

## 7 Referências Bibliográficas

ALMEIDA, ANDRÉ ATANÁSIO MARANHÃO. **Aplicação de Algoritmos de Agrupamento a Dados Biológicos**, Trabalho de conclusão de curso - Universidade Federal de Alagoas, 2004.

BALASKO, BALAZS; ABONYI, JANOS; FEIL, BALAZS. **Fuzzy Clustering and Data Analysis Toolbox**, 2005. Disponível em <<http://www.fmt.vein.hu/softcomp/fclusttoolbox/FuzzyClusteringToolbox.pdf>>. Acesso em abr. de 2005.

BAO, HO TU. **Automatic Clustering Detection**. Disponível em <<http://www.netnam.vn/unescocourse/knowledge/51.htm>>. Acesso em out. de 2004.

BERKHIN, PAVEL. **Survey of Clustering Data Mining Techniques**, 2002. Disponível em <<https://umdrive.memphis.edu/vphan/public/berkhin-survey.pdf>>. Acesso em out. de 2004.

COOK, PERRY. **Minimum-Distance Classifiers**, 2004. Disponível em <[http://www.cs.princeton.edu/courses/archive/fall05/cos436/Duda/PR\\_simp/min\\_dist.htm](http://www.cs.princeton.edu/courses/archive/fall05/cos436/Duda/PR_simp/min_dist.htm)>. Acesso em out. de 2004.

COSTA, JOSÉ ALFREDO FERREIRA. **Classificação Automática e Análise de Dados por Redes Neurais Auto-Organizáveis**, Campinas, SP, dezembro de 1999. Disponível em <[http://www.dee.ufrn.br/~alfredo/index\\_tese.html](http://www.dee.ufrn.br/~alfredo/index_tese.html)>. Acesso em nov. de 2004.

DONI, MARCELO VIANA. **Análise de Cluster: Métodos Hierárquico e de Particionamento**, Universidade Presbiteriana Mackenzie, 2004. Disponível em <<http://meusite.mackenzie.com.br/rogerio/tgi/2004Cluster.PDF>>. Acesso em jun. de 2005.

DUDA, RICHARD O.; HART, PETER E.; STORK, DAVID G. **Pattern Classification**, John Wiley & Sons, 2001.

EVERITT, BRIAN S.; LANDAU, SABINE; LEESE, MORVEN. **Cluster Analysis**, A Hodder Arnold Publication; 4th edition, 2001.

EVSUKOFF, ALEXANDRE G. **Introdução às Técnicas de Data Mining**, 2005. Disponível em <<http://www.incc.br/verao/verao05/arquivos/Intro%20Data%20Mining%201.pdf>>. Acesso em mar. de 2005.

FISHER, R. A. **The use of multiple measurements in taxonomic problems**, Annual Eugenics, vol. 7, Part II, pp 179-188, 1936.

FUNG, GLENN. **A Comprehensive Overview of Basic Clustering Algorithms**, 2001. Disponível em <<http://www.cs.wisc.edu/~gfung/clustering.pdf>>. Acesso em set. de 2004.

HAMERLY, GREGORY JAMES. **Learning structure and concepts in data through data clustering**, University Of California, San Diego, A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Computer Science and Engineering, 2003. Disponível em <<http://cs.baylor.edu/~hamerly/papers/thesis.pdf>>. Acesso em nov. de 2004.

KAINULAINEN, JUKKA. **Clustering Algorithms: Basics and Visualization**, 2002. Disponível em <<http://www.niksula.cs.hut.fi/~jkainula/pdfs/clustering.pdf>>. Acesso em nov. de 2004.

KAUFMAN, LEONARD; ROUSSEEUW, PETER J. **Finding Groups in Data: An Introduction to Cluster Analysis**. 4.ed. New York: Wiley, 1990.

KLEIN, C.H. et al. **Hipertensão arterial na Ilha do Governador**, Rio de Janeiro: II-Prevalência, Cad. Saúde Pública (Reports in Public Health), 11:389-394, 1995.

KOERICH, ALESSANDRO. **Aprendizagem de Máquina - Aprendizagem Não Supervisionada**, 2005. Disponível em <<http://www.ppgia.pucpr.br/~alekoe/AM/2005/>>. Acesso em mar. de 2005.

KUGLER, MAURÍCIO; JÚNIOR, J. T.; LOPES, H. S. **Desenvolvimento de uma Rede Neural LVQ em Linguagem VHDL para Aplicações em Tempo-Real**, VI Congresso Brasileiro de Redes Neurais, 2003. Disponível em <<http://www.cpgei.cefetpr.br/~hslopes/publicacoes/2003/cbrn2003c.pdf>>. Acesso em ago. de 2005.

JÚNIOR, F. N.; DANTAS, M. C.; SANTOS, P. M. **ImageSelector: um classificador de imagens**, Universidade Federal de Pernambuco, 2002.

MATLAB<sup>®</sup> REFERENCE. **Fuzzy Logic Toolbox**, 2005. Disponível em <<http://www.mathworks.com/access/helpdesk/help/toolbox/fuzzy/index.html>>. Acesso em jul. de 2005.

MATTEUCCI, MATTEO. **A Tutorial on Clustering Algorithms**. Disponível em <[http://www.elet.polimi.it/upload/matteucc/Clustering/tutorial\\_html/hierarchical.html](http://www.elet.polimi.it/upload/matteucc/Clustering/tutorial_html/hierarchical.html)>. Acesso em jun. de 2005.

NEVILLE, JENNIFER; ADLER, MICAH; JENSEN, DAVID. **Clustering relational data using attribute and link information**, Proceedings of the Text Mining and Link Analysis Workshop, Eighteenth International Joint Conference on Artificial Intelligence, 2003. Disponível em <<http://kdl.cs.umass.edu/papers/neville-et-al-textlink2003.pdf>>. Acesso em out. de 2004.

NOVAES, URIAS DA ROSA. **Agrupamento de Dados Através de Algoritmos SWARM**, Tese apresentada a COPPE/UFRJ, 2002. Disponível em <[http://www.coc.ufrj.br/teses/mestrado/inter/2002/teses/NOVAES\\_UR\\_02\\_t\\_M\\_int.pdf](http://www.coc.ufrj.br/teses/mestrado/inter/2002/teses/NOVAES_UR_02_t_M_int.pdf)>. Acesso em out. de 2004.

RASHID, TARIQ. **Clustering**, 2005. Disponível em <[http://www.cs.bris.ac.uk/home/tr1690/documentation/fuzzy\\_clustering\\_initial\\_report/node11.html](http://www.cs.bris.ac.uk/home/tr1690/documentation/fuzzy_clustering_initial_report/node11.html)>. Acesso em ago. de 2005.

REED, RUSSELL D.; MARKS II, ROBERTO J. **Neural Smithing – Supervised Learning in Feedforward Artificial Neural Networks**, Bradford Book, The MIT Press, 1999.

SILVA, ALICE. **Análise Classificatória**, 2005, Universidade Nova de Lisboa - Faculdade de Ciências e Tecnologia. Disponível em <<http://ferrari.dmat.fct.unl.pt/services/AnaliseDados/Cluster.pdf>>. Acesso em mar. de 2005.

SIMÕES, RODRIGO. **Complexos Industriais no Espaço: Uma Análise de Fuzzy Cluster**, Cedeplar/UFMG, 2004. Disponível em <<http://www.anpec.org.br/encontro2004/artigos/A04A112.pdf>>. Acesso em jan. de 2005.

SIMON, JULIAN. **Hierarchical Clustering**. Disponível em <[http://www.resample.com/xlminer/help/HClst/HClst\\_intro.htm](http://www.resample.com/xlminer/help/HClst/HClst_intro.htm)>. Acesso em out. de 2004.

STRUYF, ANJA et al. **Clustering in an Object-Oriented Environment**, Department of Mathematics and Computer Science, U.I.A., Universiteitsplein 1, B-2610 Antwerp, Belgium, 1996. Disponível em <<http://www.jstatsoft.org/counter.php?id=4&url=v01/i04/paper/clus.pdf>>. Acesso em fev. de 2005.

VIANA, MARCELO. **Análise de Cluster: Métodos Hierárquicos e de Particionamento**, 2004. Disponível em <<http://meusite.mackenzie.com.br/rogerio/tgi/2004Cluster.PDF>>. Acesso em maio de 2005.

WARD, JOE H. **Hierarchical Grouping to optimize an objective function**, Journal of American Statistical Association, 58(301), 236-244, 1963. Disponível em <<http://iv.slis.indiana.edu/sw/data/ward.pdf>>. Acesso em ago. de 2005.

WINIDAMS. **Cluster Analysis**, 2005a. Disponível em <<http://stat.caspur.it/Manuali/WinIdams/HtmlEN/E2clusfi.htm>>. Acesso em abril de 2005.

WINIDAMS. **Partitioning Around Medoids (PAM)**, 2005b. Disponível em <[http://www.unesco.org/webworld/idams/advguide/Chapt7\\_1\\_1.htm](http://www.unesco.org/webworld/idams/advguide/Chapt7_1_1.htm)>. Acesso em abril de 2005.

YURAS, GABRIEL. **Minimum-Distance Classifiers**. Disponível em <<http://www.profc.udec.cl/~gabriel/tutoriales/rsnote/cp11/cp11-6.htm>>. Acesso em out. de 2004.

## ANEXO - Guia de Instalação e Uso do Aplicativo

### Guia de Instalação

Siga os passos abaixo para instalar corretamente o aplicativo.

1. Copie os arquivos abaixo para a pasta onde se deseja instalar o programa.
  - Cluster Analysis.jar
  - Setup.bat
2. Execute o arquivo *Setup.bat*.
3. Caso tenha o aplicativo R<sup>®</sup> instalado, edite o arquivo *Cluster Analysis.cfg* como se segue:

Rcmd: <entrar com caminho onde se encontra o executável Rcmd.exe>

### Guia de Utilização

#### ▪ Base de Dados

Para dar início o uso do aplicativo, primeiramente é necessário definir a base de dados e as variáveis com que se deseja trabalhar. Para isso siga os seguintes passos:

1. Vá para tela *Dados*.
2. Clique no menu *Arquivo / Carregar Base de Dados* ou simplesmente clique no botão *Carregar Base de Dados*.
3. Selecione o arquivo referente a sua base de dados. O arquivo deve estar no seguinte formato:

```
<Nome da Coluna 1><TAB><Nome da Coluna 2><TAB><Nome da Coluna 3>...
<Valor 1.1><TAB><Valor 2.1><TAB><Valor 3.1>...
<Valor 1.2><TAB><Valor 2.2><TAB><Valor 3.2>...
<Valor 1.3><TAB><Valor 2.3><TAB><Valor 3.3>...
...
```

Figura 60: Modelo de arquivo de base de dados

4. Os nomes das variáveis da base de dados selecionada, referenciado como 'Nome da Coluna x' no modelo acima, deverão estar listados em *Variáveis da Base de Dados*. A lista *Variáveis para a Clusterização* deverá conter as variáveis que serão usadas no processo de agrupamento de dados, para isso, faça uso dos botões <, >, <<, >>.

No final dessa etapa, na tela *Dados*, a lista *Variáveis da Base de Dados* conterà as variáveis da base de dados que não foram selecionadas para o processo de agrupamento de dados, enquanto que a lista *Variáveis para a Clusterização* conterà as variáveis definidas no passo anterior que serão usadas no processo de agrupamento de dados.

- *Geração de Agrupamentos*

Esse é o módulo destinado à escolha do número de grupos e do método de agrupamento de dados a ser utilizado.

Para se definir os parâmetros para a geração de agrupamentos siga os passos abaixo.

1. Vá para a tela *Geração de Agrupamentos*.
2. Escolha o número de agrupamentos no campo *Número de Agrupamentos*. O número de agrupamentos deve estar entre 2 e 30.
3. Escolha um dos métodos de agrupamento de dados listados em *Métodos de Clusterização* como podem ser observados na Figura 61.

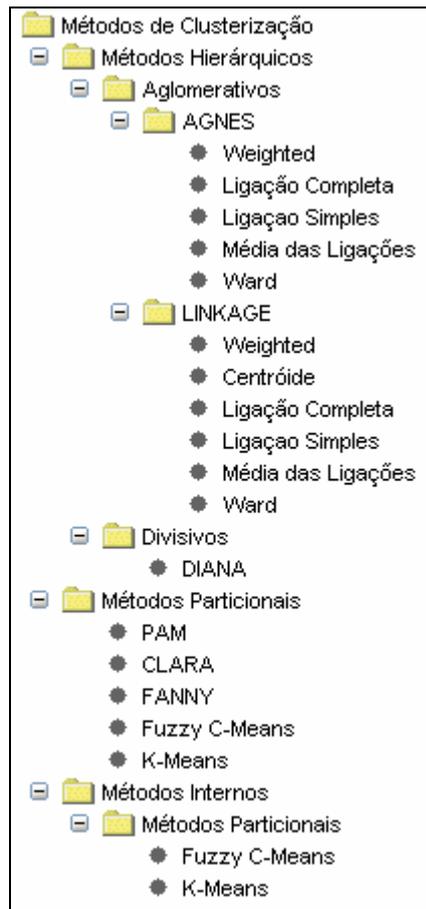


Figura 61: Métodos de Agrupamento de Dados

▪ *Geração de Resultados*

Esse módulo é responsável basicamente por três operações:

1. Seleção das variáveis que serão apresentadas nas médias dos resultados obtidos em cada agrupamento.
2. Gráfico comparativo entre os valores de duas variáveis por agrupamento (opcional).
3. Aplicação do método de agrupamento de dados e na geração dos resultados.

Siga os passos abaixo a fim de dar início ao processo de agrupamento de dados.

1. Vá para a tela *Resultados*.
2. Selecione em *Variáveis da Base de Dados* as variáveis que serão utilizadas na geração dos resultados, para isso faça uso dos botões <, >.

<<, >>. Em *Variáveis para Resultados* deverá conter as variáveis selecionadas.

3. (Opcional) Clique na caixa *Gráfico Comparativo* e selecione em *Variável 1 (Base)* e *Variável 2* as variáveis que serão utilizadas no gráfico comparativo. Mais detalhes sobre esse gráfico será abordado na seção seguinte.
4. Clique no botão *Aplicar Método e Gerar Resultados* para dar início ao processo de agrupamento de dados.

Ao final do último passo, uma nova janela será apresentada contendo as planilhas e gráficos para a análise de resultados.

Pode-se rodar o aplicativo inúmeras vezes parametrizadas de formas diferentes, e com isso ter várias janelas de análise de resultados abertas com o intuito de comparar os resultados obtidos em cada rodada.

#### ▪ *Análise de Resultados*

O aplicativo fornece duas tabelas e até três gráficos que auxiliam a análise de resultados.

- Tabela de Pertinências

Para fazer o uso do filtro siga os seguintes passos:

1. Clique no botão *Filtro*. A tela abaixo deverá ser exibida.



Figura 62: Filtro

2. Selecione no primeiro campo o agrupamento a qual se deseja aplicar o filtro.

3. Selecione uma das operações abaixo que se deseja aplicar sobre o agrupamento selecionado:
  - i. Operação *maior, maior ou igual, menor, menor ou igual, igual*.
  - ii. Intervalo de valor
4. Clique em *Filtrar*.

Pode-se aplicar o filtro sobre o resultado obtido na operação anterior. Para remover o filtro aplicado sobre os dados basta clicar em *Remover Filtro*.

- Tabela de Médias
- Silhueta
- Dendograma (apenas para métodos hierárquicos)
- Gráfico Comparativo (opcional)