelos Analíticos: Avaliação e Validação





# Modelos Preditivos

Para desenvolver os diferentes tipos de modelos de análises preditivas, temos um conjunto de tarefas, técnicas e algoritmos.

Cada uma das diferentes **tarefas** tem diferentes **técnicas** de abordagem e as técnicas um conjunto de **algoritmos** que podem ser utilizados.

Entre as tarefas preditivas temos: **classificação**, associação, agrupamentos, previsão temporal entre as principais.

# **Classificação**

# Modelos Preditivos: Classificação

O objetivo da classificação, é analisar os dados e desenvolver uma **descrição** ou **modelo** para descobrir um **relacionamento** entre os **atributos previsores** e o **atributo meta**.

A tarefa da classificação, é caracterizada por uma **boa definição das classes**, adquirida em um conjunto de exemplos **pre-classificados** (dados de treino).

# Modelos Preditivos: Classificação



1. Um conjunto de treino com exemplos rotulados é usado para treinar o classificador.

Training Data

2. Um algoritmo de aprendizagem é executado para induzir um classificador a partir do conjunto de treino



Train the Machine Learning Algorithm

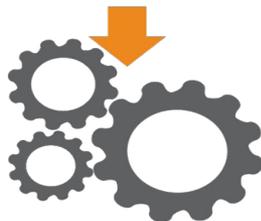


Model

3. Uma vez construído o classificador, este pode ser usado para classificar futuros exemplos



Input Data



Machine Learning Algorithm



Prediction



Evaluate

# Modelos Preditivos: Classificação

**Introdução:** Definição, objetivos, tarefas e características da classificação;

**Abordagem Simbólica:** classificação baseado na IA simbólica (heurística): árvore de decisão, teoria da informação, algoritmos ID3 e C4.5;

**Abordagem Estatística:** Classificadores Bayesianos (Naive Bayes), K-Vizinhos mais próximos (k-Nearest Neighbor);

**Abordagem Biológica:** classificador baseado na IA biológica: redes neurais e algoritmos genéticos.

# **Abordagem Simbólica**

# Predição - Classificação - Arvore de Decisão

São um **método de aprendizagem supervisionado** que constrói árvores de classificação a partir de exemplos.

Algoritmos : ID3, C4.5, (Quinlan), CART (Breiman)

Os métodos baseados em árvores, dividem o espaço de entrada em **regiões disjuntas** para construir uma **fronteira de decisão**.

As regiões são escolhidas baseadas em técnicas **heurísticas** onde a cada passo os algoritmos selecionam a variável que provê a **melhor separação de classes**.

# Predição - Classificação - Árvore de Decisão

Cada **vértice** (nodo) corresponde a um **atributo**, e cada **aresta** da árvore a um **valor possível** do atributo.

Uma **folha** da árvore corresponde ao valor esperado da **decisão** segundo os dados de treino utilizados.

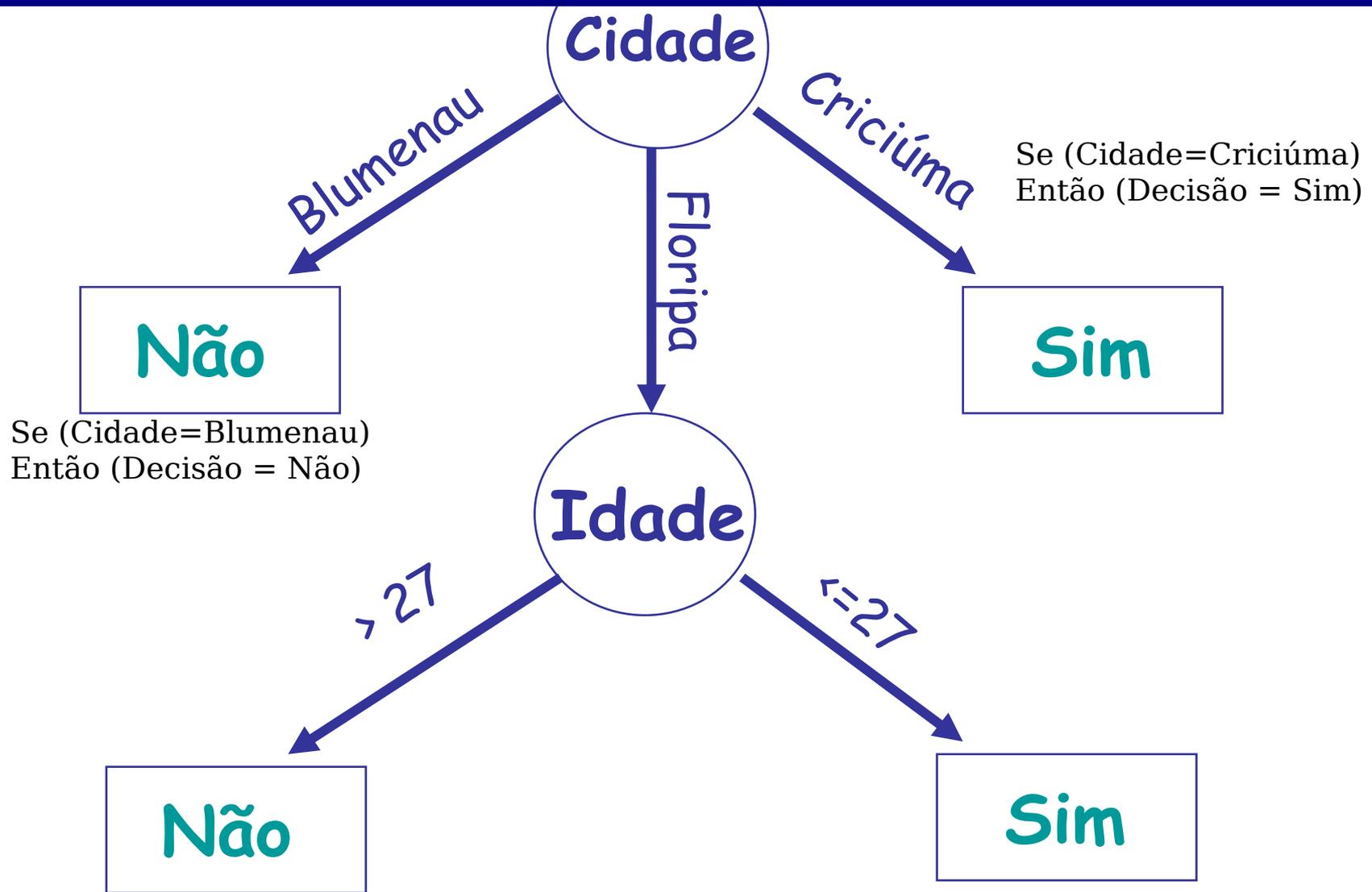
Cada **vértice** é **associado** ao **atributo** mais **informativo** que ainda não tenha sido considerado.

Para medir o **nível de informação** de um atributo se utiliza o conceito de **entropia da Teoria da Informação**.

# Predição - Classificação - Arvore de Decisão

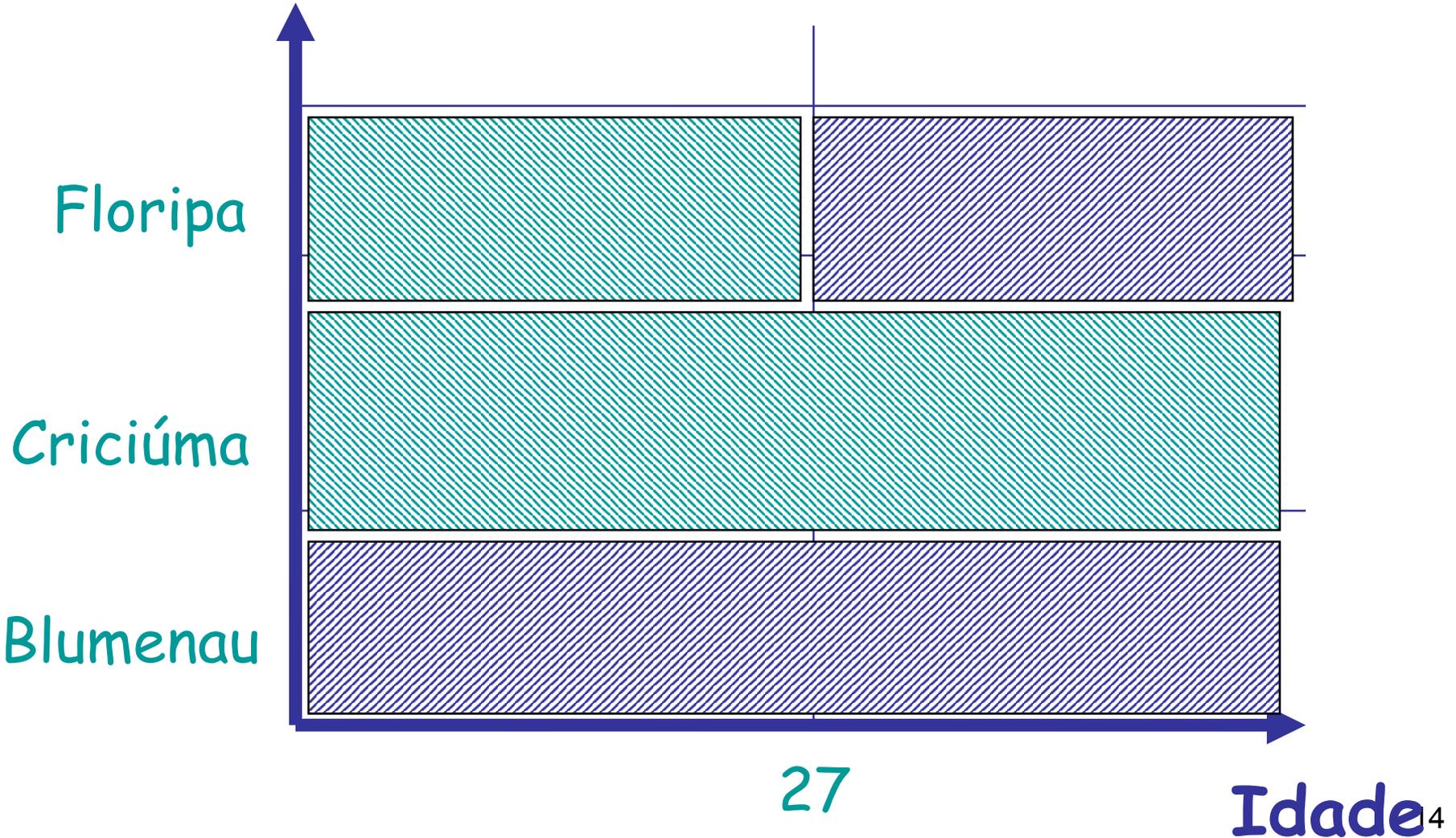
ID	Sexo	Cidade	Idade	Comrou Produto X ?
1	M	Floripa	25	→ S
2	M	Criciuma	21	→ S
3	F	Floripa	23	→ S
4	F	Criciuma	34	→ S
5	F	Floripa	30	→ N
6	M	Blumenau	21	→ N
7	M	Blumenau	20	→ N
8	F	Blumenau	18	→ N
9	F	Floripa	34	→ N
10	M	Floripa	55	→ N

# Predição - Classificação - Arvore de Decisão



# Predição - Classificação - Arvore de Decisão

Cidade  SIM  NÃO



## Algoritmo ID3

**ID3**, é um algoritmo que construí uma **árvore de decisão** sob as seguintes premissas:

Cada **vértice** (nodo) corresponde a um **atributo**, e cada **aresta** da árvore a um **valor possível** do atributo.

Uma **folha** da árvore corresponde ao valor esperado da **decisão** segundo os dados de treino utilizados.

A **explicação** de uma determinada decisão está na **trajetória** da raiz a folha representativa desta decisão.

## Algoritmo ID3

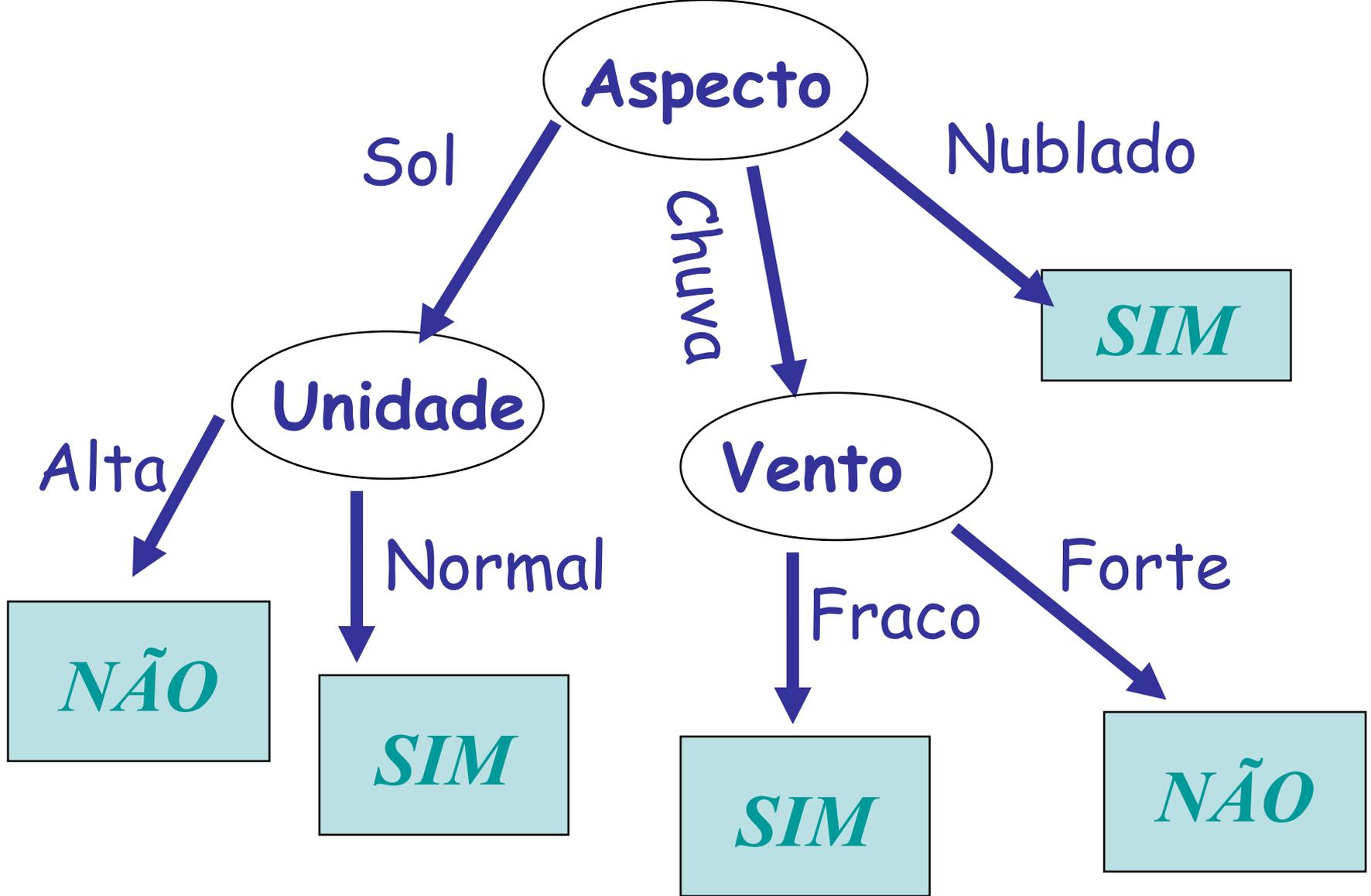
Cada **vértice** é **associado** ao **atributo** mais **informativo** que ainda não tenha sido considerado.

Para medir o **nível de informação** de um atributo se utiliza o conceito de **entropia da Teoria da Informação**.

Menor o valor da entropia, menor a incerteza e mais utilidade tem o atributo para a classificação.

# Predição - Classificação - Arvore de Decisão

<b>Dia</b>	<b>Aspecto</b>	<b>Temperatura</b>	<b>Umidade</b>	<b>Vento</b>	<b>Decisão</b>
1	Sol	Quente	Alta	Fraço	N
2	Sol	Quente	Alta	Forte	N
3	Nublado	Quente	Alta	Fraço	S
4	Chuva	Agradável	Alta	Fraço	S
5	Chuva	Fria	Normal	Fraço	S
6	Chuva	Fria	Normal	Forte	N
7	Nublado	Fria	Normal	Forte	S
8	Sol	Agradável	Alta	Fraço	N
9	Sol	Fria	Normal	Fraço	S
10	Chuva	Agradável	Normal	Fraço	S
11	Sol	Agradável	Normal	Forte	S
12	Nublado	Agradável	Alta	Forte	S
13	Nublado	Quente	Normal	Fraço	S
14	Chuva	Agradável	Alta	Forte	N



Qual é a decisão se: Aspecto = Sol, Temperatura = Agradavel;  
Umidade = Normal e Vento = Fraco?

# Predição - Classificação - Árvore de Decisão

## Algoritmo C4.5

O **C 4.5** é uma extensão do **ID3**: Construí árvores de decisão, com valores **desconhecidos** para alguns **atributos**.

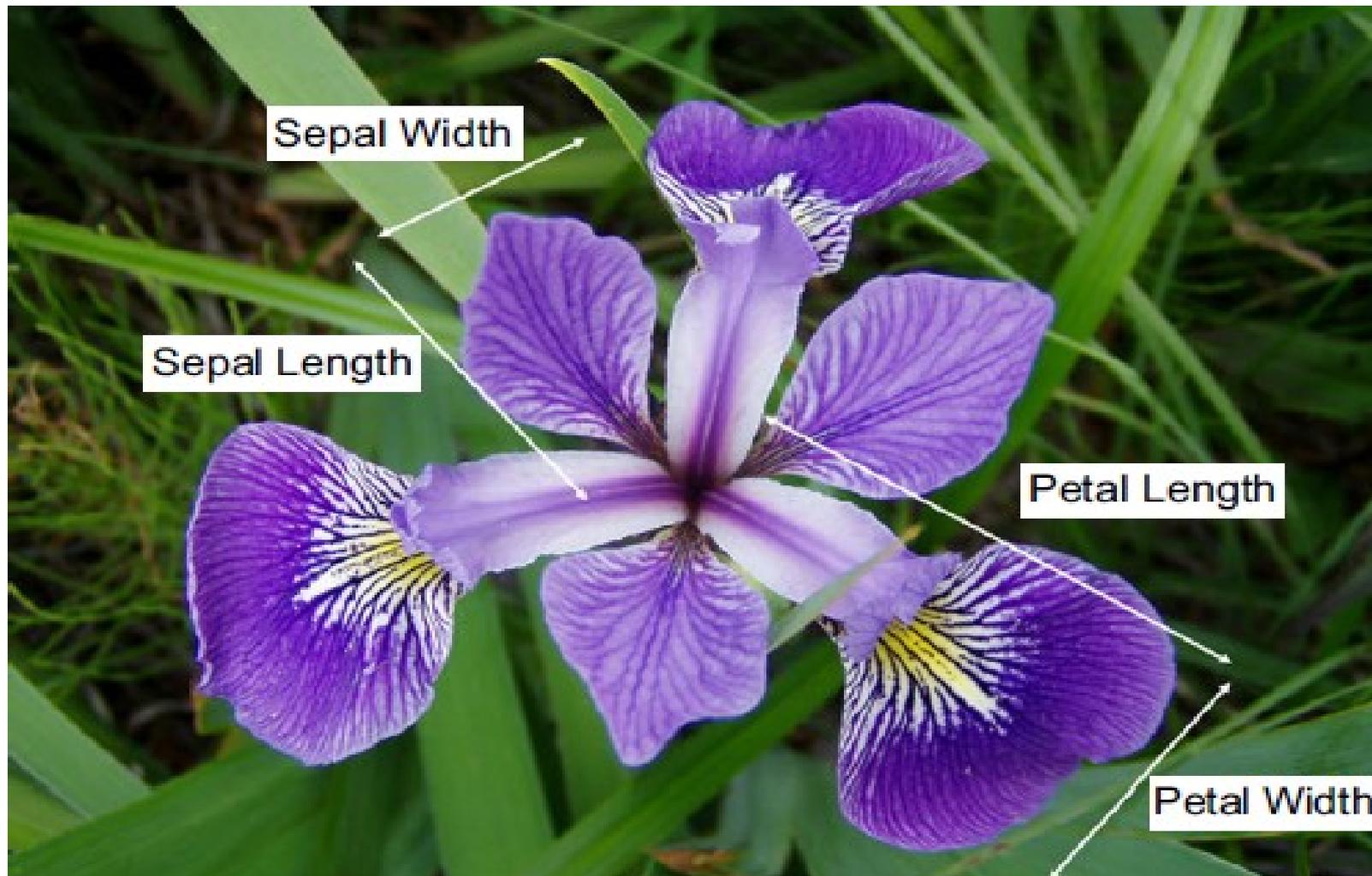
Trabalha com **atributos** que apresentam **valores contínuos**.

Utiliza o conceito de **poda (pruning)** de árvores.

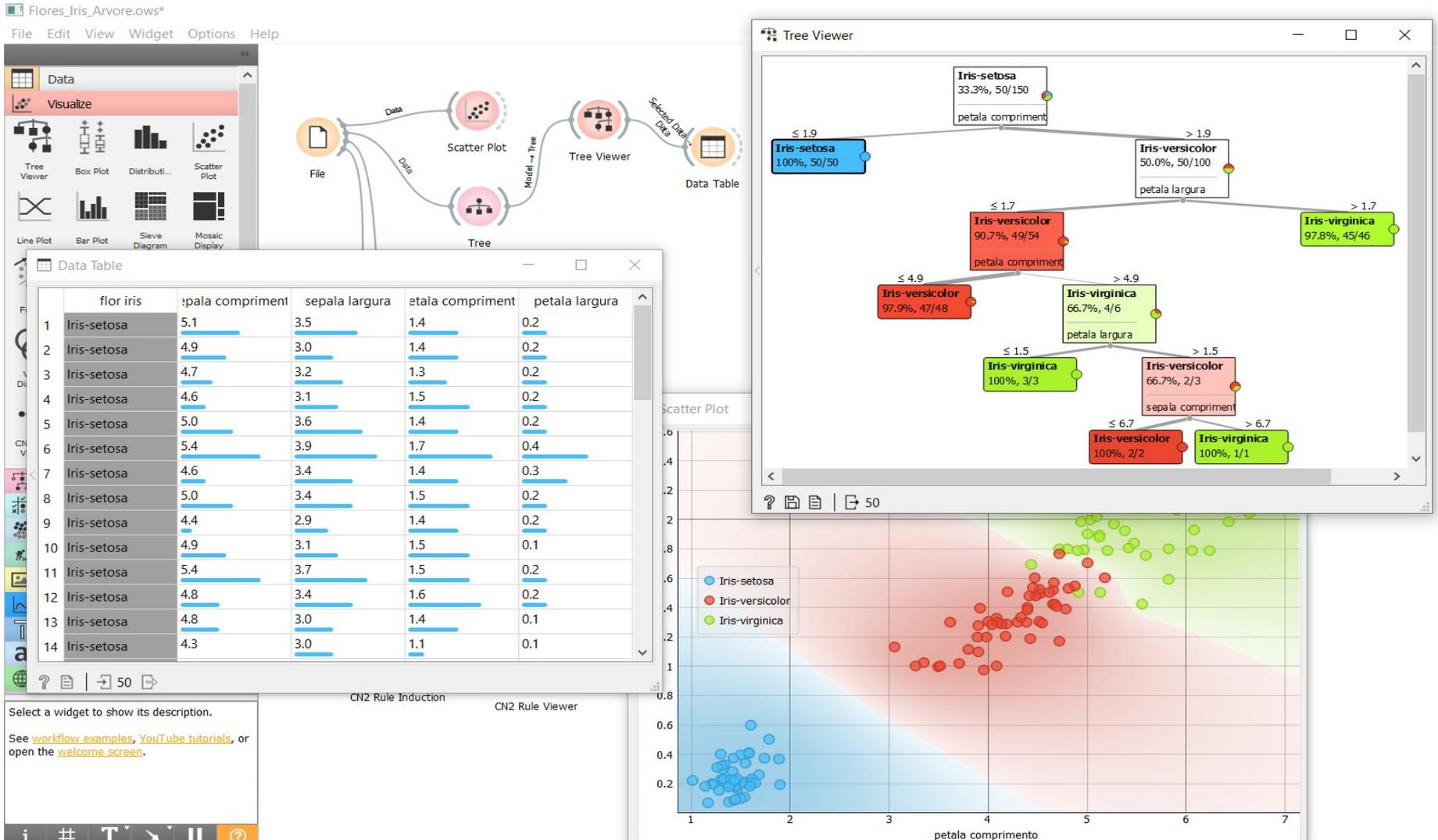
Quando existem **atributos desconhecidos** para alguma variável, os mesmos são considerado como uma nova categoria.

Quando existem variáveis com **atributos contínuos**, o algoritmo cria intervalos segundo as alterações na variável de decisão.

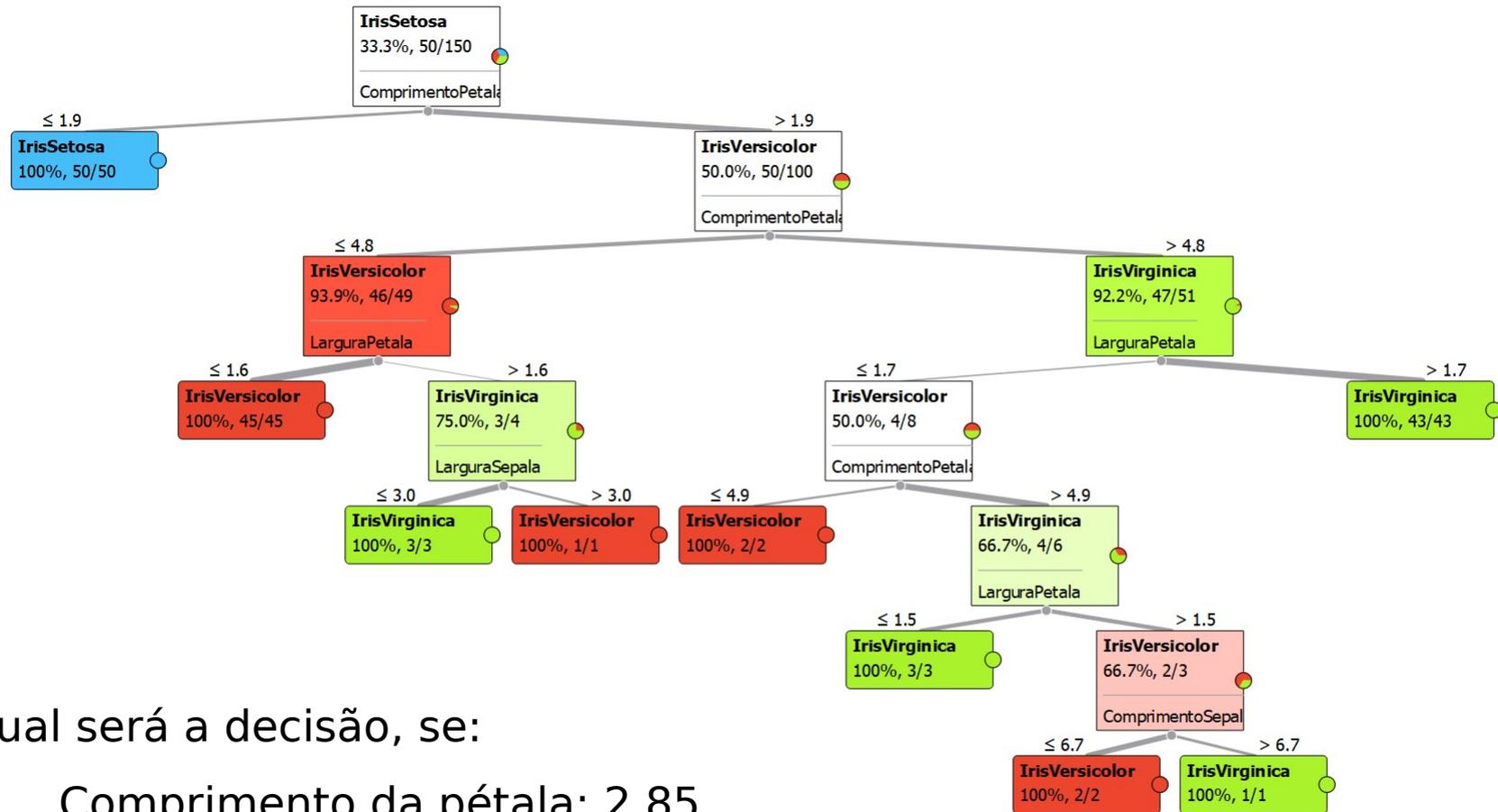
# Classificação de Flores Iris



# Classificação de Flores Iris



# Classificação de Flores Iris



Qual será a decisão, se:

Comprimento da pétala: 2,85

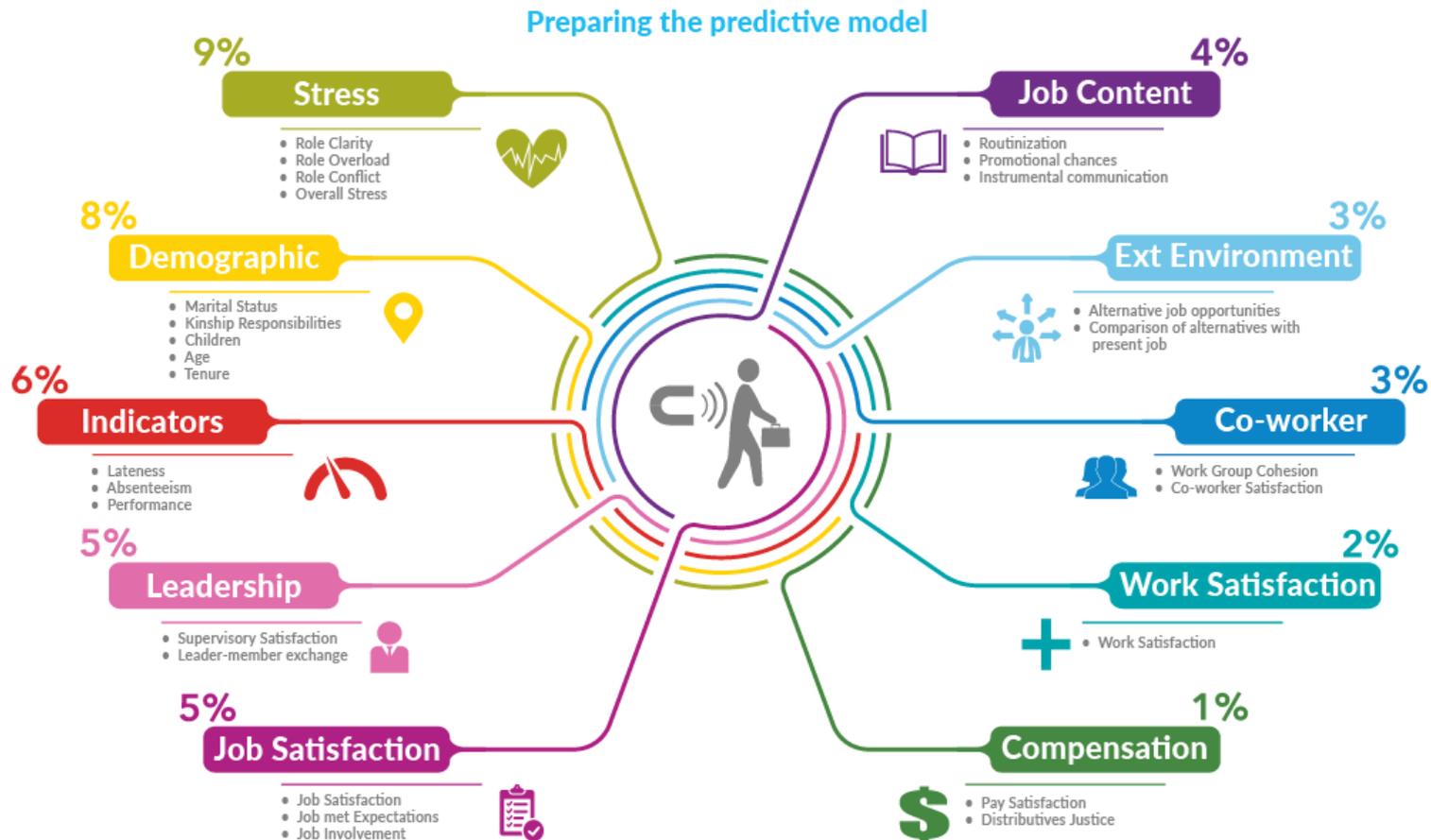
Comprimento da sépala: 4,6

Largura da pétala: 2,15

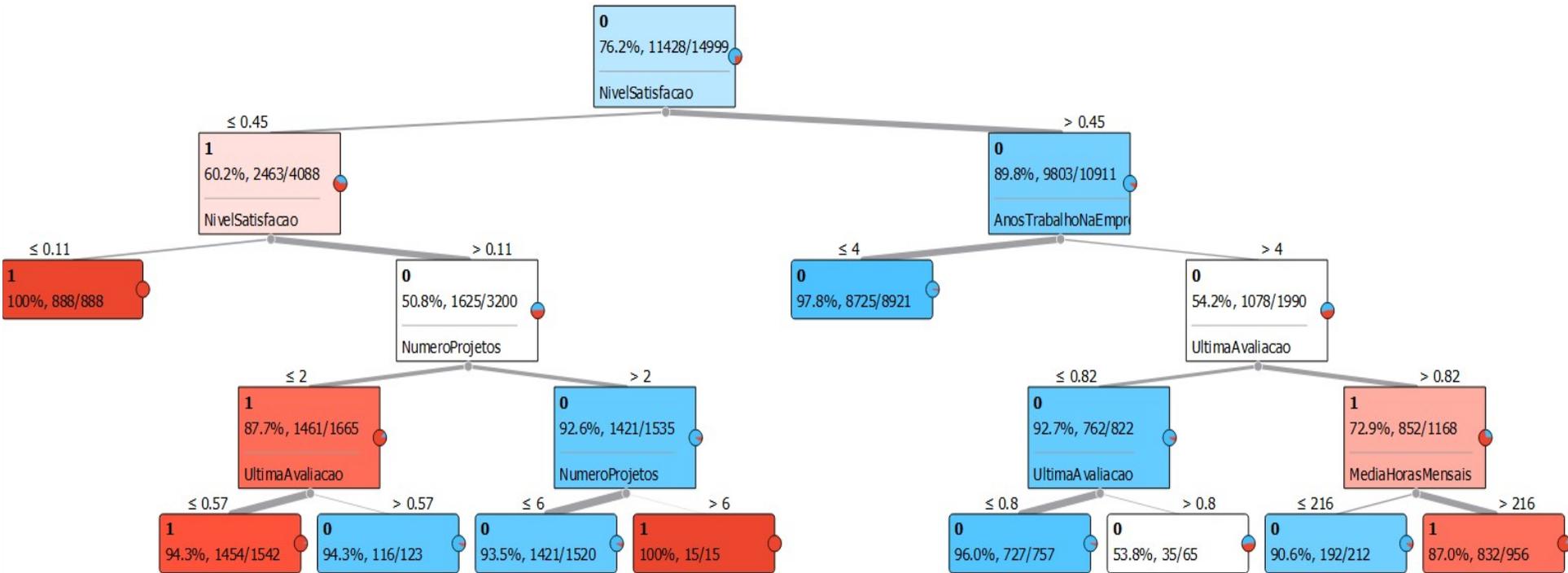
Largura da sépala: 0,9

# Turnover de funcionários

## DRIVERS OF EMPLOYEE TURNOVER



# Turnover de funcionários



# Sistemas de Apoio a Decisão

(Inteligência nos Negócios - *Business Inteligente*)



## Sistemas de Informação

Aran Bey Tcholakian Morales, Dr. Eng.

Apostila 7: Análises de Dados: Modelos Preditivos