

## Laboratório - 3 - Arquivo: jogar\_volei.arff

Os dados são relativos a informações meteorológicas levantadas em dias de jogos de volei realizados ou não. São 14 dados sobre umidade, temperatura, vento, dentre outros, que refletem a atitude dos jogadores em jogar ou não a partida de acordo com o tempo.

**Pergunta:** Caso você acordasse pela manhã e desejasse jogar volei, qual seria sua decisão baseada nos dados informados.

**Resposta:** Minere os dados através do algoritmo de classificação **EM** e estabeleça uma decisão de acordo com os atributos disponibilizados.

### Passo a Passo

Siga o Passo a Passo e analise os resultados:

Desenvolvido com o objetivo de criar agrupamento de dados, conhecida como técnicas de agrupamento (clustering) que explorem semelhanças entre padrões e agrupem os padrões parecidos em categorias ou grupos. Essa mineração de dados criará a estrutura de um conjunto de dados por dois rótulos **cluster0** e **cluster1** interprete os dados obtidos.

Vamos minerar a seguinte base de dados:

Tabela condicoes: DECISÕES SOBRE JOGAR OU NÃO VOLEI

PREVISAO	TEMPERATURA	UMIDADE	VENTO	JOGAR
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

### **PASSO 1 - Minerando com o software WEKA**

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

O resultado é o seguinte:

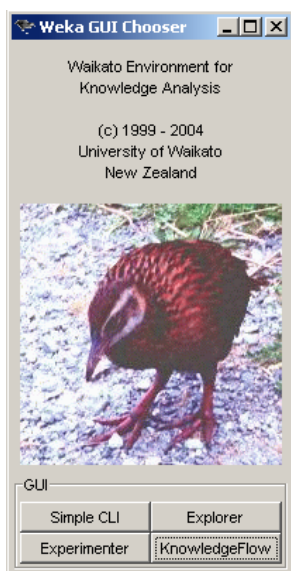
```
@relation weather.symbolic
```

```
@attribute previsao {ensolarado, nublado, chuvoso}  
@attribute temperatura {quente, brando, frio}  
@attribute umidade{alta, normal}  
@attribute vento {ventando, nao_ventando}  
@attribute jogar {sim,nao}
```

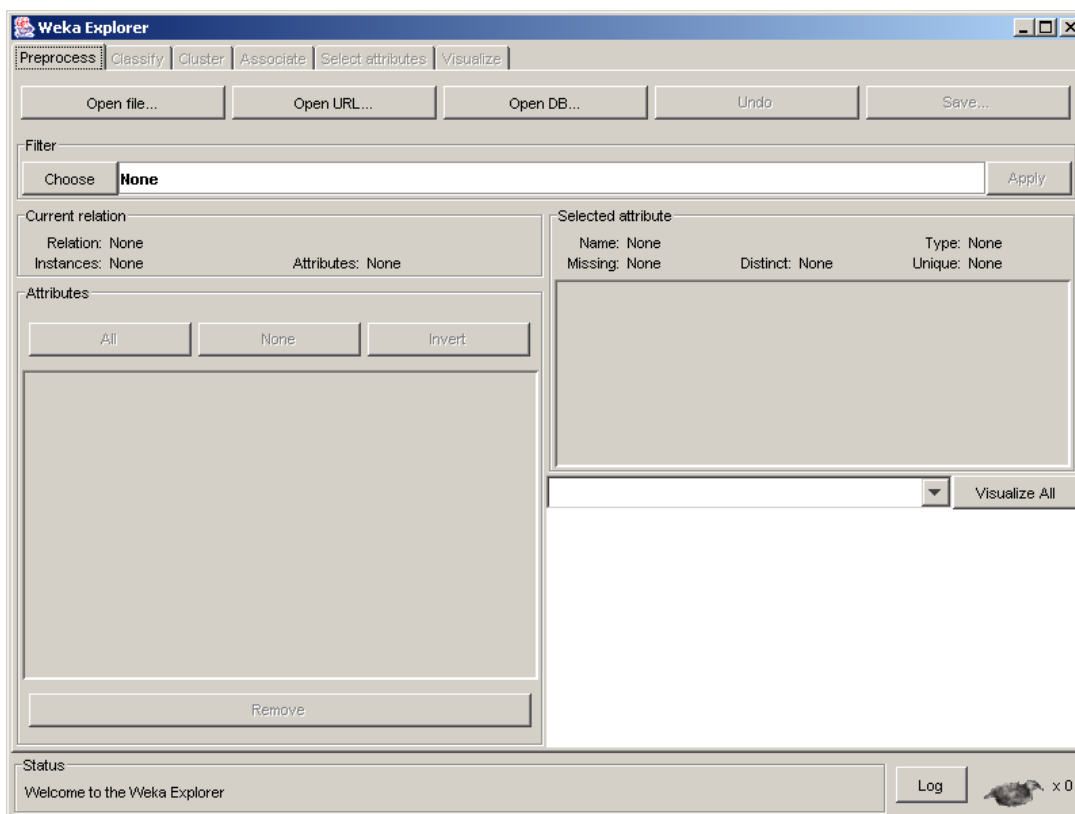
```
@data  
ensolarado,quente,alta,nao_ventando,nao  
ensolarado,quente,alta,ventando,nao  
nublado,quente,alta,nao_ventando,sim  
chuvoso,brando,alta,nao_ventando,sim  
chuvoso,frio,normal,nao_ventando,sim  
chuvoso,frio,normal,ventando,nao  
nublado,frio,normal,ventando,sim  
ensolarado,brando,alta,nao_ventando,nao  
ensolarado,frio,normal,nao_ventando,sim  
chuvoso,brando,normal,nao_ventando,sim  
ensolarado,brando,normal,ventando,sim  
nublado,brando,alta,ventando,sim  
nublado,quente,normal,nao_ventando,sim  
chuvoso,brando,alta,ventando,nao
```

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

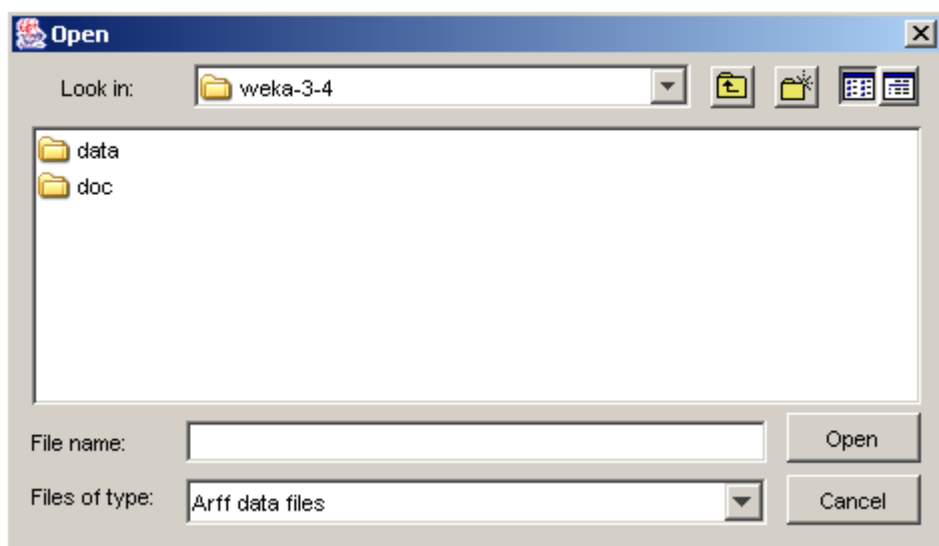
Salvamos como arquivo de texto. Com o nome condicoes.arff.  
Abrimos o WEKA, apresentará a tela abaixo:



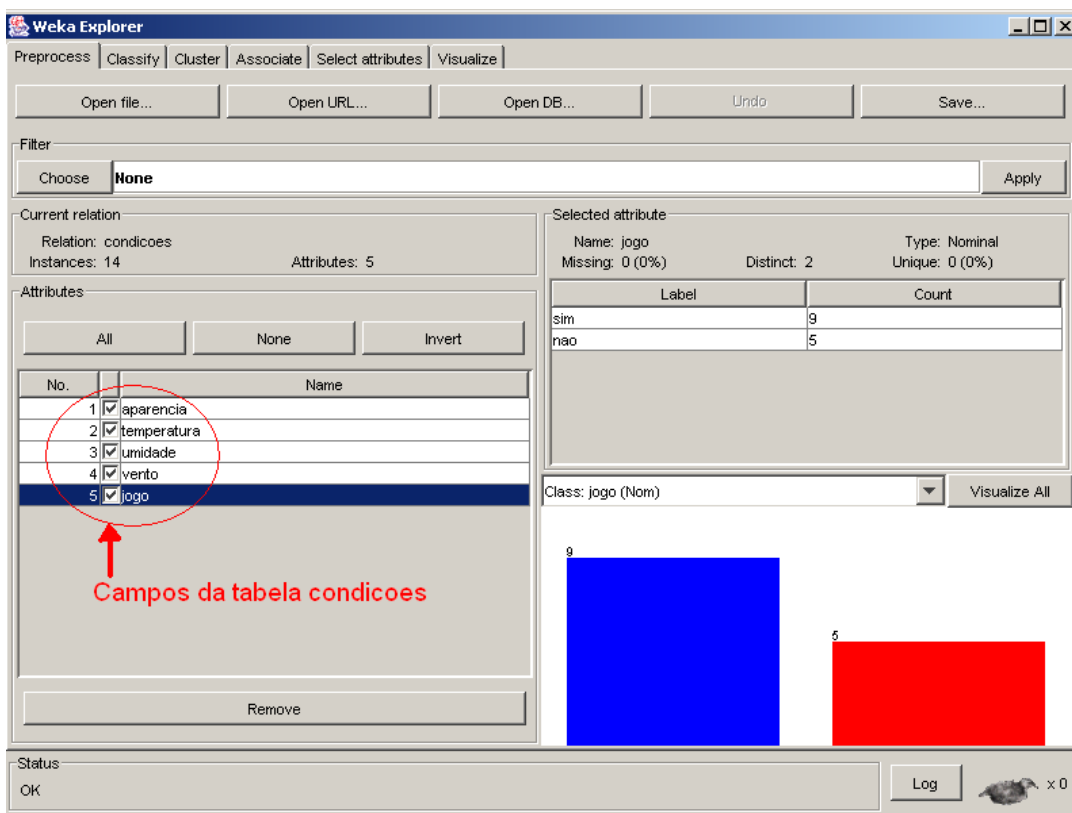
Clicamos no botão “Explorer”, logo será aberto a seguinte janela:



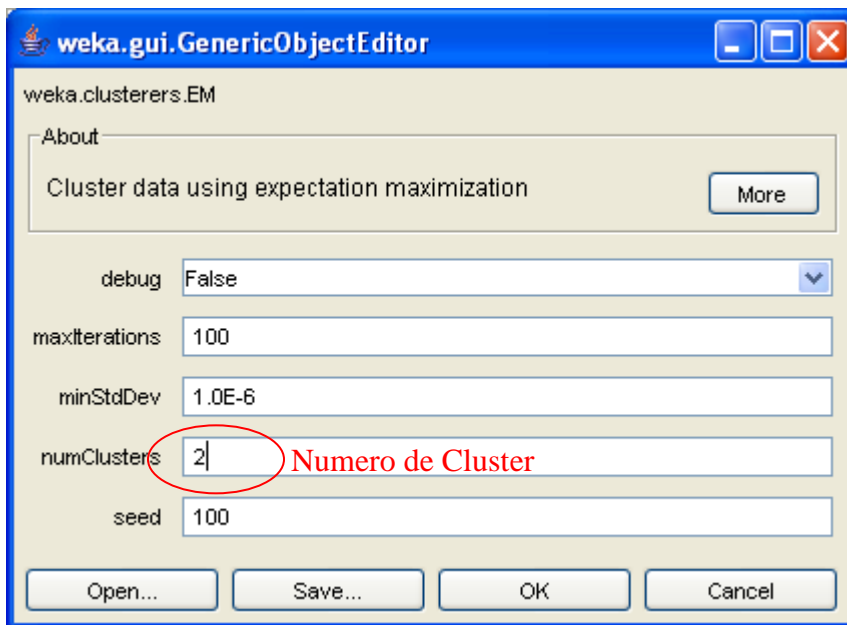
Clicamos no botão “Open file...”, abrirá uma janela de abertura de arquivo:



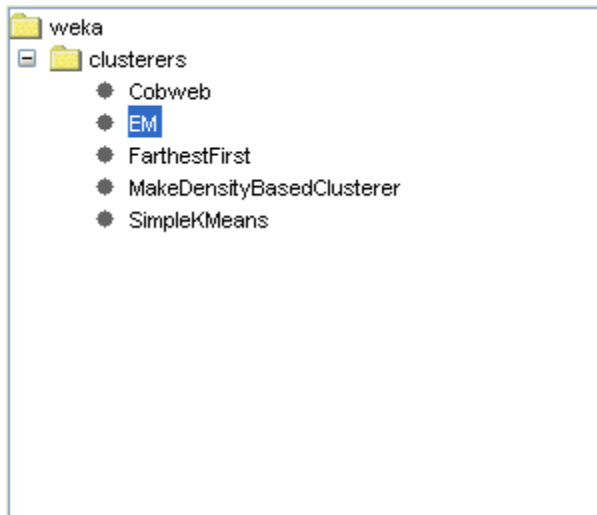
Colocamos a localização do arquivo condições.arff e clicamos no botão “**open**”, a janela WEKA Explorer mostra os campos, marcamos todos os campos para serem minerados:



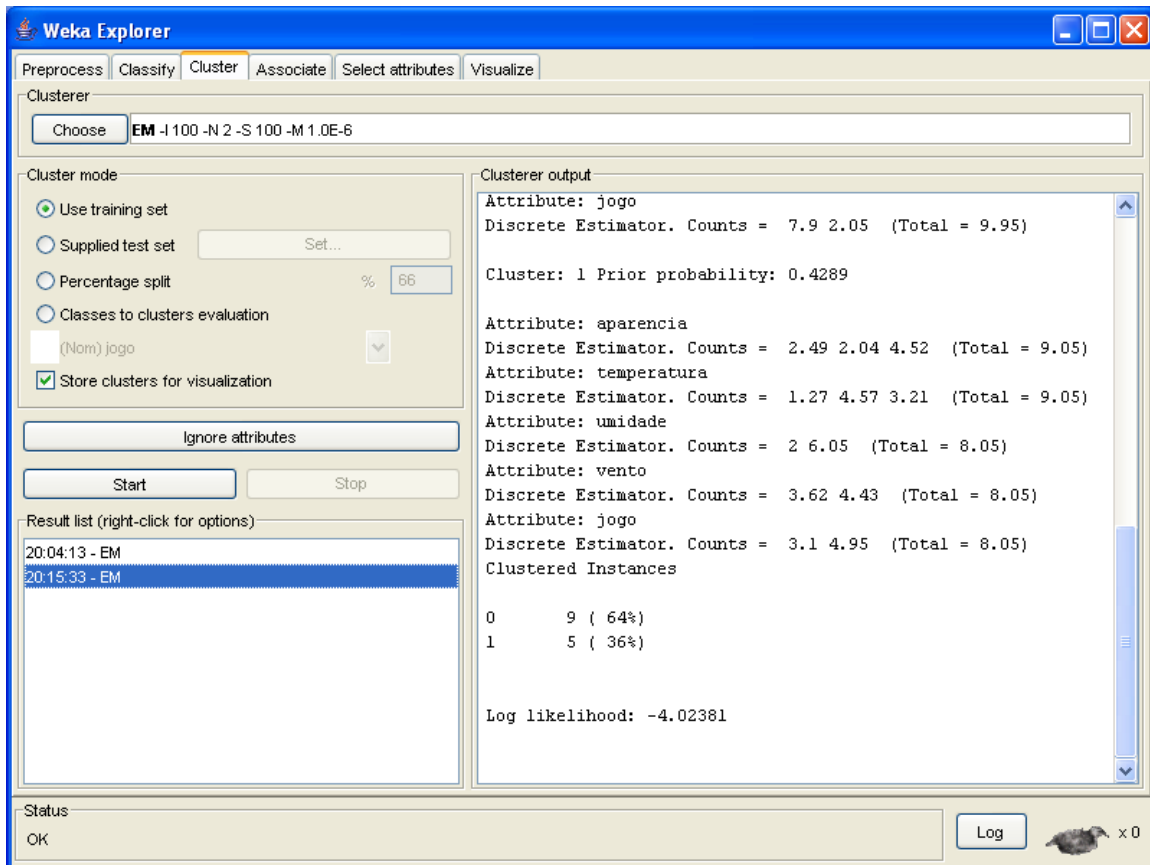
Logo após clicamos na guia **Cluster**, Clique sobre o nome do algoritmo **EM**:



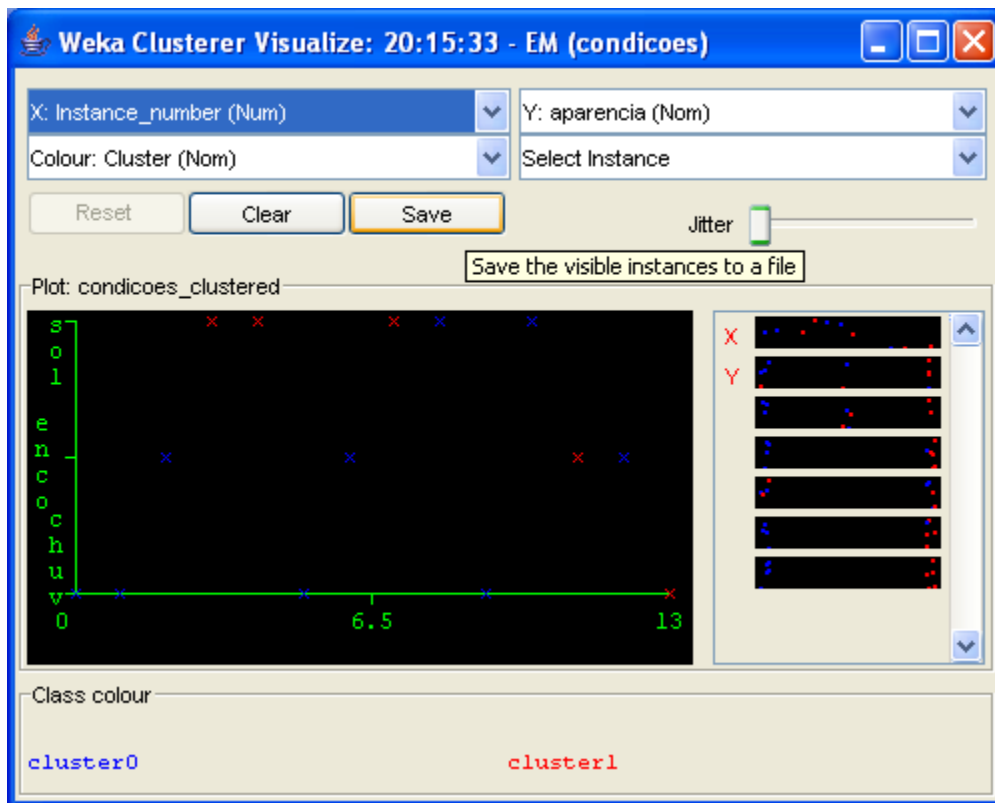
Mude o número de cluster para **2**, para selecionar o tipo do algoritmo que foi implementado clicamos no botão “**choose**”, em nosso caso foi o algoritmo **EM**, para isso selecione dentro da pasta weka, a pasta **Cluster**. Como mostra a figura abaixo:



Agora que selecionamos o tipo do algoritmo, vamos clicar em “**use training test**”, para dizer que estamos na fase de treinamento do método de cluster, como mostra a figura:



Analizamos o **Log likelihood**, quando **maior**(valores absolutos) for esse valor, melhor será a definição dos grupos, logo após clicamos no botão “direito”, sobre o “Result List” e solicitamos para exibir o gráfico dos cluster e em seguida **salvar** os dados.



Para melhor visualização dos cluster, abra-os em um editor de texto:  
Abaixo é visto o arquivo gerado com a separação dos cluster:

```
@attribute Instance_number numeric
@attribute previsao {ensolarado,nublado,chuvoso}
@attribute temperatura {quente,brando,frio}
@attribute umidade {alta,normal}
@attribute vento {ventando,nao_ventando}
@attribute jogar {sim,nao}
@attribute Cluster {cluster0,cluster1}

@data
0,ensolarado,quente,alta,nao_ventando,nao,cluster1
1,ensolarado,quente,alta,ventando,nao,cluster1
2,nublado,quente,alta,nao_ventando,sim,cluster0
3,chuvoso,brando,alta,nao_ventando,sim,cluster0
4,chuvoso,frio,normal,nao_ventando,sim,cluster0
5,chuvoso,frio,normal,ventando,nao,cluster0
6,nublado,frio,normal,ventando,sim,cluster0
```

7,ensolarado,brando,alta,nao\_ventando,nao,cluster1  
8,ensolarado,frio,normal,nao\_ventando,sim,cluster0  
9,chuvoso,brando,normal,nao\_ventando,sim,cluster0  
10,ensolarado,brando,normal,ventando,sim,cluster0  
11,nublado,brando,alta,ventando,sim,cluster1  
12,nublado,quente,normal,nao\_ventando,sim,cluster0  
13,chuvoso,brando,alta,ventando,nao,cluster1