

Redes Neurais



Cruzeiro do Sul Virtual
Educação a distância

Material Teórico



Estudos de Caso com Redes Neurais

Responsável pelo Conteúdo:

Prof. Dr. Alberto Messias da Costa Souza

Revisão Textual:

Prof. Me. Luciano Vieira Francisco



- Estudo de Caso com o *Software Weka*;
- Segundo Estudo de Caso com Weka;
- Anexo I.

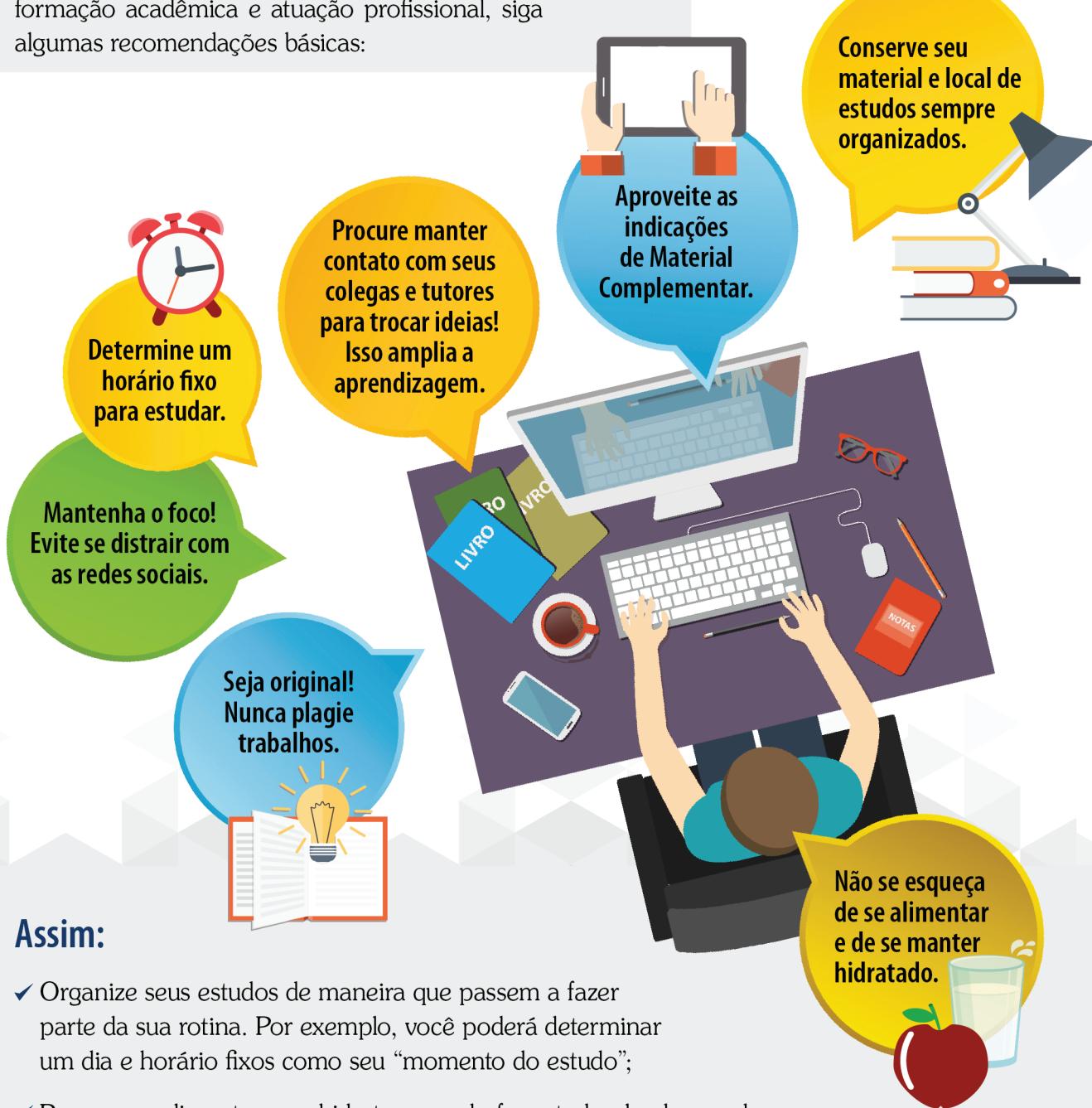


OBJETIVO DE APRENDIZADO

- Conhecer dois estudos de caso utilizando o *software Weka* e as bases de dados *Íris Dataset*, assim como uma base de dados na qual se pretende predizer o diagnóstico de câncer de mama.

Orientações de estudo

Para que o conteúdo desta Disciplina seja bem aproveitado e haja maior aplicabilidade na sua formação acadêmica e atuação profissional, siga algumas recomendações básicas:



Assim:

- ✓ Organize seus estudos de maneira que passem a fazer parte da sua rotina. Por exemplo, você poderá determinar um dia e horário fixos como seu “momento do estudo”;
- ✓ Procure se alimentar e se hidratar quando for estudar; lembre-se de que uma alimentação saudável pode proporcionar melhor aproveitamento do estudo;
- ✓ No material de cada Unidade, há leituras indicadas e, entre elas, artigos científicos, livros, vídeos e sites para aprofundar os conhecimentos adquiridos ao longo da Unidade. Além disso, você também encontrará sugestões de conteúdo extra no item **Material Complementar**, que ampliarão sua interpretação e auxiliarão no pleno entendimento dos temas abordados;
- ✓ Após o contato com o conteúdo proposto, participe dos debates mediados em fóruns de discussão, pois irão auxiliar a verificar o quanto você absorveu de conhecimento, além de propiciar o contato com seus colegas e tutores, o que se apresenta como rico espaço de troca de ideias e de aprendizagem.

Estudo de Caso com o Software Weka

Para o nosso estudo de caso utilizaremos o software *Weka*, que pode ser consideravelmente útil na criação do modelo de dados e, posteriormente, na sua aplicação em grande escala.

Assim, na janela inicialmente aberta *GUI Chooser*, clique no botão *Explorer* para acessar o programa *Weka* propriamente dito.

Utilizaremos uma base de dados já existente e comumente conhecida, denominada *IrisDataSet*, que você encontrará facilmente na *internet* ou dentro da própria pasta de instalação do *Weka*, ou seja, na pasta *data*.

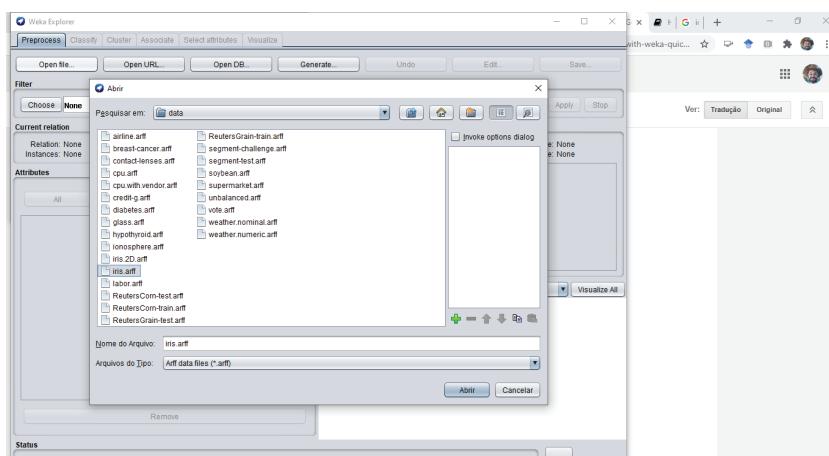


Figura 1 – Selezionando a base de dados *íris.arff*

Fonte: Acervo do Conteudista

Clique no botão *Open File*, navegue até a pasta de instalação do *Weka*, depois na pasta *Data* e selecione o arquivo *íris.arff*. Esta base possui 150 linhas ou instâncias de dados com 4 atributos sobre a flor de íris, sendo comumente utilizada na literatura para problemas de classificação.

O conjunto de dados contém 3 classes de 50 instâncias, cada, referindo-se a um tipo de planta de íris. Uma classe é linearmente separável das outras 2, e estas linearmente separáveis entre si. O atributo previsto é a classe da planta de íris.

O objetivo é criar uma rede neural que classifique uma flor de íris como uma das três espécies (setosa, versicolor ou virgínica) com base em 4 valores numéricos (comprimento e largura da sépala e comprimento e largura da pétala).



Sépala: É uma estrutura em forma de folha.

Após o carregamento do arquivo, o *Weka* inicialmente as divide em classes, conforme pode-se observar na Figura 2:

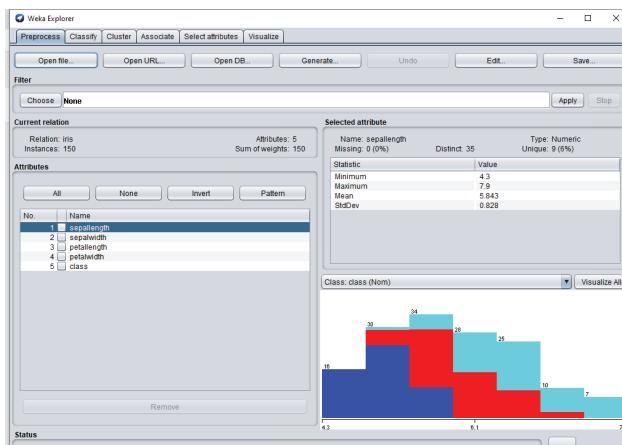


Figura 2 – Carregamento do arquivo *íris.arff*

Fonte: Acervo do Conteudista

Após o carregamento dos dados, podemos escolher a técnica ou o conjunto de técnicas a serem utilizadas no experimento; nesse sentido, observe que o *software Weka* possui implementações de diversos algoritmos para a mineração de dados, fornecendo também a *Application Programming Interface (API)* em *Java*, caso você queira utilizar em seus programas *Java*.

Segue a Figura 3, que ilustra a seleção na guia de *classificação e*, em seguida, *Function e* a seleção de *MultlayerPerceptron*:

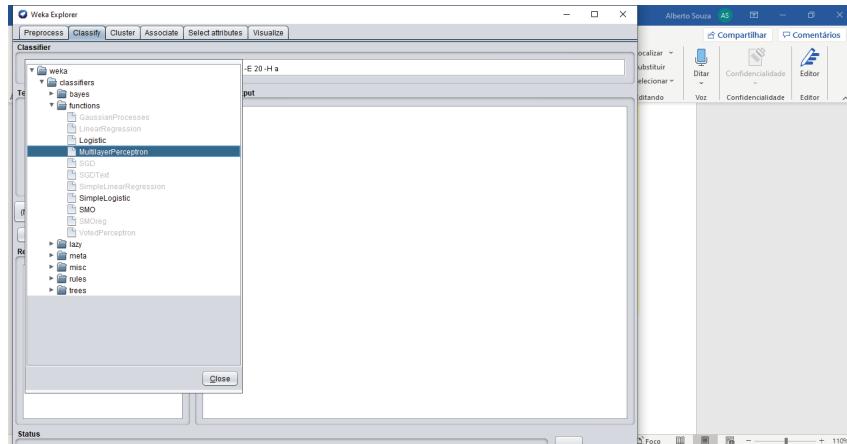


Figura 3 – Seleção de algoritmo ou função a ser utilizada no experimento no *Weka*

Fonte: Acervo do Conteudista

Em seguida já é possível executar o experimento, mesmo com as opções padrões do *Weka*; para isto, você deve clicar no botão *Start* – conforme pode-se observar na Figura 4:

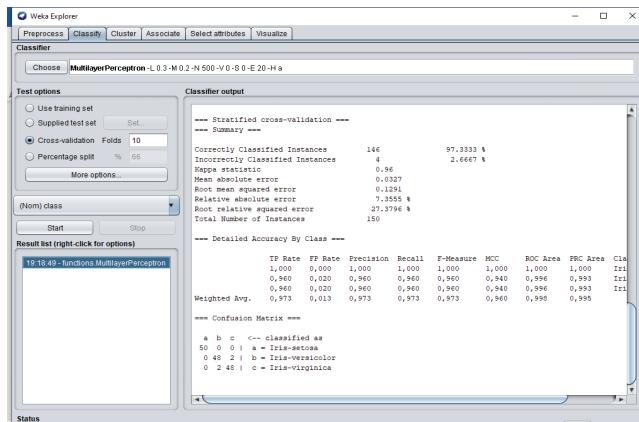


Figura 4 – Resultado da execução e treinamento da rede neural no Weka

Fonte: Acervo do Conteudista

Note, na Figura 4, que podemos verificar as saídas da rede neural, bem como os resultados de validação.

Observe também que 146 das 150 instâncias foram classificadas corretamente e que o erro quadrado do modelo gerado foi de 0,1291, correspondendo a um erro relativo de 7,3555%.

Para o experimento foi utilizada a técnica de validação cruzada, ou *cross validation*, de 10% (Figura 4), que consiste em treinar a rede com 10% do conjunto e depois executar os testes e a validação para o restante das instâncias. A partir destes testes com o restante da rede, verifica-se a quantidade de instâncias classificadas corretamente e as que foram classificadas de maneira incorreta.

Outra forma de verificação de assertividade do modelo gerado é o que se pode observar na Figura 4, por meio dos valores de:

- **True Positive (TP)**: ou verdadeiro positivo, corretamente classificado como verdadeiro;
- **True Negative (TN)**: ou verdadeiro negativo, corretamente classificado como falso;
- **False Positive (FP)**: ou falso positivo, erroneamente classificado como positivo;
- **False Negative (FN)**: ou falso negativo, erroneamente classificado como negativo.

Com esses valores consegue-se medir a precisão e acurácia do modelo.

Utilizando as medidas mencionadas, pode-se criar a curva *Receiver Operating Characteristic* (ROC), que ilustra a assertividade do modelo, onde a área abaixo da curva (AUC) aponta graficamente a precisão do modelo gerado.

Weka já calcula tudo isso, exibindo na tela de resultados do algoritmo, conforme observa-se na Figura 4.

Outra saída importante do modelo gerado corresponde às informações de pesos dadas as entradas de cada um dos neurônios da rede, conforme pode-se observar na Figura 5:

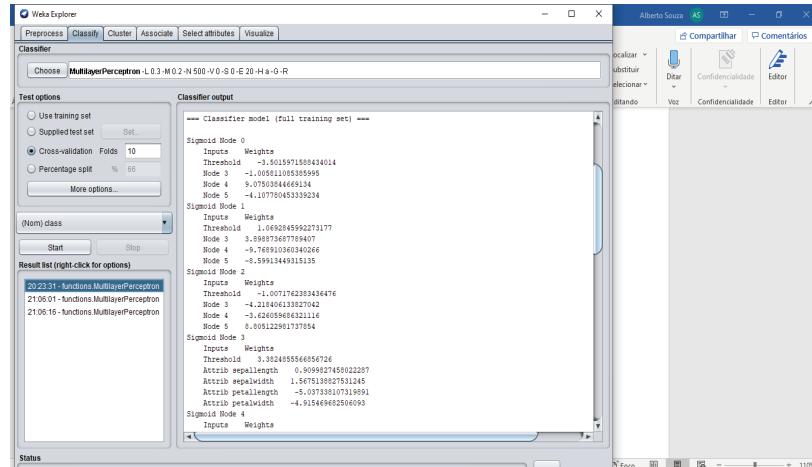


Figura 5 – Saída com os pesos para cada um dos neurônios do modelo

Fonte: Acervo do Conteudista

Observe que o modelo gerado estabelece peso aos atributos de entrada de cada um dos neurônios: inicialmente, com os atributos diretos *e*, em seguida, com as saídas dos neurônios da camada anterior; embora isto possa aparecer invertido na saída, você conseguirá observar quais entradas cada sigmoide ou neurônio possui, como *node* (nó da rede ou neurônio), ou *attrib* (atributo da base).

Outras opções de parametrização do *MultilayerPerceptron* poderão ser editadas ao se clicar na barra que tem o nome do algoritmo – abrindo-se a janela exibida na Figura 6:

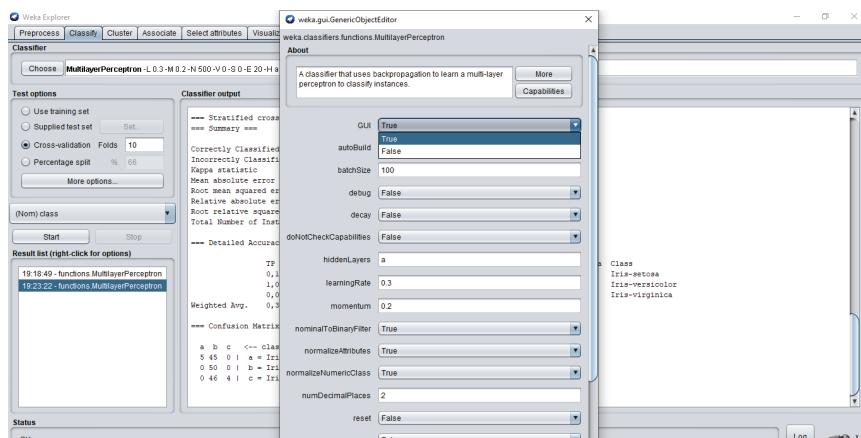


Figura 6 – Opções de parametrização da função *MultilayerPerceptron*

Fonte: Acervo do Conteudista

Através da tela exibida na Figura 6, pode-se editar os parâmetros da rede neural como, por exemplo, o número de camadas ocultas da rede em *hiddenLayers*, a taxa de atualização ou o salto dos pesos dos neurônios a cada iteração do algoritmo, ou se é para exibir graficamente a rede neural do modelo, ao selecionar a opção *true* em *GUI*.

Ao editar os parâmetros da rede, deve-se clicar em *OK* e, em seguida, *Start* para novamente executar o experimento. Deixando selecionada a opção para exibir a rede gerada, o software mostrará a rede conforme a Figura 7:

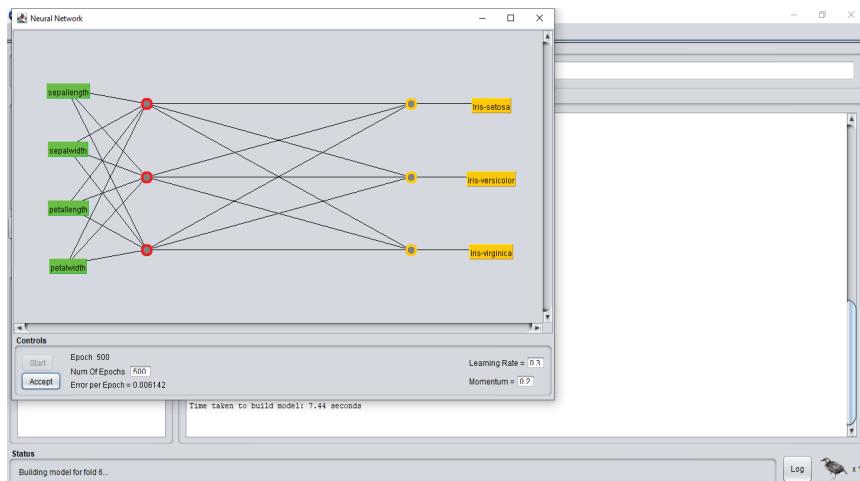


Figura 7 – Rede neural gerada para o experimento

Fonte: Acervo do Conteudista

Observe que a rede gerada possui como entradas os valores de cada um dos atributos e a saída é cada uma das classes existentes no conjunto de dados.

Outra opção de visualização interessante do Weka é a saída gráfica da classificação gerada, conforme pode-se observar na Figura 8:

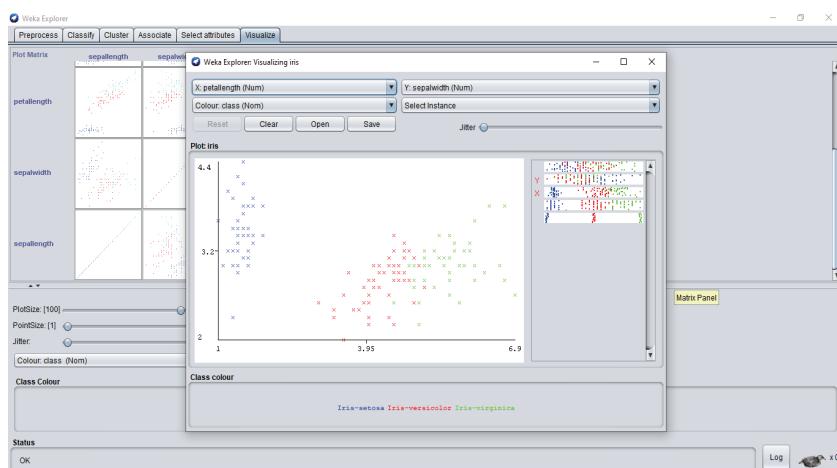


Figura 8 – Saída gráfica do modelo de classificação gerado

Fonte: Acervo do Conteudista

Para se chegar a esta visualização você pode clicar na aba denominada *Visualize* e, em seguida, clicar em alguns dos gráficos gerados, ao se plotar um atributo *versus* outro como, por exemplo, o exibido na Figura 8, os atributos *sepalwidth* por *petallength*.

Segundo Estudo de Caso com Weka

Outra base de dados interessante de se trabalhar é a de câncer de mama, no arquivo *breast-cancer.arff*, igualmente presente na pasta de bases de dados do Weka.

Cada instância representa detalhes médicos de pacientes e amostras de seu tecido do tumor, de modo que a tarefa é prever se a paciente tem câncer de mama ou não.

Existem 9 variáveis de entrada, todas nominais. Os melhores resultados são da ordem de 75% de precisão.

Segue a Figura 9, que ilustra o melhor resultado obtido no experimento, com 72.3776% de precisão:

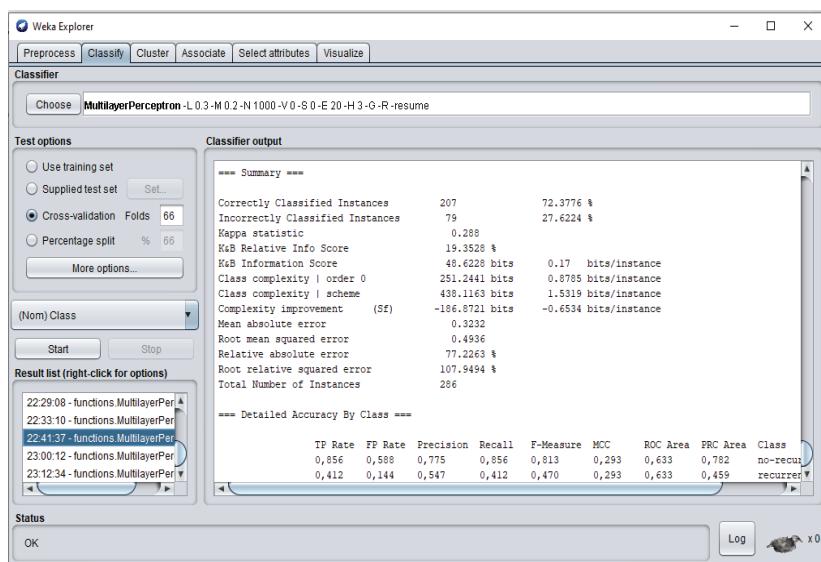


Figura 9 – Melhor resultado obtido com a base de dados *breast-cancer.arff*

Fonte: Acervo do Conteudista

Este melhor resultado foi obtido ao executar a rede neural com 3 camadas de neurônios ocultos e 1.000 iterações no algoritmo, *learning rate* ou taxa de aprendizagem de 0,2 e 66% de parametrização de base de treinamento para a validação cruzada.

Segue a Figura 10, que ilustra a parametrização com as 3 camadas ocultas de execução no campo *hiddenLayers* e a taxa de aprendizagem no campo *learningRate*:

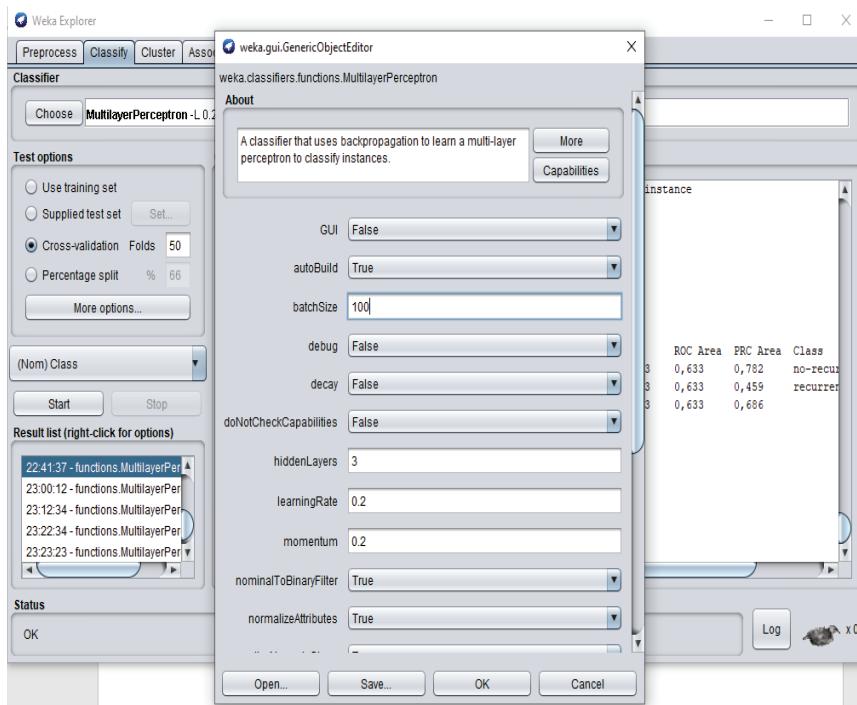


Figura 10 – Parametrização do número de camadas ocultas da rede neural

Fonte: Acervo do Conteudista

Outra parametrização interessante é o número de iterações da rede ou *epochs*, no campo *trainingTime*, conforme ilustrado na Figura 11:

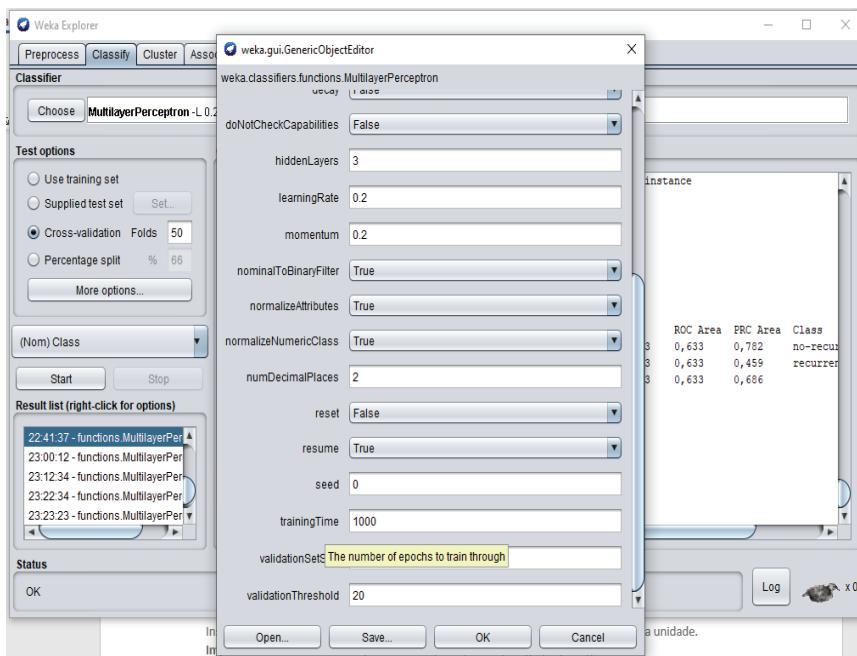


Figura 11 – Parametrização do número de iterações da rede

Fonte: Acervo do Conteudista

Conforme se observa neste experimento, o importante é avaliar os diversos resultados obtidos com as alterações de parametrizações de número de camadas ocultas da rede neural, iterações do algoritmo, conjunto de treinamento e validação cruzada, além da taxa de aprendizagem da rede.

Cabe destacar ainda a necessidade de executar esse tipo de experimento para uma base de dados menor e em um *software* como o *Weka* em menor escala, de modo a conseguir melhor modelagem, a fim de posteriormente partir para a execução ou implementação real para todo o conjunto de dados da organização, utilizando outros recursos computacionais como, por exemplo, tecnologias de *big data*.

Anexo I

Vamos ao tutorial de instalação do *software Weka* no *Windows*, o qual usaremos no estudo de caso desta Unidade:

1. Acesse o site da *Weka*, fazendo uma pesquisa ou diretamente em:
<https://bit.ly/3jxkc9C>
2. Clique no link *Download*.



Figura 1
Fonte: Acervo do Conteudista

3. É importante que você tenha instalado a *Java Virtual Machine (JVM)*; então, clique no executável ou descompacte o arquivo, caso esteja usando *Linux*, para iniciar a instalação. Você será direcionado ao site *Source Forge*.
4. Clique na opção *Executar* após a finalização do *download* para iniciar o programa de instalação:

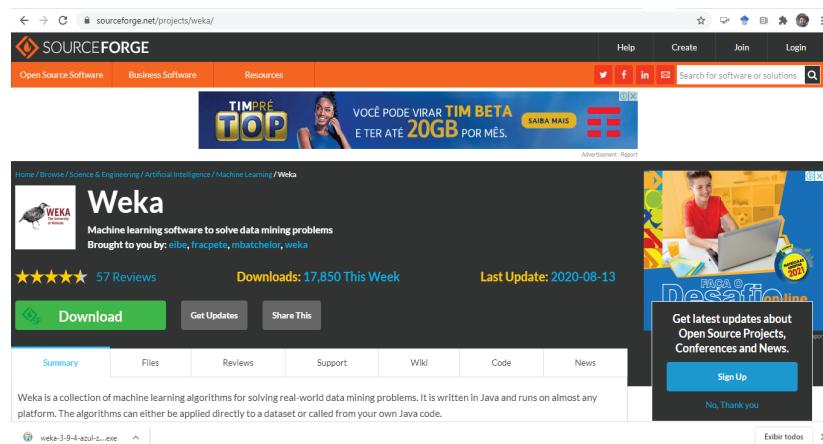


Figura 2
Fonte: Acervo do Conteudista

5. A instalação possui sete minicaixas de diálogo, de modo que você poderá aceitar todos os padrões e apenas clicar nas telas:



Figura 3
Fonte: Acervo do Conteudista

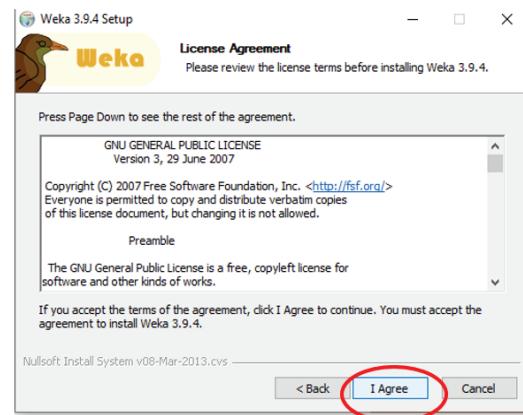


Figura 4
Fonte: Acervo do Conteudista

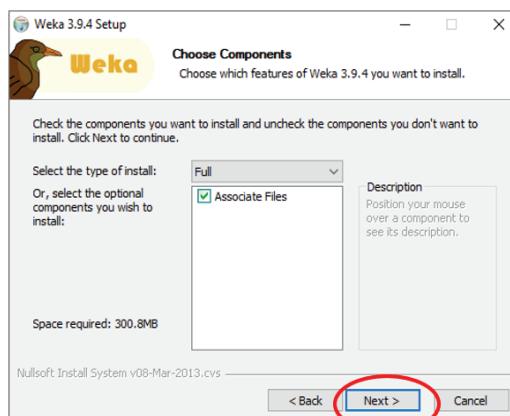


Figura 5
Fonte: Acervo do Conteudista

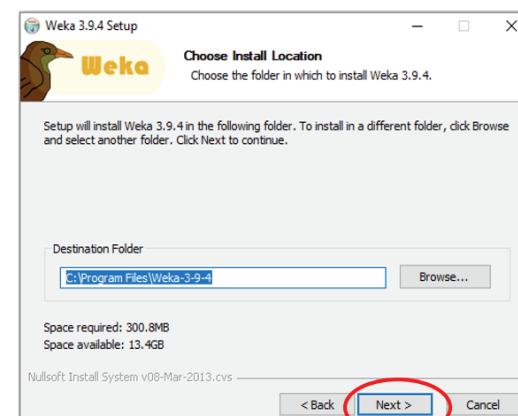


Figura 6
Fonte: Acervo do Conteudista

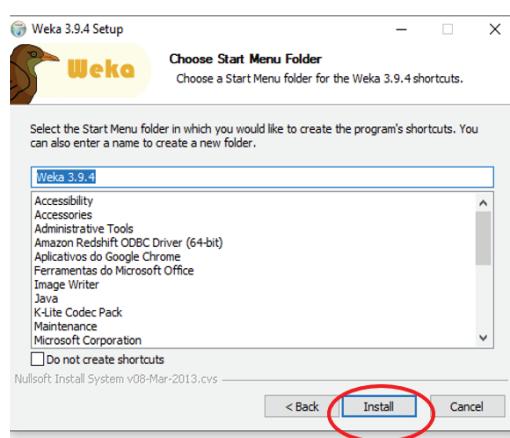


Figura 7
Fonte: Acervo do Conteudista

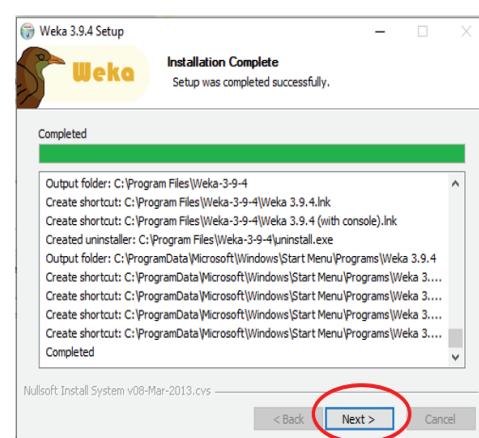


Figura 8
Fonte: Acervo do Conteudista



Figura 9

Fonte: Acervo do Conteudista

Caso você tenha chegado aqui, significa que finalizou a instalação do software *Weka* em seu sistema operacional *Windows*.

Se, contudo, utilizar *Linux*, após descompactar o arquivo “baixado”, deverá executar o script de instalação usando o comando: `./weka.sh`

Mas se *MacOS* for o seu sistema operacional, deverá executar o arquivo de imagem “baixado”.

Material Complementar

Indicações para saber mais sobre os assuntos abordados nesta Unidade:

Sites

Neural Network

Execução e teste visual de redes neurais.

<https://bit.ly/2DkffS5>

Machine Learning Repository

Diversas bases de dados de testes e experimentos.

<https://bit.ly/2QHE5hz>

Leitura

Data Analysis based on Data Mining Algorithms using Weka Workbench

Um artigo que discute o experimento com a base de dados de câncer de mama.

<https://bit.ly/31NMCGo>

Referências

- DUDA, R. O.; HART, P. E. *Pattern classification and scene analysis.* [S.l.]: John Wiley & Sons, 1973.
- FISHER, R. A. *The use of multiple measurements in taxonomic problems.* *Annual Eugenics*, v. 7, n. 2, p. 179-188, 1936.
- MICHALSKI, R. S. et al. *The multi-purpose incremental learning system AQ15 and its testing application to three medical domains.* In: NATIONAL Conference on Artificial Intelligence, 5., 1986, Philadelphia, PA, USA. *Proceedings...* Philadelphia, PA, USA: Morgan Kaufmann, 1986. p. 1.041-1.045.
- WILEY, J. *Contributions to mathematical statistics.* New York: [s.n.], 1950.
- WITTEN, I.; FRANK, E.; HALL, M. *Morgan Kaufmann series in data management systems.* 3^{ed}. Amsterdam: Morgan Kaufmann, 2011.



Cruzeiro do Sul
Educacional