

## Sistemas Inteligentes – Aula 03

### Classificação e Árvores de Decisão

#### Como Gerar uma Árvore de Decisão Usando o Weka

A ferramenta Weka (Weka, 2013) permite gerar Árvores de Decisão de forma automática ou interativa. Vamos mostrar a geração automática a partir de um arquivo de entrada.

**Passo 1** - Uma forma rápida de criar o arquivo de entrada tipo “.arff” a partir de uma planilha de dados “.xls” (Fig. 3.13(a)) é salvá-la no formato “.csv” (Fig. 3.13(b)), depois abrir este arquivo num editor de texto, acrescentar algumas palavras-chave e salvar novamente (Fig. 3.14(a)). Saia do editor de texto e mude manualmente a extensão do arquivo de “.csv” para “.arff” (Fig. 3.14(b)). Note que as Fig. 3.14(a) e (b) são idênticas, mas a extensão dos arquivos é diferente.

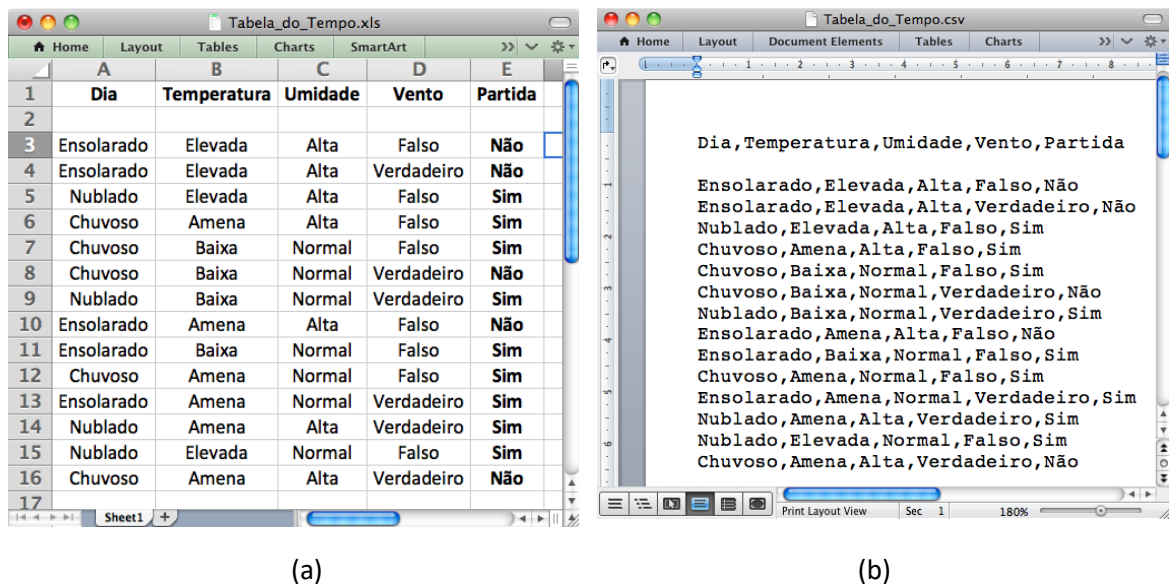


Fig. 3.13 – Tabela do Tempo (a) formato “.xls” e (b) formato “.csv”.

Na unidade anterior, sobre como gerar Regras de Associação no Weka, foi explicado que o Weka aceita trabalhar diretamente o formato “.csv”, sem a necessidade de acrescentar palavras-chaves. Mas, em alguns casos a criação do arquivo “.arff” se justifica. Em caso de dúvida sobre esta questão, consulte o material da unidade anterior.

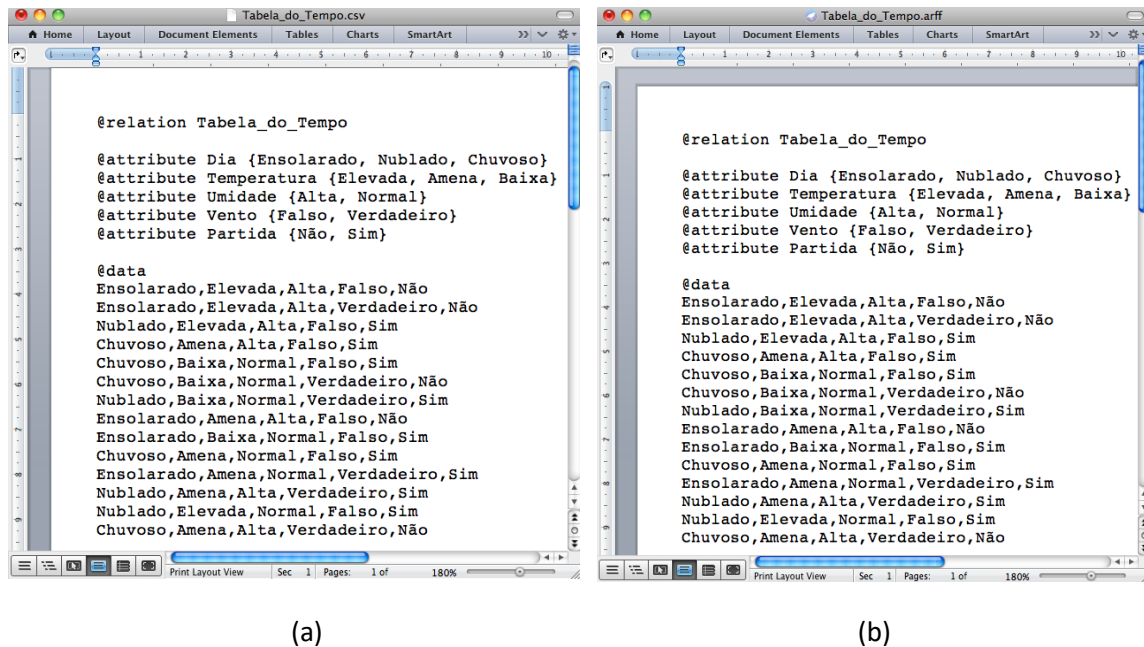


Fig. 3.14 – Arquivo Tabela\_do\_Tempo (a) “.csv” com Palavras-Chave, e (b) Arquivo “.arff”.

**Passo 2** - Dando dois cliques sobre o ícone do arquivo “Tabela\_do\_Tempo.arff” ele se abrirá dentro do Weka, mostrando uma figura semelhante à Fig. 3.15. Alternativamente, é possível primeiramente abrir o Weka Explorer, depois “Open file...” e localizar o arquivo “Tabela\_do\_Tempo.arff”.

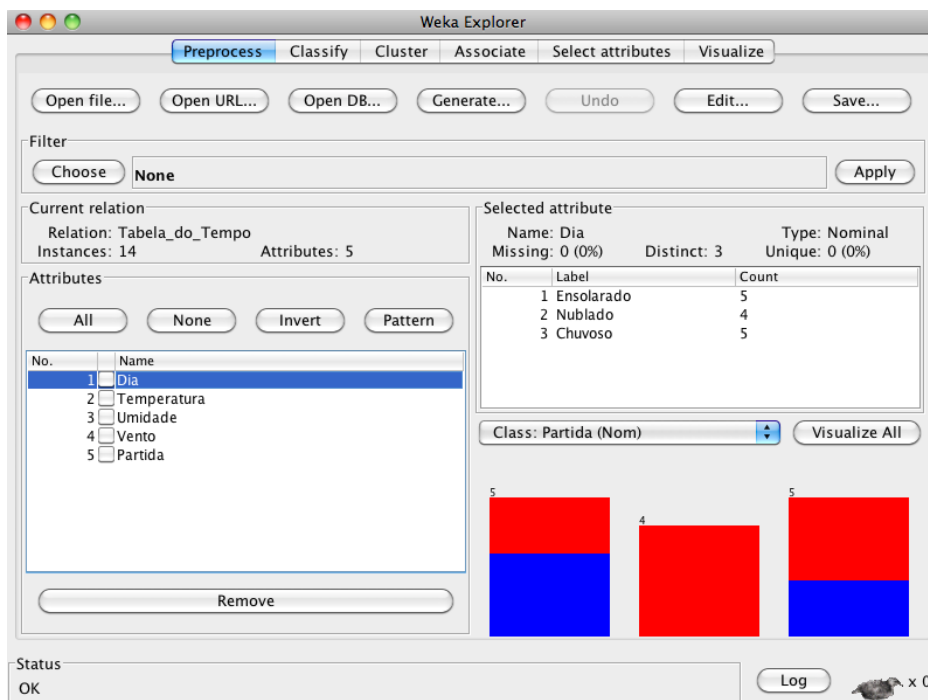


Fig. 3.15 – Arquivo “Tabela\_do\_Tempo.arff” aberto no Weka

Note que na seção “Attributes” aparecem os cinco atributos da Tabela do Tempo na ordem em que foram declarados. Na Fig. 3.15 aparecem ainda os detalhes da composição de um dos atributos selecionados, neste caso “Dia”, mas qualquer um dos quatro atributos restantes pode ser selecionado.

**Passo 3** – Com o arquivo “Tabela\_do\_Tempo.arff” aberto, clique na aba “Classify”, localizada da parte superior esquerda da janela do Weka Explorer. A seguir, clique em “Choose” para escolher o algoritmo de classificação. Primeiro clique em “trees”, depois em “J48” (Fig. 3.16). O algoritmo “J48” é uma implementação mais recente do “ID3” e, além disso, tem também a vantagem de permitir, dentro do Weka, visualizar a Árvore de Decisão construída.

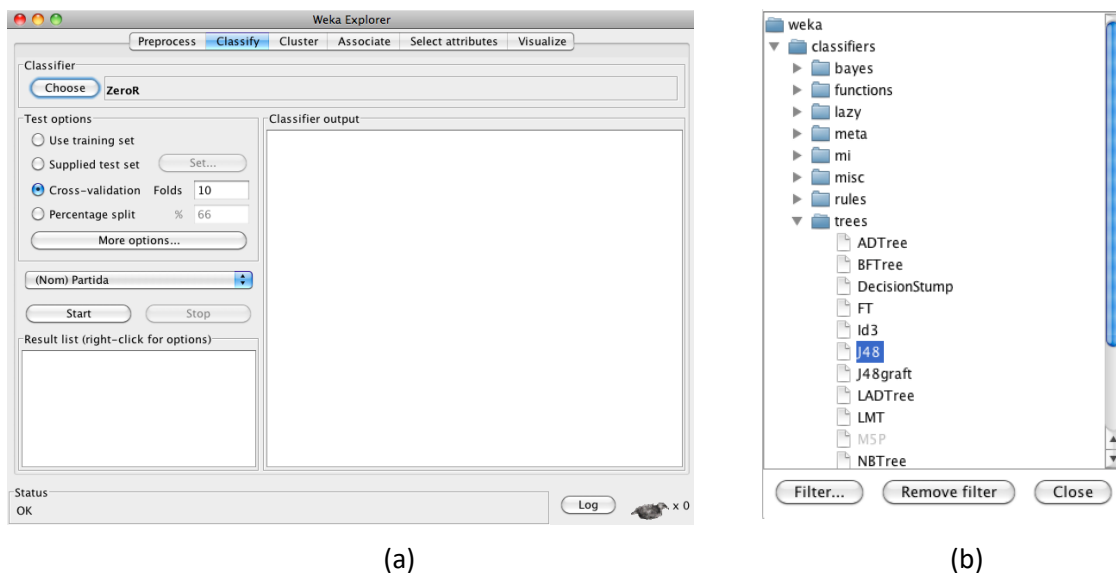


Fig. 3.16 – Aba “Classify” com a Opção (a) “Choose” para escolher (b) o menu de algoritmos.

Uma vez escolhido o algoritmo “J48” é possível conferir e alterar alguns parâmetros. Clicando com o botão esquerdo do mouse sobre “J48” abre-se um menu com todos os valores *default*. Inicialmente vamos trabalhar com estes valores.

**Passo 4** – Ainda dentro da aba “Classify”, em “Test options” escolha “Use training set” e clique no botão “Start”, um pouco abaixo. A opção “Use training set” significa que o mesmo **Conjunto de Treinamento** usado para gerar a Árvore de Decisão será usado para testar os resultados. Se o Conjunto de Exemplos da Base de Dados não for inconsistente, geralmente a taxa de acerto com esta opção deve ser de 100%. Mais adiante, na próxima unidade de estudo, veremos que há outras formas mais interessantes de testar a robustez e a qualidade do Modelo gerado.

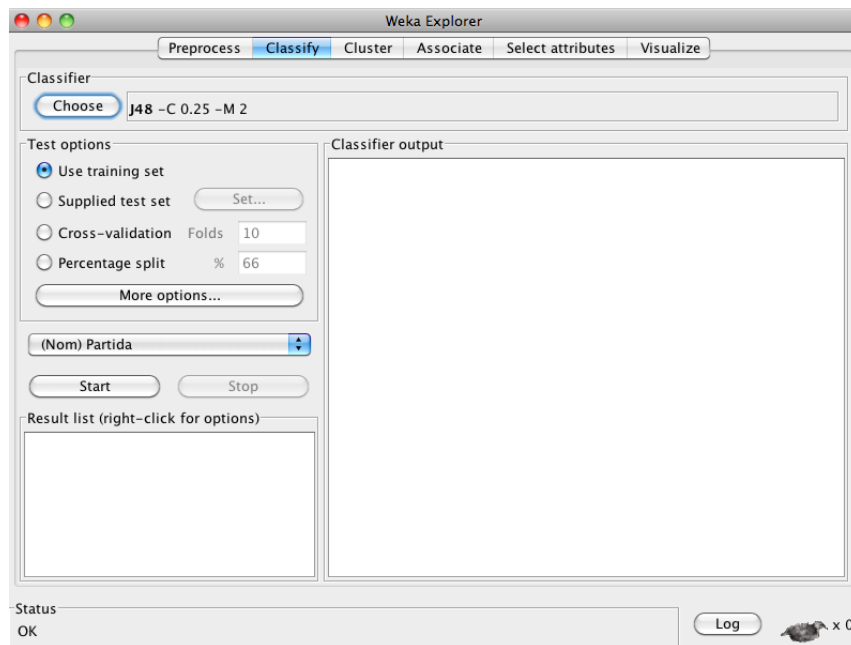


Fig. 3.17 – A Escolha da Opção de Teste “Use training set”.

**Passo 5** – Ao disparar o processo de treinamento com o algoritmo “J48” aparecem na região direita da tela (“Classifier output”) os resultados desejados (Fig. 3.18). A Árvore de Decisão aparece na forma textual, mas pode ser vista na forma gráfica. Na parte inferior da tela, aparecem o número e a porcentagem de exemplos classificados corretamente, a Acurácia (ou Precisão) por Classe e a Matriz de Confusão.

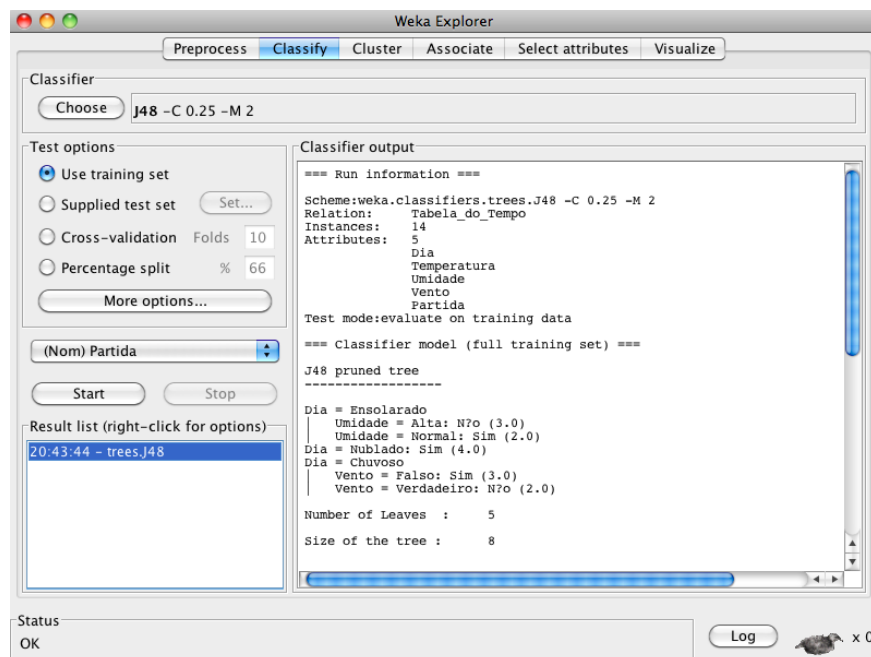
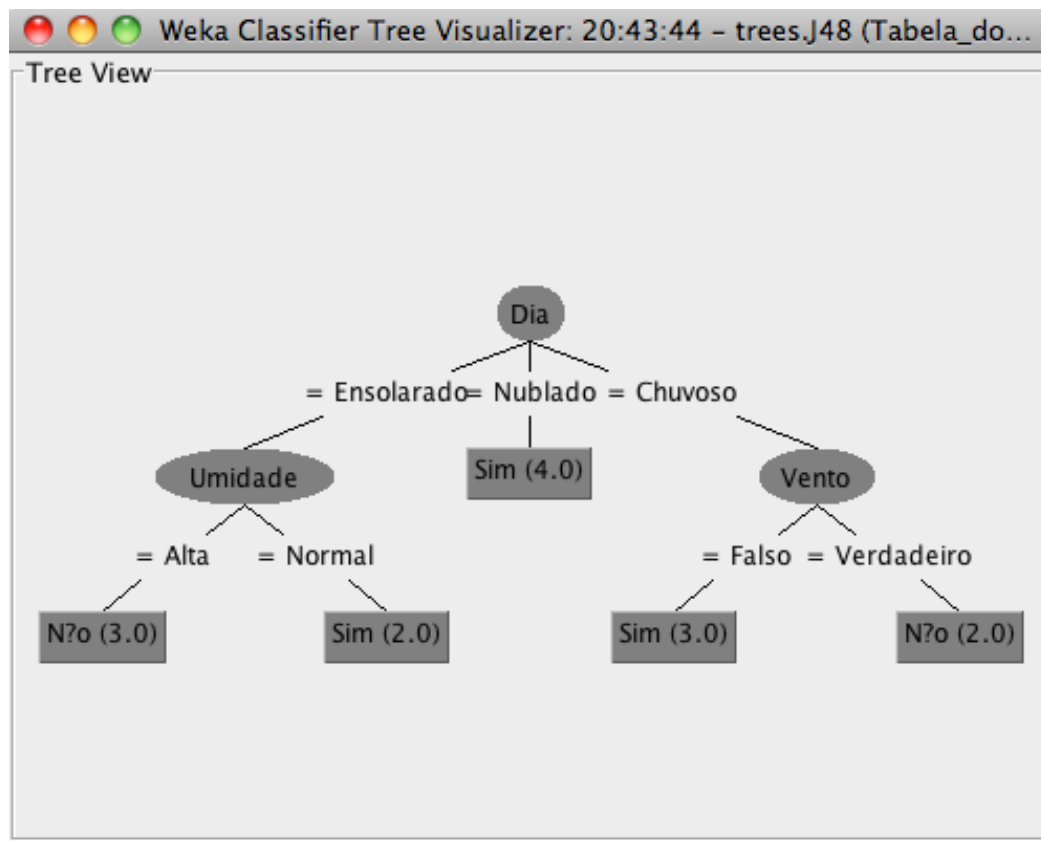


Fig. 3.18 – Resultado do Processo de Treinamento e Indução da Árvore de Decisão.

Se você tiver interesse em saber como cada uma dos exemplos foi classificado, clique na aba “More options...” e depois habilite “Output predictions”. Clique em “Start” novamente.

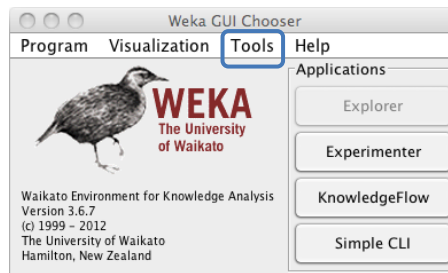
**Passo 6** – Clicando com o botão direito do mouse na região inferior esquerda, onde se lê “Result list (right-click for options)”, mais precisamente sobre a faixa azul “trees.J48”, é possível visualizar graficamente o resultado, escolhendo “Visualize tree” (Fig 3.19).



**Fig. 3.19 – Representação Gráfica da Árvore de Decisão.**

Os números entre parênteses em cada nó folha da Fig. 3.19 indicam quantos exemplos chegaram até esta folha. Somando-se estes números, verifica-se que 14 exemplos foram testados nesta simulação.

**Passo 7** – Se você quiser saber exatamente quais Exemplos foram classificados em quais classes, primeiramente abra o arquivo “Tabela\_do\_Tempo.arff” no editor do Weka, que pode ser acessado pelo seguinte caminho: interface “Weka GUI Chooser”, depois “Tools”, e depois “ArffViewer” (Fig. 3.20).



**Fig. 3.20 – Interface “Weka GUI Chooser”.**

Quando a janela “ARFF-Viewer” se abrir, localize o arquivo “Tabela\_do\_Tempo.arff” clicando primeiramente em “File” e depois em “Open...”.

Com o editor do Weka aberto, clique no atributo “Umidade”, ou qualquer outro atributo, para ordenar os Exemplos de acordo com os valores que este atributo pode assumir.

No.	Dia Nominal	Temperatura Nominal	Umidade Nominal	Vento Nominal	Partida Nominal
1	Ensolarado	Elevada	Alta	Falso	N?o
2	Ensolarado	Elevada	Alta	Verdadeiro	N?o
3	Nublado	Elevada	Alta	Falso	Sim
4	Chuvoso	Amena	Alta	Falso	Sim
8	Ensolarado	Amena	Alta	Falso	N?o
12	Nublado	Amena	Alta	Verdadeiro	Sim
14	Chuvoso	Amena	Alta	Verdadeiro	N?o
5	Chuvoso	Baixa	Normal	Falso	Sim
6	Chuvoso	Baixa	Normal	Verdadeiro	N?o
7	Nublado	Baixa	Normal	Verdadeiro	Sim
9	Ensolarado	Baixa	Normal	Falso	Sim
10	Chuvoso	Amena	Normal	Falso	Sim
11	Ensolarado	Amena	Normal	Verdadeiro	Sim
13	Nublado	Elevada	Normal	Falso	Sim

**Fig. 3.21 – Tabela do Tempo Ordenada pelos Valores de “Umidade”.**

Note que com este editor, é possível alterar os valores dos atributos, clicando sobre o campo a ser alterado.

Obs.: O editor do Weka também pode ser acessado no “Weka Explorer”, escolhendo a aba “Preprocess” e depois “Edit...”.

**Passo 8** – Para testar novos Exemplos no Classificador sem sair da interface “Weka Explorer”, há a opção “Supplied test set”. Vamos supor que os quatro novos Exemplos da Tab. 3.16 de nosso material de Teoria devam ser testados no “Weka Explorer”.

Primeiramente crie um arquivo “.arff” com os quatro Exemplos extras da Tab. 3.16 (basta abrir o arquivo “Tabela\_do\_Tempo.arff” no editor do Weka, fazer as modificações e salvar com um novo nome, digamos “Teste.arff”).

Carregue no “Weka Explorer” o **Conjunto de Treinamento** “Tabela\_do\_Tempo.arff” da forma usual, i.e., clicando no “Preprocess” da barra superior e depois em “Open file...”. A seguir clique em “Classify”, escolha em “Choose” o algoritmo desejado, digamos “J48”, e em “Test options” escolha “Supplied test set”. Pressionando a tecla “Set”, aparece a opção “Open file...” com a qual é possível carregar o **Conjunto de Teste** “Teste.arff”.

Em “More options” habilite a opção “Output predictions” e dispare o programa com a opção “Start”. Na seção “Classifier output”, devem aparecer as quatro predições buscadas.

### Referência Bibliográfica

Weka. The Waikato University. In <http://www.cs.waikato.ac.nz/ml/weka/> . Acessado em 03.03.13.

WITTEN, I. H. & FRANK, E. **Data Mining: Practical Machine Learning Tools and Techniques**. Second Edition. Amsterdam: Morgan Kaufmann Publishers, 2005.