

# Aprendizagem de Máquina através de Combinação de Classificadores em Bases de Dados da Área da Saúde

Lucelene Lopes<sup>1</sup>, Edson Emílio Scalabrin<sup>2</sup>, Caroline Leitão Riella<sup>2</sup>, Patrícia Manfre<sup>2</sup>

<sup>1</sup> Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)

<sup>2</sup> Pontifícia Universidade Católica do Paraná (PUCPR)

**Resumo** - Este artigo compara a taxa de acerto de um método baseado em classificador simples de árvore de decisão (algoritmo C4.5), com dois outros métodos de combinação de classificadores (BAGGING e BOOSTING) em bases de dados da área da saúde. Especificamente, os três métodos são aplicados a 15 bases de dados da área da saúde, onde são analisadas as curvas de aprendizagem sobre amostras aleatórias geradas de cada base de dados. Os métodos BAGGING e BOOSTING são aplicados usando diferentes números de classificadores e a taxa de acerto é adquirida usando a técnica de validação cruzada. A contribuição deste trabalho está em recomendar o método mais eficaz de aprendizagem de máquina para base de dados da área da saúde em geral.

**Palavras-chave:** Aprendizagem de Máquina, Curvas de Aprendizagem, C4.5, BAGGING, BOOSTING

**Abstract** - This paper compares the accuracy of a simple classifier method based on decision tree (C4.5 algorithm) with two ensemble classifier methods (BAGGING and BOOSTING) for health data bases. Specifically, the three methods are applied to 15 health data bases where the learning curves are analysed over random samples of the original data. BAGGING and BOOSTING methods are applied using different numbers of classifiers and the accuracy is computed using a cross-validation technique. As for this work contribution, an analysis of the experiments is presented leading to the recommendation of the most efficient machine learning method to health databases in general.

**Keywords:** Machine Learning, Learning Curve, C4.5, BAGGING, BOOSTING

## Introdução

Atualmente a maior parte dos problemas de tomada de decisão não tem por desafio o tratamento numérico, mas a transformação de dados e informações em conhecimento, principalmente quando as bases de dados dizem respeito à área da saúde.

Tais bases, em geral, possuem grande número de atributos (variáveis), pequeno número de ocorrências (instâncias) e grande número de valores ausentes, tornando diversos dados redundantes e irrelevantes do ponto de vista da aprendizagem de máquina.

Devido a essas características, o uso de técnicas de descoberta do conhecimento nos parece essencial.

Em particular, técnicas baseadas na geração de classificadores simbólicos (explícitos), bem como, na combinação de classificadores visando à obtenção de conhecimentos com alta taxa de acerto, contribui de maneira significativa à tomada de decisão na área da saúde.

Apesar de não estar focado em bases da área da saúde, Kotsianti e Kanellopoulous [1] fazem um trabalho semelhante de comparação dos métodos BAGGING [2] e BOOSTING [3], mas também

consideram um outro método chamado DAGGING (Disjoint-Sample Aggregation) [4].

Além de comparar a aplicação destes três métodos em bases genéricas verificando a taxa de acerto de cada um deles, o trabalho de Kotsianti e Kanellopoulous propõe uma combinação destes classificadores e defende que este novo método combinado possui melhor taxa de acerto que cada um dos métodos individualmente.

Este trabalho é bastante importante em relação a esse artigo, pois, ele mostra a atualidade do tema de comparação de métodos de aprendizagem de máquina que combinam classificadores. Isto mostra que não há consenso na área sobre qual método é o mais adequado, mas sim apenas algumas conclusões empíricas.

A falta de consenso da área sobre o melhor método, aparece também no trabalho de Quilan [5] que faz comparações entre C4.5 [6], BAGGING e BOOSTING sobre 27 bases genéricas usando a taxa de acerto como parâmetro de comparação.

Neste trabalho de Quilan são feitas algumas considerações quanto ao tempo de processamento que é maior quando se usa combinação de classificadores.

Tanto nos trabalhos de Kotsianti e Kanellopoulous quanto no de Quinlan, fica claro que a taxa de acerto do método C4.5 é inferior ao dos métodos BAGGING e BOOSTING .

No entanto, ambos trabalhos são inconclusivos no que diz respeito a comparação entre BAGGING e BOOSTING .

Na verdade, de acordo com diversos autores da área [7], [8], [9], [10] cada base de dados pode apresentar um comportamento distinto tendo para algumas bases em BAGGING, para outras em BOOSTING , o método mais preciso de classificação.

Tendo isto em mente, a hipótese de trabalho deste artigo é que para um conjunto específico de bases de dados, as bases da área da saúde, pode-se afirmar que um destes dois métodos é mais adequado.

Este trabalho tem como objetivo a experimentação dos três métodos de aprendizagem de máquina: C4.5 , BAGGING e BOOSTING sobre um conjunto de bases de dados da área da saúde com o intuito de verificar as sua taxa de acerto.

Esta análise é feita em dois momentos: primeiro são analisadas as curvas de aprendizagem obtidas por cada um dos métodos para cada uma das bases; e posteriormente, os métodos BAGGING e BOOSTING são analisados novamente observando o impacto da variação do número de classificadores na precisão dos métodos.

## Metodologia

O estudo investigativo levou em conta 15 bases de dados da área da saúde Quadro 1, que foram tomadas do repositório de dados da escola de computação e informática da Universidade da Califórnia - Irvine [11].

ID	Database	instances	attributes	classes
01	arrhythmia	452	280	2
02	audiology	226	70	24
03	breast cancer	286	10	2
04	wisconsin breast cancer	699	10	2
05	wdbc	569	31	2
06	kr vs kp	3196	37	2
07	pima diabetes	768	9	2
08	cleveland 14 heart disease	303	14	2
09	echo cardiogram	132	12	3
10	heart statlog	270	14	2
11	hepatitis	155	20	2
12	hyper hormonal	2800	30	4
13	hypothyroid	3772	30	4
14	post operative	90	9	3
15	sick euthyroid	3163	26	2

Quadro 1. Características das bases

Pode-se observar que algumas destas bases de dados possuem um grande número de atributos, por exemplo, a base de dados arrhythmia com 280 atributos. Tem-se ainda, para esta base de dados, o complicador dado pelo pequeno número de instâncias comparativamente ao número de

atributos (452). Tal desproporcionalidade pode ser entendida como uma especificidade da área da saúde, visto que os dados não são, na sua grande maioria, capturados de forma automática e também devido à subjetividade intrínseca da área.

Destacamos que todos os experimentos deste trabalho foram testados no pacote de software WEKA versão 3.4 [12].

Os experimentos para obtenção das curvas de aprendizagem foram realizados utilizando o método de amostragem *Resample* [13], [14] que retorna um vetor contendo 10 valores reais, onde cada um deles corresponde à taxa de acerto de 10 combinações de classificadores obtidos a partir de 10 amostras aleatórias da mesma base de dados.

Cada uma destas amostras tem tamanhos distintos que correspondem a 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% até uma amostra de tamanho 100% que contém toda a base de dados original.

No caso dos métodos BAGGING e BOOSTING , utiliza-se 10 iterações, ou seja, uma combinação de classificadores composta de 10 classificadores individuais.

Desta forma, são geradas três curvas de aprendizagem para cada uma das 15 bases, mostrando a evolução dos métodos C4.5, BAGGING e BOOSTING com o incremento do tamanho da amostra.

Os experimentos sobre o impacto do número de classificadores consistiu em gerar, para cada base de dados de um conjunto valores de taxa de acerto, variando significativamente o número de classificadores (ou iterações) de 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180 até 200.

Evidentemente, esta variação só é possível para os métodos baseados em combinação de classificadores, logo apenas para os métodos BAGGING e BOOSTING e não para o C4.5.

O interesse aqui é verificar por meio de um conjunto de experimentos e análise estatística dos resultados através do teste de Friedman [15], qual é o impacto do número de classificadores na eficiência dos métodos BAGGING e BOOSTING .

## Resultados

### Curvas de Aprendizagem

Os quadros 2, 3 e 4 representam respectivamente os dados gerados para curva de aprendizagem de acordo com o tamanho da amostra para todas as bases de dados. Nota-se que nos três quadros falta o valor para amostra 10% para a base 14. Isso ocorre por se tratar de uma base de dados com um número muito pequeno de instâncias.

Para quase todas as bases de dados, pôde-se observar que o aumento do tamanho da amostra

também influencia no incremento da taxa de acerto.

Entretanto, percebeu-se que, para 14 das 45 curvas de aprendizagem (bases 01, 03, 05, 09, 10, 11, 14 para C4.5, bases 03, 10, 11 para BAGGING e bases 03, 05, 09, 12 para BOOSTING), a taxa de acerto para as amostras de tamanho 100% foram menores que aqueles obtidos com algumas amostras parciais.

Por exemplo, o método BOOSTING aplicado sobre uma amostra de 20% da base de dados *echo cardiogram* (09), assim como o método BAGGING aplicado sobre uma amostra de 10% da base *heart statlog* (10), têm-se uma excelente taxa de acerto.

Como exemplo contrário, o método BOOSTING aplicado sobre a amostra de 20% da base *breast cancer* (03) não tem o mesmo comportamento.

Em geral, observa-se para quase todas as bases de dados, que existe uma forte variação da taxa de acerto para amostras de diferentes origens.

Portanto, conclui-se que existem amostras de mesmo tamanho que são mais adequadas e outras menos adequadas em função da base de dados.

De acordo com o gráfico 1, percebe-se que o método BOOSTING apresenta as maiores taxas de acerto na média para as amostras de mais do que 20%. Isso percebe-se também no Quadro 5, onde estão as médias das amostras.

No Quadro 6, verifica-se pelas médias calculadas para cada base e para cada método que os métodos BAGGING e BOOSTING foram superiores ao C4.5 em 14 bases. Percebe-se ainda uma supremacia do método BOOSTING que foi mais eficiente em 12 destas bases.

Percebe-se ainda mais resumidamente no Quadro 7, que na média geral, BOOSTING possui a maior taxa de acerto.

Este resultado não é surpresa, pois os métodos BAGGING e BOOSTING são métodos mais sofisticados e teoricamente deveriam ter uma melhor aprendizagem devido à combinação de classificadores. No entanto, o melhor desempenho de BOOSTING é uma observação inesperada.

ID	Tamanho da Amostra									
	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
01	87	80	70	77	82	85	88	89	89	87
02	77	69	69	70	81	75	78	82	81	83
03	75	75	74	75	76	70	73	73	78	75
04	90	96	98	96	96	97	97	97	97	96
05	93	93	98	99	99	99	99	99	99	99
06	92	97	98	99	99	99	99	99	99	99
07	75	78	72	73	75	75	77	78	84	84
08	77	77	76	71	77	76	80	83	82	83
09	69	85	59	69	73	68	73	79	72	75
10	93	70	79	76	79	75	84	86	86	87
11	53	71	87	87	83	83	89	85	88	87
12	97	98	99	99	99	99	99	99	99	99
13	98	99	99	100	100	99	99	100	100	100
14		33	56	50	49	61	60	67	57	63
15	97	97	96	97	97	98	98	98	98	98

Quadro 2. Percentuais de acerto para C4.5

ID	Tamanho da Amostra									
	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
01	80	77	78	84	82	89	90	89	91	91
02	77	73	75	77	81	77	83	84	84	85
03	79	74	69	74	78	73	75	75	78	78
04	90	94	97	97	97	97	98	97	98	98
05	98	95	95	97	96	96	97	98	97	98
06	93	97	99	99	99	99	99	99	99	99
07	70	80	77	81	78	80	84	84	87	87
08	83	77	70	78	81	82	85	84	84	87
09	69	69	59	67	71	70	76	73	72	79
10	89	80	74	81	85	83	88	87	87	88
11	73	84	87	87	83	86	89	86	91	89
12	98	98	99	99	99	99	99	99	99	99
13	98	99	100	100	100	100	99	100	99	100
14		39	56	61	56	70	63	68	67	70
15	97	97	96	97	98	98	98	98	98	99

Quadro 3. Percentuais de acerto para BAGGING

ID	Tamanho da Amostra									
	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
01	82	83	76	80	86	89	90	91	90	93
02	77	73	75	80	81	86	89	89	92	92
03	79	58	73	72	76	78	82	86	87	83
04	91	97	98	96	98	99	99	98	99	98
05	96	95	95	96	95	97	97	97	98	97
06	95	99	99	100	100	100	100	100	100	100
07	72	82	77	82	80	82	84	87	88	89
08	80	67	70	77	80	86	87	89	88	89
09	54	85	62	67	80	72	83	81	81	83
10	89	80	80	78	85	83	88	88	89	91
11	53	77	89	90	83	81	87	90	91	91
12	98	99	99	99	99	99	99	100	100	99
13	98	98	99	99	100	100	100	100	100	100
14		39	63	69	69	80	76	72	74	84
15	98	97	96	97	98	99	99	99	99	99

Quadro 4. Percentuais de acerto para BOOSTING

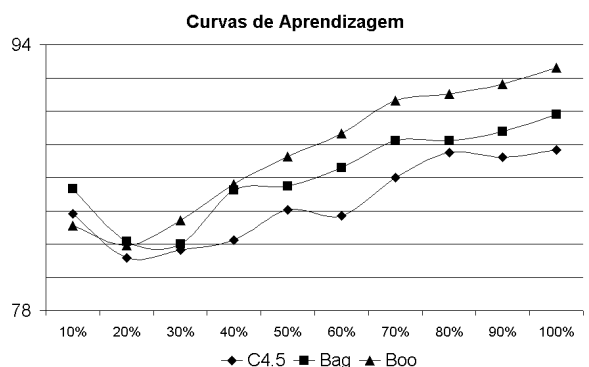


Gráfico 1. Curvas de Aprendizagem Média para os 3 Métodos

Método	Tamanho das Amostras									
	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
C4.5	83.8	81.2	81.7	82.3	84.1	83.7	86.0	87.5	87.2	87.7
BAGGING	85.3	82.1	82.0	85.2	85.5	86.6	88.2	88.2	88.8	89.8
BOOSTING	83.1	81.9	83.4	85.6	87.3	88.6	90.6	91.1	91.6	92.6

Quadro 5. Médias das Amostras

Método	Bases de Dados														
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
C4.5	83.4	76.5	74.4	96.0	95.5	98.0	77.1	78.2	72.2	81.5	81.3	98.7	99.4	55.1	97.4
BAGGING	85.1	79.6	75.3	96.3	96.7	98.2	80.8	81.1	70.5	84.2	85.5	98.8	99.5	61.1	97.6
BOOSTING	86.0	83.4	77.4	97.3	96.3	99.3	82.3	81.3	74.8	85.1	83.2	99.1	99.4	69.6	98.1

Quadro 6. Médias por base para cada Método

Método	Média
C4.5	84.31
BAGGING	86.02
BOOSTING	87.50

Quadro 7. Média Geral para cada Método

## Impacto do Número de Classificadores

Os quadros 8 e 9 apresentam respectivamente os dados gerados com variação do número de classificadores para BAGGING e BOOSTING. Uma primeira observação dos valores permite verificar que os dados encontrados são semelhantes para ambos os métodos.

Em geral, as variações em percentual são numericamente pequenas, mas representam um aumento da precisão com o aumento do número de classificadores.

Num olhar mais minucioso sobre os resultados das bases de dados observa-se que para BAGGING a maioria das bases (11 delas) houve algum aumento da taxa de acerto, porém pouco relevante.

Apenas para as bases *arrhythmia* (01) e *echo cardiogram* (09) houve um aumento da taxa de acerto acima de 1%.

Já o método BOOSTING apresenta um aumento mais significativo na taxa de acerto e este aumento acompanha o número de classificadores em 9 das 15 bases de dados.

Os casos mais significativos são para as bases de dados: *arrhythmia* (01), *pima diabetes* (07), *cleveland14heart disease* (08) e *hepatitis* (11).

Número de Classificadores															
ID	10	20	30	40	50	60	70	80	90	100	120	140	160	180	200
1	82	83	84	83	84	84	84	84	84	84	84	84	84	84	84
2	81	81	81	81	81	81	81	81	81	81	81	81	81	81	81
3	73	73	73	73	73	73	73	73	73	73	73	73	73	73	73
4	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96
5	95	95	95	96	96	96	96	96	96	96	96	96	96	96	96
6	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99
7	76	76	76	76	76	76	76	76	76	76	76	76	76	76	76
8	79	80	80	79	80	80	80	80	80	80	80	80	80	80	80
9	61	61	62	62	63	62	62	63	62	63	62	63	63	63	63
10	81	81	81	81	81	81	81	81	81	81	81	82	82	82	82
11	81	81	81	81	82	82	81	81	81	81	81	81	81	82	82
12	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99
13	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99
14	69	69	69	69	69	69	69	69	69	69	69	69	69	69	69
15	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98

Quadro 8. Valores de taxa de acerto para BAGGING

Número de Classificadores															
ID	10	20	30	40	50	60	70	80	90	100	120	140	160	180	200
01	81	82	83	83	83	83	83	84	84	84	84	84	84	84	84
02	85	85	84	85	85	85	85	85	85	85	85	85	85	84	85
03	67	67	67	66	66	66	66	66	66	66	66	66	66	66	66
04	96	96	96	97	97	97	97	97	97	97	97	97	97	97	97
05	96	97	97	97	97	97	97	97	97	97	97	97	97	97	97
06	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
07	72	73	73	73	74	74	74	74	74	74	74	74	74	74	74
08	79	80	80	80	80	80	80	81	81	81	81	81	81	81	81
09	61	61	60	62	61	62	62	62	62	62	62	62	62	62	62
10	79	80	80	80	80	80	81	81	80	80	80	80	80	80	80
11	82	83	84	83	84	84	84	84	85	85	85	85	85	85	85
12	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99
13	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99
14	55	56	56	56	56	56	56	56	56	56	56	56	56	56	56
15	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98

Quadro 9. Valores de taxa de acerto para BOOSTING

Aprofundando a análise dos resultados, aplica-se o teste de Friedman, para estes conjunto de dados (Quadro 10). Este teste indica se há significância estatística da variação da taxa de acerto para os métodos BAGGING e BOOSTING. Neste Quadro temos para cada número de classificadores as médias e desvios encontrados sobre as taxas de acerto para todas as bases de dados, bem como os índices de *p-valor* inferiores

ao nível de significância de 0,05 (confiabilidade de 95%).

Este último valor permite rejeitar estatisticamente a hipótese nula pela qual um valor de taxa de acerto obtido pertence a um mesmo conjunto de dados, ou seja, admite-se a hipótese alternativa de que existem diferenças relevantes entre os valores.

Nota-se que esta conclusão por si só não permite afirmar uma progressão na taxa de acerto em função do aumento do número de classificadores. Por outro lado, ela permite afirmar que os conjuntos de valores são de um ponto de vista estatístico, diferentes entre si.

Esta conclusão, somada à observação do aumento numérico da taxa de acerto verificado em algumas bases, permite deduzir que de fato existe uma progressão nos valores.

Note-se ainda que pelos *p-valores* calculados é possível observar que a variação nas bases de dados é significativa (*p-valor* inferior a  $10E-5$ ).

Cabe salientar que esta existência de variação significativa indica de fato um crescimento da taxa de acerto. No entanto, de um ponto de vista numérico este aumento é muito pequeno (cerca de um ponto percentual apenas).

N.Class	BAGGING		BOOSTING	
	Média	Desvio	Média	Desvio
10	84.53	12.43	83.12	14.73
20	84.80	12.33	83.67	14.47
30	84.81	12.34	83.73	14.61
40	84.89	12.24	83.89	14.41
50	85.02	12.13	83.96	14.46
60	84.99	12.19	84.02	14.43
70	85.00	12.14	84.04	14.41
80	85.05	12.09	84.14	14.37
90	85.03	12.14	84.11	14.45
100	85.03	12.07	84.11	14.44
120	85.04	12.10	84.19	14.42
140	85.11	12.06	84.24	14.36
160	85.09	12.08	84.25	14.37
180	85.10	12.06	84.21	14.40
200	85.10	12.08	84.16	14.47
p-valor	0.00000		0.00000	

Quadro 10. Teste de Friedman

## Discussão e Conclusões

Na área da saúde, a seleção adequada do método de descoberta de conhecimento é essencial. Em geral, nesta área, as bases de dados possuem um grande número de atributos, um pequeno número de ocorrências e grande frequência de valores ausentes. Frente a tais características, a utilização de técnica de aprendizagem de máquina baseada na geração e combinação de classificadores mostrou-se uma abordagem mais adequada. Pôde-se verificar, a partir da geração e análise de várias curvas de aprendizagem, que o método BOOSTING tem uma eficiência superior aos demais, a saber: C4.5 e BAGGING. Entretanto, cabe aqui uma ressalva para destacar que para casos pontuais o método

BAGGING apresentou uma taxa de acerto semelhante e em raras situações chegou a ser numericamente superior aos resultados obtidos pelo método BOOSTING . Com relação ao C4.5, o método BAGGING foi sempre superior.

Em geral, sobre o impacto do número de classificadores, através do teste estatístico de Friedman pode-se dizer que o aumento do número de classificadores para os métodos BAGGING e BOOSTING gera um impacto positivo sobre as bases de dados. Dentro do escopo da análise aqui desenvolvida verifica-se que existe um aumento da taxa de acerto com o aumento do número de classificadores.

Em conclusão, recomenda-se a utilização do método BOOSTING com tantos classificadores quanto possível para aumentar a taxa de acerto obtida. Obviamente, tal recomendação é válida quando as características da base de dados de treinamento se assemelham aquelas das bases de dados de teste utilizadas no contexto deste trabalho.

## Referências

- [1] KOTSIANTI, S.B.; KANELLOPOULOS, D. Combining Bagging, BOOSTING and Dagging for Classification Problems. **Knowledge-Based Intelligent Information and Engineering Systems**, LNAI 4693, Springer, 2007.
- [2] BREIMAN, L. Bagging predictors. **Machine learning**, 24(2), 123-140, Springer, 1996.
- [3] FREUND, Y.; SCHAPIRE, R.E. Experiments with a new BOOSTING algorithm. **Proceedings of the 13<sup>th</sup> International Conference on Machine Learning**, 148-156, Morgan Kaufmann, 1996.
- [4] TING, K.M., WITTEN, I.H. Stacking Bagged and Dagged Models. In: **Fourteenth International Conference on Machine Learning**, San Francisco, CA, pp. 367-375, 1997.
- [5] QUINLAN, J.R. Bagging, BOOSTING and C4.5. **Proceedings of the Thirteenth National Conference on Artificial Intelligence and Eighth Innovative Applications of Artificial Intelligence Conference, AAAI 96, IAAI 96**, August 4-8, 1996, Portland, Oregon - Volume 1. AAAI Press / The MIT Press, 1996.
- [6] QUINLAN, J.R. **C4.5: Programs for machine learning**. San Francisco: Morgan Kaufman, 1993.
- [7] DIETTERICH T.G. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, BOOSTING , and Randomization. **Machine Learning**, 40(2): 139-157, Springer, 2000.
- [8] EKLUND, P.W; HOANG, A. **A performance Survey of Public Domain Supervised Machine Learning Algorithms**, 2001.  
Disponível em <http://citeseer.ist.psu.edu/142129.html> Acesso em 01 Agosto 2008.
- [9] FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. **From data mining to knowledge discovery. advances in knowledge discovery & Data Mining**,1-34, 1996.
- [10] BAUER, E.; KOHAVI, R. An Empirical Comparison of Voting Classification Algorithms: Bagging, BOOSTING, and Variants. **Machine Learning**, 36(1/2): 105-139. Springer, 1999.
- [11] **UCI: Machine Learning Repository Content Summary**. Disponível em <http://www.ics.uci.edu/~mllearn/MLSummary.html> Acesso em 01 Agosto 2008.
- [12] SCUSE, D.E.; REUTEMANN, P. **Weka Experimenter Tutorial for Version 3.4**, Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/> - Acesso em 01 de Agosto de 2008.
- [13] LIU, H.; YU, L. Toward integrating feature selection algorithms for classification and clustering. **IEEE Transactions on Knowledge and Data Engineering**, 17(4): 491-502, 2005.
- [14] DIAMANTIDIS, N.A.; KARLIS, D.; GIANKOUMAKIS, E.A. Unsupervised stratification of cross-validation for accuracy estimation. **Artificial Intelligence**, 116: 1-16, 2000.
- [15] DEMSAR, J. Statistical Comparisons of Classifiers over Multiple Data Sets. **Journal of Machine Learning Research**, 7: 1-30, 2006.

## Contato

Lucelene Lopes – PPGCC - FACIN - PUCRS  
Av. Ipiranga 6681 – prédio 32  
90.619-900 - Porto Alegre - RS  
[lucelene@gmail.com](mailto:lucelene@gmail.com)