



Instituto Federal de Educação, Ciência e Tecnologia da Bahia  
Curso de Tecnologia em Análise e Desenvolvimento de Sistemas  
Prof.: Grinaldo Lopes de Oliveira  
Disciplina: Banco de Dados 2

### Atividade de Laboratório

Membros da Dupla: \_\_\_\_\_

#### Objetivo do Laboratório:

Fixar conceitos relativos à **Mineração de Dados** aprendido em sala de aula através de uma abordagem prática utilizando o programa WEKA escrito em linguagem JAVA. Neste laboratório teremos a oportunidade de exercitar a técnica de mineração de dados intitulado classificação e associação;

### INICIANDO OS MOTORES



#### Iniciando as Atividades

Este laboratório deverá ser feito **OBRIGATORIAMENTE** em dupla. Iniciaremos nossas atividades praticando nossos conhecimentos através de um programa que procura representar o conhecimento acerca de uma árvore genealógica.

Inicialmente, execute o programa WEKA instalado em seu computador. Pressione o botão EXPLORER logo em seguida.



O objetivo deste laboratório é gerar regras de classificação. Iremos utilizar a implementação do algoritmo id3 desenvolvido por Ross Quinlan.

Esse protótipo cria a estrutura de uma árvore de decisão e interpreta a árvore transformando-a em regras.

Vamos minerar a base de dados abaixo. Ela representa um conjunto de dados coletado para decidir se haverá uma partida de futebol.

<u>aparência</u>	<u>temperatura</u>	<u>umidade</u>	<u>vento</u>	<u>jogo</u>
Chuvoso	Frio	Normal	Falso	Sim
Chuvoso	Agradável	Alta	Falso	Sim
Encoberto	Quente	Alta	Falso	Sim
Sol	Quente	Alta	Verdade	Nao
Sol	Quente	Alta	Falso	Nao
Chuvoso	Frio	Normal	Verdade	Nao
Encoberto	Frio	Normal	Verdade	Sim
Sol	Agradável	Alta	Falso	Nao
Sol	Frio	Normal	Falso	Sim
Chuvoso	Agradável	Normal	Falso	Sim
Sol	Agradável	Normal	Verdade	Sim
Encoberto	Agradável	Alta	Verdade	Sim
Encoberto	Quente	Normal	Falso	Sim
Chuvoso	Agradável	Alta	Verdade	nao

Primeiro transformaremos a tabela em um arquivo texto no formato ARFF, pois é o tipo de arquivo que o WEKA consegue interpretar.

O resultado é o seguinte:

```
@relation condicoes
```

```
@attribute aparencia { chuvoso,encoberto,sol }
@attribute temperatura { frio,agradável,quente }
@attribute umidade { normal,alta }
@attribute vento { falso,verdade }
@attribute jogo { sim,nao }
```

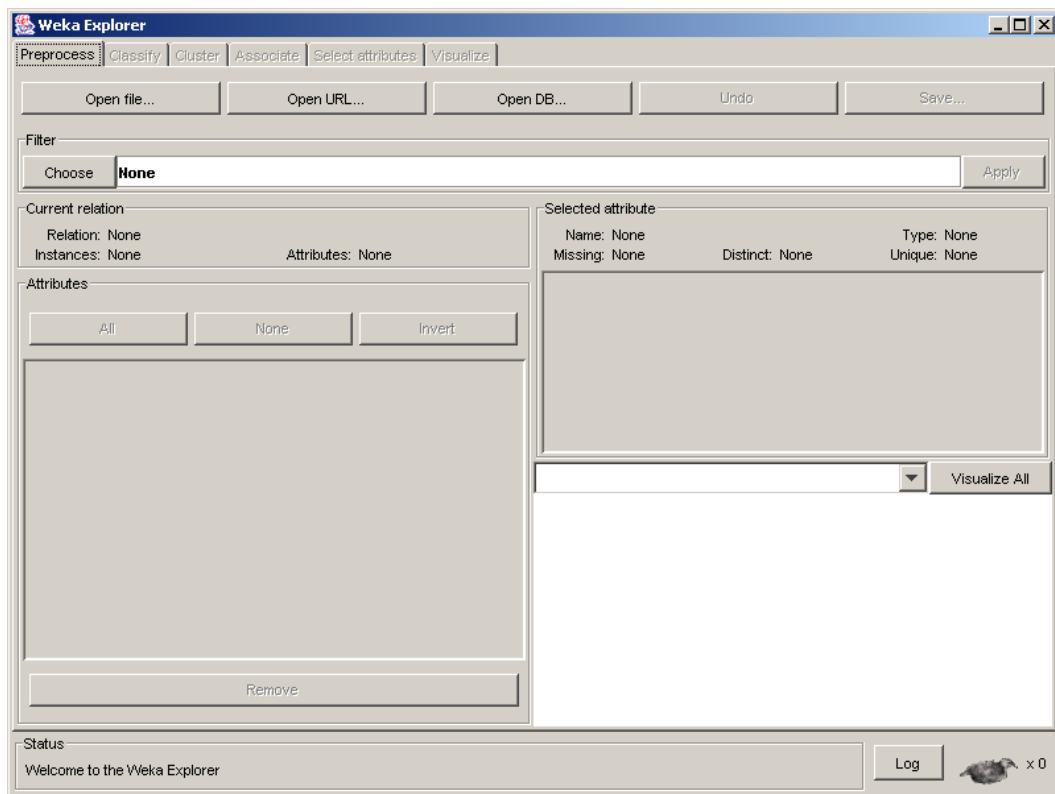
```
@data
```

```
chuvoso,frio,normal,falso,sim
chuvoso,agradável,alta,falso,sim
encoberto,quente,alta,falso,sim
sol,quente,alta,verdade,nao
sol,quente,alta,falso,nao
chuvoso,frio,normal,verdade,nao
```

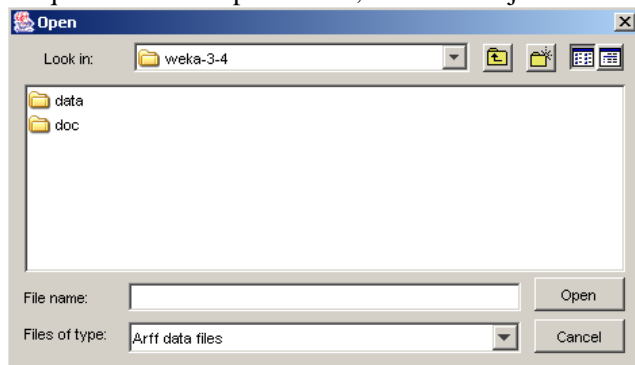
encoberto,frio,normal,verdade,sim  
sol,agradável,alta,falso,nao  
sol,frio,normal,falso,sim  
chuvoso,agradável,normal,falso,sim  
sol,agradável,normal,verdade,sim  
encoberto,agradável,alta,verdade,sim  
encoberto,quente,normal,falso,sim  
chuvoso,agradável,alta,verdade,nao

**Salvamos como arquivo de texto. Com o nome condicoes.arff.**

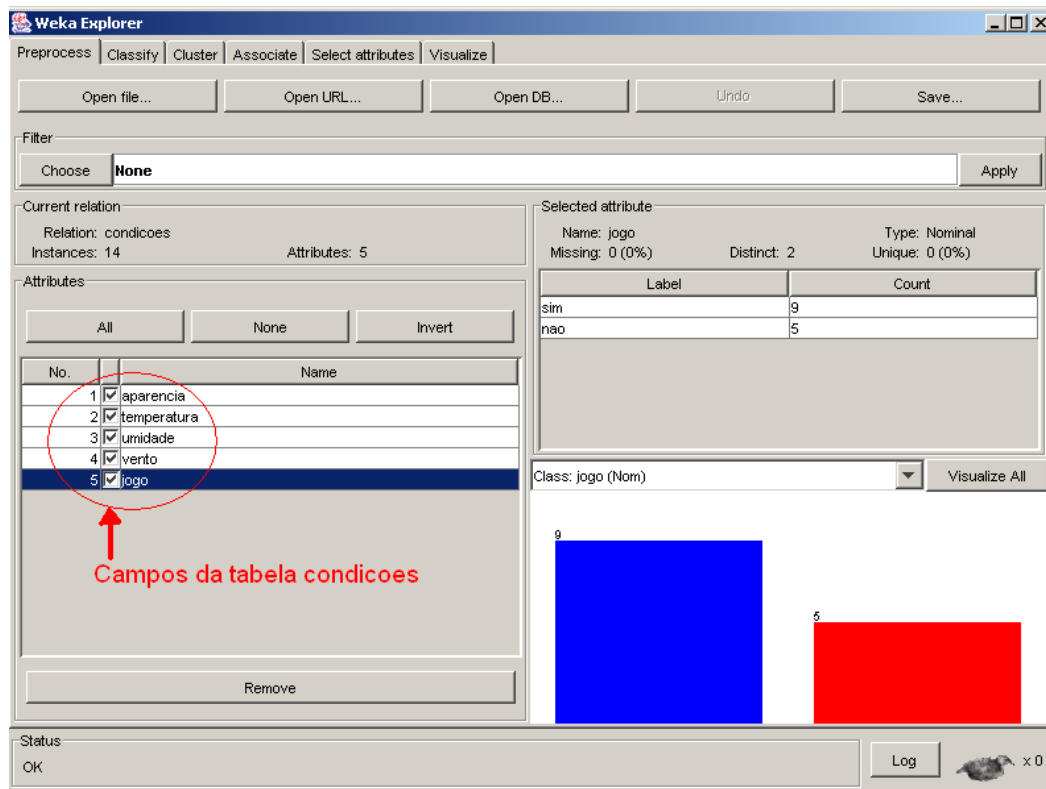
Clicamos no botão “Explorer”, logo será aberto a seguinte janela:



Clique no botão “Open file...”, abrirá uma janela de abertura de arquivo:



Coloque a localização do arquivo condições.arff e clique no botão “open”, a janela WEKA Explorer mostra os campos, marcamos todos os campos para serem minerados:



Localize no WEKA a informação de quantos atributos ele possui.

---

Veja no WEKA quantas instâncias existe no conjunto de dados digitado.

---

Clique no botão Visualize All. Explique o que significa as barras azuis e vermelhas. **Dica: Veja na caixa de seleção qual é o atributo classe (class).**

---

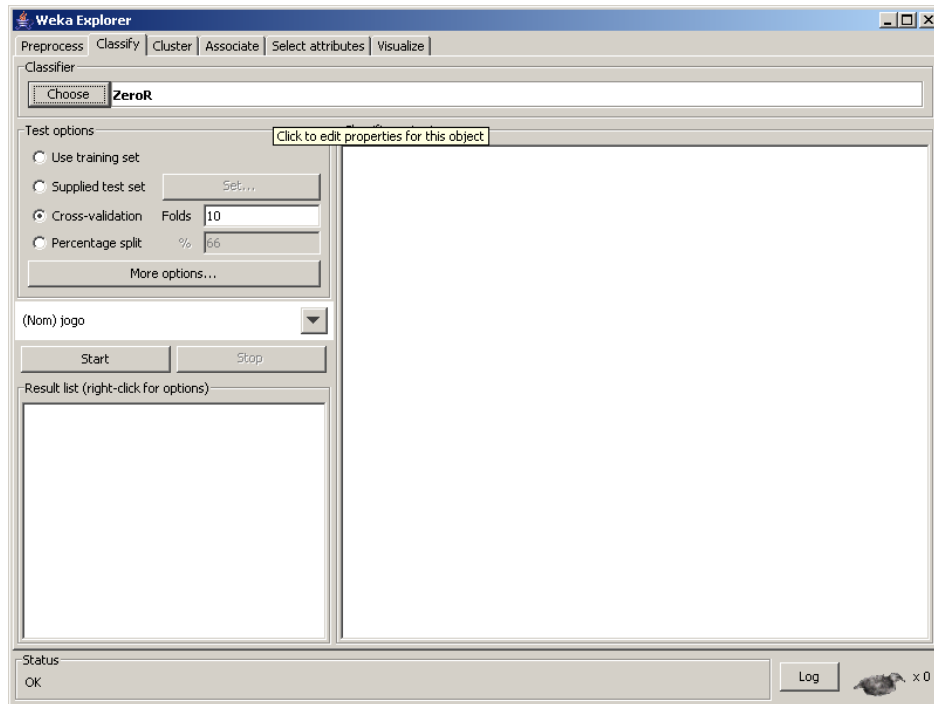


---



---

Vamos agora classificar os dados, clique na guia *classify*, será mostrado a janela abaixo:



Em sua opinião para que serve uma mineração de dados que visa classificá-lo a partir de um atributo alvo (também chamado de atributo-classe)?

---

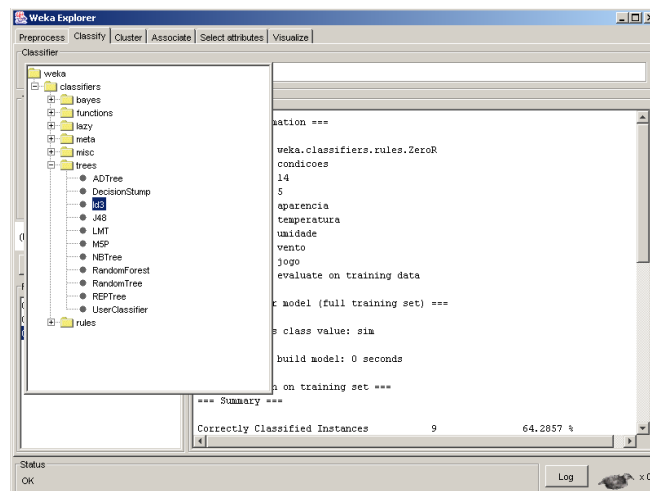


---

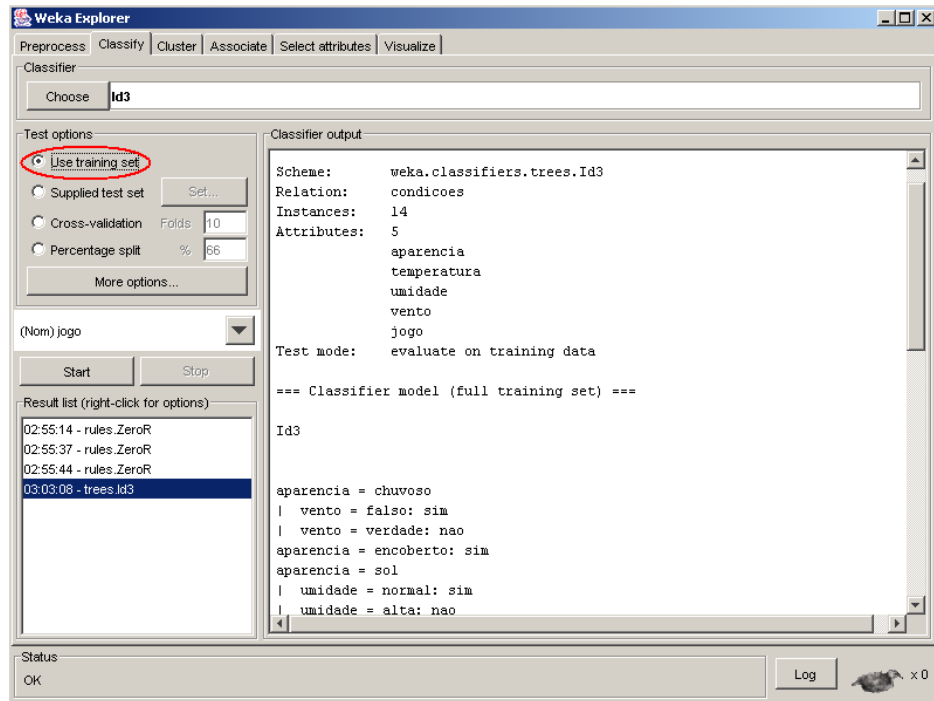


---

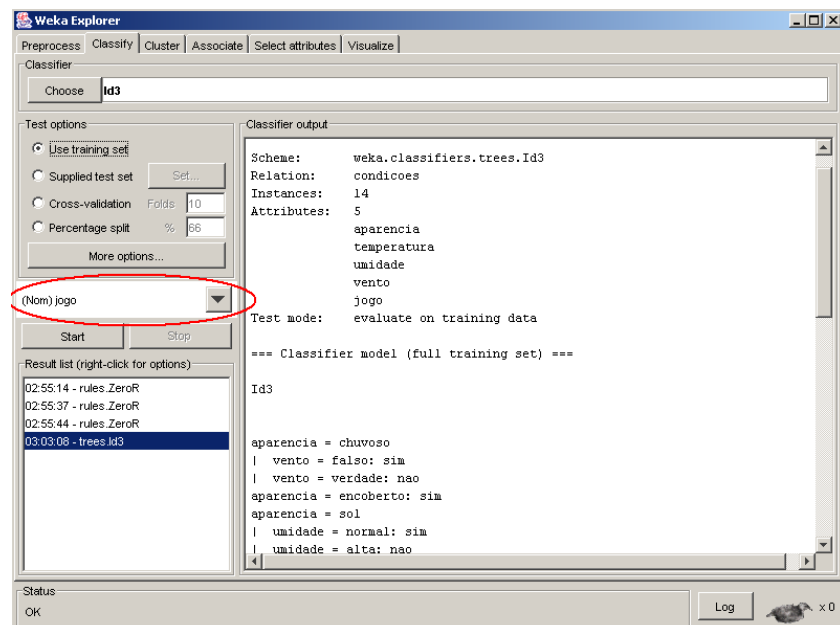
Para selecionar o tipo do algoritmo que foi implementado, clique no botão “*choose*”. Em nosso caso o algoritmo a ser utilizado será o ID3. Para isto, selecione dentro da pasta *weka*, a pasta *classifiers*, dentro dela o nó *trees* e logo após *id3*. Como mostra a figura abaixo:



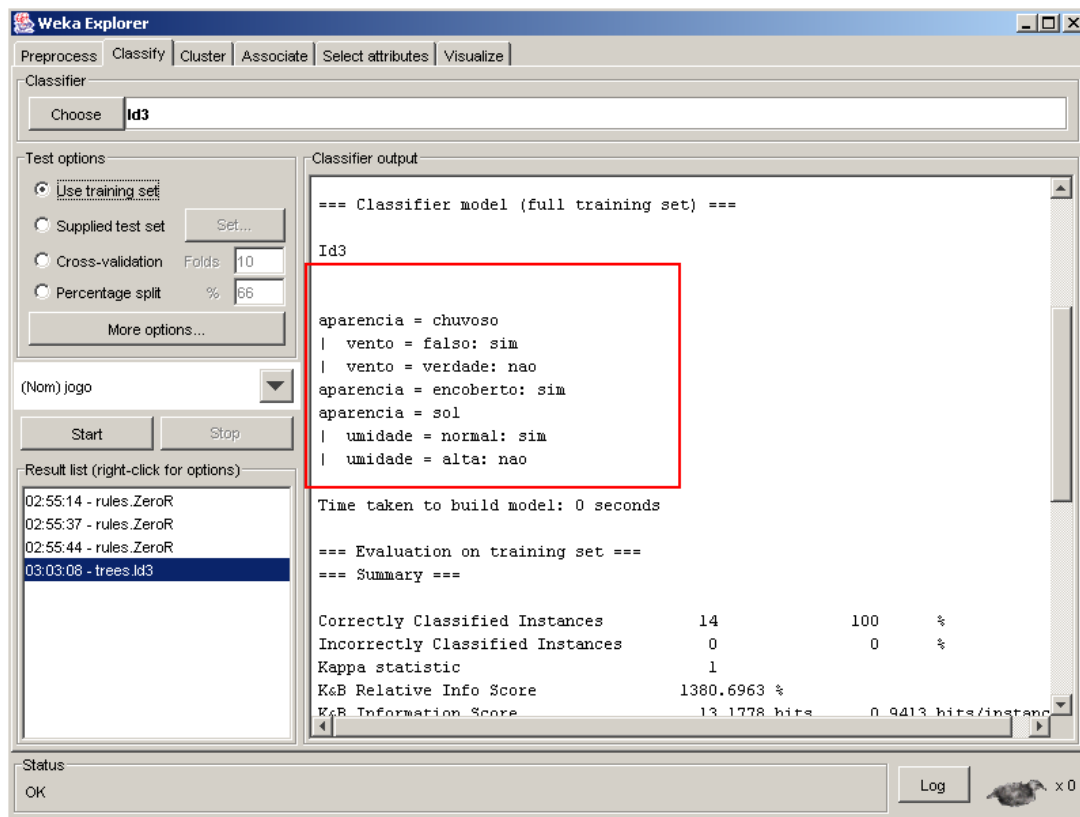
Agora que selecionamos o tipo do algoritmo, vamos clicar em “*use training test*”, para dizer que estamos na fase de treinamento do método de classificação (o mesmo conjunto de treinamento será utilizado para teste do classificador), como mostra a figura:



Selecione o atributo classe, que em nosso caso é *jogo*, como mostra a figura abaixo:



Clique no botão “start”, o weka irá minerar mostrando na parte direita da janela as seguintes regras:



Localize no WEKA o percentual de instâncias corretamente classificadas. Em sua opinião, a saída do algoritmo gerou uma boa classificação?

A partir da árvore gerada, informe uma condição para que não haja jogo de futebol.

É importante que deve ser analisado a estatística de Kappa, que informa se os dados estão bem classificados ou não, Kappa > 0,80 indica uma boa classificação dos dados.

Quanto foi o percentual Kappa do classificador?



Clique no botão “**choose**”. Vamos agora utilizar o algoritmo J48 Para isto, selecione dentro da pasta **weka**, a pasta **classifiers**, dentro dela o nó **trees** e logo após **J48**.

## TER PÊLOS

## TER DENTES

## TER PENAS

TENTE MINERAR OS DADOS RETIRANDO ATRIBUTOS. CLIQUE NA ABA **PREPROCESS** E RETIRE ALGUNS ATRIBUTOS.

[illegible]





### Descobrimos regras de associação.

Vamos descobrir alguns fatos que se repetem nestes dados. Clique na ABA ASSOCIATE e execute o algoritmo **APRIORI**.

Pesquise na internet para que serve as regras de associação.

---

---

---

---

---

---

---

---

---

---

Escreva abaixo um pouco do que você descobriu aplicando as regras de associação no arquivo ZOO.ARF.

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

### QUESTÃO DESAFIO

Consiga uma base de dados qualquer e aplique as técnicas de classificação e associação. Escreva o que você achou durante a mineração de dados. Esta resposta deve contemplar uma explicação sobre a base de dados e a explicação dos conhecimentos aprendidos a partir da mineração de dados.