UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE – UFRN
DEPARTAMENTO DE INFORMÁTICA E MATEMÁTICA APLICADA – DIMAP
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS E COMPUTAÇÃO – PPgSC

Novas Abordagens para Configurações Automáticas dos Parâmetros de Controle em Comitês de Classificadores

Diego Silveira Costa Nascimento

Natal – RN

Dezembro, 2014

Diego Silveira Costa Nascimento

Novas Abordagens para Configurações Automáticas dos Parâmetros de Controle em Comitês de Classificadores

Tese apresentada à banca examinadora como parte dos requisitos necessários para formação no doutoramento em Sistema e Computação pela Universidade Federal do Rio Grande do Norte.

Orientador:
Dra. Anne Magály de Paula Canuto

Coorientador:
Dr. André Luís Vasconcelos Coelho

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE – UFRN
DEPARTAMENTO DE INFORMÁTICA E MATEMÁTICA APLICADA – DIMAP
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS E COMPUTAÇÃO – PPGSC

Natal – RN

Dezembro, 2014

DIEGO SILVEIRA COSTA NASCIMENTO

Novas Abordagens para Configurações Automáticas dos Parâmetros de Controle em Comitês de Classificadores

Esta Tese foi julgada adequada para a obtenção do título de doutor em Ciência da Computação e aprovado em sua forma final pelo Programa de Pós-Graduação em Sistemas e Computação do Departamento de Informática e Matemática Aplicada da Universidade Federal do Rio Grande do Norte.

Prof.ª Dr.ª Anne Magaly de Paula Canuto – UFRN
Orientador

Prof. Dr. Uirá Kulesza – UFRN
Coordenador do Programa

Banca Examinadora

Prof.ª Dr.ª Anne Magaly de Paula Canuto – UFRN
Presidente

Prof. Dr. Bruno Motta de Carvalho – UFRN

Prof. Dr. André Carlos Ponce de Leon Ferreira de Carvalho – USP

Prof. Dr. André Luís Vasconcelos Coelho – UNIFOR

Prof. Dr. Araken de Medeiros Santos – UFERSA

Dezembro, 2014

Este trabalho é dedicado à minha família.

Agradecimentos

A Deus pela dádiva da vida e a Nossa Senhora de Fátima por iluminar e abençoar meus caminhos.

Aos meus pais, Petrúcio e Arlete, que sempre estiveram próximos, com amor e carinho, em todas as etapas de minha vida. Vocês foram os meus primeiros professores, e sem dúvidas, tudo isso não poderia ter acontecido.

Aos meus familiares, a minha irmã Andreza, as tias Alice e Araci, pelo apoio e presença constantes.

À minha esposa Ana Iris, por ter me incentivado na retomada das pesquisas e me dado forças para que eu pudesse concluir esse árduo doutorado.

À professora Anne pela oportunidade singular em fazer parte do Dimap, aceitando-me como aluno no doutoramento, com atenção e paciência.

Ao professor André, pelas contribuições significativas e diferenciadas na minha formação acadêmica tanto no mestrado quanto no doutorado. Pelos seus exemplos éticos, pela seriedade acadêmica e busca intelectual clássica, nos quais pretendo me espelhar sempre.

Aos que fazem a Universidade Federal do Rio Grande do Norte, em especial, a todos os professores e funcionários do Dimap, pela dedicação diária para fazer um departamento cada dia melhor.

Não poderia deixar de agradecer também, às instituições de ensino nas quais fui aluno, e que sem dúvidas, deixaram inestimáveis contribuições para minha longa caminhada: a Creche Escola Divino Espírito Santo, Escolinha Mundo Encantado, Colégio Cristo Rei, Centro de Estudos Superiores de Maceió (CESMAC), Universidade Federal do Ceará e Universidade de Fortaleza (Unifor).

A todos os professores que me ensinaram, citando, é claro, a Professora da Disciplina de

Ciências, Solange Maria Lemes de Campos, pela motivação inicial da investigação científica.

Aos Professores participantes da banca de avaliação, André Carlos Ponce de Leon Ferreira de Carvalho, Bruno Motta de Carvalho e Araken de Medeiros Santos, pelas inestimáveis críticas construtivas para melhoria da tese.

E a todos que contribuíram de forma direta ou indiretamente para realização deste trabalho.

A todos vocês, meu **muito obrigado**!

Epígrafe

"Sabemos como é a vida: num dia dá tudo certo e no outro as coisas já não são tão perfeitas assim. Altos e baixos fazem parte da construção do nosso caráter. Afinal, cada momento, cada situação, que enfrentamos em nossas trajetórias é um desafio, uma oportunidade única de aprender, de se tornar uma pessoa melhor. Só depende de nós, das nossas escolhas...

Não sei se estou perto ou longe demais, se peguei o rumo certo ou errado. Sei apenas que sigo em frente, vivendo dias iguais de forma diferente. Já não caminho mais sozinho, levo comigo cada recordação, cada vivência, cada lição. E, mesmo que tudo não ande da forma que eu gostaria, saber que já não sou a mesma de ontem me faz perceber que valeu a pena. Procuro ser uma pessoa de valor, em vez de procurar ser uma pessoa de sucesso.
O sucesso é só consequência."

Albert Einstein

Resumo

Significativos avanços vêm surgindo em pesquisas relacionadas ao tema de Comitês de Classificadores. Os modelos que mais recebem atenção na literatura são aqueles de natureza estática, ou também conhecidos por *ensembles*. Dos algoritmos que fazem parte dessa classe, destacam-se os métodos que utilizam reamostragem dos dados de treinamento: *Bagging*, *Boosting* e *Multiboosting*. A escolha do tipo de arquitetura e dos componentes a serem recrutados não é uma tarefa trivial, e tem motivado, ainda mais, o surgimento de novas propostas na tentativa de se construir tais modelos de forma automática e, muitas delas, são baseadas em métodos de otimização. Muitas dessas contribuições não têm apresentado resultados satisfatórios quando aplicadas a problemas mais complexos ou de natureza distinta. Em contrapartida, a tese aqui apresentada propõe três novas abordagens híbridas para construção automática em *ensembles* de classificadores: Incremento de Diversidade, Função de Avaliação Adaptativa e Meta-aprendizado para a elaboração de sistemas de configuração automática dos parâmetros de controle para os modelos de *ensemble*. Na primeira abordagem, é proposta uma solução que combina diferentes técnicas de diversidade em um único arcabouço conceitual, na tentativa de se alcançar níveis mais elevados de diversidade em *ensemble*, e com isso, melhor o desempenho de tais sistemas. Já na segunda abordagem, é utilizado um algoritmo genético para o *design* automático de *ensembles*. A contribuição consiste em combinar as técnicas de *filtro* e *wrapper* de forma adaptativa para evoluir uma melhor distribuição do espaço de atributos a serem apresentados aos componentes de um *ensemble*. E por fim, a última abordagem, que propõe uma nova técnica de recomendação de arquitetura e componentes base em *ensemble*, via técnicas de meta-aprendizado tradicional e multirrótulo. De forma geral os resultados são animadores, e corroboram com a tese de que ferramentas híbridas são uma poderosa solução na construção de *ensembles* eficazes em problemas de classificação de padrões.

Palavras-chave: *Bagging*, *Boosting*, *Multiboosting*, Incremento de Diversidade em *Ensembles*, Função de Avaliação Adaptativa, Meta-aprendizado, Meta-aprendizado Multirrótulo.

Abstract

Significant advances have emerged in research related to the topic of Classifier Committees. The models that receive the most attention in the literature are those of the static nature, also known as ensembles. The algorithms that are part of this class, we highlight the methods that using techniques of resampling of the training data: Bagging, Boosting and Multiboosting. The choice of the architecture and base components to be recruited is not a trivial task and has motivated new proposals in an attempt to build such models automatically, and many of them are based on optimization methods. Many of these contributions have not shown satisfactory results when applied to more complex problems with different nature. In contrast, the thesis presented here, proposes three new hybrid approaches for automatic construction for ensembles: Increment of Diversity, Adaptive-fitness Function and Meta-learning for the development of systems for automatic configuration of parameters for models of ensemble. In the first one approach, we propose a solution that combines different diversity techniques in a single conceptual framework, in attempt to achieve higher levels of diversity in ensembles, and with it, the better the performance of such systems. In the second one approach, using a genetic algorithm for automatic design of ensembles. The contribution is to combine the techniques of filter and wrapper adaptively to evolve a better distribution of the feature space to be presented for the components of ensemble. Finally, the last one approach, which proposes new techniques for recommendation of architecture and based components on ensemble, by techniques of traditional meta-learning and multi-label meta-learning. In general, the results are encouraging and corroborate with the thesis that hybrid tools are a powerful solution in building effective ensembles for pattern classification problems.

Keywords: Bagging, Boosting, Multiboosting, Increment of Diversity on Ensemble, Adaptive-fitness Function, Meta-learning, Multi-label Meta-learning.

Sumário

Lista de Abreviações **xi**

Lista de Figuras **xiv**

Lista de Tabelas **xvi**

1 Introdução **1**

1.1 Motivações 3

1.2 Justificativas 4

1.3 Objetivos 5

1.4 Contribuições da Pesquisa 6

1.5 Organização do Documento 7

2 Aprendizado de Máquina Supervisionado **8**

2.1 Comitês de Classificadores 9

2.1.1 *Ensembles* de Classificadores 12

2.1.2 Construção dos Componentes 15

2.2 Classificação Multirrótulo 19

2.2.1 Abordagem Independente de Algoritmo 20

2.2.2 Abordagem Dependente de Algoritmo 21

2.2.3 Medidas de Avaliação 22

2.3 Meta-aprendizado 24
2.3.1 Caracterização de Bases de Dados 27
2.3.2 Medidas de Avaliação 30
2.3.3 Formas de Sugestão 30
2.4 Síntese do Capítulo 31

3 Construção Automática em Comitês **34**
3.1 Trabalhos Relacionados 35
3.1.1 Escolha dos Componentes de um Comitê 35
3.1.2 Métodos de Otimização para a Seleção de Componentes 37
3.1.3 Meta-aprendizado para a Seleção de Componentes 37
3.2 Novas Abordagens para Construção Automática em Comitês 38
3.2.1 Primeira Abordagem: Algoritmo Genético para Incremento de Diversidade 39
3.2.2 Segunda Abordagem: Algoritmo Genético com Função de Avaliação Adaptativa 42
3.2.3 Terceira Abordagem: Recomendação via Meta-aprendizado 44
3.3 Síntese do Capítulo 49

4 Experimentos Computacionais **51**
4.1 Problemas de Classificação 51
4.2 Algoritmos de Aprendizagem 52
4.3 Algoritmos de Classificação Multirrótulo 55
4.4 Algoritmo Genético 56
4.4.1 Extração das Meta-características 56
4.5 Avaliação de Algoritmos 57

4.6 Síntese do Capítulo 58

5 Resultados **59**

5.1 Desempenho dos Algoritmos de *Ensembles* Tradicionais 59

5.2 Primeira Abordagem: Algoritmo Genético para Incremento de Diversidade . . . 64

5.3 Segunda Abordagem: Algoritmo Genético com Função de Avaliação Adaptativa . 73

5.4 Terceira Abordagem: Recomendação via Meta-aprendizado 76

5.4.1 Meta-aprendizado tradicional 77

5.4.2 Meta-aprendizado Multirrótulo 78

5.5 Síntese do Capítulo 81

6 Conclusões **83**

6.1 Limitações da Proposta 84

6.2 Trabalhos Futuros 85

Referências Bibliográficas **86**

Apêndice A – Problemas de Classificação **96**

Apêndice B – Configurações dos Algoritmos de Aprendizagem **99**

Apêndice C – Resultados Empíricos **102**

Apêndice D – Extração de Meta-características **124**

Lista de Abreviações

AG – Algoritmo Genético

AMAA – Assimetria Média Absoluta dos Atributos

BD – Boa Diversidade

CCP – Coeficiente de Correlação de *Pearson*

CMA – Curtose Média dos Atributos

DS – Decision *Stump*

DP – Desvio Padrão

DT – *Decision Table*

EA – Entropia dos Atributos

EC – Entropia de Classe

ECON – Entropia Conjunta

FM – *F-Measure*

HL – *Hamming-loss*

IM – Informação Mútua

KNN – *k - Nearest Neighbors*

LP – *Label Powerset*

LVFR – Linhas com Valores Faltosos Relativos

LVFT – Linhas com Valores Faltosos Totais

MAFM – Macro-*averaged F-Measure*

MD - Má Diversidade

MDA – Média Aritmética

ME – Mistura de Especialista

MIFM – Micro-*averaged F-Measure*

MLP – *Multi-layer Perceptron*

ML-KNN – *Multilabel k - Nearest Neighbors*

NA – Número de Atributos

NC – Número de Classes

NE – Número de Exemplos

NAN – Número de Atributos Numérico

NAO – Número de Atributos com *Outliers*

NAS – Número de Atributos Simbólico

NB – *Naïve Bayes*

N. Comp. – Número de Componentes

RAKEL – *Random k - Labelsets*

RAKELD – *Random k - Labelsets Disjoint*

RBF – *Radial Basis Function*

REP – *Reduced Error Pruning*

RL – *Ranking-loss*

ROC – *Receiver Operator Characteristic*

RT – REP *Tree*

SMO – *Sequential Minimal Optimisation*

SVM – *Support Vector Machine*

VFR – Valores Faltante Relativo

VFT – Valores Faltante Total

WEKA – Waikato *Environment for Knowledge Analysis*

Lista de Figuras

1 Três razões pelas quais *ensembles* têm um melhor desempenho que um preditor único (Adaptada de Dietterich (2000a)). 11

2 Arquitetura geral para um modelo de *ensemble* (NASCIMENTO, 2009). 13

3 Processo de criação de um *ensemble* via algoritmo *Bagging* (NASCIMENTO, 2009). 16

4 Processo de criação de um *ensemble* via algoritmo *Boosting* (NASCIMENTO, 2009). 17

5 Processo de criação de um *ensemble* via algoritmo *MultiBoosting* (NASCIMENTO, 2009). 19

6 Estágios adotados para incremento dos níveis de diversidade em *ensemble*. 40

7 Codificação do cromossomo para seleção de características aplicado à base "credit-a". 41

8 Extração de meta-características. 45

9 Testes para mapeamento. 45

10 Mapeamento tradicional. 46

11 Mapeamento tradicional para escolha da arquitetura. 46

12 Mapeamento tradicional para escolha dos componentes homogêneos. 46

13 Mapeamento multirrótulo. 47

14 Mapeamento multirrótulo para escolha da arquitetura. 48

15 Mapeamento multirrótulo para escolha dos componentes homogêneos. 48

16 Mapeamento multirrótulo para escolha dos componentes heterogêneos. 49

17 Meta-aprendiz. 49

18 Partição dos dados. 57

19 Valores de erro médio para *Boosting*. 60

20 Valores de erro médio para *Bagging*. 61

21 Valores de erro médio para *Multiboosting*. 61

22 Valores de erro médio para *Bagging* quando aplicado à base “anneal 2”. 62

23 Valores de erro médio para *Bagging* quando aplicado à base “breast-cancer 2”. . . 63

24 Valores de erro médio para *Bagging* quando aplicado à base “iris 4”. 63

25 Valores de erro médio para *Bagging* e *Boosting* quando aplicado à base “credit-a”. 64

26 Valores de erro médio para *Multiboosting* quando aplicado à base “haberman”. . 64

27 Erro médio de generalização produzido por cada estágio da abordagem integrada. 69

28 Erro médio de generalização produzido por cada estágio da abordagem integrada (continuação). 70

29 Diagrama de diversidade-erro para base “sick” usando *ensemble* com *Bagging* padrão (topo à esquerda), *ensemble* gerada com seleção de filtro (topo à direita), *ensemble* gerada com seleção de filtro mais reamostragem (base à esquerda) e *ensemble* gerada com seleção de filtro, reamostragem e componentes heterogêneos (base à direita). 71

30 Diagrama de diversidade-erro para base “waveform” usando *ensemble* com *Bagging* padrão (topo à esquerda), *ensemble* gerada com seleção de filtro (topo à direita), *ensemble* gerada com seleção de filtro mais reamostragem (base à esquerda) e *ensemble* gerada com seleção de filtro, reamostragem e componentes heterogêneos (base à direita). 72

Lista de Tabelas

1 Medidas de informação do projeto Statlog 28

2 Medidas de informação adicionais do projeto Metal 29

3 Níveis de diversidade para estatística-Q, Boa e Má Diversidades 66

4 Resultados do teste Nemenyi para estatística-Q 68

5 Resultados do teste Nemenyi para *Boa* Diversidade 68

6 Resultados do teste Nemenyi para *Má* Diversidade 68

7 Resultados do teste de Friedman para o erro médio de generalização 68

8 Resultados do teste de Nemenyi para o erro médio de generalização 68

9 Erro médio (%) e desvio padrão para o genético original 74

10 Erro médio (%) e desvio padrão para o genético proposto 74

11 Valores de t-teste comparando genético proposto com genético original 75

12 Valores de t-teste comparando genético proposto com *Bagging* 76

13 Valores de t-teste comparando genético original com *Bagging* 76

14 Medidas de diversidade 77

15 Valores de erro médio e desvio padrão dos meta-aprendizes para recomendação de arquitetura de *ensemble* 78

16 Valores de erro médio e desvio padrão dos meta-aprendizes para recomendação dos componentes homogêneos para *Bagging*, *Boosting* e *Multiboosting* 78

17 Recomendação multirrótulo para tipos de arquiteturas 79

18 Recomendação multirrótulo para tipos de componentes homogêneo para *Bagging* 80

19 Recomendação multirrótulo para tipos de componentes homogêneos para *Boosting* 80

20 Recomendação multirrótulo para tipos de componentes homogêneos para *Multiboosting* 80

21 Recomendação multirrótulo para tipos de componentes heterogêneos para *Bagging* 81

22 Recomendação multirrótulo para tipos de componentes heterogêneos para *Boosting* 81

23 Recomendação multirrótulo para tipos de componentes heterogêneos para *Multiboosting* 81

24 Conjunto de problemas de classificação 97

25 Características de atributos, instâncias, classes e valores faltantes para os problemas de classificação 98

26 Distribuição das classes 98

27 Configurações dos parâmetros de controle para o algoritmo RBF 99

28 Configurações dos parâmetros de controle para o algoritmo J48 99

29 Configurações dos parâmetros de controle para o algoritmo SMO 100

30 Configurações dos parâmetros de controle para o algoritmo *Naïve Bayes* 100

31 Configurações dos parâmetros de controle para o algoritmo IB*k* 100

32 Configurações dos parâmetros de controle para o algoritmo REP *Tree* 101

33 Configurações dos parâmetros de controle para o algoritmo OneR 101

34 Configurações dos parâmetros de controle para o algoritmo PART 101

35 Configurações dos parâmetros de controle para o algoritmo *Decision Table* 101

36 Valores de erro médio e desvio padrão para *Bagging* quando aplicado às bases "anneal", "breast-cancer" e "bupa" 103

37 Valores de erro médio e desvio padrão para *Bagging* quando aplicado às bases "car", "colic" e "credit-a" 104

38 Valores de erro médio e desvio padrão para *Bagging* quando aplicado às bases "diabetes", "gaussian3" e "glass" 105

39 Valores de erro médio e desvio padrão para *Bagging* quando aplicado às bases "haberman", "heart-c" e "hepatitis" 106

40 Valores de erro médio e desvio padrão para *Bagging* quando aplicado às bases "ionosphere", "iris" e "segment" 107

41 Valores de erro médio e desvio padrão para *Bagging* quando aplicado às bases "sick", "sonar" e "vehicle" 108

42 Valores de erro médio e desvio padrão para *Bagging* quando aplicado às bases "vote" e "waveform-5000" 109

43 Valores de erro médio e desvio padrão para *Boosting* quando aplicado às bases "anneal", "breast-cancer" e "bupa" 110

44 Valores de erro médio e desvio padrão para *Boosting* quando aplicado às bases "car", "colic" e "credit-a" 111

45 Valores de erro médio e desvio padrão para *Boosting* quando aplicado às bases "diabetes", "gaussian3" e "glass" 112

46 Valores de erro médio e desvio padrão para *Boosting* quando aplicado às bases "haberman", "heart-c" e "hepatitis" 113

47 Valores de erro médio e desvio padrão para *Boosting* quando aplicado às bases "ionosphere", "iris" e "segment" 114

48 Valores de erro médio e desvio padrão para *Boosting* quando aplicado às bases "sick", "sonar" e "vehicle" 115

49 Valores de erro médio e desvio padrão para *Boosting* quando aplicado às bases "vote" e "waveform-5000" 116

50 Valores de erro médio e desvio padrão para *Multiboosting* quando aplicado às bases "anneal", "breast-cancer" e "bupa" 117

51 Valores de erro médio e desvio padrão para *Multiboosting* quando aplicado às bases “car”, “colic” e “credit-a” 118

52 Valores de erro médio e desvio padrão para *Multiboosting* quando aplicado às bases “diabetes”, “gaussian3” e “glass” 119

53 Valores de erro médio e desvio padrão para *Multiboosting* quando aplicado às bases “haberman”, “heart-c” e “hepatitis” 120

54 Valores de erro médio e desvio padrão para *Multiboosting* quando aplicado às bases “ionosphere”, “iris” e “segment” 121

55 Valores de erro médio e desvio padrão para *Multiboosting* quando aplicado às bases “sick”, “sonar” e “vehicle” 122

56 Valores de erro médio e desvio padrão para *Multiboosting* quando aplicado às bases “vote” e “waveform-5000” 123

57 Extração de meta-características para as bases “anneal”, “breast-cancer” e “bupa” 125

58 Extração de meta-características para as bases “anneal”, “breast-cancer” e “bupa” 126

59 Extração de meta-características para as bases “car”, “colic” e “credit-a” 127

60 Extração de meta-características para as bases “car”, “colic” e “credit-a” 128

61 Extração de meta-características para as bases “diabetes”, “gaussian3” e “glass” . 129

62 Extração de meta-características para as bases “diabetes”, “gaussian3” e “glass” . 130

63 Extração de meta-características para as bases “haberman”, “heart-c” e “hepatitis” 131

64 Extração de meta-características para as bases “haberman”, “heart-c” e “hepatitis” 132

65 Extração de meta-características para as bases “ionosphere”, “iris” e “segment” . 133

66 Extração de meta-características para as bases “ionosphere”, “iris” e “segment” . 134

67 Extração de meta-características para as bases “sick”, “sonar” e “vehicle” 135

68 Extração de meta-características para as bases “sick”, “sonar” e “vehicle” 136

69 Extração de meta-características para as bases “vote” e “waveform-5000” 137

70 Extração de meta-características para as bases “vote” e “waveform-5000” 137

Capítulo 1

Introdução

No campo da Inteligência Artificial, mais especificamente em Aprendizado de Máquina, novas abordagens estão sempre sendo discutidas, objetivando construir algoritmos que possuam a capacidade de aprender sobre um domínio restrito, bem como, que apresentem autonomia em solucionar um novo problema de forma inteligente. Como um passo nesse sentido, Comitês de Classificadores recorrem à estratégia de se fundirem as decisões oriundas de vários classificadores independentes, na busca de se conseguir uma decisão final que seja potencialmente mais eficaz (HANSEN; SALAMON, 1990).

A primeira razão de combinar múltiplos classificadores está em obter maiores ganhos de acurácia (HANSEN; SALAMON, 1990). Outra razão está relacionada à dificuldade de aplicação de algoritmos de indução a conjuntos de dados extremamente grandes. Muitos desses algoritmos são computacionalmente complexos e acabam não sendo aplicáveis devido às limitações dos equipamentos computacionais disponíveis, essas limitações estão geralmente associadas à memória e tempo de processamento. Uma solução para este problema é o particionamento horizontal do conjunto de dados original em pequenos fragmentos. Em seguida, treina-se um modelo preditivo numa parte menor tratável e finalmente combina-se as predições de todos os modelos resultantes (TSOUMAKAS; ANGELIS; VLAHAVAS, 2005).

É bastante comum encontrar na literatura o termo *Ensembles* de Classificadores (GUAN *et al.*, 2014; NETO; CANUTO, 2014; KARIM; FARID, 2014). Essa é uma subdivisão, por assim dizer, do conceito de Comitês de Classificadores que está relacionada à combinação de modelos de aprendizado de máquina via uma concepção estática para resolver um único problema de forma redundante. É intuitivo que a combinação de múltiplos classificadores para resolver um único problema é vantajosa. No entanto, para que a abordagem de *ensembles* seja capaz de promover melhorias de acurácia, cada classificador, também conhecido por *componente*, deve apresentar bom desempenho individual, ao mesmo tempo em que deve exibir comportamento diverso em relação aos demais para que se consiga alcançar níveis elevados de acurácia. Tomando um exemplo prático de um comitê de pessoas, caso todos os indivíduos tenham as mesmas opiniões e pontos de vista sobre o assunto em questão, dificilmente surgirão discussões que contribuam para uma melhora significativa da decisão final tomada por este comitê. Logo, se um *ensemble* for formado por componentes iguais, consequentemente, cada componente apresentará a mesma resposta, acarretando desta forma, um aumento apenas do custo computacional, sem resultados práticos de ganhos de generalização.

Podemos encontrar alguns métodos disponíveis de geração de *ensembles* capazes de induzir a diferentes formas de diversidade dos componentes (KOTSIANTIS, 2014; KOTSIANTIS; PINTELAS, 2004), dentre eles, alguns conhecidos são: *Bagging* (BREIMAN, 1996a), *Boosting* (FREUND; SCHAPIRE, 1996) e *MultiBoosting* (WEBB, 2000). Esses dizem respeito a técnicas estatísticas poderosas que lançam mão do conceito de redistribuição dos dados (*Bootstrapping* (EFRON; TIBSHIRANI, 1993)) como meio de gerar componentes diversos a serem agregados por voto majoritário ou voto ponderado.

Modelos que são derivados de diferentes execuções de um mesmo algoritmo de aprendizagem são comumente chamados de *ensembles homogêneos* (NIKULIN; BAKHARIA; HUANG, 2013; OLIVEIRA; LUDERMIR, 2011; VINAY; RAO; KUMAR, 2011; BIAN; WANG, 2007). Tais modelos podem ser inferidos mediante alterações dos parâmetros de controle do próprio algoritmo de aprendizagem ou através de manipulação das instâncias de treinamento, atributos de entrada ou códigos das classes de saída (DIETTERICH, 2000a). Por outro lado, modelos que são derivados de diferentes algoritmos de aprendizagem são chamados de *ensembles heterogêneos* (ZHAO; JIANG; XU, 2011, 2010; TSOUMAKAS; ANGELIS; VLAHAVAS, 2005; CANUTO *et al.*, 2005). Tais modelos podem ser inferidos mediante a combinação de algoritmos

de aprendizado de máquina de diferentes paradigmas, como por exemplo: árvores de decisão, redes neurais, máquinas de vetores-suporte, entre outros.

Pesquisas na linha de *ensembles* têm sido apresentadas ao longo dos últimos anos. E muitas das contribuições apresentam novas técnicas para geração automática dos modelos de *ensembles* (SOARES; ANTUNES; ARAúJO, 2013; BAGHERI; QIGANG; ESCALERA, 2013; ZHAO; JIANG; XU, 2011, 2010; NASCIMENTO, 2009; CANUTO *et al.*, 2005; SANTANA *et al.*, 2010) que vão desde a escolha dos componentes, à seleção de características e à escolha dos métodos de combinação, entre outras. E os métodos usualmente aplicados para geração automática dos modelos de *ensembles* são as técnicas de otimização e as soluções bio-inspirados.

1.1 Motivações

É inquestionável que, para todas as áreas do conhecimento humano, por melhor que seja um profissional, quanto ao seu domínio prático ou teórico, não é possível que esse tenha conhecimento completo da sua área de atuação. Isso é facilmente justificado porque o legado de informações é muito extenso, e a cada dia antigos conhecimentos são aprimorados e novos são descobertos. Logo, quando se deseja decidir sobre assuntos melindrosos, sejam eles de impacto social ou econômico, pessoas com domínio no assunto reúnem-se na forma de *comitês* a fim de se capitalizar das várias experiências individuais de cada membro, na busca de se adotar a melhor solução para o caso em questão.

A tentativa constante em se alcançar a melhor solução para um problema, a partir da exploração das potencialidades individuais de um grupo dinâmico e comprometido com um objetivo comum, não é recente. Trata-se de um assunto que recebe muita atenção por profissionais de Administração e suas sub-áreas, cuja denominação usual é *Brainstorming* [1] (OSBORN, 1953). Em ciências políticas, o modelo de júri proposto por Condorcet (1758) foi concebido no final do Século XVIII para estudar sob que condições uma democracia como um todo seria mais efetiva que qualquer uma de suas partes constituintes. Já na área de previsão de séries econométricas, tem sido cada vez mais advogada a tese de que melhores resultados podem ser alcançados mediante a combinação das previsões feitas por diferentes técnicas (CLEMEN, 1989). Final-

[1] Palavra em inglês cuja tradução é "tempestade mental". É uma metodologia de exploração de ideias, visando à obtenção das melhores soluções a partir de um grupo de pessoas.

mente, na disciplina de Engenharia de *Software*, é comum se lançar mão de múltiplas versões redundantes de mesmo código com o intuito de se atingir níveis mais seguros de tolerância a falhas (ECKHARDT; LEE, 1985).

No campo de *ensembles* de classificadores, reunir um conjunto de algoritmos de aprendizado para construir um comitê, pode não ser uma tarefa trivial. Isso porque, deve-se levar em consideração aspectos referentes a: escolha dos tipos de arquitetura (*Bagging*, *Boosting* ou *Multiboosting*) e escolha dos tipos de componentes (árvores de decisão, redes neurais artificiais, classificadores baseados em vizinhança, tabelas de decisões, classificadores baseados em máquinas de vetores-suporte, entre outros), uma vez que estudos nessa linha (ZHAO; JIANG; XU, 2011, 2010; KOTSIANTIS, 2014; NASCIMENTO; COELHO, 2009a, 2009b; NASCIMENTO, 2009; CANUTO *et al.*, 2007) vêm confirmando que encontrar o arranjo de *ensemble* mais adequado para um problema de classificação, é importante para se obter ganhos significativos de generalização.

De forma a contribuir com o tema, partimos de três hipóteses na tentativa de se alcançar melhores resultados de acurácia. Na primeira hipótese, propomos uma abordagem que busca elevar os níveis de diversidade de um *ensemble* incrementalmente, a partir de diferentes técnicas de diversidade. Com isso acreditamos que a cada passo de incremento do nível de diversidade, o erro médio do *ensemble* irá diminuir. Na segunda hipótese, acreditamos que podemos melhorar o erro médio através da seleção de características combinados os métodos de avaliação de filtro e *wrapper*. E na terceira hipótese, a recomendação dos parâmetros de controle em *ensembles* pode ser realizadas mediante uma nova abordagem via meta-aprendizado.

Resumidamente, essas são as motivações nas quais apresentamos nesta tese, três novas abordagens híbridas para a configuração automática dos parâmetros de controle em *ensembles* que serão avaliadas e testadas quando aplicadas em diferentes problemas de classificação de padrões.

1.2 Justificativas

A manifestação da inteligência pressupõe a aquisição, o armazenamento e a inferência de conhecimento. Para que o conhecimento possa ser armazenado, é essencial que se possa representá-lo (FERNANDES, 2005). Para isso, podemos contar com uma quantidade significativa de algo-

ritmos de aquisição automática do conhecimento. Muitos deles, vêm obtendo êxito considerável em vários domínios de atuação, tais como engenharia, medicina, economia, entre outros. Esses algoritmos variam nos seus objetivos, na disponibilidade dos dados de treinamento, nas estratégias de aprendizagem e na linguagem que empregam para representação do conhecimento. Embora cada algoritmo possua sua particularidade, todos eles aprendem através de buscas em espaço de estados possíveis, para encontrar uma generalização aceitável (TSOUMAKAS; ANGELIS; VLAHAVAS, 2005). Porém, deve-se observar qual o melhor método a ser utilizado, visto que não existe um algoritmo que seja bom o suficiente para qualquer tipo de problema – *No Free Lunch* (WOLPERT; MACREADY, 1997). O presente projeto insere-se no contexto de aprendizado automático, na medida em que passa a contemplar novas estratégias híbridas em Comitês de Classificadores, as quais serão empregadas em tarefas não-triviais de classificações de padrões. É importante motivar mais uma vez que *ensembles* têm sido bastante explorado nos últimos anos no contexto da Estatística, Aprendizado de Máquina, Reconhecimento de Padrões e Descoberta de Conhecimento, por se tratar de uma abordagem simples e capaz de aumentar a capacidade de generalização de soluções baseadas em aprendizado indutivo (COELHO, 2004, 2006).

1.3 Objetivos

A proposta desta pesquisa tem como cerne de investigação estudar alguns dos métodos para construção de *ensembles*, levando-se em considerações diferentes aspectos de construção, buscando obter respostas às seguintes questões:

- É possível elevar os níveis de diversidade em *ensemble* de forma a se alcançar melhores índices de desempenhos a partir da combinação de diferentes técnicas de diversidade?
- As técnicas de avaliação de indivíduos em Algoritmo Genético via filtro ou *wrapper*, quando aplicadas juntas, pode influenciar na evolução de um possível melhor indivíduo?
- Meta-aprendizado, tem apresentando sucesso em recomendação de algoritmos, tal ideia apresenta resultados satisfatório para recomendação de parâmetros de controle em *ensemble*?

1.4 Contribuições da Pesquisa

As contribuições alcançadas com as pesquisas foram:

- A proposta de uma nova abordagem que combina diferentes técnicas de diversidade em um único arcabouço conceitual, na tentativa de se alcançar níveis de generalização em *ensemble* mais elevados;
- A proposta de uma segunda nova abordagem para configuração automática em *ensembles* que utiliza as técnicas de filtro e *wrapper*, para seleção do espaço de atributos a serem treinados pelos componentes individuais em *ensembles*;
- A proposta de uma terceira nova abordagem para recomendação automática dos parâmetros de controle em *ensembles* via meta-aprendizado tradicional e multirrótulo, no que se refere à:
  - Escolha da arquitetura;
  - Escolha dos tipos de componentes homogêneos;
  - Escolha dos tipos de componentes heterogêneos;

Como contribuição literária, publicamos vários dos resultados alcançados durante as investigações em importantes meios científicos, tanto nacionais quantos internacionais:

1. NASCIMENTO, D.S.C.; CANUTO, A.M.P.; COELHO, A.L.V.. **An Empirical Analysis of Meta-learning for The Automatic Choice of Architecture and Components in Ensemble Systems**, *Brazilian Conference on Intelligent Systems*, São Paulo, 2014.
2. NASCIMENTO, D.S.C; COELHO, A.L.V.; CANUTO, A.M.P.. **Integrating Complementary Techniques for Promoting Diversity in Classifier Ensembles: a Systematic Study**, Neurocomputing, 2014.
3. CANUTO, A.M.P.; NASCIMENTO, D.S.C.. **A Genetic-based Approach to Features Selection for Ensembles Using a Hybrid and Adaptive Fitness Function**, *World Congress on Computational Intelligence*, Brisbane, 2012.

4. NASCIMENTO, D.S.C.; CANUTO, A.M.P.; COELHO, A.L.V.. **Combining Different Ways to Generate Diversity in Bagging Models: an Evolutionary Approach**, *International Joint Conference on Neural Networks*, San Jose, 2011.

1.5 Organização do Documento

Este documento está organizado em 6 capítulos, incluindo este. Cada um dele, permite ao leitor uma compreensão didática de todas as etapas necessárias para o entendimento das contribuições apresentadas nesta tese, e estão dispostos conforme a seguir:

- No Capítulo 2, são discutidas a fundamentação teórica no que se refere à área de Aprendizado de Máquina, bem como, os principais algoritmos de classificação de padrões aqui utilizados, e a fundamentação teórica necessária para o entendimento dos conceitos de Comitês de Classificadores. O capítulo também traz uma revisão teórica sobre Aprendizado Multirrótulo e a fundamentação teórica sobre Meta-aprendizado, bem como uma descrição das fases para construção de modelos de recomendação de algoritmos.
- O Capítulo 3 traz discussões sobre as contribuições propostas e as etapas que seguidas para a consolidação da contribuição da tese;
- O Capítulo 4 detalha os experimentos computacionais utilizados;
- O Capítulo 5 apresenta as discursões sobre os resultados obtidos através das novas contribuições da pesquisa proposta; e
- Por fim, a tese é concluída no Capítulo 6, onde são apresentados as conclusões finais e os trabalhos futuros.

Capítulo 2

Aprendizado de Máquina Supervisionado

Aprendizado de Máquina é uma área de estudo destinada a estudar e desenvolver sistemas capazes de aprender sobre um determinado domínio específico, e a partir do aprendizado adquirido, tomar decisões de forma autônoma (FACELI *et al.*, 2011; REZENDE, 2003; MITCHELL, 1997). Já o termo supervisionado vem a contribuir com a ideia de que tais sistemas inteligentes, discutidos aqui, aprendem auxiliados por um elemento supervisor.

Mitchell (1997) sugere uma organização das contribuições em Aprendizado de Máquina Supervisionado segundo alguns paradigmas: simbólico, estatístico, baseados em exemplos, conexionista e evolutivo. No paradigma de aprendizado simbólico, a ideia geral consiste em construir representações simbólicas através da análise de conceitos representativos de um conjunto de dados. Existem dois tipos de métodos que atendem a esse paradigma, os algoritmos baseados em árvores de decisão (WITTEN; FRANK, 2005; QUINLAN, 1993) e os algoritmos baseados em regras (FRANK; WITTEN, 1998; KOHAVI, 1995; HOLTE, 1993). Já para o paradigma estatístico, são utilizados modelos estatísticos para encontrar uma boa aproximação do conceito induzido. Vários desses métodos são paramétricos, assumindo alguma forma de modelo, e então encontrando valores apropriados para os parâmetros do modelo a partir dos dados. Um método representativo deste paradigma é o *Naïve Bayes* (JOHN; LANGLEY, 1995). Para o paradigma

baseado em exemplos, a ideia geral consiste em usar a experiência passada, acumulando casos e tentando descobrir, por analogia, soluções para outros problemas (FERNANDES, 2005). Um método representativo desse paradigma é o algoritmo *k*-NN, baseado nos conceitos de vizinhança e aprendizado local. No paradigma conexionista, a ideia geral consiste em construir modelos matemáticos simplificados inspirados no modelo biológico do sistema nervoso. O conhecimento é representado como padrões de atividades em redes de pequenas unidades de processamento individuais. Métodos representativos desses domínio são as redes neurais artificiais (HARPHAM; DAWSON; BROWN, 2004) e máquinas de vetores-suporte (PLATT, 1999). E o paradigma evolutivo, que é derivado do modelo biológico de aprendizado (GOLDBERG, 1989). A ideia toma como base que as populações evoluíram na natureza de acordo com os princípios de seleção natural e sobrevivência dos mais aptos, postulados em Darwin (1859). Os Algoritmos Genéticos estabelecidos por Holland (1975) são métodos adaptativos que podem ser usados para resolver problemas de busca e otimização. Outros temas em Aprendizado de Máquina são discutidos nas Seções 2.1, 2.2 e 2.3.

2.1 Comitês de Classificadores

Ao invés de se dedicar ao estudo de desenvolver algoritmos de aprendizado de máquina individual superajustados para resolver um problema de classificação em particular (NIKULIN; BAKHARIA; HUANG, 2013), a ideia de Comitês de Classificadores consiste em reunir um conjunto de algoritmos de aprendizado de máquina para resolver um problema de forma conjunta. Desta forma, espera-se que o resultado alcançado pelo grupo de classificadores seja potencialmente melhor quando comparado ao resultado produzido individualmento por um único algoritmo de aprendizado de máquina.

Hansen e Salamon (1990) sugerem ainda a utilização desta abordagem mesmo no caso do comitê inicialmente apresentar um desempenho igual ou inferior a um algoritmo de classificação único. Isso porque um melhor desempenho ainda pode ser alcançado mediante uma combinação mais apropriada dos parâmetros de configuração ou provendo treinamentos individuais com diferentes subconjuntos de dados para cada algoritmo. Outros autores também justificam os benefícios da utilização de comitês de classificadores. Como podemos encontrar em Dietterich (2000a), existem três razões fundamentais pelas quais se deve utilizar essas técnicas, conforme

ilustradas na Figura 1.

Um algoritmo de aprendizagem opera realizando uma busca no espaço H de hipóteses, na tentativa de alcançar uma função objetivo f desejada. O primeiro motivo é *estatístico*. Ele acontece quando o tamanho do conjunto de dados de treinamento é inferior ao tamanho do espaço de hipóteses. Sem um conjunto suficiente de dados, o algoritmo de aprendizagem pode encontrar diferentes hipóteses em H. Na Figura 1 (a) o modelo é treinado pelo subconjunto formado apenas pelas observações restritas à marcação linear interna. Portanto, várias hipóteses individuais dão a mesma acurácia para base de treinamento. Utilizando comitês com bons classificadores, o algoritmo pode realizar uma média das diferentes respostas obtidas pelos classificadores, reduzindo, dessa forma, o risco de escolher hipóteses errôneas. O segundo motivo é *computacional*. Muitos algoritmos trabalham no refinamento de uma busca limitada, restringindo-se apenas a um ótimo local. Em casos onde exista um conjunto de dados de treinamento de tamanho e qualidade suficientes e considerando o problema estatístico ausente, pode ser muito difícil, em termos computacionais, que o algoritmo de aprendizado encontre a melhor hipótese. Em redes neurais artificiais, isso ocorre principalmente pelo fato do treinamento ser *NP-Difícil* (BLUM; RIVEST, 1988). Na Figura 1 (b), conforme a linha tracejada, mostra-se a dificuldade e os diversos caminhos para um algoritmo de classificação encontrar a melhor hipótese em H. Com um comitê, a busca é iniciada em diferentes pontos, permitindo uma melhor aproximação da função objetivo não conhecida. E por fim, há o motivo *representacional*. Em muitos algoritmos de aprendizado, uma função f verdadeira não pode ser representada fora do conjunto de hipóteses em H. Na Figura 1 (c), mostra-se a dificuldade de generalização além dos dados treinados. Comitês prometem evitar que a representação fique restrita ao conjunto finito de hipóteses.

Outras situações em que a utilização de comitês de classificadores são justificáveis incluem o melhor entendimento por parte do projetista do sistema, visto que se utilizam comumente algoritmos de aprendizado mais simples ou padrão como aprensentados na literatura, e a questão de tolerância a falhas, porque módulos corrompidos podem ser substituídos ou eliminados sem danos maiores para o sistema como um todo (COELHO, 2004).

Mesmo apresentando essas justificativas, para se obter os melhores resultados da utilização de *ensembles*, faz-se necessário que todos os algoritmos de classificação tenham perícia na resolução de um tipo de problema específico, e que cada um deles se mostre diverso em relação aos demais, pois desta forma, surgirá algum tipo de contribuição para a composição da decisão final.

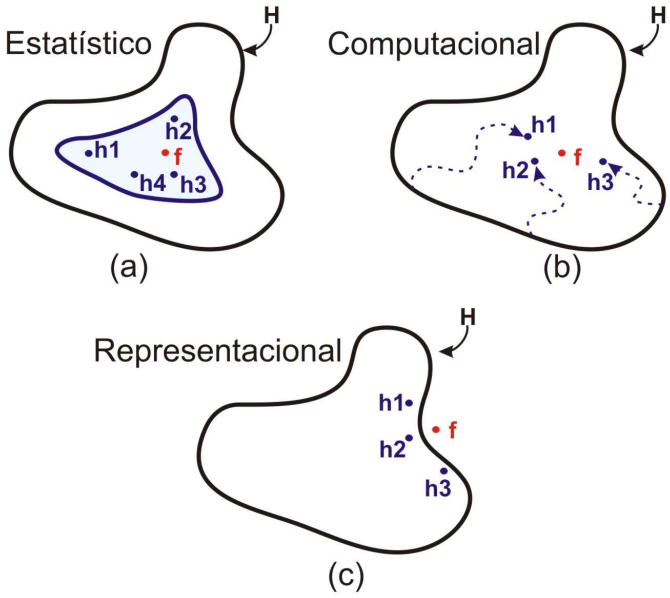

Figura 1: Três razões pelas quais *ensembles* têm um melhor desempenho que um preditor único (Adaptada de Dietterich (2000a)).

Neste caso a diversidade se refere à variação do erro apresentado por cada componente de forma a tratar as novas instâncias de entrada.

Embora já possamos encontrar várias pesquisas nesta linha, existem certas divergências quanto à taxonomia a ser adotada para o tema. No trabalho de Haykin (2001), o autor sugere uma subdivisão das abordagens de Comitês de Classificadores em relação ao tipo de estrutura:

1. Estruturas estáticas – As respostas dos diversos estimadores são combinadas por meio de um mecanismo que não é influenciado pelo novo problema de classificação a ser apresentado como entrada do sistema, de forma que não é realizada uma escolha de quais algoritmos serão recrutados para resolver o novo problema. Neste grupo são encontrados os métodos de *ensembles*, como, por exemplo, *Bagging*, *Boosting* e *MultiBoosting* (KHOSHGOFTAAR; HULSE; NAPOLITANO, 2011; JAIN; KULKARNI, 2012; SU *et al.*, 2011); e

2. Estruturas dinâmicas – O novo problema de classificação está adiante envolvido na atuação do mecanismo que integra as saídas dos especialistas individuais em uma saída global. Em outras palavras, precisa-se determinar quais especialistas são melhores para tratar o novo

problemas. Neste grupo é encontrada a arquitetura de mistura de especialistas (ME) (YUKSEL; WILSON; GADER, 2012).

Outros autores como Kuncheva (2000), Valentini e Masulli (2002) consideram os vários modelos apenas como variações do conceito de *ensemble*.

Estudos em *Ensemble* vêm sendo aplicados com sucesso na resolução de problemas em áreas de pesquisa distintas, como por exemplo: reconhecimento de caracteres (SRIMANY *et al.*, 2014; NABIHA; NADIR, 2012; MAO, 1998; HANSEN; LIISBERG; SALAMON, 1992), análise de imagens (CHEN; ZHAO; LIN, 2014; KRUPKA *et al.*, 2014; CHERKAUER, 1996), diagnóstico médico (MCLEOD; VERMA; ZHANG, 2014; PAPADANIIL; HADJILEONTIADIS, 2014; ZHOU *et al.*, 2000), regressão de função (HASHEM; SCHMEISER, 1995; LIMA; COELHO; ZUBEN, 2002) e predição de séries temporais (STEPNICKA; STEPNICKOVA; BURDA, 2014; INOUE; NARIHISA, 2004).

2.1.1 *Ensembles* de Classificadores

Ensemble é um paradigma de aprendizado em que um grupo finito de propostas alternativas para a solução de um dado problema, denominados componentes do *ensemble*, é empregado em conjunto na proposição de uma única solução para o problema (SOLLICH; KROGH, 1996). A Figura 2 ilustra uma arquitetura geral para um modelo de *ensemble*. A partir dessa estrutura, cada novo padrão de entrada é tratado de forma redundante por diferentes classificadores (redes neurais, árvores de decisão, etc), os quais produzem individualmente suas respostas. Esses resultados, então, são coletados e passados para um módulo de combinação, para dar origem à decisão consensual final. Para se combinar as respostas individuais, é comumente utilizado, em problemas de classificação de padrões, o *voto majoritário* ou *ponderado*; já em problemas de regressão, a *média simples*.

Sempre que se menciona o termo *ensemble*, a questão da *diversidade* passa a ser obrigatoriamente relevante (WANG; YAO, 2013; KUNCHEVA; WHITAKER, 2003). Isso se deve, principalmente, ao fato da *diversidade* ser um fator primordial para que o modelo resultante do *ensemble* proporcione ganho de generalização. Para isso, cada componente tem que apresentar um bom desempenho quando aplicado isoladamente. Mas por outro lado, um componente

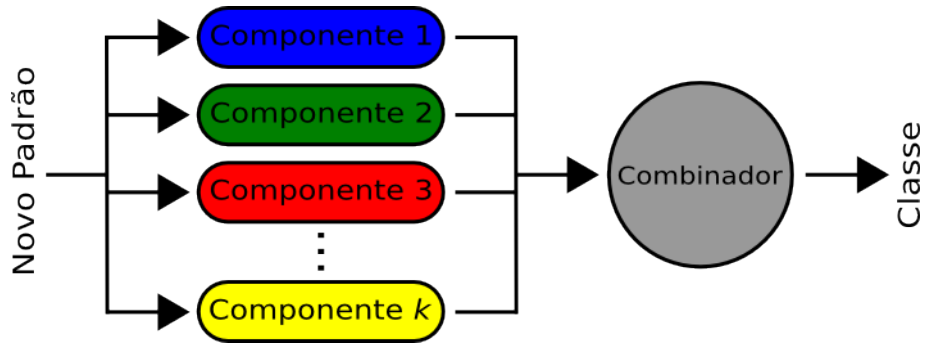

Figura 2: Arquitetura geral para um modelo de *ensemble* (NASCIMENTO, 2009).

deve apresentar dissimilaridade do erro entre as demais em regiões distintas do problema, de forma que a diversidade das respostas possam contribuir ou influenciar na melhor hipótese. É notável, que se todos os componentes forem treinados de forma semelhante, todos por sua vez, apresentarão também o mesmo comportamento, acarretando apenas, em um aumento no custo computacional, sem resultados práticos de incremento de desempenho (NASCIMENTO, 2009).

Existem vários métodos propostos para se obter a diversidade em *ensembles*. Na literatura as mais citadas são (NASCIMENTO; COELHO; CANUTO, 2014; BROWN; KUNCHEVA, 2010; WANG; YAO, 2009; KUNCHEVA; WHITAKER, 2003):

- Métodos que manipulam os dados de treinamento, ou seja, variam a apresentação dos dados via estratégias de reamostragem;
- Métodos que atuam sobre o ponto de partida no espaço de hipóteses, ou seja, variam o ponto de partida da busca no espaço de hipóteses; e
- Métodos que manipulam a arquitetura dos componentes, ou seja, variam a arquitetura de cada componente de forma que diferentes conjuntos de hipóteses estejam acessíveis a cada um deles;
  - *Ensembles* heterogêneos: cada componente é construído a partir de diferentes algoritmos de aprendizagem (CANUTO *et al.*, 2005, 2007).

Diferentes métricas de avaliação do grau de diversidade em *ensembles* já foram propostas, podendo estas pertencerem a duas classes distintas (KUNCHEVA; WHITAKER, 2003): Métricas baseadas em medidas par-a-par, que consistem em tomar a média de uma dada métrica de

distância calculada sobre partes de componentes do *ensemble*, e métricas baseadas em medidas sobre todo o grupo, que se baseiam em entropia ou na correlação de cada componente com a saída média do grupo.

Entre as métricas par-a-par, encontra-se a Estatística-Q. Para sua aplicação, a saída de cada classificador k é representada por um vetor binário n-dimensional $V = \{v_{k1}, v_{k2}, ..., v_{kn}\}$, sendo n correspondente ao tamanho do conjunto de exemplos D. Se a instância D_i for classificada corretamente, então $V_{ki} = 1$, caso contrário, $V_{ki} = 0$. O grau de divergência entre dois classificadores k e k' pode ser mensurado mediante a Equação (2.1). O valor de N^{11} corresponde ao número de amostras classificadas corretamente por k e k'. N^{00} é o número de amostras classificadas incorretamente por k e k'. N^{10} é o número de amostras classificadas corretamente por k e incorretamente classificadas pelo classificador k'. Já N^{01} corresponde ao número de amostras classificadas incorretamente por k e corretamente classificadas pelo classificador k'. A Estatística-Q assumirá valores entre $[-1, 1]$.

$$Q_{k,k'} = \frac{N^{11}N^{00} - N^{01}N^{10}}{N^{11}N^{00} + N^{01}N^{10}} \tag{2.1}$$

Já a diversidade do *ensemble* é dada pela média sobre todos os possíveis acoplamentos entre pares de componentes (COELHO, 2004), sendo calculado como na Equação (2.2).

$$Q(ensemble) = \frac{2}{K(K-1)} \sum_{k=1}^{K-1} \sum_{k'=k+1}^{K} Q(k,k') \tag{2.2}$$

Entre as métricas não par-a-par, a mais recente delas toma como base a decomposição do erro do *ensemble* mediante voto majoritário em três termos: acurácia individual, e outras duas que se referem a *boa* e *má* diversidades (BROWN; KUNCHEVA, 2010). Esse termo tem sido aplicado em *ensembles* baseados no número de votos no momento em que uma decisão é construída. As métricas de *Boa* e *Má* diversidades são apresentadas formalmente nas Equações (2.3) e (2.4), respectivamente.

$$BD = \frac{1}{k} \sum_{i \in P_c} w_i \tag{2.3}$$

$$MD = \frac{1}{k} \sum_{i \in P_w} v_i \tag{2.4}$$

Nessas equações, k representa o número de classificadores de um *ensemble*, v_i representa o número de votos corretos e w_i o número de votos incorretos. Finalmente, P_c e P_w representam o números de padrões que são classificados corretamente e incorretamente, respectivamente, pelo *ensemble*. Em outras palavras, podemos dizer que a *boa* diversidade é medida pelo número de votos incorretos dos componentes quando o *ensemble* acerta, enquanto que a má diversidade é medida pelo número de votos corretos dos componentes quanto o *ensemble* erra. Logo, valores elevados obtidos pela *boa* diversidade indicam redução de erro do *ensemble*, em contrapartida, valores elevados de *má* diversidade indicam um aumento na taxa de erro do *ensemble*.

2.1.2 Construção dos Componentes

A construção dos componentes do *ensemble* preocupa-se em como todos os preditores serão construídos de forma a manterem um nível significativo de diversidade entre eles. Dentre os vários métodos publicados e consagrados na literatura, os algoritmos *Bagging* e *Boosting*, propostos por Breiman (1996a) e Freund e Schapire (1996), respectivamente, são os mais mencionados. Também podemos encontrar pesquisas recentes que utilizam como objeto de estudo os algoritmos *MultiBoosting* proposto por Webb (2000) e *Stacking* proposto por Wolpert (1992), estes por sua vez, são variações dos algoritmos de *ensemble* propostos originalmente. De forma geral, todos eles têm apresentado sucesso na melhoria da acurácia de determinados classificadores quando aplicados a conjuntos de testes reais e artificiais.

O algoritmo *Bagging* é baseado na ideia de *Bootstrap Aggregating* (EFRON; TIBSHIRANI, 1993). Ele provê a diversidade, lançando-se mão do conceito de redistribuição aleatória dos dados. Ou seja, para um conjunto de dados de treinamento D de tamanho n, é gerado, por reamostragem uniforme, um sub-conjunto de dados D' de tamanho n. Desse modo, uma instância do conjunto de dados de treinamento tem uma probabilidade de $1-(1-1/n)^n$ de ser selecionada, permitindo assim que 63.2% do novo conjunto de dados seja composto de amostras únicas e o restante de valores duplicados (BAUER; KOHAVI, 1998). Não há praticamente nenhuma chance dos conjuntos de dados gerados serem idênticos.

Mesmo garantindo a diversidade através da reamostragem, Breiman (1996a) ainda sugere para o melhor desempenho do *ensemble* via algoritmo *Bagging*, a utilização de estimadores instáveis, como por exemplo: modelos neurais, árvores de decisão, árvores de regressão, e modelos de regressão linear (HASTIE; TIBSHIRANI; FRIEDMAN, 2001). Esses algoritmos são considerados instáveis pelo fato de serem sensíveis a pequenas mudanças nos dados de treinamento. A Figura 3 ilustra graficamente a execução do processo de criação de um *ensemble* homogêneo via algoritmo *Bagging*.

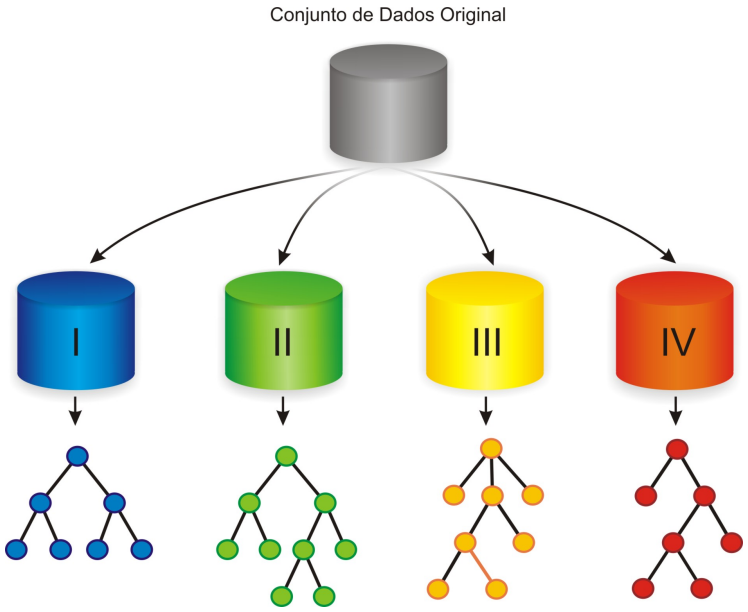

Figura 3: Processo de criação de um *ensemble* via algoritmo *Bagging* (NASCIMENTO, 2009).

Bagging tem se mostrado estável quando aplicado em conjunto de dados ruidosos, e também tem se mostrado atrativo no que se refere à questão de diminuição da variância (COELHO, 2004; KOTSIANTIS; PINTELAS, 2004). Um outro ponto a receber atenção com a utilização desta técnica é que a construção dos componentes pode ser realizada paralelamente, podendo ser beneficiada pelos recursos provenientes da computação distribuída.

O algoritmo *Boosting*, concebido por Schapire (1990) e também referenciado por Breiman (1998) como *Arcing*[1], é semelhante ao *Bagging* no quesito reamostragem dos dados. Porém, apresenta um diferencial sutil. Ele não utiliza um conjunto de dados de treinamento via reamostragem aleatória uniforme. Isso caracteriza, portanto, que a distribuição de probabilidade

[1]Derivado de *adaptive resampling and combining*.

associada a cada um dos novos conjuntos passa a ser ajustada adaptativamente e é dita ser uma representação viesada[2] da distribuição original (COELHO, 2004). À medida que casos de treinamento forem preditos incorretamente, os mesmos passam a ter maior probabilidade de serem integrados a um novo conjunto de dados para geração do próximo componente, de modo que os últimos estimadores na sequência, geralmente, deverão arcar com regiões mais complicadas do espaço de atributos (COELHO, 2004).

No algoritmo *AdaBoost.M1* (FREUND; SCHAPIRE, 1996, 1999), para um conjunto de dados de treinamento D de tamanho n, cada instância d recebe um peso inicial $w_i = 1/n$. O primeiro classificador é treinado com todas as amostras. Em seguida, é testado utilizando as mesmas instâncias. Às amostras que forem classificadas erroneamente, é acrescido o valor do seu peso, e para as classificadas corretamente, os pesos são mantidos. O classificador que acertar as instâncias mais difíceis receberá um peso maior. Esse procedimento ocorrerá até se completar K componentes (valor informado previamente), ou quando $\varepsilon_k \leq 0$ ou $\varepsilon_k \geq 0,5$, sendo ε_k o erro agregado. O resultado final é obtido por um voto ponderado de todos os componentes.

A Figura 4 ilustra graficamente a execução do processo de criação de um *ensemble* homogêneo via algoritmo *Boosting*.

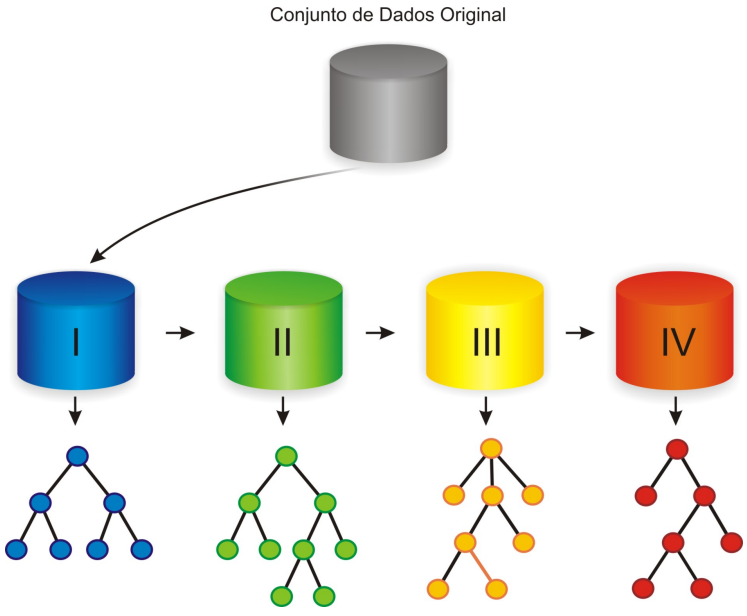

Figura 4: Processo de criação de um *ensemble* via algoritmo *Boosting* (NASCIMENTO, 2009).

[2]Citados em alguns artigos como *Overrepresentation*.

O algoritmo *Boosting*, assim como *Bagging*, apresenta uma redução significativa em termos da variância; porém, o seu desempenho quando aplicado a dados ruidosos geralmente não é bom (KHOSHGOFTAAR; HULSE; NAPOLITANO, 2011; KOTSIANTIS, 2014; DIETTERICH, 2000b). Isso já era esperado, visto que, como os estimadores individuais são treinados hierarquicamente (COELHO, 2004) ao passo que são construídos, os mesmos vão se especializando em áreas do espaço de atributos que provavelmente dispõem de valores corrompidos, levando à perda de acurácia. Por outro lado, *Boosting* geralmente apresenta, quando o problema de ruído é ausente, ganhos significativos em termos de controle do *bias* (COELHO, 2004; KOTSIANTIS; PINTELAS, 2004).

Contrastando as vantagens e limitações de cada algoritmo mencionado, um dos pontos positivos mais citados para *Bagging* é a redução da variância, ao passo que para *Boosting* é a redução tanto da variância como do *bias*. Já um ponto negativo bastante questionado para *Boosting* é o fato dele se mostrar sensível a conjuntos de dados ruidosos. Na tentativa de se combinar as vantagens e ao mesmos tempo tentar minimizar as limitações desses métodos, Webb (1998) propôs a ideia de se combinar *Bagging* e *Boosting* para a formação de subcomitês, cuja denominação atribuída foi a de Múltiplos *Boosting* (ZHENG; WEBB, 1998).

O algoritmo *MultiBoostAB* (WEBB, 2000) é uma técnica que combina *AdaBoost* e *Wagging*. *Wagging* é uma variação do *Bagging* que utiliza instâncias de treinamento d com diferentes pesos gerados a partir de um valor aleatório, conforme a Equação (2.5):

$$Poisson(d) = -log(\frac{Random(1...999)}{1000}) \tag{2.5}$$

Todo o processo de construção e combinação dos votos é semelhante ao *AdaBoost*; apenas os pesos das instâncias é que são calculados aleatoriamente inicialmente para cada subcomitê.

A Figura 5 ilustra graficamente a execução do processo de criação de um *ensemble* homogêneo via algoritmo *MultiBoosting*.

MultiBoosting tende a apresentar maior acurácia em relação ao *Bagging*, pelo fato de prover o incremento da diversidade e ainda ampliar a independência dos membros do comitê (JAIN; KULKARNI, 2012). Ao mesmo tempo, mostra-se mais estável que *Boosting*; isso porque *MultiBoosting* quebra o processo de criação dos componentes em processos menores, no caso, em

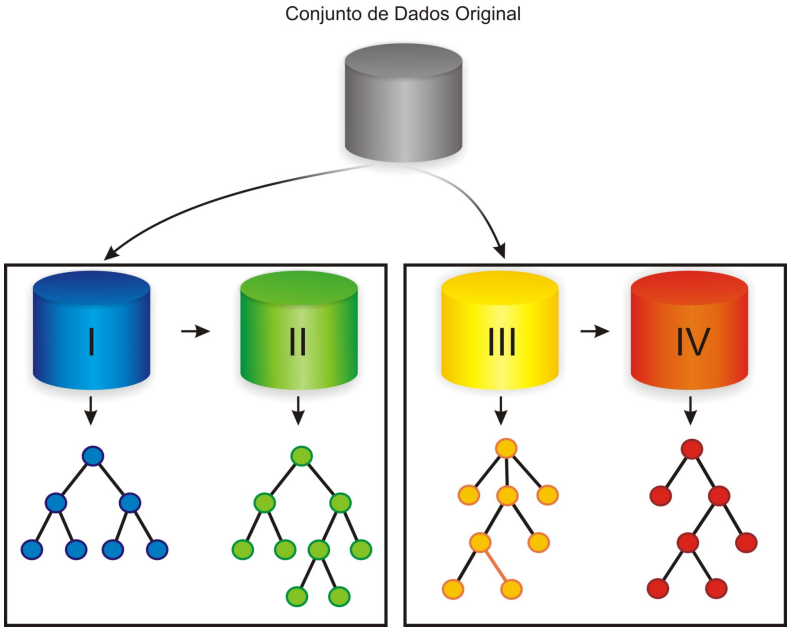

Figura 5: Processo de criação de um *ensemble* via algoritmo *MultiBoosting* (NASCIMENTO, 2009).

subcomitês (ZHENG; WEBB, 1998). Assim como *Bagging*, *MultiBoosting* é propício de ser utilizado de forma distribuída.

2.2 Classificação Multirrótulo

Os algoritmos de Aprendizado de Máquina tradicionais associam uma única classe a cada exemplo (problemas unirrótulo). Em outras palavras, um classificador é treinado em um conjunto de exemplos D, em que cada exemplo d_n é associado a uma única classe y_n de um conjunto Y de classes disjuntas, tal que $|Y| \geq 2$. Por outro lado, existe um grupo de problemas reais de classificação, conhecidos como problemas de classificação multirrótulo, que nesse caso, os exemplos estão associados a um conjunto de classes L, tal que $L \subseteq Y$, ou seja, cada exemplo pode ser associado a mais de uma classe simultaneamente, de forma que as classes não são disjuntas (FACELI *et al.*, 2011; SANTOS, 2012).

Na literatura, podem ser encontrados diferentes métodos para tratar problemas de classificação multirrótulo (ZHANG; ZHOU, 2014). Em algumas desses métodos, problemas multirrótulo são quebrados em vários problemas unirrótulo. Assim, classificadores unirrótulo podem ser com-

binados para viabilizar o tratamento de problemas de classificação multirrótulo. Por outro lado, há métodos que resultam da modificação de algoritmos de classificação unirrótulo, de modo que, através da adaptação de seus mecanismos internos, torna-se possível a sua utilização em problemas de classificação multirrótulo. Adicionalmente, novos mecanismos podem ser desenvolvidos para tratar especificamente problemas de classificação multirrótulo (TSOUMAKAS; KATAKIS; VLAHAVAS, 2010a).

Há duas importantes abordagens a serem utilizadas para tratar um problema de classificação multirrótulo: abordagem independente do algoritmo e abordagem dependente do algoritmo. Na primeira abordagem, problemas de classificação multirrótulo são tratados utilizando qualquer algoritmo de classificação tradicional. Para isso, basta que o problema multirrótulo original seja transformado em um conjunto de problemas de classificação unirrótulo. Já na segunda abordagem, novos algoritmos são propostos para tratar problemas de classificação multirrótulo como um todo, em uma única etapa. Tais algoritmos podem ser desenvolvidos especificamente para classificação multirrótulo ou serem baseados em técnicas de classificação convencionais, como máquinas de vetores-suporte ou árvores de decisão (FACELI *et al.*, 2011).

"A decisão entre aplicar ou não a transformação sobre os dados ainda é alvo de discussão, pois alguns pesquisadores acreditam que, ao transformar os dados multirrótulo, estariam sendo perdidas informações relevantes de correlação entre os rótulos. A vantagem dos classificadores adaptados para o caso multirrótulo estaria justamente na possibilidade de capturar tais correlações para realizar com mais precisão as predições. Entretanto, ainda não há uma conclusão consensual sobre qual seria a melhor abordagem a ser adotada para todos os casos" (COSTA, 2012; COSTA; COELHO, 2011).

2.2.1 Abordagem Independente de Algoritmo

Duas soluções estão disponíveis na literatura para a abordagem independente de algoritmo, que são: *Label Powerset* (KATAKIS; TSOUMAKAS; VLAHAVAS, 2008) e *Random k-Labelsets* (RA*k*EL) (TSOUMAKAS; KATAKIS; VLAHAVAS, 2010b).

No *Label Powerset* (LP), cada subconjunto diferente de rótulos de L é considerado como uma única classe da nova tarefa de classificação unirrótulo. Desse modo, um classificador unirrótulo $C : X \rightarrow P(L)$ é treinado, onde $P(L)$ é o *Powerset* de L, contendo todos os subconjuntos de

rótulos possíveis. Assim, dada uma nova instância, o classificador unirrótulo *C* retorna como saída a classe mais provável, que neste caso é um conjunto de rótulos.

Uma das vantagens do LP é que as correlações entre os rótulos são consideradas. Contudo, é suscetível ao fato de, no caso de haver um número muito grande de subconjuntos de rótulos, o número de rótulos de uma classe pode crescer exponencialmente, resultando em muitas classes com poucos exemplos associados, aumentando o custo computacional do LP e diminuindo a acurácia dos classificadores. Além disso, o LP só pode prever confiavelmente conjuntos de rótulos (*labelsets*) observados no conjunto de treinamento. Esta é uma importante limitação, uma vez que novos *labelsets* tipicamente aparecerão em instâncias de teste (SANTOS, 2012).

No trabalho de Tsoumakas, Katakis e Vlahavas (2010b) é apresentada uma solução para a limitação do LP. Desta forma, são consideradas as correlações entre rótulos, de forma a evitar o problema de suscetibilidade à ocorrência de muitas classes com poucos exemplos do LP. A solução apresentada é chamada de RA*k*EL (derivado do inglês, *Random k-labelsets*).

No RA*k*EL é construído um comitê de classificadores LP, onde cada classificador é treinado, usando um diferente subconjunto aleatório de *labelsets*. Assim, pode-se afirmar que no RA*k*EL, os classificadores unirrótulo, além de considerar as correlações entre rótulos, são aplicados em subtarefas com um número gerenciável de rótulos e número adequado de exemplos por rótulo (NASIERDING; TSOUMAKAS; KOUZANI, 2009; TSOUMAKAS; KATAKIS; VLAHAVAS, 2010a).

2.2.2 Abordagem Dependente de Algoritmo

Podemos também citar duas soluções para a adaptação do algoritmos disponíveis na literatura, que são: Árvore de Decisão (CLARE; KING, 2001) e *k* vizinhos mais próximos (ZHANG; ZHOU, 2005).

O algoritmo de indução a árvore de decisão mais conhecido é o C 4.5, e, no trabalho de Clare e King (2001), esse algoritmo foi adaptado de modo a viabilizar a manipulação de dados multirrótulo. No algoritmo original, os nós da árvore de decisão são definidos através de uma medida de entropia. Desta forma, uma adaptação no cálculo de entropia foi sugerida de forma a possibilitar seu uso em problemas multirrótulo conforme demonstrada na Equação (2.6).

$$Entropia(D'') = -\sum P(\lambda_j) * logP(\lambda_j) + q(\lambda_j) * logq(\lambda_j) \quad (2.6)$$

Onde D'' representa o conjunto de exemplos multirrótulo, $P(\lambda_j)$ representa a frequência relativa da classe λ_j e q(λ_j) é $1 - P(\lambda_j)$. Essa adaptação permite a utilização de nós-folhas da árvore para representar conjuntos de rótulos. Quando um nó-folha, alcançado na classificação de um exemplo, contém um conjunto de classes, uma regra é produzida para cada classe.

O algoritmo ML-*k*NN (ZHANG; ZHOU, 2005)(derivado do inglês *Multilabel k Nearest Neighbors*), é uma adaptação do algoritmo *k*-NN para dados multirrótulo. Nessa adaptação, utiliza-se o princípio do máximo a *posteriori* para determinar o conjunto de rótulos da instância de teste, baseado em probabilidades a *priori* e a *posteriori* para a frequência de cada rótulo nos vizinhos mais próximos. Em essência, o ML-*k*NN usa o algoritmo *k*-NN independente para cada rótulo λ. Desse modo, o ML-*k*NN busca os vizinhos mais próximos para a instância de teste, considerando as instâncias que são rotuladas ao menos com λ como positivas e as demais instâncias como negativas. Assim, o que diferencia este método do *k*-NN original é justamente o uso de probabilidades a *priori*. Adicionalmente, o ML-*k*NN tem a capacidade de produzir um *ranking* dos rótulos como saída.

2.2.3 Medidas de Avaliação

Diferentemente da classificação unirrótulo, em que um exemplo é classificado de maneira certo ou errado, na classificação multirrótulo, um exemplo pode ser classificado de maneira parcialmente certo ou parcialmente errado. Esses casos acontecem quando um classificador atribui corretamente a um exemplo pelo menos uma das classes a que ele pertence, mas também não atribui ao exemplo uma ou mais classes às quais ele pertence. Pode acontecer também de o classificador atribuir a um exemplo uma ou mais classes às quais ele não pertence (FACELI *et al.*, 2011). Para a discussão que segue, considere n o número de instâncias e L o conjunto de rótulos. Dada uma instância d_i, Y_i denota o conjunto verdadeiro de rótulos e Z_i denota o conjunto de rótulos preditos por um certo classificador multirrótulo.

Hamming-loss (HL) (ZHANG; ZHOU, 2014) é uma medida de avaliação multirrótulo baseada em exemplo. A avaliação é feita com base nas diferenças médias entre os conjuntos de

rótulos verdadeiros e os preditos sobre todos os exemplos do conjunto de dados de avaliação. Essa medida varia de zero a um, e é dada pela Equação (2.7) (COSTA, 2012). Sendo $\oplus$ o valor de operação XOR da lógica *booleana*.

$$HL = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i \oplus Z_i|}{|L|} \tag{2.7}$$

F-Measure (FM) (ZHANG; ZHOU, 2014) é outra medida multirrótulo baseda em exemplos, e varia de zero a um (valor ótimo), e representa uma média harmônica das medidas precisão e revocação. A primeira indica a fração de rótulos corretos no conjunto predito, segundo a Equação (2.8), enquanto que a segunda representa a fração de rótulos corretos que foram preditos, segundo a Equação (2.9). A medida FM pode ser então obtida para todo um conjunto de instâncias através da Equação (2.10). Sendo $\cdot$ o valor de operação AND da lógica *booleana*.

$$p = \frac{|Y_i \cdot Z_i|}{|Z_i|} \tag{2.8}$$

$$r = \frac{|Y_i \cdot Z_i|}{|Y_i|} \tag{2.9}$$

$$FM = \frac{2 \times (p \times r)}{p + r} = \frac{1}{n}\sum_{i=1}^{n}\frac{2 \times |Y_i \cdot Z_i|}{|Z_i| + |Y_i|} \tag{2.10}$$

Macro-*averaged F-Measure* (MAFM) (ZHANG; ZHOU, 2014) é outra medida de avaliação multirrótulo, mas nesse caso baseada em rótulo. A ideia é decompor o processo em avaliações separadas para cada rótulo e realizar subsequentemente uma média dessas avaliações. Essa medida constitui uma média de valores da medida FM aplicada para cada rótulo. Considerando os valores para verdadeiro positivo (VP), falso positivo (FP), verdadeiro negativo (VN) e falso negativo (FN), a medida MAFM tem o valor calculado pela Equação (2.11).

$$MAFM = \frac{1}{|L|}\sum_{\lambda=1}^{|L|} FM(VP\lambda, FP\lambda, VN\lambda, FN\lambda) \tag{2.11}$$

Micro-*averaged F-Measure* (MIFM) (ZHANG; ZHOU, 2014) também é uma medida base-

ada em rótulo. Corresponde a um valor de FM obtido dos valores de somatório de cada VP, FP, VN e FN como apresentado na Equação (2.12).

$$MIFM = FM(\sum_{\lambda=1}^{|L|} VP\lambda, \sum_{\lambda=1}^{|L|} FP\lambda, \sum_{\lambda=1}^{|L|} VN\lambda, \sum_{\lambda=1}^{|L|} FN\lambda) \tag{2.12}$$

Ranking-loss (RL) (ZHANG; ZHOU, 2014) é uma medida baseada em *ranking*. Nessa métrica é realizado um ranqueamento dos rótulos, sendo que os rótulos mais relevantes recebem o mais alto posto, ao passo que os menos relevantes recebem os postos mais baixo (COSTA, 2012). Em outras palavras, expressa o número de vezes que rótulos irrelevantes obtiveram um melhor posicionamento no ranqueamento do que rótulos relevantes. Tal medida varia de zero (valor ótimo) a um, e é descrita pela Equação (2.13).

$$RL = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{|Y_i||\overline{Y_i}|}|(\lambda_a, \lambda_b) : r_i(\lambda_a) > r_i(\lambda_b), (\lambda_a, \lambda_b) \in Y_i \times \overline{Y_i}| \tag{2.13}$$

Onde $\overline{Y_i}$ é o conjunto complementar de Y_i em relação a L e $r_i(\lambda)$ é o posto no *ranking* assumido pelo rótulo λ.

2.3 Meta-aprendizado

O objetivo de Meta-aprendizado é associar o desempenho dos algoritmos de aprendizagem com características dos problemas em que são aplicados (GUERRA; PRUDêNCIO; LUDERMIR, 2007), ou seja, está relacionado ao aprendizado através da experiência, quando diferentes algoritmos de aprendizado de máquina são apropriados para um problema particular (SOARES, 2004). Neste contexto, Meta-aprendizado difere da forma de aprendizagem exibida anteriormente no Capítulo 2, quanto ao escopo do nível de adaptação (VILALTA; DRISSI, 2002). Enquanto a aprendizagem de máquina tradicional trabalha sobre um conjunto de dados por vez, a aprendizagem no meta-nível é baseada no acúmulo de experiência do desempenho de múltiplas aplicações de um algoritmo de aprendizado (SOUZA, 2010). Essa abordagem, no entanto, surge com uma tentativa de explorar o comportamento dos algoritmos de aprendizado quando aplicados a diferentes problemas de classificação de padrões, na tentativa de melhorar dinamicamente

a eficiência desses algoritmos (VILALTA; DRISSI, 2002).

Embora ainda não exista um consenso final quanto à utilização dos termos utilizados para área de Meta-aprendizado, segundo Vilalta e Drissi (2002) e citado em Soares (2004), o termo meta-aprendizado pode ser organizado em quatro categorias de acordo com diferentes objetivos. A primeira categoria refere-se à adaptação continua de modelos, ou seja, uma abordagem que inclui a adaptação contínua do *bias*, não apenas entre os problemas, mais sim duranta a construção da solução. Essa é a abordagem que mais tem sido utilizada, e também vem sendo referenciada como *Transferência de Conhecimento* ou *Aprender a Aprender*. A segunda categoria refere-se à construção de modelos, no qual, são combinados partes que representam diferentes *bias* dentro de um modelo coerente. A terceira categoria refere-se à combinação de modelos, ou seja, incluem métodos nos quais objetivam gerar predições baseadas na combinação de um conjunto de modelos alternativos. E por última, a quarta categoria, que se refere à seleção de algoritmos, que incluem métodos nos quais predizem a performance relativa de um conjunto de algoritmos de um dado problema.

Parte desta pesquisa, por sua vez, apresenta uma nova abordagem que auxilie na configuração automática de comitês de classificadores em problemas de classificação de padrões, na qual, explora os benefícios da técnica de seleção de algoritmos e escolha dos parâmetros de controle. Em termos práticos, tal técnica poderia ajudar na criação de sistemas que fornecessem ao usuário sugestões sobre que algoritmos selecionar para cada problema de classificação. Esses sistemas, por sua vez, precisam realizar uma atividade de mapeamento entre as descrições dos conjuntos de dados e as medidas de desempenho dos algoritmos. Para isso, Kalousis (2002) sugere a organização de tais atividade nas seguintes etapas: caracterização das bases de dados, que é responsável por extrair as propriedades dos conjunto de dados; medidas de avaliação, que é responsável por identificar as medidas de desempenho em relação as sugestões fornecidas; forma de sugestões, apresentar qual a forma de sugestão fornecida no sistema; e construção do método de sugestões, que é responsável por mapear as propriedades dos conjuntos de dados com as medidas de avaliação realizadas.

Cada vez mais, um número significativo de propostas vêm sendo apresentadas na tentativa de melhor resolver os problemas de aprendizado de máquina, e mais especificamente, os problemas relacionados à classificação de padrões. Essas novas soluções, em sua maioria, são bastante específicas para um domínio de aplicação e demandam tempo e elevados recursos computacionais.

Então, escolher qual a melhor estratégia a ser adotada em tais tarefas não é trivial, e requer um vasto conhecimento técnico por parte do projetista de sistema. Muitos ainda são os esforços para desenvolver um mecanismo automático de se relacionar o *bias* indutivo de cada algoritmo de aprendizado de máquina à morfologia dos dados em questão. Entretanto, uma formalização para a seleção de algoritmos tem sido apresentada inicialmente por Rice (1975) e citado em Souza (2010) como:

Para uma determinada instância de problema $x \in P$, com características $f(x) \in F$, encontre o mapeamento $S(f(x))$ no espaço de algoritmos A, tal que o algoritmo selecionado $\alpha \in A$ maximize o mapeamento de desempenho $y(\alpha(x)) \in Y$.

Considerando-se o espaço P de problemas, A de algoritmos e Y de desempenhos, a solução trivial para lidar com a situação emprega a interação de aconselhamento especializado com experimentação computacional custosa. Entre outras deficiências, essa abordagem não é capaz de determinar automaticamente o S, impedindo um aproveitamento sistemático de conhecimento acumulado no passado para a resolução de problemas semelhantes no futuro. Pela utilização de Meta-aprendizado, é possível aprender tal mapeamento, propiciando a construção de sistemas de recomendação eficientes e efetivos.

O processo de construção de um sistema de recomendação de algoritmos pode ser descrito em algumas etapas. Na primeira etapa, que representada o espaço P na definição de Rice, é realizada a aquisição de um conjunto apropriado de problemas que sejam representativos daqueles para os quais a recomendação posterior será realizada. Em seguida, duas etapas são aplicadas para cada elemento de P: a avaliação dos algoritmos em A e a extração de características segundo as medidas em F. Idealmente, a caracterização dos problemas deve ser preditiva quanto ao comportamento dos algoritmos. Associando-se essas duas informações para cada problema, obtém-se um meta-exemplo, formado por meta-características de entrada e meta-atributos alvo Y, respectivamente. Ao conjunto dos meta-exemplos, dá-se o nome de meta-dados. Para induzir o mapeamento S entre meta-atributos de entrada e meta-atributo alvo, aplica-se um algoritmo de aprendizado de máquina, também citado por meta-aprendiz. Com isso, o meta-aprendiz utiliza o meta-conhecimento para realizar a recomendação de algoritmos.

De forma geral, tais sistemas podem ser formulados segundo três fases de elaboração (KALOUSIS, 2002):

- Caracterização de bases de dados;
- Medidas de avaliação; e
- Formas de sugestão.

2.3.1 Caracterização de Bases de Dados

Na fase de caracterização de bases de dados, o esforça está em encontrar ou elaborar medidas que descrevam os dados na forma no qual eles estão organizados, objetivando desta forma, fornecer informação morfológicas dos mesmos, para que possam ser aplicados à técnica de Meta-aprendizado. Isso é possível, por temos o conhecimento a priori do comportamento de tais algoritmos quando aplicados a problemas de classificações com diferentes particularidades na natureza de seus dados. Por exemplos, sabe-se que treinar uma rede neural sobre um conjunto de dados com atributos de natureza categórica pode incorrer em perda de precisão. Enquanto isso, treinar árvores de decisão sobre dados de natureza numérica, pode gerar modelos de regras recursivos com elevada repetição de atributos, o que acaba por acarretar problemas de generalização do modelo induzido. Sabe-se também, que os algoritmos baseados em vizinhança não operam satisfatoriamente na presença de atributos irrelevantes, em contrapartida, os algoritmos baseados em máquinas de vetores-suporte, possuem mecanismos internos de ponderação de atributos que os tornam mais robustos a essas limitações. E ainda, temos os modelos de algoritmos de aprendizado baseados em estatística, que não apresentam resultados satisfatórios quando aplicados a conjunto de dados com atributos altamente redundantes.

Em outras palavras, a importância em Meta-aprendizado está em prover diretrizes de como relacionar um algoritmo de aprendizado a diferentes domínios de aplicação, de forma com que se obtenha uma melhor performance (VILALTA; DRISSI, 2002). Em geral o método mais utilizado para avaliar tais algoritmos é a acuidade, mas, segundo Giraud-Carrier, Vilalta e Brazdil (2004), outras medidas podem ser igualmente importante, como por exemplo: complexidade computacional, expressividade, compactação, compreensibilidade, representação do conhecimento, entre outros. Desta forma, pesquisas em caracterização de dados para a construção de meta-características podem ser organizada em três áreas complementares:

- Caracterização direta;

- Caracterização baseada em *landmarking*; e
- Caracterização por propriedades de modelos.

Na caracterização direta, existem três principais tipos de medidas usadas até agora para a construção de meta-atributos que podem ser utilizados para fornecer sugestões quanto à qual algoritmo é mais apropriado para um específico conjunto de treinamento, vide Tabelas 1 e 2. O primeiro tipo, que representa medidas simples, fornece informações gerais das bases de dados; o segundo tipo fornece informações estatísticas que são baseadas nas propriedades dos conjuntos de dados que são usados; e por fim, o terceiro tipo que é baseado na teoria da informação.

Um dos primeiros esforços sistemáticos e em larga escala para tentar relacionar as medidas que caracterizam as bases de dados e o desempenho dos algoritmos foi empreendido no projeto Statlog (MICHIE; SPIEGELHALTER; TAYLOR, 1994). Mais recentemente, o projeto Metal (www.metal-kdd.org) visou o desenvolvimento de ferramentas que auxiliem o usuário a selecionar uma combinação adequada de técnicas de pré-processamento, classificação e regressão. Outras contribuições promissoras para a caracterização direta podem ser encontradas em (KALOUSIS, 2002; SOARES, 2004; SOUZA, 2010).

Tabela 1: Medidas de informação do projeto Statlog

Tipo	Descrição
Simples	Número de padrões. Número de atributos. Número de classes. Número de atributos binários.
Estatística	Razão média entre desvio padrão dos atributos. Correlação média absoluta entre atributos por classe. Primeira correlação canônica. Proporção de variância explicada pelo 1° discriminante canônico. Assimetria média absoluta dos atributos. Curtose média dos atributos.
Informação	Entropia normalizada das classes. Entropia média dos atributos. Informação mútua média entre classe e atributos. Razão sinal/ruído.

Na caracterização baseada em *Landmarking*, são utilizadas informações sobre o desempenho de um conjunto de algoritmos de classificação, denominados *landmarkers*, para os conjuntos de

Tabela 2: Medidas de informação adicionais do projeto Metal

Tipo	Descrição
Simples	Número de atributos nominais. Número de atributos numéricos.
Estatística	Números de atributos com *outliers*. Estatística *M* de *Box*. Graus de liberdade da Estatística *M*. Valor de Lambda de Wilk. Estatística *V* de Barlett.
Informação	Entropia conjunta de classe e atributos.

dados do repositório (PFAHRINGER; BENSUSAN; GIRAUD-CARRIER, 2000). O *landmarking* é utilizado para determinar a proximidade de uma base de dados em relação a outras, através da similaridade de desempenho dos *landmarkers*. Com isso, forma-se uma vizinhança de áreas de competência, onde bases de dados podem ser representadas. Espera-se que bases de dados de natureza semelhante pertençam às mesmas áreas de competência e, por conseguinte, sejam adequadas à aplicação dos mesmos algoritmos de classificação (SOUZA; JúNIOR, 2011).

Diferentes medidas de desempenho podem ser utilizadas como meta-atributos, como por exemplo: precisão, revocação e área sob a curva ROC. Outras abordagens mais sofisticadas também foram investigadas em Brazdil *et al.* (2009).

Na caracterização via modelos, os meta-atributos gerados para um conjunto de dados são representados por propriedades de um ou mais modelos induzidos utilizando tal conjunto. Segundo Maimon e Rokach (2005), há diversas vantagens neste tipo de caracterização, dentre as quais destacam-se:

- a base de dados é resumida em uma estrutura que contém informações sobre a complexidade e desempenho do modelo; e
- a representação dos dados nessa forma pode servir de base para explicar o desempenho do algoritmo de aprendizado.

Quando modelos são utilizados para caracterizar conjunto de dados, ocorre uma mudança no espaço de busca do processo de Meta-aprendizado, que passa do espaço de objetos para o espaço de modelos. Como o novo espaço permite uma busca eficiente no espaço de hipóteses, espera-

se uma sumarização maior e mais eficiente do conjunto de dados originais, levando a melhores meta-atributos (FACELI *et al.*, 2011).

Dentre os algoritmos de classificação mais utilizado para caracterização via modelos, temos a árvore de decisão. Como meta-atributo, diversas medidas de uma árvore de decisão podem ser utilizadas, como por exemplo: o número de nós folhas, formato da árvore, profundidade máxima da árvore e grau de balanceamento da árvore. De acordo com Bensusan, Giraud-Carrier e Kennedy (2000), há evidências empíricas que sugerem importantes conexões entre as propriedades das bases de dados e as estruturas da árvores de decisão não podadas.

2.3.2 Medidas de Avaliação

Para o processo de selecionar qual o algoritmo deve ser aplicado ao conjunto de dados, faz-se necessário que medidas de avaliação sejam empregadas, de forma a tentar garantir que o melhor modelo foi selecionado, levando-se em consideração o desempenho mais satisfatório para o problema em questão. Várias medidas de desempenho podem ser empregadas para avaliação dos algoritmos, como por exemplo: acurácia preditiva, taxa de classificações incorretas, revocação, precisão, medida F, área sob a curva ROC, custo computacional para etapas de treinamento/teste, quantidade de memória necessária, complexidade do modelo induzido e facilidade de interpretação do modelo.

2.3.3 Formas de Sugestão

De acordo com Kalousis (2002), as sugestões podem ser apresentadas de três formas:

- Sugestão do melhor algoritmo;
- Sugestão de um grupo com os melhores algoritmos; e
- Sugestão de um *ranking* dos melhores algoritmos.

A primeira abordagem consiste em fornecer apenas o melhor algoritmo, ou seja, aquele que produza, supostamente, o modelo mais apropriado para uma dada tarefa ou conjunto de dados, segundo algum critério. Uma crítica a essa abordagem refere-se à eventual não disponibilidade

do algoritmo em um dado momento, impossibilitando sua aplicação. Outro aspecto negativo é que não há garantias de que o algoritmo sugerido produzirá, de fato, o resultado mais adequado.

A segunda abordagem é mais flexível, indicando, dentre os algoritmos considerados, aqueles mais promissores para um dado problema. Ou seja, é uma alternativa em tentar fornecer como sugestão o algoritmo com melhor desempenho esperado e também aqueles cujos desempenhos não sejam inferiores ao melhor por uma dada margem.

A terceira abordagem exibe os algoritmos em ordem de preferência com relação ao conjunto de dados. É realizada então uma sugestão na forma de *ranking*. Essa abordagem é interessante pois permite que o sistema de Meta-aprendizado seja desenvolvido sem o conhecimento prévio de quantos algoritmos serão experimentados pelo usuário. Uma desvantagem é a ausência de informações absolutas quanto ao desempenho dos algoritmos, pois apenas a ordem dos mesmo é apresentada.

2.4 Síntese do Capítulo

O objetivo principal deste capítulo é permitir que o leitor se familiarize sobre o tema de Aprendizado de Máquina. Foi discutido o conceito de aprendizado indutivo supervisionado para resolver problemas de classificação de padrão (TAVARES; LOPES; LIMA, 2007).

Na sequência, mencionamos cinco paradigmas de aprendizado de máquina: *simbólico*, *estatístico*, *baseado em exemplos*, *conexionista* e *evolutivo* (MITCHELL, 1997). No paradigma simbólico foram discutidos os algoritmos baseados em árvores de decisão, dentre eles: J48; REP *Tree* e *Decision Stump*; e os algoritmos baseados em regras: OneR, PART e *Decision Table*. Para o paradigma de aprendizado estatístico foi discutido o algoritmo *Naïve Bayes* que é baseado em estatística *bayesiana*. Já para o paradigma baseado em exemplos foi apresentado o algoritmo IB*k* que segue os conceitos de vizinhança e aprendizado local. Para o paradigma conexionista, os métodos aplicados são baseados em algoritmos de função numérica não-linear, dentre eles: redes neurais RBF e máquinas de vetores-suporte. E por fim, o paradigma evolutivo que menciona conceitos sobre Algoritmos Genéticos.

Permitimos ao leitor entender, também, de forma geral os conceitos relacionados à Comitês de Classificadores. Discutimos o tema *ensembles* (HANSEN; SALAMON, 1990), bem como

uma métrica de cálculo de níveis de diversidade dos componentes mediante Estatística-*Q*, *Boa* e *Má* Diversidade. Dos métodos de construção adotados, foram mencionados os algoritmos *Bagging*, *Boosting* e *MultiBoosting*.

Este capítulo permitiu ao leito entender uma nova abordagem que está relacionada ao Aprendizado de Máquina para resolver problemas de classificação de dados multirrótulo. Dentro dessa nova perspectiva, um conjunto de classes possíveis está associada aos padrões a serem treinados. Para isso, duas categorias de soluções são apresentas: aquelas nas quais independem do algoritmo de aprendizado de máquina e aquelas que apresentam adaptações ou ajustes dos algoritmos de aprendizado de máquina para revolverem problemas de classificação multirrótulo. Para aquelas soluções que independem do algoritmo, discutimos o funcionamento das técnicas *Label Powerset* e *Random k Labelsets* com a solução RA*k*EL. E para aquelas soluções dependente, discutimos as técnicas em árvores de decisão e *k* vizinhos mais próximos com o algoritmo ML-*k*NN.

Na sequência, apresentamos as métricas de avaliação de desempenho de um algoritmo de aprendizado multirrótulo, sendo elas baseadas em exemplos, rótulo e *ranking*. Para as métricas baseadas em exemplos fazemos uma discursão das métricas *Hamming-loss* e *F-measure*. Para as métricas baseadas em rótulos fazemos uma discursão Macro-*averaged F-Measure* e Micro-*averaged F-Measure*. E por último, fazemos uma discursão sobre a métrica *Ranking-loss*, que é baseada em *Ranking*.

Este capítulo, também, permitiu ao leitor se familiarizar sobre o tema de Meta-aprendizado. O objetivo de utilizar essa técnica deve-se principalmente a tentativa de explorar o conhecimento obtido através da análise de várias atividades de treinamento dos algoritmos de aprendizado de máquina, na tentativa de escolher dinamicamente o modelo mais adequado para resolver problemas de classificação, com também, propor de forma eficiente as configurações ideais para os parâmetros de controle (VILALTA; DRISSI, 2002).

Foram mencionadas em detalhes as três fases de elaboração de tais sistemas formulados segundo Kalousis (2002): Caracterização de bases de dados, Medidas de avaliação, Formas de sugestão e Métodos de construção de sugestão. Na primeira fase, são discutidas as técnicas de caracterização de dados que consiste em identificar e extrair propriedades que, possivelmente, afetam o desempenho dos algoritmos. Na segunda fase, são discutidas as medidas de avaliação

que são importantes para determinar qual algoritmo utilizar em determinado conjunto de dados. E por fim, a terceira fase, que discute sobre as possíveis formas de sugestões para escolha do melhor algoritmo a ser utilizado pelo usuário.

Todos os algoritmos de aprendizagem aqui estudados serviram de objeto de estudo, e serão mencionados no Capítulo 3, onde discutiremos em detalhes a contribuição da pesquisa.

Capítulo 3

Construção Automática em Comitês

A proposta desta pesquisa está em apresentar três novas abordagens para construção eficientes de *ensembles* de classificadores. A primeira proposta está relacionada a combinação de técnicas de diversidade em um único arcabouço conceitual. A ideia é utilizar as técnicas de seleção de características, reamostragem e componentes heterogêneos em *ensembles* na tentativa de se construir modelos que apresentem melhores níveis de precisão.

A segunda abordagem, apresenta uma nova solução híbrida em Algoritmos Genético para seleção dos espaço de atributos a serem apresentados a diferentes componentes de um *ensemble*. A ideia é combinar as técnicas de avaliação de indivíduos via filtro ou *wrapper* de forma adaptativa durante o processo de evolução de uma população de indivíduos (soluções em potencial).

A terceira proposta está relacionada em apresentar uma nova abordagem para recomendação dos parâmetros de controle em *ensembles* de classificadores via uma metodologia de meta-aprendizado. Esses parâmetros estão relacionados ao tipo de arquitetura escolhida e aos possíveis tipos de componentes base, sejam eles homogêneos ou heterogêneos.

3.1 Trabalhos Relacionados

Com já foi mencionado em capítulos anteriores, nos últimos anos, houve um significativo aumento das pesquisas em Comitês de Classificadores, e evidentemente junto a esse crescimento, notórias contribuições foram apresentadas sempre visando melhorar a diversidade dos componentes e, desta forma, melhorar a eficiência de tais sistemas.

De forma abrangente, as primeiras contribuições se dedicavam à construção de métodos para geração de *ensembles*, tentando sempre, treinar um conjunto de componentes a partir de diferentes visões dos dados de treinamento. Dentro desse cenário, surgiram os algoritmos que são objeto de estudo nesta proposta: *Bagging* (BREIMAN, 1996a), *Boosting* (SCHAPIRE, 1990) e *Multiboosting* (WEBB, 2000). Em um segundo momento, podendo-se assim dizer, começaram a surgir propostas mais elaboradas para geração de *ensembles*, principalmente naquelas que exploravam as técnicas de modelos bio-inspirados.

A maior motivação em seguir a ideia de elaborar novos modelos para construção de *ensembles* se deve ao trabalho de Kuncheva e Whitaker (2003), no qual os autores defendem a tese de que melhores resultados de generalização só serão alcançados mediante uma configuração que apresente bons índices de diversidade dos componentes.

As subseções 3.1.1, 3.1.2 e 3.1.3, a seguir, organizam os trabalhos relacionais de acordo com o objetivo de contribuição, respectivamente, tratam da escolha dos componentes de um comitê, dos métodos de otimização para a seleção de componentes e sobre meta-aprendizado para a seleção de componentes.

3.1.1 Escolha dos Componentes de um Comitê

Alguns trabalhos têm apresentado soluções para escolha, de forma manual, dos componentes heterogêneos em *ensemble* na tentativa de se alcançar melhores desempenhos. No trabalho de Woods, Jr. e Bowyer (1997), os autores utilizaram como componentes redes neurais, classificadores do tipo *k-nearest neighbour*, árvores de decisão e classificadores quadráticos *bayesianos*. Para cada componente, utilizaram a estimativa de sua precisão local no espaço de atributos para escolher qual desses componentes seriam responsáveis pela resposta para uma nova entrada. Em Wang, Jones e Partridge (2000), os autores utilizaram como componentes redes neurais e

árvores de decisão, chegando à conclusão de que os melhores desempenhos são obtidos com um número de redes neurais maior que o de árvores de decisão, mas com pelo menos uma árvore no *ensemble*. Já Tsoumakas, Angelis e Vlahavas (2005) propõem o uso de um modelo estatístico de seleção de componentes em um *ensemble* submetidos à combinação por voto ponderado. Foram utilizados dez tipos de indutores, entre eles tabela de decisão, classificadores baseados em regras (JRip e PART), classificadores baseados em vizinhança (IB*k* e *K**), *Naïve Bayes*, máquinas de vetores-suporte via algoritmo SMO, redes neurais do tipo RBF e MLP.

No trabalho de Soares *et al.* (2006), os autores utilizaram como componentes redes neurais do tipo MLP, redes neurais do tipo RBF, classificadores *Naïve Bayes*, máquinas de vetores-suporte e classificadores baseados em aprendizado de regras proposicionais, e propuseram duas técnicas de seleção de componentes (baseadas em algoritmo de agrupamento e *k-nearest neighbours* respectivamente) que buscam não apenas reduzir o erro do *ensemble*, mas também aumentar a diversidade de seus componentes. Em Canuto *et al.* (2005), foi realizado um estudo em relação aos níveis de desempenhos e de diversidade de modelos híbridos e não-híbridos de *ensemble*. Foram utilizados indutores de redes neurais artificiais (MLP), *Naïve Bayes*, *k-nearest neighbours*, máquinas de vetores-suporte e *Fuzzy* MLP. Os valores de desempenho também são analisados via cinco técnicas de combinação, entre métodos de fusão e métodos de seleção.

Em uma outra pesquisa, Canuto *et al.* (2007) realizaram uma investigação da influência na escolha de tipos de componentes, tamanho e diversidade de modelos de *ensemble* híbridos e não-híbridos. Foram utilizados sete tipos de indutores: redes neurais artificiais do tipo MLP; *Fuzzy* MLP; redes neurais do tipo RBF; máquinas de vetores-suporte; (*k-nearest neighbours*), árvore de decisão J48, e o algoritmo baseado em regras proposicionais JRip. Finalmente, o trabalho de Bian e Wang (2007) avalia níveis de diversidade e acuidade de tipos genéricos de *ensembles* homogêneos e heterogêneos. Foram utilizados dez tipos de indutores: árvores de decisão C4.5 e *Random Forest*, *Naïve Bayes* e *Bayes network*, redes neurais do tipo MLP e RBF, máquinas de vetores-suporte, classificadores baseados em vizinhança IB*k* e classificadores baseados em regras PART e *Decision table*. No trabalho de Canuto *et al.* (2005, 2007), os autores realizaram vários estudos em relação à escolha de tipos de componentes e métodos de combinação para modelos híbridos e não-híbridos, porém, a escolha dos componentes é realizada manualmente.

No geral, todos os trabalhos apresentam novas soluções para composição de componentes heterogêneos em *ensembles*, por outro lado, nenhum faz um estudo mais aprofundado sobre a

combinação de diferentes técnicas de diversidade de forma gradativa em um único arcabouso conceitual, como é apresentado nesta tese.

3.1.2 Métodos de Otimização para a Seleção de Componentes

Outros trabalhos, fazem a investigação de técnicas de otimização para a escolha automática dos componentes heterogêneos em *ensembles*. Como por exemplo, no trabalhos de Langdon, Barrett e Buxton (2002), de uma forma geral, os autores apresentam diferentes métodos para seleção de componentes, diferindo do aqui proposto. Os métodos propostos neste trabalho realizam conjuntamente as etapas de geração e seleção (poda). Langdon, Barrett e Buxton (2002) utilizaram redes neurais e árvores de decisão em seus experimentos, mas aplicaram Programação Genética (KOZA, 1992) para evoluir uma regra de combinação dos indivíduos. No trabalho de Nascimento e Coelho (2009a), os autores fazem uma nova proposta de um algoritmo evolutivo para *Bagging* com componentes heterogêneos. Foram utilizados 10 diferentes tipos de componentes (RBF, J48, SMO, *Naïve Bayes*, IB*k*, REP *Tree*, *Decision Stump*, OneR, PART e *Decision Table*) a serem escolhidos mediante a aplicação de algoritmos genéticos. Em um outro trabalho de Nascimento e Coelho (2009b), os autores apresentam um novo algoritmo evolutivo para *Boosting* com componentes heterogêneos. No trabalho de Nascimento *et al.* (2011), os autores utilizam algoritmos genéticos para evoluir um modelo de *ensemble* que combina componentes heterogêneos e seleção de características em um único modelo conceitual. No trabalho de Nascimento (2009), o autor realiza a comparação de duas técnicas, uma evolutiva ou construtiva, para seleção de componentes heterogêneos em *Bagging*, *Boosting* e *Multiboosting*.

Nesta tese é então apresentada uma nova abordagem que combina as técnicas de filtro e *wrapper* para evoluir diferentes espaços de atributos a serem apresentados aos componentes de um *ensemble*.

3.1.3 Meta-aprendizado para a Seleção de Componentes

Outros trabalhos dizem respeito à utilização de meta-aprendizado para recomendação de algoritmos. No trabalho proposto por (GUERRA; PRUDêNCIO; LUDERMIR, 2007), os autores investigaram essa técnica para predizer o desempenho dos algoritmos *Backpropagation* e

Levenberg-Marquardt usados no treinamento de redes MLP. No trabalho de Gomes *et al.* (2010), os autores utilizaram Meta-aprendizado e técnicas de busca para seleção automática das configurações dos parâmetros de controle de máquina de vetores-suporte. No trabalho de (SOUZA, 2010), investiga a utilização de técnicas de Meta-aprendizado para recomendação de algoritmos a serem aplicados a problemas de expressão gênica. Em (BONISSONE, 2012), por exemplo, é apresentado uma abordagem para criar modelos customizados de *ensembles* sobre demanda, intitulado *Lazy* Meta-*learning*. A ideia geral é criar *ensembles* no qual meta-informação é usada para calibrar dinamicamente o bias e os pesos dos componentes. Já no trabalho de (PARENTE; CANUTO; XAVIER, 2013), os autores combinaram a ideia de meta-aprendizado a *ensembles*, com o objetivo de ajudar na recomendação de números de componentes e a escolha da arquitetura em *ensembles*. A pesquisa aqui apresentada difere dos trabalhos já realizados no sentido de fazer recomendação multirrótulo automática do tipo de arquitetura, bem como, os componentes base (homogêneos e heterogêneos) em *ensembles* de classificadores. No trabalho de (NETO; CANUTO, 2014) os autores apresentaram uma abordagem para recomentação dos componentes heterogêneos de um *ensemble* via meta-aprendizado. A diferença do trabalho aqui apresentado é que os autores não utilizaram um meta-aprendiz multirrótulo para recomendação dos componentes heterogêneos.

Apresentamos nesta tese uma nova abordagem em meta-aprendizado para recomendação dos parâmetros de controle em *ensembles*, uma na qual chamamos de Meta-aprendizado tradicional e um outra que chamamos de Meta-aprendizado Multirrótulo.

3.2 Novas Abordagens para Construção Automática em Comitês

As subseções 3.2.1, 3.2.2 e 3.2.3 descrevem, respectivamente, as novas abordagens propostas para Algoritmo Genético para Incremento de Diversidade, Algoritmo Genético com Função de Avaliação Adaptativa e Recomendação via Meta-aprendizado.

3.2.1 Primeira Abordagem: Algoritmo Genético para Incremento de Diversidade

Recentemente, investigamos novas estratégias para combinação de técnicas de diversidade em *ensemble*. Essa pesquisa tem produzido estruturas heterogêneas para *Bagging* e *Boosting* (COELHO; NASCIMENTO, 2010; NASCIMENTO; COELHO, 2009b). Em ambos os casos, a ideia era combinar duas diferentes técnicas de diversidade, permitindo o recrutamento de algoritmos de aprendizagem distintos para indução de *ensembles* em base de dados reamostrados via *Bagging* (BREIMAN, 1996a) e *Boosting* (SCHAPIRE; FREUND, 2012), respectivamente.

Nesta abordagem, o objetivo é incrementar os níveis de diversidade em *ensemble* em três estágios (NASCIMENTO *et al.*, 2011), conforme exibido na Figura 6. Cada estágio capitaliza da técnica produzida no estágio anterior.

No primeiro estágio, a diversidade é promovida através da seleção de características e a redistribuição através dos componentes de *ensemble*. Seleção de características reduz a dimensionalidade das características de uma base original, mantendo os atributos mais discriminativos e excluindo aqueles que são redundantes (GUYON *et al.*, 2006). Comumente, essa tarefa é realizada por alguma função de otimização (EIBEN; SMITH, 2007), e, para nossa proposta, um algoritmo genético (AG) foi adotado para seleção de um (quase) ótimo subconjunto de características a ser associado a diferentes componentes.

Nesse AG, um cromossomo binário representa um possível subconjunto de características. A representação binária é apropriada em problemas de otimização combinatória (EIBEN; SMITH, 2007), tal como o típico caso de seleção de características investigada nesta tese. De fato, outros trabalhos relacionados tem feito o uso dessa solução em seleção de características no contexto de *ensembles* heterogêneos (SANTANA *et al.*, 2010). Para nossa proposta, o tamanho do cromossomo é $K \times F$, onde K representa o número de componentes e F denota a cardinalidade de características originais da base. Nesse cromossomo, a primeira parte de F *bits* representa o subconjunto de características para ser associado ao classificador k_1, enquanto que o segundo F *bits* corresponde ao subconjunto de características associadas ao classificador k_2, e assim por diante. A Figura 7 ilustra como um cromossomo individual é particularmente codificado para a base "credit-a", com 15 características, conforme apresentados no Apêndice A, logo, neste caso, $|F| = 15$ e $|k| = 12$. Nesse exemplo, as características #2, 3, 6, 9, 12, 14 e 15 são associadas para

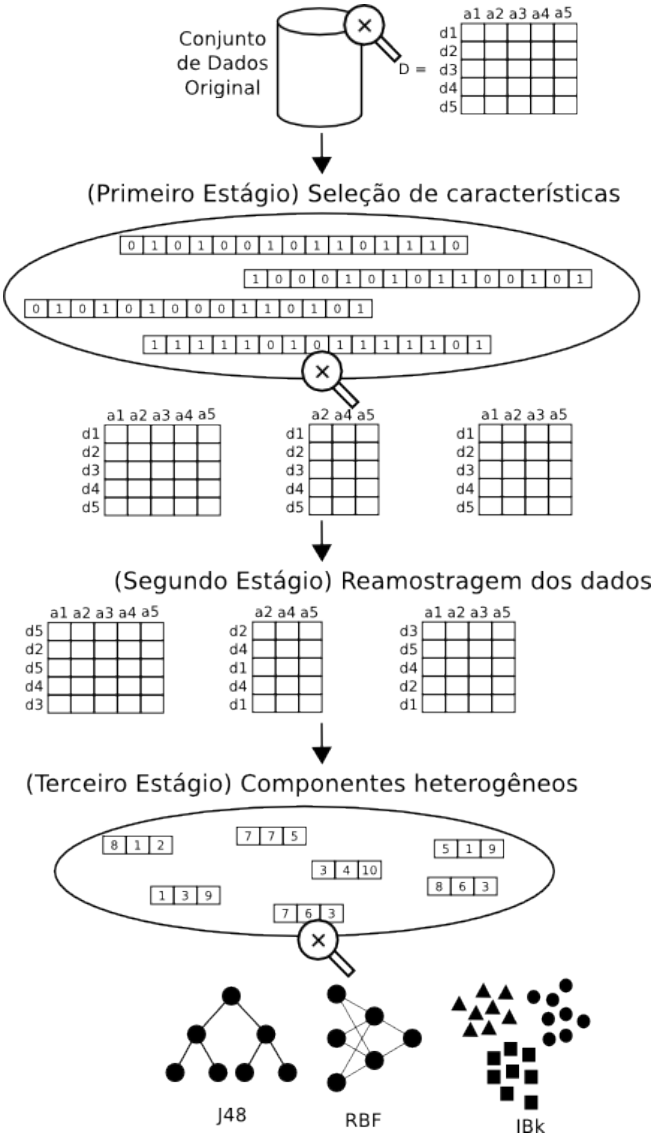

Figura 6: Estágios adotados para incremento dos níveis de diversidade em *ensemble*.

o primeiro classificador k_1, enquanto que apenas as características #4, 6 e 15 são associadas ao classificador k_{12}.

Durante a primeira fase, uma vez que os componentes que irão compor o conjunto final não tenham sido selecionados, no entanto, um método baseado em filtro de seleção de característica é empregado (GUYON *et al.*, 2006), configurado com um Coeficiente de correlação de Pearson (CCP) (CHEN; POPOVICH, 2002) como critério de otimização (ou seja, função de aptidão do GA). A fim de definir formalmente o CCP, considere duas variáveis aleatórias X e Y, cada

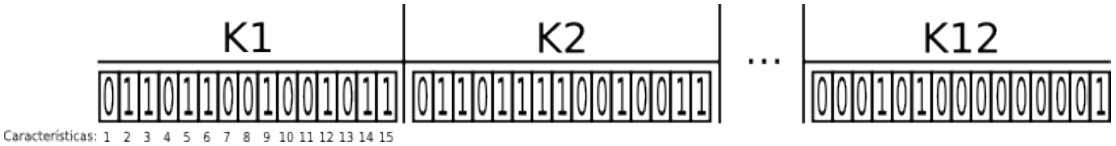

Figura 7: Codificação do cromossomo para seleção de características aplicado à base “credit-a”.

amostra de tamanho n. Então o CCP, pode ser definido pela Equação (3.1).

$$CCP = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}. \tag{3.1}$$

Em outras palavras, CCP é uma medida não-paramétrica de correlação, que avalia o quão boa uma função monotônica arbitrária poderia descrever a relação entre duas variáveis aleatórias, sem fazer quaisquer suposições sobre a distribuição dos seus valores. Assim, esta medida parece ser muito apropriada para capturar a correlação das características de qualquer padrão de entrada.

No nosso caso, CCP é utilizado para medir os níveis de intra-correlação, isto é, a correlação entre as características atribuídas a um único componente. A ideia é escolher as características para um classificador que não possuem correlação. A CCP é calculada separadamente para cada componente e em seguida a média é calculada para produzir um único resultado. Quanto maior a média, maior tende a ser a diversidade entre os componentes induzidos.

Na segunda fase da nossa abordagem, reamostragem de dados é empregada como o mecanismo para elevar os níveis de diversidade entre os classificadores. Isto é feito usando réplicas de *bootstrap* do conjunto de dados de treinamento como em *Bagging* (BREIMAN, 1996b), cada réplica que está sendo gerada aleatoriamente. Cada novo conjunto de dados tem o mesmo número de instâncias de formação do conjunto de dados original (N_{tr}); no entanto, algumas instâncias aparecerão várias vezes, enquanto outras não vão aparecer. O tamanho efetivo é menor, e os conjuntos de dados tendem a sobrepor-se significativamente. Cada conjunto de dados derivado é usado para treinar um classificador, e os recursos associados a cada conjunto de dados são aqueles selecionados pelo AG da etapa anterior da metodologia.

Finalmente, o terceiro estágio emprega um segundo AG (EIBEN; SMITH, 2007) como seletor de componente. Cada componente pode ser induzido por diferentes algoritmos de aprendizado para uma base derivada pela estratégia anterior, produzindo arquiteturas heterogêneas (CA-

NUTO *et al.*, 2007). Neste estudo, assim como em Coelho e Nascimento (2010), nós consideramos $M = 10$ diferentes algoritmos de aprendizado (WITTEN; FRANK, 2005): *Naïve Bayes* (NB), baseado em estatística *bayesiana*; redes neurais RBF e máquinas de vetores-suporte via SMO algoritmo, baseada em função não-linear; J48 e REP *Tree* (RT), baseado em árvores de decisão; IBk, baseado em aprendizado por vizinhança; e *Decision Stump* (DS), OneR, PART e *Decision Table* (DT), baseados em regras.

Nesse segundo AG, os indivíduos são codificados como vetor de valores inteiros (referente ao terceiro estágio da Figura 6). Esses valores são representados entre 1 e 10, indicando, respectivamente, os métodos de classificação listado anteriormente SMO, NB, IB*k*, RT, RBF, J48, DS, OneR and PART. Através da validação cruzada estratificada de 10-*fold*, os modelos de *ensembles* heterogêneos são induzidos. Os valores de erro médio são utilizados como valor de aptidão.

Do ponto de vista teórico, a eficácia da proposta de três estágios de diversidade também pode ser analisada, considerando o trabalho de Geman, Bienenstock e Doursat (1992), de acordo com o erro esperado de um algoritmo de aprendizagem de uma função alvo e um conjunto de treinamento de tamanho particular, podem ser decomposto em termos de *bias*, variância e ruído.

3.2.2 Segunda Abordagem: Algoritmo Genético com Função de Avaliação Adaptativa

A ideia do novo algoritmo genético é semelhante a discutida para incremento de diversidade; por outro lado, a contribuição está relacionada na utilização de uma função de aptidão adaptativa que explora duas técnicas de avaliação de *fitness*, uma baseada em filtro e uma outra baseada em *wrapper*. A motivação nessa solução deve-se a observação em muitos experimentos computacionais, nos quais, percebeu-se que melhores resultados são alcançados quando aplicados à técnica *wrapper*; por outro lado, o tempo de execução para experimentos mais complexos era melhor (menor tempo) quando aplicado à técnica de filtro. Isso é facilmente justificado, porque geralmente no *wrapper* é utilizado, como resultado de avaliação, o erro médio obtido da validação cruzada de uma algoritmo de aprendizado. Já no filtro, são utilizados como resultado de avaliação, por exemplo, cálculos estatísticos aplicados diretamente sobre o conjunto de treinamento. Logo o objetivo é conseguir, de forma concomitante, alcançar os benefícios de ganhos de desempenho e custo computacional proporcionado por cada técnica quando aplicadas em separado.

A Equação (3.2) apresenta a adaptação de uma função de avaliação que usa as técnicas de filtro e *wrapper*.

$$fit_{(i)} = WF_{(i)} * Filter_{(i)} + WW_{(i)} * Wrapper_{(i)} \tag{3.2}$$

Onde $WF_{(i)}$ é um peso associado à função de filtro $Filter_{(i)}$. No mesmo contexto, $WW_{(i)}$ corresponde a um peso associado a função *wrapper* $Wrapper_{(i)}$. A função de filtro é a mesma utilizada na versão do algoritmo genético mencionado anteriormente. Já para a função *wrapper* foi utilizado o nível de acuidade em *ensemble*. Portanto, como já mencionado, o cálculo de acuidade de um *ensemble* é complexo em termos computacionais. Ao fazer isso, precisamos decidir aspectos importantes, principalmente relacionados com a utilização da função de avaliação para aqueles indivíduos que não passaram por avaliação de *wrapper* e a proporção da população que passa pela avaliação *wrapper*. Em relação à primeira questão, a função de aptidão dos indivíduos que não passaram pela avaliação *wrapper* foi baseada apenas na avaliação do filtro, tendo $WF_{(i)} = 1$ e $WW_{(i)} = 0$. Em relação à segunda questão, essa proporção é definida de forma dinâmica e utilizando o seguinte procedimento:

1. Defina a proporção inicial de população que passa por avaliação *wrapper*, $prop$, a 50%;
2. Escolha os indivíduos para avaliação *wrapper*. Essa escolha pode ser feita aleatoriamente ou tendo em consideração um critério. No nosso caso, foi utilizada a última vez que o ascendente do indivíduo foi avaliado via *wrapper*. A ideia é dar altas probabilidades para indivíduos cujo ascendente não passaram por avaliação de *wrapper*;
3. Criar duas listas de classificação (*rankging*), baseado no critério de filtro e uma com base no critério *wrapper*;
4. Compare os dois *ranking* e então faça:
   (a) Incremente $prop$ se a o valor de similaridade entre os dois *ranking* é maior que o máximo, *max*;
   (b) Decremente $prop$ se o valor de similaridade dos dois *ranking* é menor que o mínimo, *min*; ou

(c) Mantenha o valor de *prop* se o valor de similaridade estiver entre o mínimo e máximo.

A ideia da configuração dinâmica de *prop* é que se o *ranking* usando filtro semelhante ao *ranking* produzido vai *wrapper*, mantêm-se a utilização de avaliação via filtro (por ser mais rápido), caso contrário, mantêm-se utilizando a avaliação o *wrapper* (por ser mais eficiente). Em nosso estudo experimental, usamos $max = 70\%$ e $min = 40\%$. Além disso, a escolha dos pesos associados aos filtros e avaliações *wrapper* pode ser um processo complexo. Portanto, neste trabalho, assumimos $WF(i) = WW_{(i)} = 0,5$. Neste caso, ambas as abordagens têm igual importância na função de *fitness*.

Para ambas as versões dos algoritmos genéticos, os principais parâmetros foram definidos da seguinte forma (após o ajuste manual): 20 como o tamanho da população; 80 % e 5 % em taxas de cruzamento e mutação, respectivamente; e 200 o número de máximo de gerações. O algoritmo genético também mantém o melhor indivíduo para a próxima geração (elitismo).

3.2.3 Terceira Abordagem: Recomendação via Meta-aprendizado

Como já se sabe, os variados arranjos de *ensembles* apresentam diferentes resultados em termos de desempenho. Desta forma, faz-se necessária, para a construção de um modelo de recomendação, a aplicação do conhecimento de Meta-aprendizado para associar os modelos que apresentaram melhor desempenho, as meta-características dos problemas na qual foram aplicados. Por outro lado, fazemos uma divisão quanto a abordagem de meta-aprendizado a ser aplicada, e adotamos as nomenclaturas: Meta-aprendizado Tradicional e Meta-aprendizado Multirrótulo. Em ambas divisões, o processo de associação é realizado em quatro etapas separadamente. (a) A primeira etapa é referente à extração de meta-características do total de 200 problemas de classificações, conforme ilustrada na Figura 8. (b) A segunda etapa é referente à associação entre desempenho dos modelos de arquitetura as meta-características dos problemas aplicados, dando suporte ao processo de recomendação do tipo de arquitetura, conforme ilustrada Figura 9. (c) Já na terceira etapa, mapeamos os desempenhos dos modelos homogêneos as meta-características dos problemas aplicados, dando suporte ao processo de recomendação dos componentes homogêneos. E por fim, (d) a quarta etapa (essa etapa só é implementada na abordagem multirrótulo), que faz o mapeamento do desempenho dos possíveis componentes he-

terogêneos as meta-características dos problemas aplicados. Essa última etapa, dá suporte ao processo de recomendação dos componentes heterogêneos.

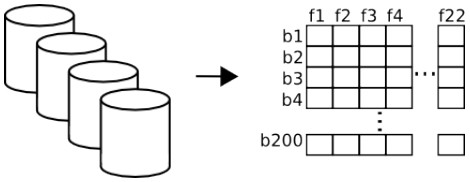

Figura 8: Extração de meta-características.

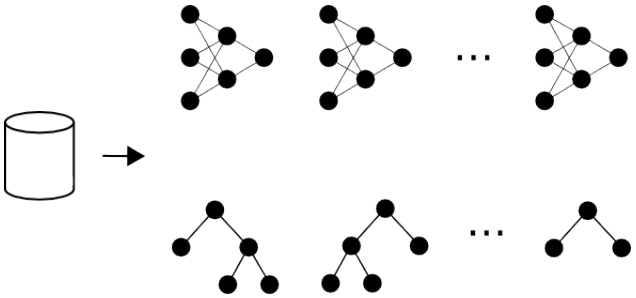

Figura 9: Testes para mapeamento.

Meta-aprendizado Tradicional

Nesta abordagem, é realizado o mapeamento entre as metas-características aos melhores modelos, levando em consideração apenas o desempenho dos algoritmos que apresentaram menor valor do erro médio, vide Figura 10.

Para implementar o mapeamento referente à arquitetura, são realizados os testes (validação cruzada de 10-*folds*) com todos as arquiteturas (1 – *Bagging*, 2 – *Boosting* e 3 – *Multiboosting*) formados por todos os tipos homogêneos de componentes base (1 – RBF, 2 – J48, 3 – SMO, 4 – NB e 5 – IBk). É realizada um média geral dos experimentos de todas as arquiteturas. Em seguida é construído um *ranking* dos erros médios. A arquitetura que apresentar menor valor é selecionada para realizar o mapeamento das meta-característica da base testada. A Figura 11 ilustra esse processo.

Para implementar o mapeamento dos componentes homogêneos, foram executadas validações cruzadas de 10-*folds*, 10 vezes sobre diferentes *seeds* de testes. Em seguida é construído um *ranking* para os tipos homogêneos de componentes base (1 – RBF, 2 – J48, 3 – SMO, 4 –

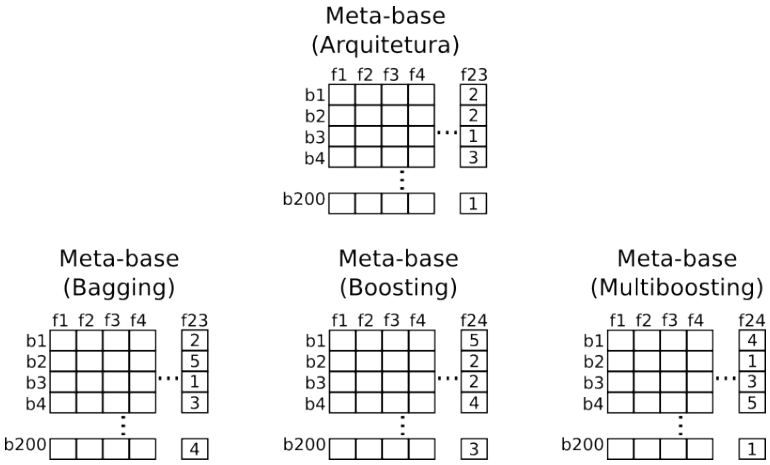

Figura 10: Mapeamento tradicional.

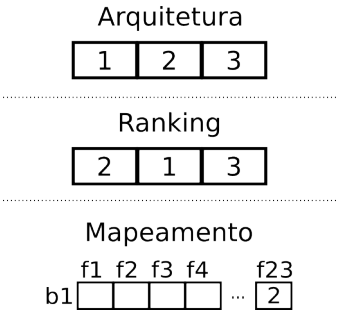

Figura 11: Mapeamento tradicional para escolha da arquitetura.

NB e 5 – IBk), com base no erro médio da validação cruzada para todos os *seeds*. O componente que apresentar menor valor é selecionado para realizar o mapeamento. A Figura 12 ilustra esse processo.

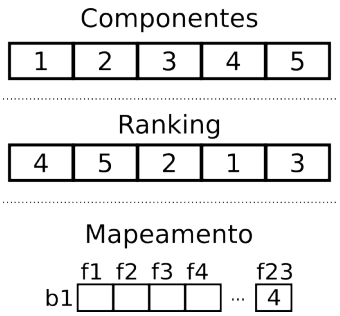

Figura 12: Mapeamento tradicional para escolha dos componentes homogêneos.

Meta-aprendizado Multirrótulo

Esta abordagem, diferentemente da abordagem tradicional, o mapeamento entre as metas-características aos melhores modelos, não é realizada levando-se em consideração apenas os valores de erro médio, mas sim, um teste estatístico que indica se a equivalência entre os desempenhos podem ou não ser rejeitados. Uma vez que os modelos apresentem resultados estatisticamente equivalentes, fazemos um mapeamento multirrótulo, vide Figura 13, considerando, desta forma, todas as soluções em potencial. Utilizamos o teste de hipótese *t*-teste bicaudal pareado, entre o melhor modelos e os demais. Esse teste estatístico tem como objetivo avaliar a equivalência entre duas medidas amostrais, supondo independência e normalização das observações – no caso, das taxas de erros (WITTEN; FRANK, 2005).

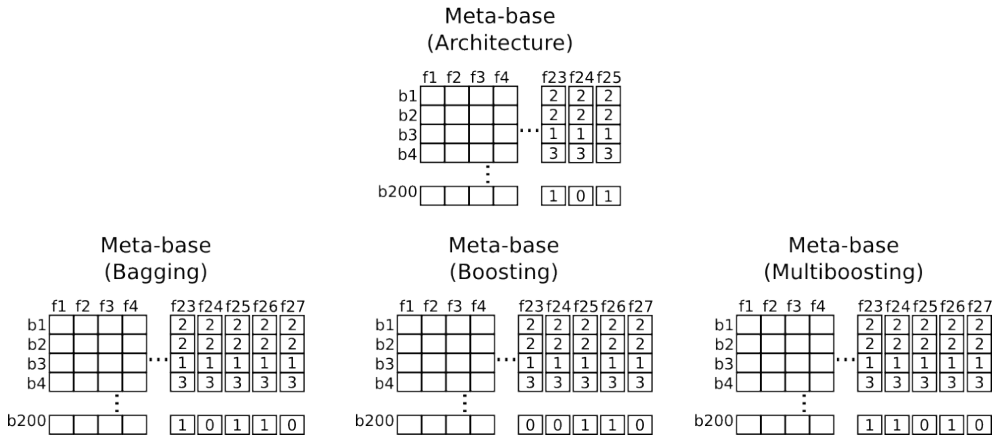

Figura 13: Mapeamento multirrótulo.

A implementação do processo de mapeamento da arquitetura multirrótulo é semelhante ao processo tradicional até a geração do *ranking*. Na sequência é calculada o teste de hipótese da melhor arquitetura com todas as demais. No mapeamento multirrótulo, o melhor valor assume valor um, enquanto que as demais, se o teste estatístico apresentar valor maior 0,5, não rejeitamos a hipótese nula H_0, então é mapeado o valor um, caso contrário, mapeamos com valor zero, visto que rejeitamos H_0. A Figura 14 ilustra esse processo.

A implementação do processo de mapeamento de componentes homogêneos segue a mesma ideia do citado para arquitetura, tanto no que se refere à construção do *ranking* quanto do mapeamento multirrótulo. A Figura 15 ilustra o processo.

A implementação do processo de mapeamento de componentes heterogêneos difere dos já

Figura 14: Mapeamento multirrótulo para escolha da arquitetura.

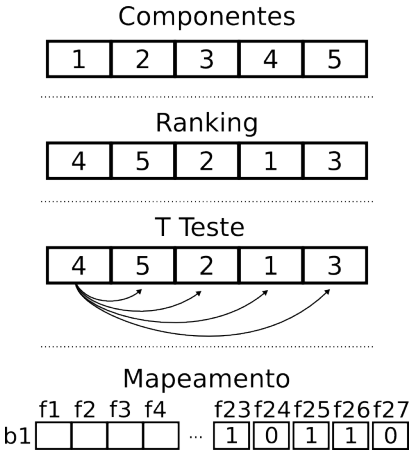

Figura 15: Mapeamento multirrótulo para escolha dos componentes homogêneos.

citados. Para a escolha dos componentes heterogêneos, é evoluída uma população de indivíduos. Cada indivíduo é uma representação inteira de 10 posições, cujo os alelos podem assumir valores entre [1,5], esse valores por sua vez indicam os tipos de componentes base heterogêneos (1 – RBF, 2 – J48, 3 – SMO, 4 – NB e 5 – IBk). Em seguida, para realização do mapeamento, é informado de forma binária, se o tipo de componente heterogêneo aparece ou não, onde um indica presença e zero a ausência. A Figura 16 ilustra o processo.

Uma vez realizado todos os mapeamentos, os meta-aprendizes são algoritmos de aprendizado conhecidos, como: rede neural *Multi-layer Perceptron* (MLP), máquina de vetores-suporte, árvore de decisão e algoritmo baseado em vizinhança. Nesse processo são criados meta-aprendizes independentes para recomendação de arquitetura, componentes homogêneos ou heterogêneos.

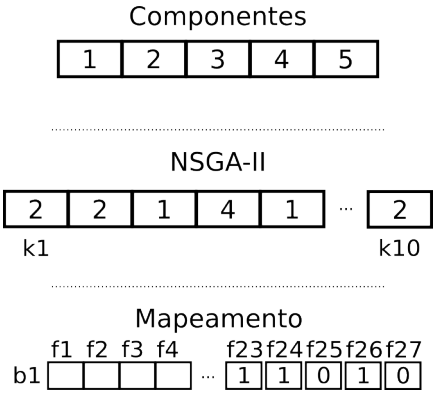

Figura 16: Mapeamento multirrótulo para escolha dos componentes heterogêneos.

Uma vez que se tenha um novo problema de classificação de padrões, todas as meta-características são extraídas, e apresentadas para cada meta-aprendiz treinado para o tipo de recomendação específica. Esses, por sua vez, farão a recomendação dos melhores parâmetros de controle do *ensemble* a ser utilizado como o novo problema apresentado. A Figura 17 ilustra o processo de recomendação.

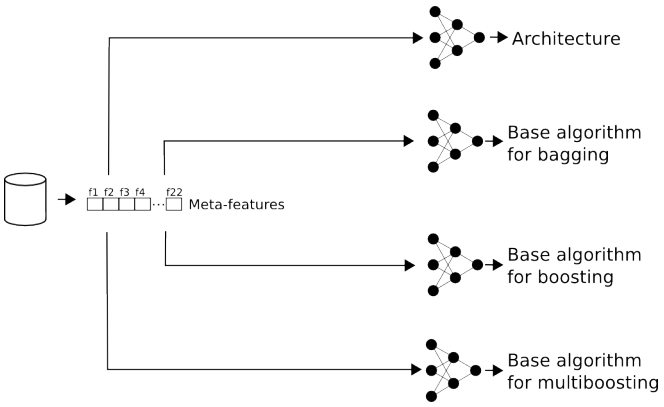

Figura 17: Meta-aprendiz.

3.3 Síntese do Capítulo

Este capítulo permitiu ao leitor entender as novas abordagens aqui apresentada. Foram destacadas as contribuições correlatas ao tema. Apresentamos três novas abordagens que são a base da contribuição desta tese. Na sequência apresentamos a nova abordagem de incremento de técnicas de diversidade em *ensembles*. Posteriormente, apresentamos uma nova abordagem via

Algoritmo Genético para seleção de atributos. E finalmente, apresentamos a abordagem para recomendação dos parâmetros de controle em *ensembles* para arquitetura e componentes homogêneos e heterogêneos.

O próximo Capítulo 4, serão apresentadas as metodologias utilizadas para realização dos experimentos computacionais, bem como, uma análise dos resultados obtidos pelas abordagens aqui mencionadas.

Capítulo 4

Experimentos Computacionais

Este capítulo é dedicado à apresentação dos experimentos computacionais de validação das novas abordagens para construção eficientes de *ensembles* apresentadas no capítulo anterior. Tais protótipos foram implementados usando linguagem de programação Java, lançando-se mão dos insumos providos pelo *framework* WEKA (WITTEN; FRANK, 2005).

4.1 Problemas de Classificação

Para fins de validação da proposta, serão conduzidos experimentos sobre 20 problemas de classificação extraídos do repositório UCI (ASUNCIóN; NEWMAN, 2007). A maioria desses problemas também já serviu de alvo de investigação em trabalhos correlatos na linha de Comitês de Classificadores (OPITZ; MACLIN, 1999; DIETTERICH, 2000b; CANUTO *et al.*, 2007). As bases de dados relativas a esses problemas, assim como uma descrição das suas propriedades em termos de número de amostras, número e tipos de atributos, número e distribuição das classes, e existência de atributos faltantes, podem ser encontradas no Apêndice A.

Para cada um dos 20 conjuntos de dados, serão gerados novos 10 conjunto de dados de forma aleatória e estratificado, totalizando 200 novos conjuntos de dados. Esse estratégia foi

adotada na tentativa de se obter um número maior de conjuntos de dados para extração das meta-características, bem como, a realização dos testes das gerações das diferentes configurações de *ensembles* via *Bagging*, *Boosting* e *MultiBoosting*.

4.2 Algoritmos de Aprendizagem

Os algoritmos de aprendizagem utilizados nesta pesquisa foram retirados do *framework* WEKA (WITTEN; FRANK, 2005). Esse *framework* é uma coleção de algoritmos de aprendizado de máquina que foi desenvolvida pela *University of Waikato* na Nova Zelândia. Esse ambiente inclui uma variedade de ferramentas para transformação de dados, tais como algoritmos de discretização, bem como implementações de métodos para mineração de dados (classificação, agrupamento, regras de associação). É válido mencionar, ainda, que esse *framework* vem sendo, recentemente, bastante adotado como base de desenvolvimento e validação de novas abordagens de aprendizado de máquina, notadamente, aquelas baseadas em comitês de classificadores (SOARES *et al.*, 2006; CANUTO *et al.*, 2007).

No estudo, foram adotados $M = 10$ diferentes algoritmos de aprendizagem representativos de cinco classes distintas de indutores retirados do repositório WEKA (WITTEN; FRANK, 2005):

- o algoritmo *Naïve Bayes* (NB) simples, baseado em estatística *bayesiana*;
- redes neurais RBF e máquinas de vetores-suporte treinadas via algoritmo SMO, que se baseiam em funções numéricas não-lineares;
- os algoritmos J48 e REP *Tree* (RT), baseados em árvores de decisão;
- o algoritmo IBk, baseado nos conceitos de vizinhança e aprendizado local; e
- os algoritmos *Decision Stump* (DS), OneR, PART e *Decision Table* (DT), os quais se baseiam em regras.

O algoritmo J48 é uma implementação em Java do algoritmo de árvore de decisão C4.5 (QUINLAN, 1993). Trata-se de uma versão melhorada do algoritmo C4.5, a versão anterior a C5.0, que

é comercial. Gera um modelo em árvore via cálculo da entropia[1] dos atributos, sendo que cada folha da árvore resultante representa uma classe e os demais nós especificam testes condicionais a serem realizados em um determinado atributo. Cada possível valor de um atributo gera um ramo na árvore.

O algoritmo REP *Tree* constrói de forma otimizada e rápida uma árvore de decisão com base no ganho de informação (QUINLAN, 1992) ordenando os atributos durante o treinamento. Os nós representam decisões que dividem os dados em dois grupos, usando, de todos os atributos ainda não considerados, o que leva a um ganho maior. As folhas da árvore representam pontos em que uma classificação é atribuída. A árvore é então podada, com recurso à técnica de *reduced error pruning* (REP), substituindo subárvores por folhas. Se essa substituição reduzir o erro esperado, minimiza-se a adaptação aos erros (*error fitting*) e reduz-se a complexidade da árvore. A nova árvore é então o modelo de classificador, pronto para ser utilizado sobre novos dados.

O algoritmo *Decision Stump* constrói uma árvore de decisão simples com um nível binário e produz classes de probabilidade (WITTEN; FRANK, 2005). Ele é baseado na ideia de ganho de informação, utilizada no algoritmo C4.5 que divide os conjuntos de dados através do atributo mais informativo. Este ganho de informação mede o erro após se ter dividido o conjunto de dados de acordo com os valores do atributo utilizado no *nó* de decisão. Posteriormente o algoritmo é avaliado mediante validação cruzada, e o valor obtido é utilizado como atributo para caracterizar o conjunto de dados.

O algoritmo OneR produz regras simples, fundamentando-se na utilização do classificador 1R (HOLTE, 1993). O modelo OneR é usado em problemas de classificação e baseia-se no pressuposto de que "frequentemente um atributo é suficiente para determinar a classe". O atributo escolhido pelo OneR é aquele que possuir menor erro. É considerado o algoritmo mais simples. Como o nome sugere, este sistema "gera" uma regra (*One Rule*). Em algumas circunstâncias é tão poderoso quanto algoritmos mais sofisticados, como o J48 e o PART.

O algoritmo PART (FRANK; WITTEN, 1998) é um indutor de modelos baseados em regras de decisão, e utiliza internamente o algoritmo C4.5. Ele constrói árvores de decisão parciais a cada iteração e transforma a melhor folha da árvore atual em uma regra. Após escolher a melhor folha, o algoritmo retira todas as instâncias que se encaixem na regra gerada pela folha para gerar

[1] A entropia é usada para estimar a aleatoriedade dos valores dos atributos para as classes que eles representam.

uma nova árvore, e, por conseguinte, uma nova regra. Esse processo é realizado iterativamente até que não se tenha mais instâncias a serem utilizadas. Como não usa a base inteira para gerar as árvores, este algoritmo, geralmente, incorre em um bom desempenho computacional quando da indução do modelo de classificação.

O algoritmo *Decision Table* é um indutor de modelos baseados em regras que representam o conhecimento na forma de tabela (KOHAVI, 1995). Uma tabela de decisão é um modelo de representação de uma função. Ela permite uma visão tabular da lógica de decisão de um problema. Indica qual o conjunto de condições que é necessário ocorrer para que um determinado conjunto de ações deva ser executado. Uma tabela de decisão é basicamente composta de: uma área de condições, onde são relacionadas as condições que devem ser verificadas para que seja executado um conjunto de ações; uma área de ações, que exibe o conjunto de ações que deve ser executado caso um determinado conjunto de condições ocorra; e regras de decisão, representadas pelas colunas, que apresentam a combinação das condições com as ações a serem executadas.

O algoritmo *Naïve Bayes* (JOHN; LANGLEY, 1995) é de cunho probabilístico, sendo baseado na estatística *bayesiana*. O algoritmo assume que os atributos são condicionalmente independentes, ou seja, a informação representada por um atributo não é informativa sobre a de nenhum outro. Diz-se que o atributo X_1 é condicionalmente independente da classe C, dado outro atributo X_2, se a distribuição de probabilidades de X_1 é independente do valor de C, dado o valor de X_2.

O algoritmo IB*k* (AHA; KIBLER; ALBERT, 1991) é uma implementação da abordagem *k* - *Nearest Neighbors* (*k*-NN), sendo classificado como um algoritmo de aprendizado "preguiçoso" (AHA, 1997), haja vista que não gera nenhum modelo a partir dos dados (utiliza apenas as instâncias em si). Este algoritmo assume que instâncias similares possuam a mesma classificação e dependam fortemente da métrica usada como distância.

Redes neurais RBF constituem um tipo popular de rede neural de três camadas (HARPHAM; DAWSON; BROWN, 2004). Cada nó da camada intermediária representa a região em torno de um ponto no espaço de entrada. Sua saída para uma instância apresentada na camada de entrada depende apenas da distância entre o seu ponto associado e a própria instância de entrada. Quanto maior a distância, maior a força de ativação do neurônio (HAYKIN, 2001). A camada intermediária utiliza uma função de ativação de base radial para converter a distância em uma

medida de similaridade. A camada de saída, por sua vez, é responsável por combinar linearmente as saídas geradas na camada intermediária e gerar a estimação final.

SMO é uma abordagem proposta por Platt (1999) para lidar, de maneira escalável, com o treinamento de máquinas de vetores-suporte (SVM). É a implementação de SVM disponível no WEKA (WITTEN; FRANK, 2005). Esse algoritmo transforma a saída do classificador SVM em probabilidades através da aplicação de uma função sigmoide padrão que não é adaptada (*fitted*) aos dados. Essa implementação não é rápida com um espaço de fatores linear, nem com dados esparsos. Substitui todos os valores que faltarem, transforma atributos nominais em binários e normaliza todos os valores numéricos.

Foram também utilizados os algoritmos de *ensembles* que são baseados em reamostragem dos dados de treinamento:

- *Bagging*;
- *Boosting*; e
- *Multiboosting*.

Esse repertório abrange indutores simples e complexos; além disso, abarca indutores notadamente instáveis (redes neurais RBF, J48) como também os de certa estabilidade (SVM) e os reconhecidamente estáveis (*Naïve Bayes*). O objetivo dessa escolha é o de avaliar se a fusão de indutores simples com complexos, instáveis com estáveis, via *Bagging*, *Boosting* e *MultiBoosting* produzem ganhos em termos de eficácia. As configurações desses algoritmos são apresentadas no Apêndice B.

4.3 Algoritmos de Classificação Multirrótulo

Os algoritmos de classificação multirrótulo foram retirados da biblioteca Mulan (TSOUMAKAS; KATAKIS; VLAHAVAS, 2010a). Essa biblioteca traz uma coleção de algoritmos implementadas em Java para a classificação de dados multirrótulo, que inclui, além dos algoritmos de classificação, funcionalidades como seleção de atributos e de cálculo de medidas de acuidade.

Mulan tem como dependência o *framework* WEKA para que se possa utilizar, por exemplo, os algoritmos de classificação único-rótulo nas transformações de dados multirrótulo. Os algoritmos utilizados na pesquisa foram RA*k*EL e o ML-*k*NN.

4.4 Algoritmo Genético

O algoritmo genético utilizado nesta pesquisa foi retirado do *framework* jMetal (DURILLO; NEBRO, 2011; DURILLO; NEBRO; ALBA, 2010) e adaptado.

Nos experimentos, os valores dos parâmetros de configuração do Algoritmo Genético adotados foram:

- Tamanho da população de 20 indivíduos;
- Probabilidade de cruzamento de 80%;
- Probabilidade de mutação de 10%; e
- Número máximo de 200 gerações.

Esses valores foram obtidos a partir de experimentos preliminares.

4.4.1 Extração das Meta-características

O processo de extração de características representativas dos dados será realizado a partir da ferramenta do projeto Metal. Esse projeto fomentou o desenvolvimento de várias abordagens relacionadas à pesquisa em Meta-aprendizado, definição de meta-atributos para caracterizar os conjuntos de dados e estudos teóricos sobre o tema.

Como foi citado na seção anterior, a pesquisa conta com um total de 20 conjuntos de dados originais. Como será realizada a geração aleatória de 10 novos conjuntos treinamento para cada problema original, contaremos com um total de 200 diferentes conjuntos de treinamento para extração dos meta-atributos. Inicialmente, serão utilizadas as medidas de caracterização de dados apresentadas nas Tabelas 1 e 2 descritas no Capítulo 2. Desta forma contaremos com uma quantidade significativa de diferentes problemas de classificação para conduzir os experimentos

e apresentar resultados mais precisos da proposta apresentada. No Apêndice D estão disponíveis os valores de todas as meta-características.

4.5 Avaliação de Algoritmos

Tendo em mente a obtenção de resultados estatisticamente significantes, para cada um dos 200 problemas de classificação, será executada uma validação cruzada 10-*fold*. Em outras palavras, cada conjunto será dividido em r partições mutuamente exclusivas (*folds*) de tamanho aproximadamente igual a n/r exemplos. Os exemplos no $(r-1)$ *folds* são usados para treinamento e a hipótese induzida é testada no *fold* remanescente. Esse processo é repetido r vezes, cada vez considerando um *fold* diferente para teste, conforme ilustrado na Figura 18.

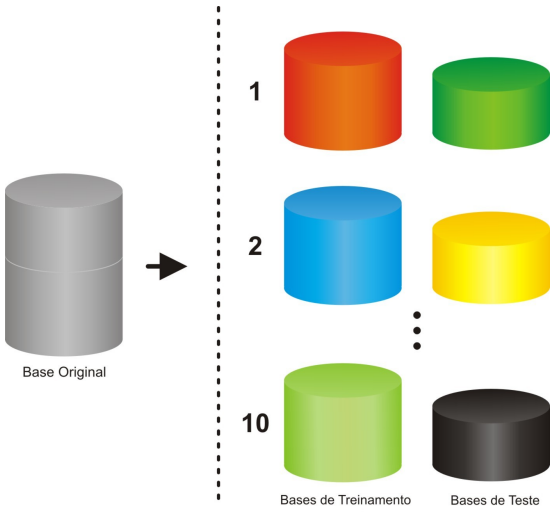

Figura 18: Partição dos dados.

O erro na validação cruzada é a média dos erros calculados em cada um dos r *folds*. Dado um algoritmo A e um conjunto de exemplos D, assume-se que D seja dividido em r partições. Para cada partição i, é induzida a hipótese h, e o erro denotado por $erro(h_i)$, $i = 1,2,\ldots,r$, é calculado. Em seguida, a Média Aritmética (MDA) e o Desvio Padrão (DP) são calculados utilizando-se as Equações 4.1 e 4.2, respectivamente.

$$MDA(A) = \frac{1}{r}\sum_{i=1}^{r} erro(h_i) \tag{4.1}$$

$$SD(A) = \sqrt{\frac{1}{r}\left[\frac{1}{r-1}\sum_{i=1}^{r}(erro(h_i) - MDA(A))^2\right]} \tag{4.2}$$

4.6 Síntese do Capítulo

Este capítulo permitiu ao leitor entender as etapas de execução dos experimentos computacionais e as ferramentas envolvidas.

O Capítulo 5 a seguir, faz uma discursão sobre os resultados alcançados de forma a corroborar com a Tese de que as abordagens de Incremento de Diversidade, Função de Avaliação Adaptativa e Meta-aprendizado são técnicas poderosas que podem ser utilizadas para configuração de *ensembles* de classificadores de forma eficiente.

Capítulo 5

Resultados

Neste Capítulo fazemos uma discursão dos resultados alcançados que estão organizados nas Seções 5.1, 5.2, 5.3 e 5.4 que descrevem, respectivamente, os resultados dos desempenho dos algoritmos de *ensembles* tradicionais, os resultados para a primeira abordagem, os resultados para a segunda abordagem e os resultados para a terceira abordagem.

5.1 Desempenho dos Algoritmos de *Ensembles* Tradicionais

Antes de discutir sobre os resultados obtidos a partir dos experimentos das novas abordagens propostas, faremos uma discursão sobre os resultados obtidos pelos métodos de *ensembles* tradicionais apresentados anteriormente no Capítulo 2.

Alguns testes preliminares foram realizados sobre 200 problemas de classificação, vide Anexo C. Como se pôde observar, os resultados são bastante diversificados, e nem sempre os comitês formados por componentes mais elaborados como redes neurais e máquinas de vetores-suporte apresentam melhor resultados de generalização. Mesmo para cada conjunto de dados original sendo gerado 10 novos conjuntos de dados aleatoriamente, os menores valores de erro médio, mediante validação cruzada, foram apresentados por *ensembles* formados por componen-

tes base homogêneos distintos. Para *Bagging*, como podemos observar na Figura 20, utilizando o problema de classificação “breast-cancer 1”, o melhor desempenho foi alcançado quando se utilizou algoritmos J48 como componente base, já para o problema “breast-cancer 2”, o melhor desempenho foi alcançado quando se utilizou algoritmos NB como componente base, enquanto que para o problema “breast-cancer 4”, o melhor desempenho foi alcançado quando se utilizou algoritmos RBF como componte base. Na mesma análise para *Boosting*, como podemos observar na Figura 19, utilizando o problema de classificação “breast-cancer 1”, o melhor desempenho foi alcançado quando se utilizou algoritmo IB*k*, como componente base, já para o problema “breast-cancer 5”, o melhor desempenho foi alcançado quando se utilizou algoritmos SMO como componente base, enquanto que para o problema “breast-cancer 9”, o melhor desempenho foi alcançado quando se utilizou algoritmos NB como componente base. E também para *Multiboosting*, como podemos observar na Figura 21, utilizando o problema de classificação “breast-cancer 1”, o melhor desempenho foi alcançado quando se utilizou algoritmos IB*k* como componente base. Já para o problema “breast-cancer 3”, o melhor desempenho foi alcançado quando se utilizou algoritmos RBF como componente base. Enquanto que para o problema “breast-cancer 9”, o melhor desempenho foi alcançado quando se utilizou algoritmo NB como componente base.

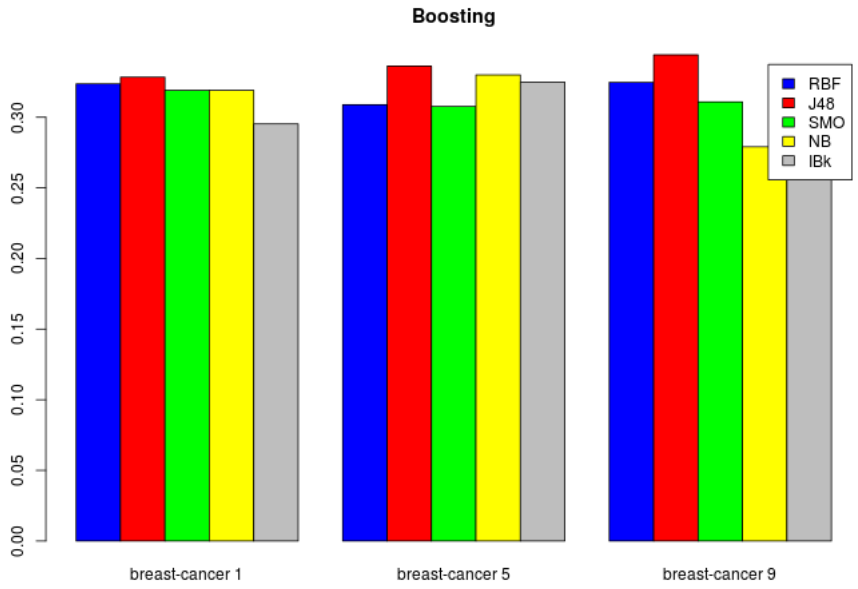

Figura 19: Valores de erro médio para *Boosting*.

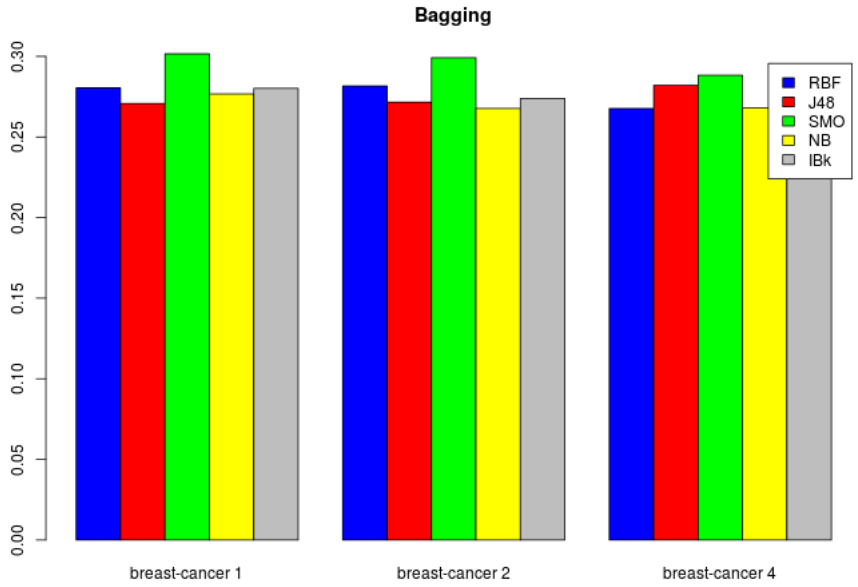

Figura 20: Valores de erro médio para *Bagging*.

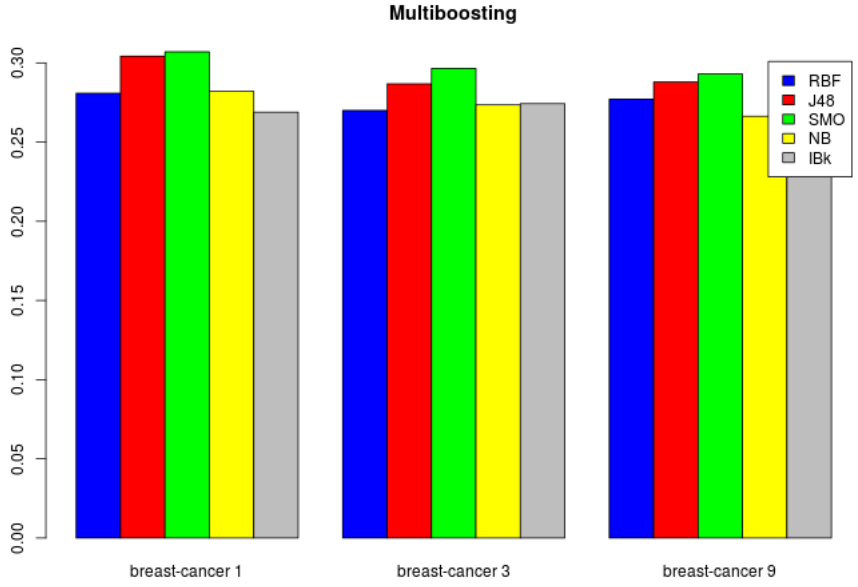

Figura 21: Valores de erro médio para *Multiboosting*.

Outros resultados interessantes, dizem respeito à escolha do algoritmo de construção de *ensembles* para um problema específico de classificação. Por exemplo, a partir das características

do algoritmo *Bagging* apresentadas no Capítulo 2, é aconselhável para tal técnica a utilização de componentes de natureza instável, por outro lado, nos problema “anneal 2”, “breast-cancer 3” e “iris 4”, respectivamente, os menores valores de erro médio foram alcançados quando utilizados modelos homogêneos formados pelos algoritmos IBk, NB e SMO, que são de natureza estável, conforme apresentados nas Figuras 22, 23 e 24. Em outras palavras, talvez para tal problema, esses algoritmos fossem inicialmente descartados pelo projetista de sistemas.

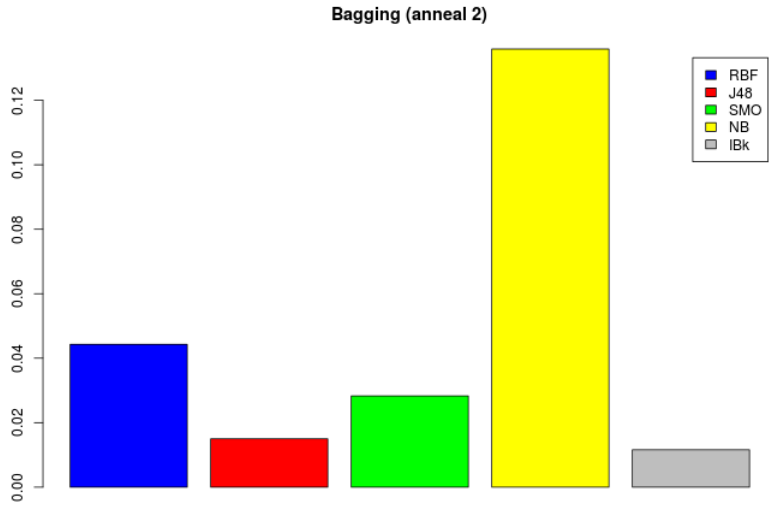

Figura 22: Valores de erro médio para *Bagging* quando aplicado à base “anneal 2”.

Em outro caso que merece destaque, o algoritmo *Boosting* homogêneo formado por componentes RBF apresentou menor erro médio quando aplicado a todas as bases “credit-a”, em relação àquele produzido por *Bagging* homogêneo que utilizava o mesmo componente base, conforme apresentação na Figura 25. Para esse caso, o *Boosting* não é aconselhado para ser utilizado em problemas que apresentam ruídos ou valores faltosos.

Outra análise faz referência ao algoritmo *MultiBoosting* homogêneo formado por componentes SMO quando aplicado ao problema “haberman”, no qual, mesmo combinando características dos algoritmos *Bagging* e *Boosting* de ganhos de variância e polarização, respectivamente, não apresentou menor erro médio, conforme se pode observar na Figura 26.

Toda essa variação de resultados já era esperada, e está diretamente ligada à contribuição que a diversidade via reamostragem proporciona. Em alguns momentos, os algoritmos mais

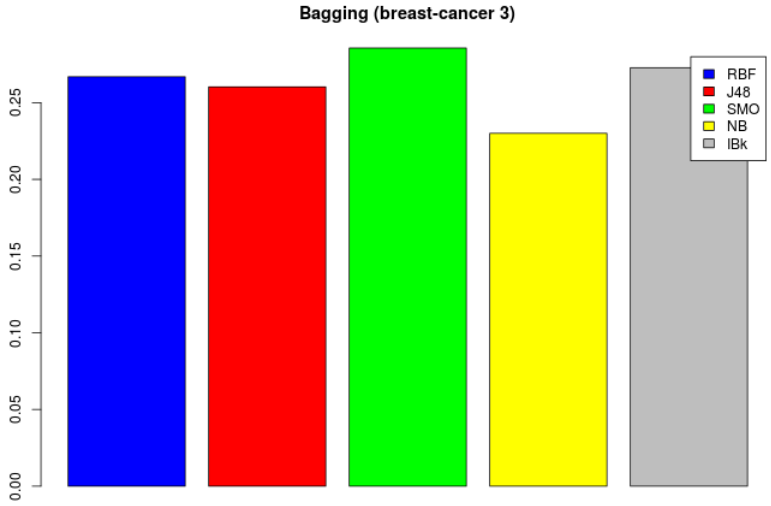

Figura 23: Valores de erro médio para *Bagging* quando aplicado à base “breast-cancer 2”.

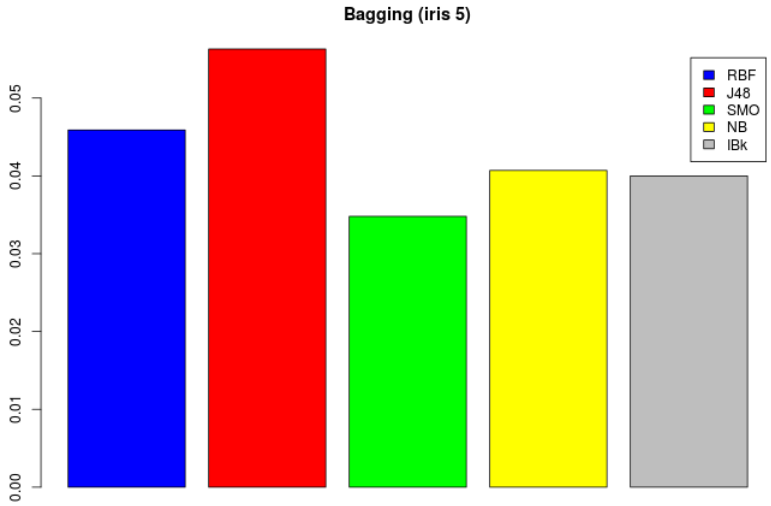

Figura 24: Valores de erro médio para *Bagging* quando aplicado à base “iris 4”.

indicados são prejudicados por causa de uma má redistribuição dos dados de treinamento, ao mesmo tempo que os algoritmos não indicados são beneficiados.

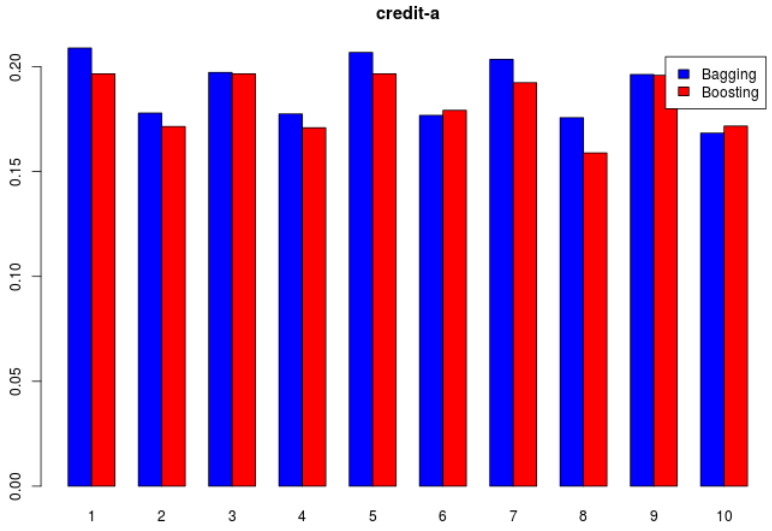

Figura 25: Valores de erro médio para *Bagging* e *Boosting* quando aplicado à base "credit-a".

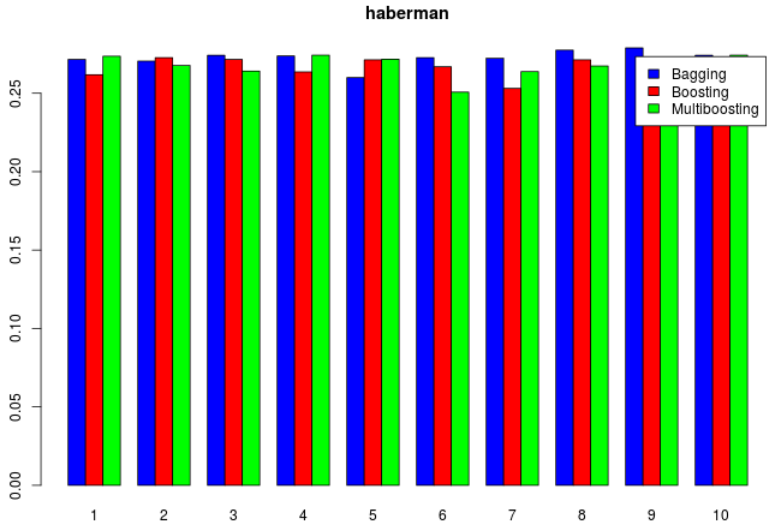

Figura 26: Valores de erro médio para *Multiboosting* quando aplicado à base "haberman".

5.2 Primeira Abordagem: Algoritmo Genético para Incremento de Diversidade

Nesta seção, vamos mostrar os resultados obtidos com os modelos de *ensemble* produzidos com a abordagem integradora proposta. Nas tabelas seguintes, o símbolo *B* refere-se ao mode-

los induzidos através de *Bagging* padrão, enquanto FS, $FS+R$ e $FS+R+HC$ se referem aos modelos induzidos nos finais da primeira, segunda, e terceira fases das abordagens propostas de incremento de diversidade, respectivamente.

A Tabela 3 apresenta os resultados de diversidade alcançados por diferentes configurações de *ensembles*, medida através de três medidas de diversidade. Nesta tabela, para todas as linhas, exceto os relacionados com a $FS+R+HC$, a média da diversidade $\pm$ desvio padrão foram calculados sobre 15 problemas de classificação, considerando conjuntos homogêneos induzidos separadamente para cada um dos 10 algoritmos de aprendizagem descritos. Na última linha de cada medida de diversidade, incluímos heterogeneidade nos modelos de *ensemble* e esta é a razão pela qual temos apenas um resultado para esta linha particular. Levando em conta estas observações, referindo-se a Tabela 3, deve-se interpretar que 0,8457, 0,6534 e 0,6351 denotam as médias de estatística-Q entregues por modelos de *ensembles* compostos de componentes de SMO e induzidos via *Bagging* padrão, no final da fase de FS, e no final da fase de $FS+R$, respectivamente. A mesma interpretação se aplica às outras métricas de diversidade.

Tabela 3: Níveis de diversidade para estatística-Q, Boa e Má Diversidades

Approaches	SMO	NB	IBk	RT	DT	RBF	J48	DS	OneR	PART
Q-statistic										
B	0,8457±0,2924	0,8983±0,2457	0,8614±0,2533	0,7650±0,2677	0,7106±0,2574	0,8364±0,2367	0,7278±0,2564	0,7020±0,4102	0,7200±0,3116	0,7068±0,2530
FS	0,6534±0,2484	0,7842±0,1574	0,5868±0,2142	0,6161±0,1907	0,6088±0,1645	0,6805±0,1505	0,5949±0,1877	0,6556±0,2075	0,5475±0,2170	0,5890±0,1936
FS + R	0,6351±0,2361	0,7519±0,2131	0,5636±0,2155	0,5508±0,2420	0,5616±0,2065	0,6281±0,2060	0,5320±0,2231	0,5194±0,3321	0,4647±0,2793	0,5345±0,2057
FS + R + HC	0,4944±0,1946									
Good Diversity										
B	0,0180±0,0163	0,0263±0,0557	0,0236±0,0208	0,0667±0,0770	0,0909±0,0879	0,0457±0,0617	0,0671±0,0749	0,0841±0,1294	0,0626±0,0635	0,0719±0,0737
FS	0,0810±0,0508	0,0809±0,0454	0,1022±0,0652	0,1085±0,0603	0,1077±0,0562	0,0987±0,0500	0,1093±0,0627	0,1192±0,0956	0,1024±0,0480	0,1108±0,0611
FS + R	0,0872±0,0515	0,0911±0,0560	0,1102±0,0658	0,1312±0,0783	0,1289±0,0742	0,1121±0,0634	0,1246±0,0761	0,1426±0,1192	0,1303±0,0604	0,1244±0,0723
FS + R + HC	0,1415±0,0718									
Bad Diversity										
B	0,0119±0,0121	0,0149±0,0206	0,0177±0,0164	0,0245±0,0235	0,0348±0,0289	0,0227±0,0246	0,0261±0,0249	0,0276±0,0326	0,0435±0,0443	0,0260±0,0230
FS	0,0309±0,0210	0,0333±0,0132	0,0329±0,0277	0,0375±0,0343	0,0366±0,0236	0,0315±0,0162	0,0335±0,0232	0,0358±0,0284	0,0593±0,0298	0,0332±0,0242
FS + R	0,0304±0,0225	0,0370±0,0232	0,0330±0,0277	0,0366±0,0305	0,0424±0,0279	0,0350±0,0248	0,0345±0,0237	0,0423±0,0229	0,0666±0,0414	0,0351±0,0264
FS + R + HC	0,0294±0,0193									

A fim de validar as observações de uma forma mais significativa, temos aplicado o teste estatístico de Friedman e o teste Nemenyi *post-hoc*, uma vez que estes testes não paramétricos são adequados para comparar o desempenho de diferentes algoritmos de aprendizagem quando aplicados a vários conjuntos de dados (para um debate aprofundado sobre esses testes estatísticos, consulte Demsar (2006)). Ambos os testes foram aplicados, considerando que todos os resultados da diversidade de uma vez (ou seja, os resultados produzidos por todos os tipos de conjuntos homogêneos/heterogêneos sobre todos os conjuntos de dados) ou estratificados por cada tipo de algoritmo de aprendizagem para os modelos homogêneos, como discutido acima.

Ao aplicar o teste de Friedman, foi possível perceber que as diferenças na diversidade apresentada pelos modelos de *ensemble* foram estatisticamente significativos (p-valor ≈ 0) para todas as três medidas de diversidade, rejeitando ao nível de 0,01 de significância a hipótese nula de equivalência.

Por outro lado, na tentativa de identificar forma de pares que os modelos diferem uns dos outros em termos de sua diversidade, o teste *post-hoc* Nemenyi foi realizado com um nível de significância de 0,05. Os resultados desses testes são apresentados nas Tabelas 4, 5 e 5.2. A aplicação do pós-teste apoia nossas observações acima, revelando que a combinação de duas ou mais técnicas para valorizar a diversidade é geralmente melhor do que usar uma técnica sozinha. É importante realçar, no entanto, que os níveis de diversidade alcançados pelos FS e $FS+R$ são compatíveis em todos os casos e em todas as medidas, a partir de um ponto de vista estatístico. Em termos quantitativos, afirma que:

- Para estatística-Q, $FS+R$ superou B em 70% dos casos, $FS+R+HC$ superou B em todos os casos, e $FS+R+HC$ superou FS em 70% dos casos;
- Para GD, $FS+R$ superou B em todos os casos, $FS+R+HC$ também superou B em todos os casos, e $FS+R+HC$ superou FS em 90% dos casos; e
- Para BD, $FS+R+H$ prevaleceu sobre B, FS e $FS+R$ em todos os casos.

Na sequência, apresentamos os resultados das taxas de erro entregues pelos modelos de *ensemble* de uma maneira como correlacionar esses resultados com os descritos acima para a diversidade, conforme apresentados nas Figuras 27 e 28. Nestas figuras, apresentamos também

Tabela 4: Resultados do teste Nemenyi para estatística-Q

Comparisons		SMO	NB	IBk	RT	DT	RBF	J48	DS	OneR	PART
FS	B	0,2532	0,3727	0,1244	0,9717	0,6451	0,2945	0,7350	0,5957	0,1701	0,9365
FS + R	B	0,1200	**0,0176**	**0,0115**	**0,0278**	0,2267	**0,0025**	**0,0041**	**0,0085**	**0,0013**	0,0616
FS + R + HC	B	**0,0001**	**0,0000**	**0,0002**	**0,0008**	**0,0005**	**0,0000**	**0,0000**	**0,0070**	**0,0016**	**0,0003**
FS + R	FS	0,9829	0,5517	0,8161	0,0888	0,8843	0,2945	0,0887	0,2237	0,3727	0,2266
FS + R + HC	FS	**0,0398**	**0,0008**	0,2267	**0,0044**	**0,0278**	**0,0024**	**0,0044**	0,1943	0,3728	**0,0041**
FS + R + HC	FS + R	0,1013	0,0617	0,7350	0,7350	0,1700	0,2944	0,7350	0,9999	1,0000	0,4595

Tabela 5: Resultados do teste Nemenyi para *Boa* Diversidade

Comparisons		SMO	NB	IBk	RT	DT	RBF	J48	DS	OneR	PART
FS	B	0,2531	0,2267	**0,0280**	0,4595	0,9364	0,6452	0,1700	0,1942	0,7350	0,4595
FS + R	B	**0,0064**	**0,0028**	**0,0000**	**0,0014**	**0,0110**	**0,0027**	**0,0001**	**0,0003**	**0,0115**	**0,0016**
FS + R + HC	B	**0,0000**	**0,0000**	**0,0000**	**0,0000**	**0,0004**	**0,0000**	**0,0000**	**0,0000**	**0,0003**	**0,0000**
FS + R	FS	0,4989	0,3727	0,1242	0,1246	0,0616	0,0888	0,1244	0,1674	0,1702	0,1244
FS + R + HC	FS	**0,0004**	**0,0013**	0,2945	**0,0043**	**0,0043**	**0,0002**	**0,0041**	**0,0499**	**0,0110**	**0,0041**
FS + R + HC	FS + R	**0,0482**	0,1243	0,9717	0,6452	0,8161	0,2945	0,6452	0,9557	0,7350	0,6452

Tabela 6: Resultados do teste Nemenyi para *Má* Diversidade

Comparisons		SMO	NB	IBk	RT	DT	RBF	J48	DS	OneR	PART
FS	B	0,1340	0,0114	0,0727	0,1032	0,2945	0,0888	0,3728	0,1437	0,4595	0,4119
FS + R	B	**0,0292**	**0,0419**	0,1942	0,1436	0,2945	**0,0617**	0,0886	0,1434	0,0612	0,4120
FS + R + HC	B	**0,0000**	**0,0000**	**0,0000**	**0,0000**	**0,0000**	**0,0000**	**0,0000**	**0,0000**	**0,0000**	**0,0000**
FS + R	FS	0,9332	0,9717	0,9714	0,9989	1,0000	0,9989	0,8843	1,0000	0,7349	1,0000
FS + R + HC	FS	**0,0014**	**0,0181**	**0,0086**	**0,0051**	**0,0027**	**0,0040**	**0,0009**	**0,0141**	**0,0005**	**0,0018**
FS + R + HC	FS + R	**0,0118**	**0,0042**	**0,0017**	**0,0031**	**0,0024**	**0,0068**	**0,0116**	**0,0136**	**0,0181**	**0,0016**

as taxas de erro produzidas por *Bagging* padrão para servir como base de comparação. Além disso, as Tabelas 7 e 8 trazer os resultados (p-valores) obtidos com a aplicação dos testes de Friedman e Nemenyi, respectivamente.

Tabela 7: Resultados do teste de Friedman para o erro médio de generalização

SMO	NB	IBk	RT	DT	RBF	J48	DS	OneR	PART
0,0090	**0,0002**	**0,0119**	**0,0258**	**0,0029**	0,0925	0,5709	**0,0001**	**0,0002**	0,1636

Tabela 8: Resultados do teste de Nemenyi para o erro médio de generalização

Comparisons		SMO	NB	IBk	RT	DT	DS	OneR
FS	B	**0,0451**	0,8843	0,1887	0,3662	0,9913	0,9711	0,9117
FS + R	B	0,1341	0,9913	**0,0130**	0,5928	0,7350	0,8129	0,9962
FS + R + HC	B	1,0000	**0,0041**	**0,0385**	0,6860	0,0617	**0,0013**	**0,0039**
FS + R	FS	0,9702	0,7350	0,7284	0,9829	0,8843	0,5457	0,8145
FS + R + HC	FS	**0,0449**	**0,0002**	0,9099	**0,0324**	**0,0278**	**0,0002**	**0,0003**
FS + R + HC	FS + R	0,1342	**0,0112**	0,9827	0,0854	**0,0025**	**0,0260**	**0,0087**

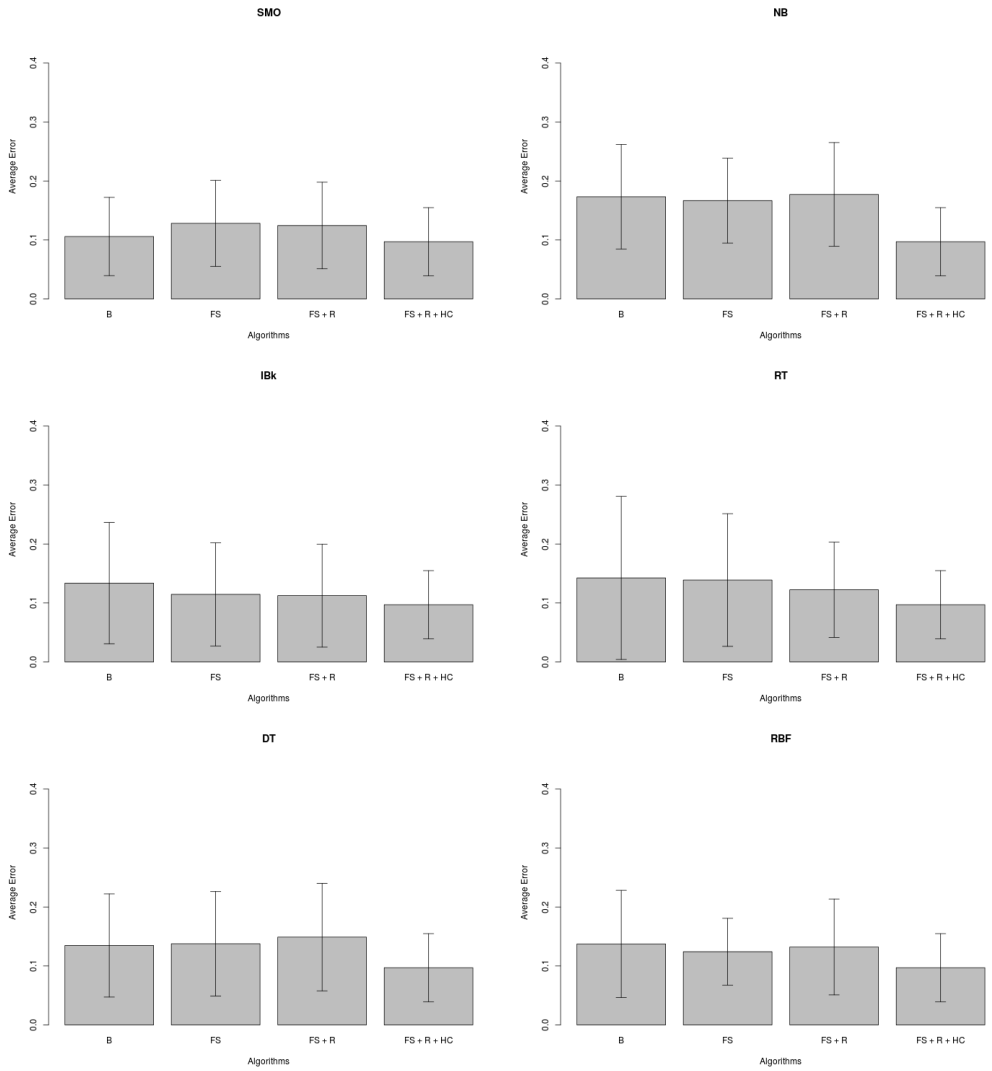

Figura 27: Erro médio de generalização produzido por cada estágio da abordagem integrada.

De forma geral, os resultados sugerem que um aumento progressivo na taxa de erro pode ocorrer em cada etapa da abordagem proposta, mas não em todas as circunstâncias. De fato, a aplicação do teste de Friedman revelou que existem diferenças significativas entre os modelos de conjunto para apenas 70% dos casos (SMO, NB, o IBk, RT, DT, DS, e Oner). Por sua vez, considerando-se a comparação aos pares alcançados pelo teste Nemenyi, $FS+R+HC$ superou B, FS e $FS+R$ em 40%, 60%, e 40% dos casos, respectivamente. Estes resultados demonstram novamente que a combinação de diferentes técnicas para promover a diversidade pode ser mais útil do que o uso de uma única técnica sozinha.

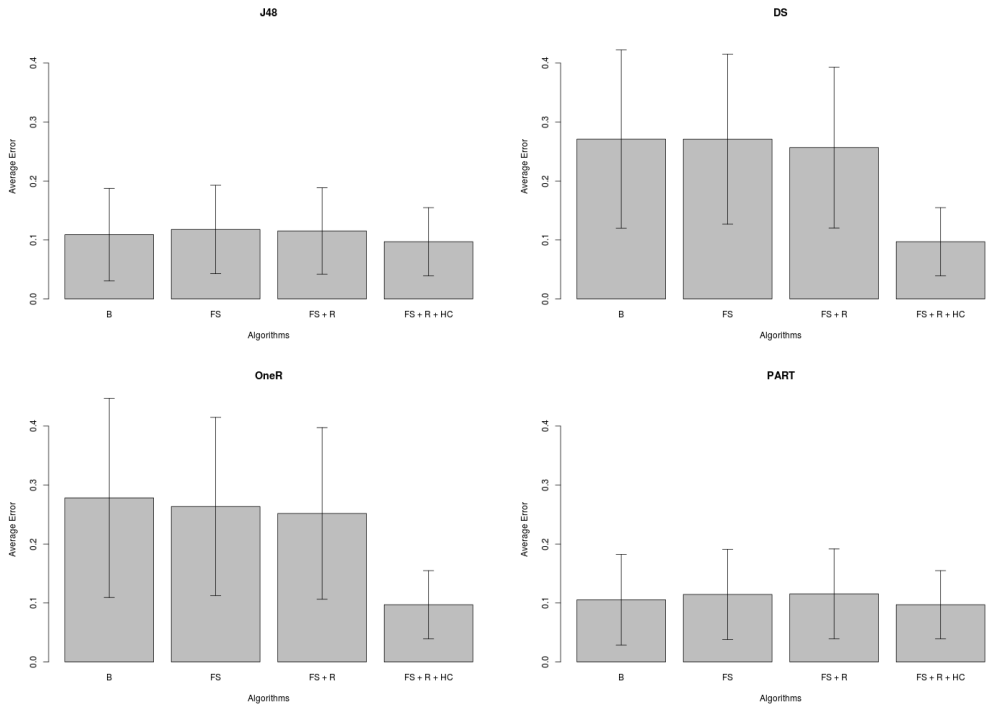

Figura 28: Erro médio de generalização produzido por cada estágio da abordagem integrada (continuação).

A fim de obter mais conhecimentos sobre o comportamento dos conjuntos de modelos produzidos por cada etapa da abordagem proposta, as Figuras 29 e 30 exibem alguns diagramas de diversidade de erros, que ajudam a visualizar a precisão da diversidade dos classificadores individuais, incluindo os modelos de *ensemble*. Os diagramas referem-se apenas aos conjuntos de dados "sick" e de "waveform", com foco nos algoritmos RN e RT para modelos homogêneos, mas eles são ilustrativos dos diagramas obtidos com o resto dos conjuntos de dados considerados. Para cada par de imagem, medimos sua precisão como a média das taxas de erro e a diversidade através da Equação (2.1), apresentada no Capítulo 2. Um diagrama de dispersão é então construído no qual cada ponto corresponde a um par de classificadores. Precisão e diversidade aumentam como os pontos se aproximam da origem. Esses diagramas são semelhantes aos κ diagramas adotadas no trabalho de Dietterich (2000b), com uma pequena modificação: Em vez de usar uma estatística de acordo com κ, foi utilizado a estatística-Q.

A partir destes gráficos, pode-se ver que os modelos formados por *Bagging* tendem a produzir nuvens mais compactas de pontos, o que indica que os classificadores são precisos, mas

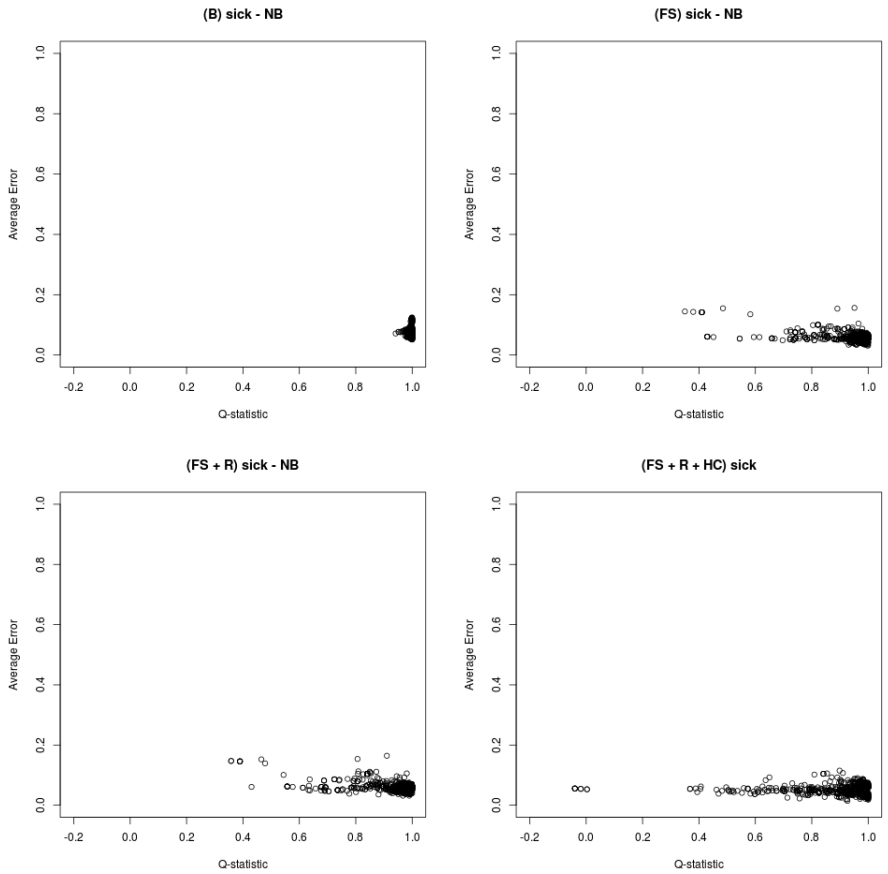

Figura 29: Diagrama de diversidade-erro para base “sick” usando *ensemble* com *Bagging* padrão (topo à esquerda), *ensemble* gerada com seleção de filtro (topo à direita), *ensemble* gerada com seleção de filtro mais reamostragem (base à esquerda) e *ensemble* gerada com seleção de filtro, reamostragem e componentes heterogêneos (base à direita).

não muito diversos, como testemunhado também por Dietterich (2000b). Por outro lado, como as etapas da abordagem proposta são progressivamente aplicadas, os diagramas resultantes tendem a se tornar mais dispersos, com os dois últimos estágios mostrando intervalos mais amplos de exatidão e graus de diversidade. É interessante observar que um comportamento semelhante foi observado por Dietterich (2000b) ao analisar o algoritmo *AdaBoost*, evidenciando a sua superioridade em relação ao *Bagging* e outra abordagem proposta pelo autor, conhecido como *Randomization*. Igualmente interessante é o fato de que, para o conjunto de dados “sick”, o aumento da diversidade do classificador obtido pelas duas últimas etapas não incorre em uma diminuição significativa na precisão dos classificadores, enquanto que para o conjunto de dados “waveform”, o mesmo comportamento não foi observado. Em ambos os casos, no entanto,

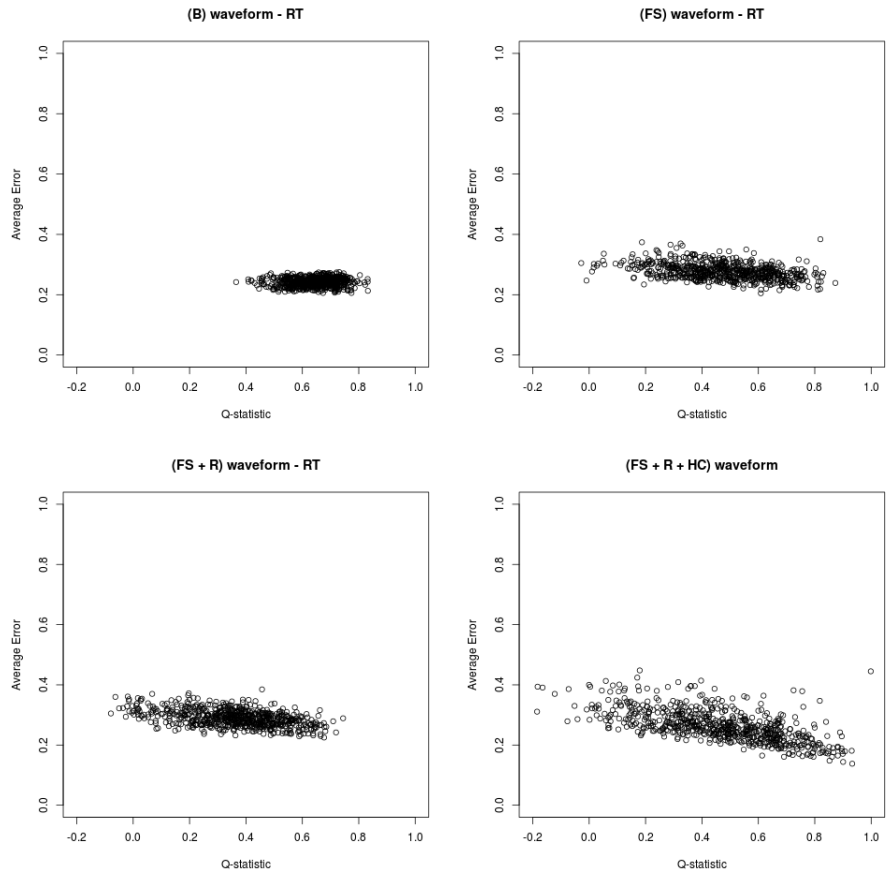

Figura 30: Diagrama de diversidade-erro para base “waveform” usando *ensemble* com *Bagging* padrão (topo à esquerda), *ensemble* gerada com seleção de filtro (topo à direita), *ensemble* gerada com seleção de filtro mais reamostragem (base à esquerda) e *ensemble* gerada com seleção de filtro, reamostragem e componentes heterogêneos (base à direita).

as taxas de erro produzido pela última etapa da abordagem foram significativamente melhores do que aquelas apresentadas pelas etapas anteriores e de *Bagging* padrão, como certificado pela aplicação dos testes estatísticos. Em geral, esse tipo de análise ajuda a entender melhor como o uso de mecanismos complementares de promoção da diversidade, quando feito corretamente, pode também resultar em aumentos em termos de generalização.

5.3 Segunda Abordagem: Algoritmo Genético com Função de Avaliação Adaptativa

Como já foi mencionado, usamos duas versões de um algoritmo genético para selecionar atributos para *ensembles*. A fim de analisar o desempenho destas técnicas, uma análise empírica é realizada. Nesta análise, avaliamos o efeito de usar essas técnicas de otimização na escolha de importantes elementos em *ensemble* (atributos). Nesta análise, vamos comparar o desempenho de ambas as versões, e com um método *Baggin* padrão, sem qualquer método de seleção de recurso. As Tabelas 9 e 10 apresentam as taxas de erro médio $\pm$ e desvio padrão para o algoritmo genético com abordagem filtro (original genético) e algoritmo genético com aptidão adaptativa (genético proposto), respectivamente. Em uma análise geral, podemos afirmar que o genético proposto tinha uma taxa de erro menor do que os outros dois métodos. Em relação ao *Bagging* padrão e original genético, eles têm taxas de erro semelhantes.

Tabela 9: Erro médio (%) e desvio padrão para o genético original

Base	RBF	J48	SMO	NB	IB*k*	RT	DS	OneR	PART	DT
Breast-cancer	5,29±3,46	4,22±1,48	3,52±2,78	6,86±4,21	3,69±1,76	5,10±2,55	9,14±4,12	9,32±3,42	**3,17±2,02**	4,22±2,78
Colic	16,85±7,29	**14,66±7,22**	17,37±8,15	18,47±7,16	19,30±5,79	15,47±6,71	17,10±5,66	16,84±6,93	16,02±7,12	16,01±5,59
Credit-a	18,26±7,49	15,07±5,43	16,66±5,81	24,20±6,11	19,13±4,73	15,07±3,56	16,23±4,03	16,95±6,52	**14,63±3,58**	14,78±3,73
Gaussian3	16,67±20,79	26,67±26,29	**0,00±0,00**	15,00±12,30	**0,00±0,00**	46,67±20,49	48,33±19,95	61,67±15,81	28,33±27,27	35,00±14,59
Heart-c	**16,14±8,40**	19,14±6,79	16,84±7,61	16,79±6,03	18,45±6,38	20,10±8,39	21,79±7,09	21,80±7,92	19,78±8,06	19,46±7,71
Image	8,62±1,64	2,77±0,84	10,82±3,57	23,20±3,23	2,68±1,13	3,77±1,23	44,81±8,74	29,83±3,38	**2,47±1,24**	6,97±1,33
Ionosphere	7,14±4,72	6,85±4,90	11,13±4,96	15,40±5,61	7,41±4,52	6,56±2,72	15,98±4,35	15,12±7,77	**6,28±4,00**	9,13±5,37
Protein	20,05±3,70	19,21±4,16	17,66±3,22	23,15±3,75	24,02±3,25	21,77±5,58	33,97±8,97	35,00±2,36	**16,97±4,70**	21,94±5,48
Segment	9,26±1,34	3,33±1,19	12,12±2,61	22,90±2,75	**2,86±0,92**	3,50±1,13	36,10±6,46	30,61±4,39	2,99±1,13	5,67±2,12
Sick	5,73±0,49	5,99±0,28	6,12±0,08	5,67±0,52	**4,72±1,33**	5,46±0,87	6,12±0,08	5,94±0,71	5,28±1,06	6,12±0,08
Sonar	22,57±5,90	19,69±5,16	21,17±7,55	29,31±4,60	**12,07±7,05**	22,10±6,73	27,38±4,35	35,59±8,87	18,26±5,38	25,38±10,38
Splice	10,35±1,93	8,43±1,30	28,66±5,34	10,36±2,76	25,29±1,97	**6,94±1,52**	46,20±3,16	26,44±5,92	8,52±2,17	9,40±2,63
Vote	8,29±2,76	8,06±3,85	**7,37±3,08**	10,84±4,71	7,83±3,15	7,38±3,79	9,23±4,69	8,97±4,41	7,38±2,90	**7,37±3,43**
Zoo	7,00±9,49	5,00±7,07	9,00±11,01	8,00±10,33	**4,00±5,16**	10,91±7,40	35,73±9,89	37,64±10,31	6,00±8,43	7,00±9,49

Tabela 10: Erro médio (%) e desvio padrão para o genético proposto

Base	RBF	J48	SMO	NB	IB*k*	RT	DS	OneR	PART	DT
Breast-cancer	5,46±3,07	3,51±1,43	**2,81±2,07**	6,69±3,72	3,69±1,29	5,97±2,51	9,84±3,33	9,49±3,34	3,52±2,04	4,58±2,67
Colic	14,95±6,65	**14,66±5,84**	16,29±7,71	19,56±7,07	17,10±5,94	16,30±7,19	16,55±10,21	19,54±8,38	16,00±6,99	16,55±7,21
Credit-a	14,64±4,35	13,33±2,72	14,64±3,71	18,40±3,42	15,36±2,29	14,35±4,07	14,34±1,74	14,92±2,65	**13,04±4,04**	13,62±2,66
Gaussian3	20,00±17,21	36,67±13,15	**0,00±0,00**	20,00±23,31	**0,00±0,00**	41,67±14,16	48,33±22,84	63,33±15,32	33,33±24,85	30,00±18,92
Heart-c	**14,83±7,00**	20,09±7,89	17,14±7,76	17,74±9,30	21,39±7,44	17,14±5,70	19,43±10,31	21,11±7,44	17,44±7,82	18,76±8,15
Image	8,49±1,60	2,60±0,86	8,05±1,46	18,18±0,98	2,60±0,96	3,85±1,16	31,95±1,79	30,26±4,98	**2,21±1,09**	6,10±1,87
Ionosphere	7,99±4,83	8,28±4,76	9,98±4,33	12,84±5,93	6,85±2,77	7,13±3,88	13,98±5,49	14,83±6,73	**6,56±3,83**	9,42±7,51
Protein	19,54±2,41	18,68±2,58	**17,50±2,95**	21,95±4,85	23,33±3,31	20,39±3,89	29,50±4,67	32,92±3,97	17,82±3,93	20,65±4,53
Segment	8,83±0,96	2,81±1,18	7,92±1,10	17,06±1,36	**2,25±0,61**	2,90±0,74	28,40±3,35	29,13±4,16	**2,25±0,73**	5,76±1,39
Sick	3,47±0,71	**1,57±0,58**	6,12±0,08	3,53±0,64	2,39±0,84	1,70±0,62	3,47±0,83	3,63±0,76	1,67±0,54	2,62±0,71
Sonar	20,19±7,16	20,69±6,43	20,69±6,43	29,36±6,67	**13,48±6,73**	27,40±8,99	27,88±8,07	34,59±8,23	15,90±7,60	22,10±8,12
Splice	6,61±1,81	5,14±0,88	17,40±1,78	6,98±2,03	20,71±2,21	**4,43±1,12**	28,81±2,01	18,75±2,58	4,93±0,52	5,85±1,96
Vote	4,38±2,98	3,70±2,25	4,38±2,76	7,61±3,13	4,61±1,90	4,60±2,86	4,60±2,41	4,37±2,74	**3,00±2,20**	5,06±2,37
Zoo	9,00±11,01	**3,00±4,83**	4,00±6,99	4,00±5,16	4,00±5,16	7,82±7,60	34,73±5,62	17,82±12,27	5,91±8,39	6,00±6,99

A fim de avaliar o ganho de desempenho das abordagens apresentadas, os resultados obtidos de erro médio para todos os algoritmos de aprendizagem quando aplicados sobre diferentes problemas de classificação, foram comparados e mostrado nas Tabelas 11, 12 e 13. Esses resultados referem-se aos *p*-valores para aplicação do teste estatístico *t*-teste sobre as abordagens. Nessas tabelas, os valores inferiores a 0,05 indicam que a hipótese nula H_0 pode se rejeitada, isto significa que as diferenças entre as médias são significativamente diferentes. A fim de facilitar a visualização dos dados, quando a primeira abordagem for melhor executada que a segunda, marcamos os valores em negrito, no contrário, nós marcamos os valores em sublinhado. No caso, não podemos rejeitar H_0, ou seja, as médias são equivalentes, apresentamos os dados sem aplicar qualquer formatação.

Tabela 11: Valores de *t*-teste comparando genético proposto com genético original

Base	RBF	J48	SMO	NB	IB*k*	RT	DS	OneR	PART	DT
Breast-cancer	0,763	**0,037**	0,308	0,834	0,996	0,140	0,508	0,841	0,594	0,552
Colic	0,152	0,996	0,440	0,219	0,339	0,639	0,809	0,195	0,991	0,728
Credit-a	0,061	0,254	0,261	**0,025**	**0,020**	0,427	0,140	0,261	**0,048**	0,344
Gaussian3	0,591	0,168	–	0,344	–	0,343	1,000	0,780	0,560	0,081
Heart-c	0,221	0,621	0,839	0,574	0,030	0,083	0,322	0,790	0,065	0,823
Image	0,757	0,623	**0,042**	**0,002**	0,802	0,834	**0,002**	0,745	0,534	**0,021**
Ionosphere	0,280	0,096	0,269	**0,030**	0,622	0,512	0,258	0,822	0,728	0,868
Protein	0,594	0,565	0,794	0,227	0,267	0,236	0,063	0,150	0,567	0,304
Segment	0,266	0,223	**0,001**	**0,000**	0,056	0,169	**0,007**	0,251	**0,049**	0,865
Sick	**0,000**	**0,000**	–	**0,000**	**0,000**	**0,000**	**0,000**	**0,000**	**0,000**	**0,000**
Sonar	0,301	0,681	0,679	0,982	0,477	0,073	0,826	0,544	0,492	0,220
Splice	**0,000**	**0,000**	**0,000**	**0,001**	**0,001**	**0,000**	**0,000**	**0,001**	**0,000**	**0,002**
Vote	**0,006**	**0,009**	**0,040**	**0,013**	**0,007**	0,088	**0,013**	**0,013**	**0,002**	0,063
Zoo	0,168	0,343	0,052	0,104	1,000	0,323	0,798	**0,002**	0,950	0,726

Finalmente, a taxa de erro dos *ensembles* também foram avaliadas em termos de três medidas de diversidade: estatística-*Q*, *Boa* e *Má* diversidades (BROWN; KUNCHEVA, 2010), conforme Tabela 14. É importante ressaltar que, de acordo com Brown e Kuncheva (2010), maior valor de *Boa* diversidade reduz o erro global do *ensemble*, ao passo que valores maiores de *Má* diversidade levam a um aumento desse erro. Em outras palavras, é desejável grandes valores para *Boa* diversidade e valores pequenos de *Má* diversidade. Para estatística-*Q*, o valor ideal são aqueles mais próximo de zero.

Na análise da Tabela 14, podemos ver que os valores para estatística-*Q* foram menores para as versões do genético original e do genético proposto. Isso significa que a utilização de seleção de características aumenta a diversidade do *ensemble*, com base na medida de estatística-*Q*. Em

Tabela 12: Valores de *t*-teste comparando genético proposto com *Bagging*

Base	RBF	J48	SMO	NB	IB*k*	RT	DS	OneR	PART	DT
Breast-cancer	0,754	0,297	0,341	0,510	0,381	0,449	0,112	0,074	0,660	0,341
Colic	**0,012**	0,736	0,397	0,239	0,147	0,568	0,269	0,168	0,138	0,862
Credit-a	**0,000**	0,213	0,706	**0,005**	**0,009**	0,605	0,343	0,394	0,066	0,229
Gaussian3	**0,024**	0,081	–	0,068	0,343	0,001	0,089	0,081	0,213	0,054
Heart-c	0,428	0,868	0,991	0,773	0,275	0,820	**0,028**	**0,019**	0,398	0,859
Image	**0,001**	0,111	0,001	**0,006**	0,323	0,033	**0,000**	**0,012**	0,679	0,008
Ionosphere	0,393	0,376	0,113	**0,002**	**0,002**	0,275	0,089	0,555	**0,008**	0,482
Protein	0,210	0,551	0,502	**0,005**	0,267	0,048	**0,000**	0,570	0,612	**0,021**
Segment	**0,000**	0,393	0,117	**0,001**	**0,025**	0,066	**0,000**	**0,006**	0,140	0,140
Sick	0,235	0,049	0,192	**0,000**	**0,000**	0,008	0,850	0,343	0,023	0,033
Sonar	0,347	0,764	0,970	0,741	0,771	0,015	0,291	0,289	0,616	0,225
Splice	0,225	0,485	0,561	0,531	**0,000**	**0,000**	**0,000**	**0,000**	0,750	0,272
Vote	0,820	0,189	0,051	**0,010**	**0,009**	0,081	0,343	–	0,168	0,191
Zoo	0,081	0,288	–	0,678	1,000	**0,000**	0,107	**0,000**	–	0,052

Tabela 13: Valores de *t*-teste comparando genético original com *Bagging*

Base	RBF	J48	SMO	NB	IB*k*	RT	DS	OneR	PART	DT
Breast-cancer	0.344	0.792	0.193	0.195	0.383	0.829	0.499	0.203	0.400	0.466
Colic	0,128	0,761	0,859	0,105	0,775	0,996	0,589	0,508	0,317	0,846
Credit-a	0,376	0,542	0,325	0,145	0,828	0,775	0,181	0,253	1,000	0,725
Gaussian3	**0,018**	0,541	–	**0,009**	0,343	0,001	0,132	0,273	0,735	0,005
Heart-c	0,733	0,696	0,859	0,764	**0,033**	0,015	0,080	0,080	0,360	0,848
Image	**0,000**	0,167	0,013	0,014	0,700	0,293	0,649	**0,002**	0,837	0,000
Ionosphere	0,141	0,109	0,625	**0,003**	**0,001**	0,259	0,276	0,501	**0,032**	0,273
Protein	0,450	0,854	0,363	0,862	1,000	0,029	**0,004**	0,383	0,350	0,212
Segment	**0,002**	0,074	0,000	0,008	0,624	0,816	**0,032**	**0,024**	0,032	0,403
Sick	0,000	0,000	0,192	**0,030**	0,009	0,000	0,000	0,000	0,000	0,000
Sonar	0,081	0,269	0,705	0,715	0,594	0,492	0,642	0,120	0,871	0,747
Splice	0,001	0,000	0,000	0,001	**0,000**	0,364	0,000	**0,002**	0,001	0,012
Vote	0,013	0,009	0,003	0,095	0,717	0,030	0,011	0,013	0,002	0,046
Zoo	0,343	0,615	0,052	0,193	1,000	**0,000**	0,405	**0,001**	0,950	0,075

relação a *Boa* e *Má* diversidades, os algoritmos genéticos, apresentaram maior *Boa* diversidade e menor *Má* diversidades. Portanto, a utilização de seleção de características também aumentou a diversidade dos *ensemble*, com base na *Boa* e *Má* diversidades.

5.4 Terceira Abordagem: Recomendação via Meta-aprendizado

Nesta seção discutiremos os resultados para recomendação dos parâmetros de controle em *ensembles* de classificadores. De forma que as discursões sejam mais didáticas, organizamos os resultados nas duas subseções 5.4.1 e 5.4.2, que descrevem, respectivamente, os resultados para a

Tabela 14: Medidas de diversidade

Bases	*Bagging*			Genético Original			Genético Proposto		
	Estat.-*Q*	*Boa*	*Má*	Estat.-*Q*	*Boa*	*Má*	Estat.-*Q*	*Boa*	*Má*
breast-cancer	0,815	0,025	**0,011**	0,765	**0,031**	0,013	**0,764**	0,025	0,012
colic	0,963	0,029	**0,015**	**0,62**	**0,11**	0,041	0,689	0,095	0,038
credit-a	0,954	0,027	**0,018**	**0,536**	**0,125**	0,051	0,594	0,108	0,041
gaussian3	**0,043**	**0,224**	**0,07**	0,299	0,156	0,071	0,3	0,143	0,073
heart-c	0,768	0,057	**0,039**	**0,43**	**0,131**	0,059	0,519	0,11	0,053
image	0,826	0,062	0,019	**0,708**	**0,122**	0,02	0,742	0,112	**0,017**
ionosphere	0,84	0,033	**0,019**	0,757	0,051	0,02	**0,752**	**0,055**	0,02
protein	0,882	0,068	**0,03**	0,81	0,1	0,037	**0,801**	**0,107**	0,036
segment	0,826	0,062	0,019	**0,708**	**0,122**	0,02	0,742	0,112	**0,017**
sick	0,962	0,006	**0,004**	**0,915**	0,011	0,015	0,921	**0,017**	0,006
sonar	0,606	**0,102**	**0,056**	**0,604**	0,099	0,059	0,626	0,096	0,057
splice	0,87	0,038	**0,021**	**0,422**	**0,175**	0,045	0,508	0,155	0,031
vote	0,879	0,01	**0,004**	**0,655**	**0,051**	0,025	0,693	0,043	0,01
zoo	**0,61**	0,02	0,014	0,683	**0,122**	0,016	0,646	0,113	**0,01**
Média	0,775	0,054	**0,024**	**0,637**	**0,100**	0,035	0,664	0,092	0,030

abordagem de meta-aprendizado tradicional e a abordagem para meta-aprendizado multirrótulo.

5.4.1 Meta-aprendizado tradicional

Baseado nos resultados discutidos na seção anterior, foi observado que não existe um padrão de comportamento referente aos parâmetros de controle em *ensembles* (arquitetura e componentes). Ou seja, em casos onde se esperava que um método apresentasse um melhor desempenho, isso acabou não acontecendo, e em alguma casos o oposto aconteceu. Pensando nisso, a ideia de propor uma nova abordagem para configuração automática em *ensemble* surgiu, ao mesmo tempo que apresenta uma alternativa diferente daquelas apresentadas que exploram os métodos bio-inspirados, tal como algoritmo genético (CANUTO; NASCIMENTO, 2012; NASCIMENTO *et al.*, 2011).

O principal objetivo nesta investigação é aprender sobre os diferentes desempenhos de diferentes *ensembles* quando aplicados a problemas de classificação de variada complexidade, e então construir meta-aprendizes que serão utilizados para recomendar uma possível melhor configuração de arquitetura e componentes base, para um novo problema apresentado. A Tabela 15 apresenta os resultados (erro médio $\pm$ desvio padrão) dos meta-aprendizes (MLP, J48, SMO e IB*k*) para recomendação de arquitetura, enquanto que a Tabela 16 apresenta os valores de erro

médio ± desvio padrão da recomendação dos componentes homogêneos para *Bagging*, *Boosting* e *Multiboosting* individualmente.

Tabela 15: Valores de erro médio e desvio padrão dos meta-aprendizes para recomendação de arquitetura de *ensemble*

Meta-aprendizes			
MLP	J48	SMO	IBk
0,271 ± 0.038	0.281 ± 0.044	0.357 ± 0.026	0.249 ± 0.040

Tabela 16: Valores de erro médio e desvio padrão dos meta-aprendizes para recomendação dos componentes homogêneos para *Bagging*, *Boosting* e *Multiboosting*

	Meta-aprendizes			
	MLP	J48	SMO	IBk
Bagging	0,145 ± 0,026	0,153 ± 0,031	0,277 ± 0,008	0,237 ± 0,035
Boosting	0,145 ± 0,027	0,139 ± 0,032	0,274 ± 0,007	0,258 ± 0,030
Multiboosting	0,133 ± 0,027	0,136 ± 0,029	0,279 ± 0,008	0,223 ± 0,033

Em uma perspectiva geral, podemos observar que os resultados obtidos pelos meta-aprendizes são promissores. Foram obtidos erros médios menores que 30% em todas as recomendações, tanto de arquitetura quanto dos componentes base. Já para a recomendação dos componentes base, o erro médio de recomendação foi menor que 15% em todos os casos.

Esses resultados são interessantes porque, vale mais uma vez reforçar, melhores desempenhos podem ser alcançados através de estudos mais avançados em meta-caractéristicas exclusivas para *ensembles*, visto que, as utilizadas nesta pesquisa são as mesmas propostas no projeto Stalog. Os resultados também corroboram com a tese de que meta-aprendizado pode ser uma ferramente poderosa a ser explorada em ferramentas de recomendação de algoritmos, e mais especificamente, é claro, em *ensembles* de classificadores.

5.4.2 Meta-aprendizado Multirrótulo

Na abordagem de recomendação dos parâmetros de controle de *ensembles* utilizando meta-aprendizado tradicional, percebemos que em alguns casos, quando não foi recomendada a melhor opção esperada, ainda a recomendação seria satisfatória, e nesse caso, não deveria ser contabilizada como um erro de recomendação. Portanto, o erro médio, talvez, não fosse a melhor alterna-

tiva para se avaliar efetivamente quão boa seria a solução. Desta forma, propomos uma abordagem inédita para recomendação dos parâmetros de controle de *ensembles* via meta-aprendizagem multirrótulo. Com essa nova abordagem, torna-se possível, tanto a recomendação de arquitetura para *Bagging*, *Boosting* ou *Multiboosting*, como os componentes base para RBF, J48, SMO, NB ou IB*k*, nesse caso, podendo eles serem homogêneos ou heterogêneos, até então limitada na abordagem tradicional.

Homogêneo

Para avaliação de recomendação multirrótulo, utilizamos de outras métricas, todas elas citadas no Capítulo 2. Na Tabela 17, são apresentados os valores de *Hamming-loss*, *F-Measure*, Macro-*averaged F-Measue*, Micro-*averaged F-Measure* e *Ranking-loss* para os meta-aprendizes para recomendação multirrótulo de arquitetura em *ensemble*. As Tabelas 18, 19 e 20 apresentam os resultados alcançados pelos meta-aprendizes para recomendação de componentes base homogêneos para *Bagging*, *Boosting* e *Multiboosting*.

Analisando os valores obtidos pela aplicação da métrica *Hamming-loss*, o maior valor obtido foi 0,2930, logo, no geral a abordagem apresentou resultados promissores, assumindo-se que o valor ótimo é zero. Já para as métricas, *F-Measure*, Macro-*averaged F-Measue* e Micro-*averaged F-Measure*, os menores valores foram 0,8882, 0,8918 e 0,6191, respectivamente. De forma geral os resultados são promissores, assumindo-se que o valor ótimo é um. E para a métrica *Ranking-loss*, o maior valor obtido foi 0,3735, e mais uma vez os resultados foram promissores, assumindo-se que o valor ótimo é zero. De forma geral, os resultados vêm a corroborar com a tese de que meta-aprendizado é uma ferramenta poderosa para recomendação automática dos parâmetros de controles de *ensembles* de classificadores.

Tabela 17: Recomendação multirrótulo para tipos de arquiteturas

Categorias	Métricas	Algoritmos Multirrótulos	
		RA*k*EL	ML-*k*NN
Exemplo	*Hamming-loss*	$0,1368 \pm 0,0559$	$0,2930 \pm 0,0702$
	F-Measure	–	–
Rótulo	Macro-*averaged F-Measure*	–	–
	Micro-*averaged F-Measure*	$0,8144 \pm 0,0705$	$0,6191 \pm 0,0888$
Ranking	*Ranking-loss*	$0,1484 \pm 0,0670$	$0,2270 \pm 0,0791$

Tabela 18: Recomendação multirrótulo para tipos de componentes homogêneo para *Bagging*

Categorias	Métricas	Algoritmos Multirrótulos	
		RA*k*EL	ML-*k*NN
Exemplo	*Hamming-loss*	$0,1358 \pm 0,0412$	$0,1400 \pm 0,0391$
	F-Measure	$0,9116 \pm 0,0304$	$0,9164 \pm 0,0260$
Rótulo	Macro-*averaged F-Measure*	$0,9206 \pm 0,0271$	$0,9216 \pm 0,0240$
	Micro-*averaged F-Measure*	$0,9236 \pm 0,0251$	$0,9238 \pm 0,0229$
Ranking	*Ranking-loss*	$0,3238 \pm 0,1110$	$0,3735 \pm 0,1294$

Tabela 19: Recomendação multirrótulo para tipos de componentes homogêneos para *Boosting*

Categorias	Métricas	Algoritmos Multirrótulos	
		RA*k*EL	ML-*k*NN
Exemplo	*Hamming-loss*	$0,0242 \pm 0,0156$	$0,0232 \pm 0,0123$
	F-Measure	$0,9865 \pm 0,0087$	$0,9871 \pm 0,0068$
Rótulo	Macro-*averaged F-Measure*	$0,9871 \pm 0,0085$	$0,9877 \pm 0,0068$
	Micro-*averaged F-Measure*	$0,9877 \pm 0,0080$	$0,9882 \pm 0,0063$
Ranking	*Ranking-loss*	–	–

Tabela 20: Recomendação multirrótulo para tipos de componentes homogêneos para *Multiboosting*

Categorias	Métricas	Algoritmos Multirrótulos	
		RA*k*EL	ML-*k*NN
Exemplo	*Hamming-loss*	$0,1663 \pm 0,0215$	$0,1642 \pm 0,0267$
	F-Measure	$0,8882 \pm 0,0195$	$0,8958 \pm 0,0184$
Rótulo	Macro-*averaged F-Measure*	$0,8918 \pm 0,0196$	$0,9001 \pm 0,0208$
	Micro-*averaged F-Measure*	$0,9023 \pm 0,0159$	$0,9073 \pm 0,0169$
Ranking	*Ranking-loss*	$0,2151 \pm 0,0399$	$0,2805 \pm 0,0798$

Heterogêneo

A proposta de recomendação multirrótulo foi ampliada para recomendar componentes heterogêneos para *Bagging*, *Boosting* e *Multiboosting*. As Tabelas 21, 22 e 23 apresentam os resultados das métricas *Hamming-loss*, *F-Measure*, Macro-*averaged F-Measure*, Micro-*averaged F-Measure* e *Ranking-loss*. O maior valor alcançado pela métrica *Hamming-loss* foi 0,2030, mais uma vez, apresentaram resultados promissores. E para as métricas, *F-Measure*, Macro-*averaged F-Measure* e Micro-*averaged F-Measure*, o menor valor foi 0,8835, 0,8767 e 0,8861, respectivamente, os resultados também são promissores. Para a métrica *Ranking-loss*, 0,3793, e por fim os resultados também se mostraram promissores. De forma geral, tais resultados vêm a corroborar

com a tese de que meta-aprendizado é uma ferramenta poderosa para recomendação automática de componentes base heterogêneos para *ensembles* de classificadores via *Bagging*, *Boosting* e *Multiboosting*.

Tabela 21: Recomendação multirrótulo para tipos de componentes heterogêneos para *Bagging*

Categorias	Métricas	Algoritmos Multirrótulos	
		RA*k*EL	ML-*k*NN
Exemplo	*Hamming-loss*	$0,1700 \pm 0,0179$	$0,1790 \pm 0,0411$
	F-Measure	$0,8954 \pm 0,0131$	$0,8934 \pm 0,0257$
Rótulo	Macro-*averaged F-Measure*	$0,8950 \pm 0,0130$	$0,8971 \pm 0,0258$
	Micro-*averaged F-Measure*	$0,9018 \pm 0,0123$	$0,9007 \pm 0,0245$
Ranking	*Ranking-loss*	$0,3793 \pm 0,0827$	$0,3365 \pm 0,0902$

Tabela 22: Recomendação multirrótulo para tipos de componentes heterogêneos para *Boosting*

Categorias	Métricas	Algoritmos Multirrótulos	
		RA*k*EL	ML-*k*NN
Exemplo	*Hamming-loss*	$0,1690 \pm 0,0437$	$0,2030 \pm 0,0329$
	F-Measure	$0,8915 \pm 0,0266$	$0,8787 \pm 0,0207$
Rótulo	Macro-*averaged F-Measure*	$0,8868 \pm 0,0303$	$0,8767 \pm 0,0225$
	Micro-*averaged F-Measure*	$0,9004 \pm 0,0260$	$0,8861 \pm 0,0203$
Ranking	*Ranking-loss*	$0,3699 \pm 0,0676$	$0,2929 \pm 0,0684$

Tabela 23: Recomendação multirrótulo para tipos de componentes heterogêneos para *Multiboosting*

Categorias	Métricas	Algoritmos Multirrótulos	
		RA*k*EL	ML-*k*NN
Exemplo	*Hamming-loss*	$0,1720 \pm 0,0399$	$0,1950 \pm 0,0341$
	F-Measure	$0,8889 \pm 0,0271$	$0,8835 \pm 0,0218$
Rótulo	Macro-*averaged F-Measure*	$0,8910 \pm 0,0255$	$0,8871 \pm 0,0212$
	Micro-*averaged F-Measure*	$0,8983 \pm 0,0253$	$0,8912 \pm 0,0211$
Ranking	*Ranking-loss*	$0,3274 \pm 0,1020$	$0,3147 \pm 0,0816$

5.5 Síntese do Capítulo

Os resultados para a abordagem de Incremento da diversidade confirmam que, a combinação de diferentes técnicas de diversidade em um único arcabouço conceitual incorre em ganhos

de precisão para os modelos de *ensemble*. Foram então combinadas as técnicas de diversidade: seleção de atributo, reamostragem aleatória dos dados e recrutamento dos componentes heterogêneos. Para cada solução, foi então avaliado o ganho de diversidade quando uma nova técnica é então inserida ao modelo, e ao passo que conseguimos melhorar o ganho de diversidade, obtivemos também uma diminuição do erro médio.

Os resultados para a abordagem de Algoritmo Genético com função de avaliação adaptativa também se mostraram promissores. A ideia é combinar as técnicas de *filtro* e *wrapper* para seleção de diferentes espaços de atributos de um conjunto de treinamento a serem apresentados a diferentes componentes de um *ensemble*. Percebemos que não só conseguimos ganho de precisão do modelo de *ensemble* resultante proporcionada pela técnica de *wrapper*, ao passo que conseguimos uma diminuição no custo computacional referente a tempo, permitida pela técnica *filtro*.

Os resultados para Meta-aprendizado confirmam que, tanto a abordagem de meta-aprendizado tradicional e a multirrótulo são promissoras no sentido de auxiliarem o projetista de sistema na construção de *ensembles* de classificadores. Embora todos os valores de avaliação não apresentassem o valor ótimo possível, isso não compromete o bom desempenho da proposta, visto que, melhores resultados podem ser alcançados, levando-se em consideração a escolha de meta-características mais adequadas para *ensembles*. Outro fator é que essa abordagem recomenda o modelo de *ensembles* (arquitetura e tipos de componentes) mais adequado para um novo problema de classificação de padrões, e uma vez escolhido o modelo, o mesmo passará por um processo de treinamento que permite, se necessário, o refinamento dos parâmetros interno dos componentes manualmente, ou a partir de outras técnicas automáticas.

Esta tese é concluída no Capítulo 6, a seguir, onde são levantadas as considerações finais, as limitações da pesquisa e os trabalhos futuros que poderão ser realizados a partir das contribuições aqui apresentadas.

Capítulo 6

Conclusões

Neste documento foram apresentados os conceitos teóricos necessários para entendimento do tema em estudo. Como se pôde observar em trabalhos correlatos, muitos são os esforços que vêm surgindo nos últimos anos na tentativa de apresentar uma solução capaz de construir e configurar automaticamente *ensembles* de classificadores. No que se refere aos resultados preliminares, constatou-se que mesmo conhecendo as características dos algoritmos de classificação, bem como a natureza dos dados que eles serão aplicados, ainda é muito difícil para o projetista de sistema prever qual será o melhor modelo a ser utilizado, restando, na maioria das vezes, uma realização de experimentos empíricos para auxiliar na escolha do melhor modelo. Desta forma, propomos três inéditas abordagens para construção de *ensembles* de forma eficiente: uma primeira que utiliza diferentes técnicas de diversidade em *ensembles*, a segunda que propõe uma nova técnica de avaliação adaptativa em algoritmos genéticos e uma terceira que utiliza meta-aprendizado para recomendação automática dos parâmetros de controle em *ensemble*.

Para a primeira abordagem, foi realizada uma extensa análise do impacto da combinação de técnicas complementares, de forma sequencial, para aumentar a diversidade em *ensembles* de classificadores. Para o melhor de nosso conhecimento, nenhum estudo detalhado como este foi realizado até agora na literatura. No geral, os resultados obtidos em uma série de problemas de

classificação conhecidos indicam que a estratégia de integração proposta pode de fato ser mais conveniente do que empregar estratégias individuais isoladamente, geralmente levando a ganhos em termos de diversidade e, às vezes, também em termos de erro de generalização.

Para a segunda abordagem, propomos um método evolutivo de seleção de características de uma base para *ensemble*. A principal contribuição da proposta é o uso da uma função híbrida de aptidão adaptativa em algoritmos genéticos. A ideia é tirar proveito das abordagens de seleção (*filtro* e *wrapper*). A fim de avaliar os benefícios do método proposto, uma análise empírica foi conduzida. Nesta análise, o método proposto foi comparado com um algoritmo genético utilizando apenas uma avaliação baseada em *filtro* e um *ensemble* padrão sem seleção de características. De forma geral, os resultados apontam para um ganho de generalização por parte da nova abordagem, bem como, ganhos computacionais de desempenho.

Para a terceira abordagem, um estudo de investigação foi então conduzido, na tentativa de se criar a abordagem de recomendação utilizando meta-aprendizado. De forma mais didática, organizamos a abordagem em dois tipos: Recomendação Tradicional (i) e Recomendação Multi-objetivo (ii). No primeiro tipo (i), aplicamos os conceitos de meta-aprendizado como está sendo utilizando na literatura para recomendação de algoritmos de aprendizagem. No nosso caso, recomendamos os tipos de arquitetura baseada em reamostragem dos dados e os tipos de componente base. Já no segundo tipo (ii), fazemos uma abordagem flexível, por assim dizer, do primeiro tipo, utilizando não apenas uma solução, mas um conjunto de soluções possíveis, via meta-aprendizado multirrótulo. De forma geral, os resultados alcançados corroboram com a Tese de que meta-aprendizado é uma poderosa ferramenta na configuração automática de *ensembles* de classificadores e que avanços nessa linha de pesquisa, contribuirá ainda mais para o surgimento de melhorias da abordagem.

6.1 Limitações da Proposta

Dentre as contribuições apresentadas nesta tese, citamos também, algumas limitações identificadas durante o fechamento do documento, tais como:

- A pesquisa não faz uma análise mais aprofundada no que se refere à relação meta-características com o processo de recomendação das arquiteturas e componentes base em *ensemble*;

- Os experimentos empíricos foram realizados com bases relativamente pequenas;
- A abordagem de recomendação para parâmetros de controle em *ensembles* trata apenas de arquitetura e tipos de componentes base, deixando de fora os métodos de combinação; e
- A realização da pesquisa utiliza apenas três arquitetura e 10 tipos de componentes base.

6.2 Trabalhos Futuros

A Tese aqui apresentada deixa como contribuição implícita uma variedade de alternativas de trabalhos futuros que podem ser realizados, podendo citar:

- Realização de um estudo aprofundado sobre meta-características que sejam mais representativas aos problemas de recomendação em *ensemble*;
- Ampliar a aplicação de meta-aprendizado na recomendação de diferentes arquitetura em *ensemble*;
- Realizar um estudo para avaliar os benefícios de se aplicar meta-aprendizado para recomendação de diferentes métodos de combinação;
- Realizar um estudo para avaliar os benefícios de se aplicar meta-aprendizado para recomendação de pesos para cada componente individualmente;
- Realizar um estudo para avaliar os benefícios de se aplicar meta-aprendizado para recomendação do número de componente combinados; e
- Realizar um estudo dos benefícios de se aplicar meta-aprendizado para recomendação de parâmetro de controle de *ensemble* em problemas de regressão.

Referências Bibliográficas

AHA, D. W. *Lazy Learning*. [S.l.]: Kluwer Academic, 1997.

AHA, D. W.; KIBLER, D.; ALBERT, M. K. Instance-based learning algorithms. *Machine Learning*, Kluwer Academic Publishers, Hingham, MA, USA, v. 6, n. 1, p. 37–66, 1991. ISSN 0885-6125.

ASUNCIóN, A.; NEWMAN, D. J. *UCI Machine Learning Repository*. `http://ics.uci.edu/~mlearn/MLRepository.html`, 2007.

BAGHERI, M. A.; QIGANG, G.; ESCALERA, S. A framework towards the unification of ensemble classification methods. In: *Proceedings of The 12th International Conference on Machine Learning and Applications*. [S.l.: s.n.], 2013. v. 2, p. 351–355.

BAUER, E.; KOHAVI, R. *An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting and Variants*. 1998.

BENSUSAN, H.; GIRAUD-CARRIER, C.; KENNEDY, C. *A Higher-order Approach to Meta-learning*. Bristol, UK, 2000.

BIAN, S.; WANG, W. On diversity and accuracy of homogeneous and heterogeneous ensembles. *Int. J. Hybrid Intell. Syst.*, IOS Press, Amsterdam, The Netherlands, The Netherlands, v. 4, n. 2, p. 103–128, 2007. ISSN 1448-5869.

BLUM, A.; RIVEST, R. L. Training a 3–node neural network is NP–complete. *Workshop on Conputacional Learning Theory*, p. 9–18, 1988.

BONISSONE, P. P. Lazy meta-learning: Creating customized model ensembles on demand. In: *Proceedings of The 2012 World Congress conference on Advances in Computational Intelligence*. Berlin, Heidelberg: Springer-Verlag, 2012. p. 1–23. ISBN 978-3-642-30686-0.

BRAZDIL, P.; GIRAUD-CARRIER, C.; SOARES, C.; VILALTA, R. *Metalearning: Applications to Data Mining*. [S.l.]: Springer, 2009.

BREIMAN, L. Bagging predictors. *Machine Learning*, Kluwer Academic Publishers, Hingham, MA, USA, v. 24, n. 2, p. 123–140, 1996. ISSN 0885-6125.

BREIMAN, L. *Bias, Variance, and Arcing Classifiers*. Berkeley, 1996.

BREIMAN, L. Arcing classifiers. *The Annals of Statistics*, v. 26, n. 3, p. 801–824, 1998.

BROWN, G.; KUNCHEVA, L. “good” and “bad” diversity in majority vote ensembles. In: GAYAR, N.; KITTLER, J.; ROLI, F. (Ed.). *Multiple Classifier Systems*. Berlin, Heidelberg: Springer, 2010. v. 5997, cap. 13, p. 124–133. ISBN 978-3-642-12126-5.

CANUTO, A. M. de P.; NASCIMENTO, D. S. C. A genetic-based approach to features selection for ensembles using a hybrid and adaptive fitness function. In: *Proceedings of The International Joint Conference on Neural Networks*. [S.l.]: IEEE, 2012.

CANUTO, A. M. P.; ABREU, M. C. C.; OLIVEIRA, L. de M.; JR., J. C. X.; SANTOS, A. de M. Investigating the influence of the choice of the ensemble members in accuracy and diversity of selection-based and fusion-based methods for ensembles. *Pattern Recognition Letters*, v. 28, n. 4, p. 472–486, 2007.

CANUTO, A. M. P.; OLIVEIRA, L. de M.; JR., J. C. X.; SANTOS, A. de M.; ABREU, M. C. C. Performance and diversity evaluation in hybrid and non-hybrid structures of ensembles. In: *Proceedings of The Fifth International Conference on Hybrid Intelligent Systems*. [S.l.: s.n.], 2005. p. 285–290.

CHEN, P.; POPOVICH, P. *Correlation: Parametric and Nonparametric Measures*. [S.l.]: Sage Publications, 2002.

CHEN, Y.; ZHAO, X.; LIN, Z. Optimizing subspace svm ensemble for hyperspectral imagery classification. *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 7, n. 4, p. 1295–1305, 2014.

CHERKAUER, J. K. Human expert level performance on a scientific image analysis task by a system using combined artificial neural networks. *Proceedings of The Workshop on Integrating Multiple Learned Models for Improving and Scaling Machine Learning Algorithms*, p. 15–21, 1996.

CLARE, A.; KING, R. D. Knowledge discovery in multi-label phenotype data. In: *Principles of Data Mining and Knowledge Discovery*. [S.l.]: Springer, 2001. v. 2168, p. 42–53. ISBN 978-3-540-42534-2.

CLEMEN, R. T. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, v. 5, n. 4, p. 559–583, 1989.

COELHO, A. L. V. *Evolução, Simbiose e Hibridismo Aplicados à Engenharia de Sistemas Inteligentes Modulares: Investigação em Redes Neurais Artificiais, Comitês de Máquinas e Sistemas Multiagentes*. Tese (Doutorado) — Universidade Estadual de Campinas, 2004.

COELHO, A. L. V.; NASCIMENTO, D. S. C. On the evolutionary design of heterogeneous bagging models. *Neurocomputing*, v. 73, p. 3319–3322, 2010.

COELHO, G. P. *Geração, Seleção e Combinação de Componentes para Ensembles de Redes Neurais Aplicadas a Problemas de Classificação*. Dissertação (Mestrado) — Universidade Estadual de Campinas, 2006.

CONDORCET, M. L. M. de. *Essai sur l'application de l'analyse á la probabilité des décisions rendues á la pluralité des voix*. [S.l.: s.n.], 1758.

COSTA, N. de C. *Seleção de Componentes em Ensembles de Classificação Multirrótulo*. Dissertação (Mestrado) — Universidade de Fortaleza, 2012.

COSTA, N. de C.; COELHO, A. L. V. Genetic and ranking-based selection of components for multilabel classifier ensembles. In: *Proceedings of The 11th International Conference on Hybrid Intelligent Systems*. [S.l.: s.n.], 2011. p. 311–317.

DARWIN, C. *On The Origin of Species, by Means of Natural Selection*. London: [s.n.], 1859.

DEMSAR, J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, v. 7, p. 1–30, 2006.

DIETTERICH, T. G. Ensemble methods in machine learning. In: *Proceedings of The First International Workshop on Multiple Classifier Systems*. London, UK: Springer-Verlag, 2000. p. 1–15.

DIETTERICH, T. G. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization. *Machine Learning*, v. 40, n. 2, p. 139–158, 2000.

DURILLO, J.; NEBRO, A.; ALBA, E. The jmetal framework for multi-objective optimization: Design and architecture. In: *CEC 2010*. Barcelona, Spain: [s.n.], 2010. p. 4138–4325.

DURILLO, J. J.; NEBRO, A. J. jmetal: A java framework for multi-objective optimization. *Advances in Engineering Software*, v. 42, p. 760–771, 2011. ISSN 0965-9978.

ECKHARDT, D. E.; LEE, L. D. A theoretical basis for the analysis of multiversion software subject to coincident errors. *Proceedings of The IEEE Transactions on Software Engineering*, IEEE Press, Piscataway, NJ, USA, v. 11, n. 12, p. 1511–1517, 1985. ISSN 0098-5589.

EFRON, B.; TIBSHIRANI, R. J. *An Introduction to the Bootstrap*. [S.l.: s.n.], 1993.

EIBEN, A. E.; SMITH, J. E. *Introduction to Evolutionary Computing*. 2nd. ed. [S.l.]: Springer, 2007.

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. L. F. de. *Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina*. [S.l.: s.n.], 2011.

FERNANDES, A. M. da R. *Inteligência Artificial: Noções Gerais*. Florianópolis - SC: [s.n.], 2005.

FRANK, E.; WITTEN, I. H. Generating accurate rule sets without global optimization. In: *Proceedings of the Fifteenth International Conference on Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998. p. 144–151. ISBN 1-55860-556-8.

FREUND, Y.; SCHAPIRE, R. E. *Experiments with a New Boosting Algorithm*. 1996.

FREUND, Y.; SCHAPIRE, R. E. A short introduction to boosting. *Japonese Society for Artificial Intelligence*, v. 14, n. 5, p. 771–780, 1999.

GEMAN, S.; BIENENSTOCK, E.; DOURSAT, R. Neural networks and the bias/variance dilemma. *Neural Computing*, MIT Press, Cambridge, MA, USA, v. 4, n. 1, p. 1–58, 1992. ISSN 0899-7667.

GIRAUD-CARRIER, C.; VILALTA, R.; BRAZDIL, P. Introduction to the special issue on meta-learning. *Mach. Learn.*, Kluwer Academic Publishers, Hingham, MA, USA, v. 54, n. 3, p. 187–193, 2004. ISSN 0885-6125.

GOLDBERG, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*. Boston, MA, USA: Addison – Wesley Longman Publishing Co., Inc., 1989. ISBN 0201157675.

GOMES, T. A. F.; PRUDêNCIO, R. B. C.; SOARES, C.; ROSSI, A. L. D.; CARVALHO, A. C. P. L. F. de. Combining meta-learning and search techniques to svm parameter selection. In: *Proceedings of The XI Brazilian Symposium on Neural Networks*. [S.l.]: IEEE, 2010. p. 79–84. ISBN 978-1-4244-8391-4.

GUAN, J.; GUO, W.; CHEN, H.; LOU, O. An ensemble of classifiers algorithm based on ga for handling concept-drifting data streams. In: *Proceedings of The Sixth International Symposium on Parallel Architectures, Algorithms and Programming*. [S.l.: s.n.], 2014. p. 282–284.

GUERRA, S.; PRUDêNCIO, R.; LUDERMIR, T. Meta-aprendizado de algoritmos de treinamento para redes multi-layer perceptron. In: *Anais do XXVII Congresso da Socieadade Brasileira da Computação*. Rio de Janeiro: Citeseer, 2007. p. 1022–1031.

GUYON, I.; GUNN, S.; NIKRAVESH, M.; ZADEH, L. A. *Feature Extraction: Foundations and Applications*. [S.l.]: Springer, 2006.

HANSEN, L. K.; LIISBERG, L.; SALAMON, P. Ensemble methods for hadwritten digit recognition. *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing*, p. 333–342, 1992.

HANSEN, L. K.; SALAMON, P. Neural network ensemble. *Proceedings of The IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 12, n. 10, October 1990.

HARPHAM, C.; DAWSON, W.; BROWN, R. A review of genetic algorithms applied to training radial basis function networks. *Neural Computing & Applications*, Springer-Verlag, London, UK, v. 13, n. 3, p. 193–201, 2004. ISSN 0941-0643.

HASHEM, S.; SCHMEISER, B. Improving model accuracy using optimal linear combinations of trained neural networks. *Proceedings of IEEE Transactions on Neural Networks*, p. 792–794, 1995.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. [S.l.: s.n.], 2001.

HAYKIN, S. *Redes Neurais, Princípios e Práticas*. Segunda. [S.l.]: Editora Bookman, 2001.

HOLLAND, J. H. *Adaptation in Natural and Artificial Systems*. [S.l.: s.n.], 1975.

HOLTE, R. C. Very simple classification rules perform well on most commonly used datasets. *Machine Learning*, Kluwer Academic Publishers, Hingham, MA, USA, v. 11, n. 1, p. 63–90, 1993. ISSN 0885-6125.

INOUE, H.; NARIHISA, H. Effective online pruning method for ensemble self-generating neural networks. In: *Proceedings of The 47th Midwest Symposium on Circuits and Systems*. [S.l.: s.n.], 2004. p. 85–88.

JAIN, K.; KULKARNI, S. Incorporating bagging into boosting. In: *Proceedings of The 12th International Conference on Hybrid Intelligent Systems*. [S.l.: s.n.], 2012. p. 443–448.

JOHN, G.; LANGLEY, P. Estimating continuous distributions in bayesian classifiers. In: *In Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. [S.l.]: Morgan Kaufmann, 1995. p. 338–345.

KALOUSIS, A. *Algorithm Selection via Meta-Learning*. Tese (Doutorado) — University of Geneva, 2002.

KARIM, M.; FARID, D. M. An adaptive ensemble classifier for mining complex noisy instances in data streams. In: *Proceedings of The International Conference on Informatics, Electronics & Vision*. [S.l.: s.n.], 2014. p. 1–4.

KATAKIS, I.; TSOUMAKAS, G.; VLAHAVAS, I. Multilabel text classification for automated tag suggestion. In: *Proceedings of the ECML/PKDD-08 Workshop on Discovery Challenge*. [S.l.: s.n.], 2008.

KHOSHGOFTAAR, T.; HULSE, J. V.; NAPOLITANO, A. Comparing boosting and bagging techniques with noisy and imbalanced data. *Proceedings of The IEEE Transactions on Systems, Man and Cybernetics*, v. 41, n. 3, p. 552–568, 2011. ISSN 1083-4427.

KOHAVI, R. The power of decision tables. In: *Proceedings of the 8th European Conference on Machine Learning*. London, UK: Springer-Verlag, 1995. p. 174–189. ISBN 3-540-59286-5.

KOTSIANTIS, S. B. Bagging and boosting variants for handling classifications problems: a survey. *The Knowledge Engineering Review*, v. 29, p. 78–100, 2014.

KOTSIANTIS, S. B.; PINTELAS, P. E. Combining bagging and boosting. *International Journal of Computational Intelligence*, v. 1, n. 4, p. 1304–2386, 2004.

KOZA, J. R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection (Complex Adaptive Systems)*. [S.l.]: The MIT Press, 1992. Hardcover. ISBN 0262111705.

KRUPKA, E.; VINNIKOV, A.; KLEIN, B.; HILLEL, A.; FREEDMAN, D.; STACHNIAK, S. Discriminative ferns ensemble for hand pose recognition. In: *Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2014. p. 3670–3677.

KUNCHEVA, L. I. Clustering-and-selection model for classifier combination. In: *Proceedings of Fourth International Conference onKnowledge-Based Intelligent Engineering Systems and Allied Technologies*. [S.l.: s.n.], 2000. v. 1, p. 185–188.

KUNCHEVA, L. I.; WHITAKER, C. J. Measures of diversity in classifier ensembles. *Machine Learning*, v. 51, p. 181–207, 2003.

LANGDON, W. B.; BARRETT, S. J.; BUXTON, B. F. Combining decision trees and neural networks for drug discovery. In: *Proceedings of The 5th European Conference on Genetic Programming*. London, UK: Springer-Verlag, 2002. p. 60–70. ISBN 3-540-43378-3.

LIMA, C. A. M.; COELHO, A. L. V.; ZUBEN, F. J. V. Ensembles of support vector machines for regression problems. *Proceedings of The IEEE International Joint Conference on Neural Networks*, p. 2381–2386, 2002.

MAIMON, O.; ROKACH, L. *The Data Mining and Knowledge. Discovery Handbook: A Complete Guide for Practitioners and Researchers*. [S.l.]: Springer, 2005.

MAO, J. A case study on bagging, boosting and basic ensembles of neural networks for ocr. *Proceedings of The IEEE International Joint Conference on Neural Networks*, p. 1828–1833, 1998.

MCLEOD, P.; VERMA, B.; ZHANG, M. Optimizing configuration of neural ensemble network for breast cancer diagnosis. In: *Proceedings of The International Joint Conference on Neural Networks*. [S.l.: s.n.], 2014. p. 1087–1092.

MICHIE, D.; SPIEGELHALTER, D. J.; TAYLOR, C. *Machine Learning, Neural and Statistical Classification*. Upper Saddle River, NJ, USA: Ellis Horwood, 1994. ISBN 0-13-106360-X.

MITCHELL, T. M. *Machine Learning*. [S.l.: s.n.], 1997.

NABIHA, A.; NADIR, F. New dynamic ensemble of classifiers selection approach based on confusion matrix for arabic handwritten recognition. In: *Proceedings of The International Conference on Multimedia Computing and Systems*. [S.l.: s.n.], 2012. p. 308–313.

NASCIMENTO, D. S.; COELHO, A. L.; CANUTO, A. M. Integrating complementary techniques for promoting diversity in classifier ensembles: A systematic study. *Neurocomputing*, v. 138, p. 347–357, 2014. ISSN 0925-2312.

NASCIMENTO, D. S. C. *Configuração Heterogênea de Ensembles de Classificadores: Investigação em Bagging, Boosting e MultiBoosting*. Dissertação (Mestrado) — Universidade de Fortaleza, 2009.

NASCIMENTO, D. S. C.; CANUTO, A. M. de P.; SILVA, L. M. M. e; COELHO, A. L. V. Combining different ways to generate diversity in bagging models: An evolutionary approach. In: *Proceedings of The International Joint Conference on Neural Networks*. [S.l.]: IEEE, 2011. p. 2235–2242.

NASCIMENTO, D. S. C.; COELHO, A. L. V. Bagging heterogêneo evolutivo: Caracterização e análise comparativa com ensembles homogêneas de redes neurais rbf. *Simpósio Brasileiro de Automação Inteligente*, 2009.

NASCIMENTO, D. S. C.; COELHO, A. L. V. Ensembling heterogeneous learning models with boosting. In: *Internacional Conference on Neural Information Processing*. [S.l.]: Lecture Notes in Computer Science, 2009. v. 5863, n. 512–519.

NASIERDING, G.; TSOUMAKAS, G.; KOUZANI, A. Z. Clustering based multi-label classification for image annotation and retrieval. In: *Proceedings of The IEEE International Conference on Systems, Man and Cybernetics.* [S.l.: s.n.], 2009. p. 4514–4519. ISSN 1062-922X.

NETO, A. A.; CANUTO, A. M. Meta-learning and multi-objective optimization to design ensemble of classifiers. In: *Brazilian Conference on Intelligent Systems (BRACIS)*. [S.l.: s.n.], 2014. p. 91–96.

NIKULIN, V.; BAKHARIA, A.; HUANG, T.-H. On the evaluation of the homogeneous ensembles with cv-passports. In: *Trends and Applications in Knowledge Discovery and Data Mining*. [S.l.]: Springer, 2013, (Lecture Notes in Computer Science, v. 7867). p. 109–120.

OLIVEIRA, J. de; LUDERMIR, T. Homogeneous ensemble selection through hierarchical clustering with a modified artificial fish swarm algorithm. In: *Proceedings of The 23rd IEEE International Conference on Tools with Artificial Intelligence*. [S.l.: s.n.], 2011. p. 177–180. ISSN 1082-3409.

OPITZ, D.; MACLIN, R. Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, v. 11, p. 169–198, 1999.

OSBORN, A. F. *Applied Imagination*. [S.l.: s.n.], 1953.

PAPADANIIL, C.; HADJILEONTIADIS, L. Efficient heart sound segmentation and extraction using ensemble empirical mode decomposition and kurtosis features. *Proceedings of The IEEE Journal of Biomedical and Health Informatics*, v. 18, n. 4, p. 1138–1152, 2014.

PARENTE, R. R.; CANUTO, A. M. P.; XAVIER, J. C. Characterization measures of ensemble systems using a meta-learning approach. In: *Proceedings of The 2013 International Joint Conference on Neural Networks*. [S.l.: s.n.], 2013. p. 1–8. ISSN 2161-4393.

PFAHRINGER, B.; BENSUSAN, H.; GIRAUD-CARRIER, C. Meta-learning by landmarking various learning algorithms. In: *Proceedings of the 17th International Conference on Machine Learning*. [S.l.]: Morgan Kaufmann, 2000. p. 743–750.

PLATT, J. C. Fast training of support vector machines using sequential minimal optimization. MIT Press, Cambridge, MA, USA, p. 185–208, 1999.

QUINLAN, J. R. Learning with continuous classes. In: . [S.l.]: World Scientific, 1992. p. 343–348.

QUINLAN, J. R. *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993. ISBN 1-55860-238-0.

REZENDE, S. O. *Sistemas Inteligentes: Fundamentos e Aplicações*. Baurueri - SP: [s.n.], 2003.

RICE, J. R. *The Algorithm Selection Problem*. Tese (Doutorado) — Purdue University, 1975.

SANTANA, L. E. A.; SILVA, L.; CANUTO, A. M. P.; PINTRO, F.; VALE, K. O. A comparative analysis of genetic algorithm and ant colony optimization to select attributes for an heterogeneous ensemble of classifiers. In: *Proceedings of The IEEE Congress on Evolutionary Computation*. [S.l.: s.n.], 2010. p. 1–8.

SANTOS, A. de M. *Investigando a Combinação de Técnicas de Aprendizado Semissupervisionado e Classificação Hierárquica Multirrótulo*. Tese (Doutorado) — Universidade Federal do Rio Grande do Norte, 2012.

SCHAPIRE, R. E. The strength of weak learnability. *Machine Learning*, v. 5, p. 197–227, 1990.

SCHAPIRE, R. E.; FREUND, Y. *Boosting: Foundations and Algorithms*. [S.l.]: The MIT Press, 2012.

SOARES, C. M. M. de O. P. *Learning Rankings of Learning Algorithms*. Tese (Doutorado) — Universidade do Porto, 2004.

SOARES, R. G. F.; SANTANA, A.; CANUTO, A. M. P.; SOUTO, M. C. P. de. Using accuracy and diversity to select classifiers to build ensembles. In: *Proceddings of The IEEE International Joint Conference on Neural Networks*. [S.l.: s.n.], 2006. p. 2289–2295.

SOARES, S.; ANTUNES, C. H.; ARAúJO, R. Comparison of a genetic algorithm and simulated annealing for automatic neural network ensemble development. *Neurocomputing*, v. 121, n. 0, p. 498–511, 2013. ISSN 0925-2312.

SOLLICH, P.; KROGH, A. Learning with ensembles: How overfitting can be useful. *Advances in Neural Information Processing Systems*, v. 8, p. 190–196, 1996.

SOUZA, A. F. de; JúNIOR, W. M. > *Atualizações em Informática 2011*. [S.l.]: Sociedade Brasileira de Computação, 2011.

SOUZA, B. F. de. *Meta-aprendizagem Aplicada à Classificação de Dados de Expressão Gênica*. Tese (Doutorado) — Universidade de São Paulo - São Carlos, 2010.

SRIMANY, A.; CHOWDHURI, S.; BHATTACHARYA, U.; PARUI, S. Holistic recognition of online handwritten words based on an ensemble of svm classifiers. In: *Proceedings of The 11th IAPR International Workshop on Document Analysis Systems*. [S.l.: s.n.], 2014. p. 86–90.

STEPNICKA, M.; STEPNICKOVA, L.; BURDA, M. Fuzzy rule-based ensemble for time series prediction: The application of linguistic associations mining. In: *Proceedings of The IEEE International Conference on Fuzzy Systems*. [S.l.: s.n.], 2014. p. 505–512.

SU, L.; YU, Z.; GUO, J.; MAO, C.; LIAO, Y. Question classification using multiboost. In: *Proceedings of The Eighth International Conference on Fuzzy Systems and Knowledge Discovery*. [S.l.: s.n.], 2011. v. 3, p. 1831–1835.

TAVARES, L. G.; LOPES, H. S.; LIMA, C. R. E. Estudo comparativo de métodos de aprendizado de máquina na detecção de regiões promotoras de genes de escherichia coli. *Anais do I Simpósio Brasileiro de Inteligência Computacional*, p. 8–11, 2007.

TSOUMAKAS, G.; ANGELIS, L.; VLAHAVAS, I. Selective fusion of heterogeneous classifiers. *Intelligent Data Analysis*, IOS Press, v. 9, n. 6, p. 511–525, 2005.

TSOUMAKAS, G.; KATAKIS, I.; VLAHAVAS, I. Mining multi-label data. In: *In Data Mining and Knowledge Discovery Handbook*. [S.l.: s.n.], 2010. p. 667–685.

TSOUMAKAS, G.; KATAKIS, I.; VLAHAVAS, L. Random k-labelsets for multilabel classification. *Proceedings of The IEEE Transactions on Knowledge and Data Engineering*, v. 23, n. 7, p. 1079–1089, July 2010.

VALENTINI, G.; MASULLI, F. Ensembles of learning machines. *Springer*, v. 2486, p. 3–19, 2002.

VILALTA, R.; DRISSI, Y. A perspective view and survey of meta-learning. *Artificial Intelligence Review*, v. 18, p. 77–95, 2002.

VINAY, K.; RAO, A.; KUMAR, G. Computerized analysis of classification of lung nodules and comparison between homogeneous and heterogeneous ensemble of classifier model. In: *Proceedings of The Third National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics*. [S.l.: s.n.], 2011. p. 231–234.

WANG, S.; YAO, X. Diversity analysis on imbalanced data sets by using ensemble models. In: *Proceedings of The IEEE Symposium on Computational Intelligence and Data Mining*. [S.l.: s.n.], 2009. p. 324–331.

WANG, S.; YAO, X. Relationships between diversity of classification ensembles and single-class performance measures. *Proceedings of The IEEE Transactions on Knowledge and Data Engineering*, v. 25, n. 1, p. 206–219, 2013.

WANG, W.; JONES, P.; PARTRIDGE, D. Diversity between neural networks and decision trees for building multiple classifier systems. In: *Proceedings of The First International Workshop on Multiple Classifier Systems*. London, UK: Springer-Verlag, 2000. p. 240–249.

WEBB, G. I. *Idealized Models of Decision Committee Performance and Their Application to Reduce Committee Error*. [S.l.], 1998.

WEBB, G. I. Multibooosting: A technique for combining boosting and wagging. *Machine Learning*, v. 40, p. 159–39, 2000.

WITTEN, I. H.; FRANK, E. *Data Mining: Pratical Machine Learning Tools and Techniques*. Second. San Francisco - CA: [s.n.], 2005.

WOLPERT, D. H. Stacked generalization. *Neural Networks*, v. 5, p. 241–259, 1992.

WOLPERT, D. H.; MACREADY, W. G. No free lunch theorems for optimization. *Proceedings of The IEEE Transactions on Evolutionary Computation*, v. 1, n. 1, p. 67–82, 1997.

WOODS, K.; JR., W. P. K.; BOWYER, K. Combination of multiple classifiers using local accuracy estimates. *Proceedings of The IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE Computer Society, Los Alamitos, CA, USA, v. 19, n. 4, p. 405–410, 1997. ISSN 0162-8828.

YUKSEL, S.; WILSON, J.; GADER, P. Twenty years of mixture of experts. *Proceedings of The IEEE Transactions on Neural Networks and Learning Systems*, v. 23, n. 8, p. 1177–1193, 2012. ISSN 2162-237X.

ZHANG, M.-L.; ZHOU, Z.-H. A k-nearest neighbor based algorithm for multi-label classification. In: *Proceedings of The IEEE International Conference on Granular Computing*. [S.l.: s.n.], 2005. v. 2, p. 718–721.

ZHANG, M.-L.; ZHOU, Z.-H. A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, v. 26, n. 8, p. 1819–1837, 2014.

ZHAO, Q.-L.; JIANG, Y.-H.; XU, M. Incremental learning by heterogeneous bagging ensemble. In: *Advanced Data Mining and Applications*. [S.l.]: Springer, 2010, (Lecture Notes in Computer Science, v. 6441). p. 1–12.

ZHAO, Q.-L.; JIANG, Y.-H.; XU, M. Incremental learning based on ensemble pruning. In: *Proceedings of The Eighth International Conference on Fuzzy Systems and Knowledge Discovery*. [S.l.: s.n.], 2011. v. 1, p. 377–381.

ZHENG, Z.; WEBB, G. I. Multiple boosting: A combination of boosting and bagging. In: *Proceedings of the 1998 International Conference on Parallel and Distributed Processing Techniques and Applications*. [S.l.]: CSREA Press, 1998. p. 1133–1140.

ZHOU, Z. H.; JIANG, Y.; YANG, Y. B.; CHEN, S. F. Lung cancer cell identification based on neural network ensembles. *Artificial Intelligence in Medicine*, v. 24, p. 25–36, 2000.

APÊNDICE A – Problemas de Classificação

A Tabela 24 exibe a descrição dos conjuntos de problemas de classificação utilizados nos testes retirados do repositório UCI (ASUNCIóN; NEWMAN, 2007), enquanto que a Tabela 25 exibe as características entre atributos, instâncias, classes e valores faltantes para os problemas de classificação. A Tabela 26 exibe as distribuições entre as classes.

Tabela 24: Conjunto de problemas de classificação

Base	Descrição
anneal	Avalia características da têmpera em metais.
breast-cancer	Busca identificar a presença de tumores em mamas, que pode ser benignos ou malignos.
bupa	Avaliação de exames de sangue que se pensa ser sensível aos problemas hepáticos que possam surgir a partir do consumo excessivo de álcool em homens.
car	Avaliação de modelos de carros.
colic	Avalia se cólicas em cavalos são provenientes de lesões cirúrgicas ou não.
credit-a	Aprovar ou não o crédito a um dado cliente mediante a análise de seu perfil de crédito.
diabetes	Diagnosticar a presença ou não de diabetes a partir de várias medidas fisiológicas e testes clínicos.
gaussian3	Descrição não encontrada.
glass	Identificar o tipo de vidro com base nas características físicas da amostra.
haberman	Analisa a sobrevida de pacientes submetidos à cirurgia de câncer de mama.
heart-c	Diagnosticar a presença ou ausência de risco de doença coronária (enfarto) a partir dos resultados de vários testes clínicos.
hepatitis	Dentre pacientes com hepatite, indicar quais deles irão falecer ou não.
ionosphere	Classificar sinais de radar provenientes da ionosfera.
iris	Predizer a classe de uma planta (Íris) de acordo com o formato de sua folha.
segment	Analisa *pixels* de imagens tiradas ao ar livre classificando os tipos de segmentos que podem ser tijolo, céu, folhagem, cimento, janela, caminho e capim.
sick	Classifica a presença ou ausência de tireoidiana.
sonar	Distinguir entre materiais metálicos ou rochosos de acordo com os níveis de intensidade de sinais de sonar enviados sob diferentes condições (ângulos, frequências, etc.)
vehicle	Classificar uma dada silhueta 2D em um dentre quatro tipos de veículos usando um conjunto de atributos físicos.
vote	Avalia os votos da Câmara de Representantes nos E.U.A. se foram realizados por um republicano ou um democrata.
waveform-5000	Identificar tipos de formas de ondas.

Tabela 25: Características de atributos, instâncias, classes e valores faltantes para os problemas de classificação

Base	Atributos	Instâncias	Classes	V. Faltantes
anneal	38	798	6	Sim
breast-cancer	9	286	2	Sim
bupa	7	345	2	Não
car	6	1728	4	Não
colic	28	368	2	Sim
credit-a	15	690	2	Sim
diabetes	8	768	2	Não
gaussian3	600	60	3	Não
glass	10	214	7	Não
haberman	3	306	2	Não
heart-c	13	303	5	Sim
hepatitis	19	115	2	Sim
ionosphere	34	351	2	Não
iris	4	150	3	Não
segment	19	2310	7	Não
sick	29	3772	2	Sim
sonar	60	208	2	Não
vehicle	18	946	4	Não
vote	16	435	2	Sim
waveform-5000	40	5000	3	Não

Tabela 26: Distribuição das classes

Base	Distribuição das classes
anneal	1% ; 11% ; 76% ; 0% ; 7% ; 4%
breast-cancer	70% ; 30%
bupa	42% ; 58%
car	70% ; 22% ; 4% ; 4%
colic	63% ; 37%
credit-a	44% ; 56%
diabetes	65% ; 35%
gaussian3	33% ; 33% ; 33%
glass	33% ; 35% ; 8% ; 0% ; 6% ; 4% ; 14%
haberman	73% ; 27%
heart-c	54% ; 46% ; 0% ; 0% ; 0%
hepatitis	21% ; 79%
ionosphere	36% ; 64%
iris	33% ; 33% ; 33%
segment	14% ; 14% ; 14% ; 14% ; 14% ; 14% ; 14%
sick	94% ; 6%
sonar	47% ; 53%
vehicle	25% ; 26% ; 26% ; 24%
vote	61% ; 39%
waveform-5000	34% ; 33% ; 33%

APÊNDICE B – Configurações dos Algoritmos de Aprendizagem

As Tabelas 27 a 35 disponibilizam as configurações dos parâmetros de controle dos algoritmos de aprendizagem que foram utilizados para realização dos experimentos. Esses algoritmos estão disponíveis no ambiente WEKA (WITTEN; FRANK, 2005).

Tabela 27: Configurações dos parâmetros de controle para o algoritmo RBF

RBF		
Parâmetro	Descrição	Valor
clusteringSeed	O *seed* aleatório para cada *k* vizinhos.	1
maxIts	Número máximo de iterações para a regressão logística a ser executada nos problemas com classe discreta.	-1
minStdDev	Define o desvio padrão mínimo para os *clusters*.	0.1
numClusters	O número de *clusters* para *k* vizinhos a serem gerados.	2
ridge	Defina o valor para o cume de regressão linear.	1.0E-8

Tabela 28: Configurações dos parâmetros de controle para o algoritmo J48

J48		
Parâmetro	Descrição	Valor
binarySplits	Se usa divisões binárias em atributos nominais na construção das árvores.	*false*
confidenceFactor	O fator de confiança utilizado para a poda.	0.25
minNumObj	O número mínimo de instâncias por folha.	2
numFolds	Determina a quantidade de dados usada para poda de erro reduzida. Um *fold* é usado para poda, e o restante para o crescimento da árvore.	3
reducedErrorPruning	Se a redução de erro da poda a ser usada é a mesma utilizada no algoritmo C.4.5.	*false*
seed	*Seed* utilizado para randomizar os dados quando a redução de erro da poda é utilizado.	1
subtreeRaising	Se considerar a sub-árvore na operação de poda.	*true*
unpruned	Se é realizada poda.	*false*
useLaplace	Se a contagem de folhas são suavizados com base em *laplace*.	*false*

Tabela 29: Configurações dos parâmetros de controle para o algoritmo SMO

SMO		
Parâmetro	Descrição	Valor
buildLogisticModels	Adéqua modelos logísticos para as saídas por estimativas de probabilidade.	*false*
c	Parâmetro de complexidade *C*.	1.0
cacheSize	O tamanho do *cache* do *kernel*.	250007
epsilon	O *epsilon* para arredondar erro.	1.0E-12
exponent	O expoente para o *kernel* polinomial.	1.0
featureSpaceNormalization	Se o recurso de normalização espacial é realizado.	*false*
filterType	Determina se os dados serão transformados.	*Normalize training data*
gamma	O valor do parâmetro de *gama* para *kernels* RBF.	0.01
lowerOrderTerms	Se ordem mais baixa dos polinomiais finais são usados.	*false*
numFolds	O número de *folds* para validação cruzada utilizada para gerar dados de treinamento para modelos logísticos.	-1
randomSeed	*Seed* Randômico para a validação cruzada.	1
toleranceParameter	Parâmetro de tolerância.	0.0010
useRBF	Se utiliza *kernel* RBF em vez de um polinômio.	*false*

Tabela 30: Configurações dos parâmetros de controle para o algoritmo *Naïve Bayes*

Naïve Bayes		
Parâmetro	Descrição	Valor
useKernelEstimator	Usar um estimador *kernel* para atributos numéricos ao invés de uma distribuição normal.	*false*
useSupervisedDiscretization	Usar discretização supervisionada para converter atributos numéricos para categóricos.	*false*

Tabela 31: Configurações dos parâmetros de controle para o algoritmo IB*k*

IB*K*		
Parâmetro	Descrição	Valor
kNN	Número de vizinhos.	1
crossValidate	Se a validação cruzada será utilizada para selecionar o melhor valor de *k*.	*false*
distanceWeighting	Obtém o método de ponderação utilizado.	*No distance weighting*
meanSquared	Se o erro quadrado médio é usado ao invéz de erro médio absoluto, ao fazer a validação cruzada para problemas de regressão.	*false*
noNormalization	Normalização do atributo.	*false*
windowSize	Obtém o número máximo de instâncias permitidas no *pool* de treinamento.	0

Tabela 32: Configurações dos parâmetros de controle para o algoritmo REP *Tree*

REP *Tree*		
Parâmetro	Descrição	Valor
maxDepth	A profundidade máxima da árvore.	-1
minNum	O peso mínimo total dos casos em uma folha.	2.0
minVarianceProp	A proporção mínima de variância em todos os dados que precisa estar presente em um nó.	0.0010
noPruning	Se a poda é realizada.	*false*
numFolds	Determina a quantidade de dados utilizados para a poda. Um *fold* é usado para poda, o restante para a aplicação das regras.	3
seed	O *seed* utilizado para randomizar os dados.	1

Tabela 33: Configurações dos parâmetros de controle para o algoritmo OneR

OneR		
Parâmetro	Descrição	Valor
minBucketSize	O tamanho mínimo de *bucket* usado para discretizar atributos numéricos.	6

Tabela 34: Configurações dos parâmetros de controle para o algoritmo PART

PART		
Parâmetro	Descrição	Valor
binarySplits	Usar divisão binária em atributos nominais ao construir as árvores parciais.	*false*
confidenceFactor	Fator de confiança utilizado na poda.	0.25
minNumObj	Valor mínimo de objetivos.	2
numFolds	Determina a quantidade de dados usados para redução de erro de poda.	3
reducedErrorPruning	Usar redução de erro de poda ao invés da técnica utilizada no algoritmo C 4.5.	*false*
seed	O *seed* utilizado para randomizar os dados quando a redução de erro da poda é utilizada.	1
unpruned	Se a poda é realizada.	*false*

Tabela 35: Configurações dos parâmetros de controle para o algoritmo *Decision Table*

Decision Table		
Parâmetro	Descrição	Valor
crossVal	Define o número de *folds* para a validação cruzada.	1
maxStale	Define o número de tabelas de decisão para considerar antes de abandonar a pesquisa.	5
useIBk	Utilizar IB*k* em vez de classe majoritária.	*false*

APÊNDICE C – Resultados Empíricos

Neste capítulo são apresentados os resultados dos experimentos empíricos realizados sobre todas as 200 bases de dados para os algoritmos *Bagging*, *Boosting* e *Multiboosting* homogêneos, todos formados por componentes de redes neurais artificiais (RBF), árvore de decisão (J48), máquinas vetores-suporte (SMO) e algoritmo baseado em vizinhança (IBk). Das Tabelas 36 a 42 são apresentados os valores de erro médio e desvio padrão para *Bagging* homogêneo. Enquanto que das Tabelas 43 a 49 são apresentados os valores de erro médio e desvio padrão para *Boosting* homogêneo. E por fim, das Tabelas 50 a 56 são apresentados os valores de erro médio e desvio padrão para *Multiboosting* homogêneo.

Tabela 36: Valores de erro médio e desvio padrão para *Bagging* quando aplicado às bases "anneal", "breast-cancer" e "bupa"

Base	RBF	J48	SMO	NB	IBk
anneal 1	0,0463±0,0072	**0,0126±0,0031**	0,0280±0,0025	0,1282±0,0029	0,0137±0,0026
anneal 2	0,0443±0,0043	0,0121±0,0020	0,0283±0,0031	0,1359±0,0026	**0,0116±0,0039**
anneal 3	0,0413±0,0067	**0,0105±0,0011**	0,0262±0,0038	0,1278±0,0026	0,0139±0,0026
anneal 4	0,0443±0,0048	**0,0120±0,0025**	0,0286±0,0032	0,1350±0,0019	0,0131±0,0034
anneal 5	0,0423±0,0059	0,0124±0,0015	0,0254±0,0024	0,1295±0,0020	**0,0118±0,0038**
anneal 6	0,0280±0,0051	**0,0127±0,0020**	0,0228±0,0024	0,1181±0,0043	0,0145±0,0023
anneal 7	0,0275±0,0036	**0,0110±0,0015**	0,0229±0,0038	0,1285±0,0055	0,0141±0,0038
anneal 8	0,0432±0,0060	0,0142±0,0024	0,0272±0,0027	0,1301±0,0021	**0,0135±0,0021**
anneal 9	0,0398±0,0046	0,0111±0,0022	0,0245±0,0032	0,1333±0,0031	**0,0110±0,0031**
anneal 10	0,0400±0,0055	**0,0110±0,0016**	0,0236±0,0033	0,1266±0,0033	0,0153±0,0026
breast-cancer 1	0,2805±0,0100	**0,2708±0,0139**	0,3016±0,0193	0,2767±0,0081	0,2802±0,0095
breast-cancer 2	0,2817±0,0106	0,2716±0,0098	0,2992±0,0117	**0,2677±0,0060**	0,2739±0,0106
breast-cancer 3	0,2669±0,0106	0,2603±0,0108	0,2856±0,0146	**0,2588±0,0074**	0,2728±0,0141
breast-cancer 4	**0,2677±0,0116**	0,2821±0,0157	0,2883±0,0156	0,2681±0,0096	0,2743±0,0187
breast-cancer 5	**0,2700±0,0092**	0,2825±0,0134	0,2938±0,0156	0,2708±0,0072	0,2907±0,0118
breast-cancer 6	**0,2907±0,0116**	0,3125±0,0149	0,3140±0,0163	0,2988±0,0093	0,3233±0,0134
breast-cancer 7	**0,2752±0,0130**	0,2903±0,0128	0,3058±0,0128	0,2752±0,0093	0,3019±0,0102
breast-cancer 8	0,2713±0,0084	**0,2694±0,0108**	0,2888±0,0097	0,2748±0,0099	0,2919±0,0129
breast-cancer 9	0,2713±0,0105	0,2628±0,0120	0,2783±0,0083	**0,2578±0,0069**	0,2833±0,0172
breast-cancer 10	0,2775±0,0108	**0,2686±0,0134**	0,3031±0,0117	0,2787±0,0076	0,2942±0,0107
bupa 1	0,3555±0,0146	**0,2923±0,0172**	0,4152±0,0055	0,4439±0,0122	0,3726±0,0102
bupa 2	0,3600±0,0115	**0,3197±0,0229**	0,4165±0,0039	0,4542±0,0083	0,3719±0,0086
bupa 3	0,3558±0,0086	**0,2932±0,0178**	0,4171±0,0066	0,4429±0,0113	0,3542±0,0166
bupa 4	0,3555±0,0105	**0,3203±0,0282**	0,4190±0,0060	0,4490±0,0127	0,3761±0,0165
bupa 5	0,3452±0,0112	**0,3110±0,0188**	0,4142±0,0038	0,4710±0,0112	0,4090±0,0095
bupa 6	0,3431±0,0145	**0,2987±0,0232**	0,4222±0,0053	0,4363±0,0116	0,4061±0,0084
bupa 7	0,3331±0,0162	**0,2871±0,0194**	0,4212±0,0073	0,4415±0,0130	0,3939±0,0146
bupa 8	0,3293±0,0166	**0,2871±0,0107**	0,4151±0,0038	0,4495±0,0140	0,3556±0,0113
bupa 9	0,3437±0,0116	**0,2958±0,0137**	0,4177±0,0047	0,4138±0,0175	0,3601±0,0103
bupa 10	0,3486±0,0208	**0,3032±0,0161**	0,4222±0,0061	0,4463±0,0178	0,3711±0,0190

Tabela 37: Valores de erro médio e desvio padrão para *Bagging* quando aplicado às bases “car”, “colic” e “credit-a”

Base	RBF	J48	SMO	NB	IBk
car 1	0,0781±0,0027	**0,0658±0,0048**	0,0684±0,0022	0,1485±0,0044	0,0677±0,0035
car 2	0,0798±0,0036	0,0738±0,0050	**0,0628±0,0030**	0,1493±0,0034	0,0663±0,0049
car 3	0,0788±0,0033	0,0758±0,0036	**0,0621±0,0023**	0,1484±0,0051	0,0687±0,0034
car 4	0,0830±0,0045	0,0713±0,0042	0,0671±0,0023	0,1498±0,0027	**0,0669±0,0046**
car 5	0,0770±0,0043	0,0666±0,0050	**0,0577±0,0022**	0,1487±0,0026	0,0630±0,0039
car 6	0,0777±0,0040	0,0745±0,0047	**0,0653±0,0024**	0,1475±0,0035	0,0678±0,0034
car 7	0,0844±0,0041	0,0754±0,0046	**0,0650±0,0026**	0,1480±0,0043	0,0713±0,0026
car 8	0,0828±0,0025	0,0758±0,0042	**0,0669±0,0029**	0,1475±0,0052	0,0670±0,0020
car 9	0,0778±0,0039	0,0645±0,0040	**0,0618±0,0027**	0,1436±0,0049	0,0637±0,0043
car 10	0,0814±0,0047	0,0739±0,0041	0,0726±0,0024	0,1463±0,0033	**0,0670±0,0028**
colic 1	0,1931±0,0069	**0,1465±0,0062**	0,1716±0,0119	0,2082±0,0058	0,2130±0,0069
colic 2	0,1864±0,0086	**0,1366±0,0065**	0,1429±0,0105	0,2027±0,0033	0,1991±0,0077
colic 3	0,2073±0,0043	**0,1508±0,0059**	0,1798±0,0067	0,2215±0,0038	0,2175±0,0043
colic 4	0,1958±0,0058	**0,1514±0,0033**	0,1801±0,0070	0,2139±0,0057	0,2148±0,0093
colic 5	0,2091±0,0094	**0,1538±0,0063**	0,1858±0,0098	0,2181±0,0060	0,2202±0,0122
colic 6	0,1906±0,0101	**0,1502±0,0047**	0,1695±0,0061	0,2012±0,0038	0,2100±0,0076
colic 7	0,1822±0,0093	**0,1353±0,0042**	0,1707±0,0098	0,2085±0,0070	0,2163±0,0078
colic 8	0,1991±0,0052	**0,1486±0,0060**	0,1743±0,0045	0,2130±0,0062	0,1973±0,0055
colic 9	0,1955±0,0085	**0,1515±0,0043**	0,1726±0,0078	0,2142±0,0064	0,2039±0,0110
colic 10	0,1991±0,0124	**0,1623±0,0066**	0,1804±0,0117	0,2178±0,0055	0,2108±0,0103
credit-a 1	0,2089±0,0070	**0,1428±0,0047**	0,1441±0,0043	0,2274±0,0039	0,1878±0,0087
credit-a 2	0,1973±0,0055	**0,1401±0,0055**	0,1478±0,0026	0,2230±0,0038	0,1915±0,0077
credit-a 3	0,2068±0,0032	**0,1452±0,0092**	0,1548±0,0047	0,2277±0,0036	0,1882±0,0082
credit-a 4	0,2035±0,0046	**0,1366±0,0083**	0,1517±0,0048	0,2243±0,0036	0,1728±0,0070
credit-a 5	0,1963±0,0028	**0,1367±0,0099**	0,1441±0,0043	0,2216±0,0027	0,1815±0,0084
credit-a 6	0,1966±0,0072	**0,1403±0,0055**	0,1475±0,0022	0,2159±0,0030	0,1868±0,0070
credit-a 7	0,1966±0,0032	**0,1403±0,0049**	0,1478±0,0028	0,2238±0,0042	0,1940±0,0063
credit-a 8	0,1966±0,0049	**0,1386±0,0057**	0,1502±0,0028	0,2298±0,0027	0,1907±0,0062
credit-a 9	0,1924±0,0045	**0,1329±0,0049**	0,1462±0,0033	0,2201±0,0027	0,1791±0,0038
credit-a 10	0,1960±0,0041	**0,1386±0,0060**	0,1481±0,0027	0,2156±0,0031	0,1834±0,0073

Tabela 38: Valores de erro médio e desvio padrão para *Bagging* quando aplicado às bases "diabetes", "gaussian3" e "glass"

Base	RBF	J48	SMO	NB	IBk
diabetes 1	0,2410±0,0069	0,2421±0,0085	**0,2259±0,0038**	0,2379±0,0050	0,2916±0,0093
diabetes 2	0,2556±0,0064	0,2498±0,0098	**0,2311±0,0054**	0,2495±0,0036	0,2894±0,0070
diabetes 3	0,2482±0,0069	0,2486±0,0040	**0,2291±0,0040**	0,2491±0,0033	0,3052±0,0058
diabetes 4	0,2476±0,0092	0,2438±0,0109	**0,2308±0,0049**	0,2452±0,0039	0,3130±0,0096
diabetes 5	0,2525±0,0084	0,2478±0,0122	**0,2323±0,0040**	0,2492±0,0042	0,2871±0,0079
diabetes 6	0,2378±0,0051	0,2326±0,0111	**0,2262±0,0023**	0,2355±0,0035	0,2926±0,0065
diabetes 7	0,2625±0,0036	0,2452±0,0121	**0,2337±0,0031**	0,2573±0,0057	0,3107±0,0089
diabetes 8	0,2527±0,0070	0,2514±0,0094	**0,2318±0,0034**	0,2475±0,0061	0,2936±0,0091
diabetes 9	0,2483±0,0061	0,2439±0,0144	**0,2262±0,0043**	0,2373±0,0049	0,2853±0,0072
diabetes 10	0,2360±0,0055	0,2399±0,0111	**0,2259±0,0040**	0,2445±0,0042	0,2864±0,0095
gaussian3 1	0,3963±0,0488	0,3148±0,0617	**0,0000±0,0000**	0,3130±0,0620	0,0074±0,0096
gaussian3 2	0,4296±0,0692	0,2870±0,0218	**0,0000±0,0000**	0,3593±0,0580	0,0074±0,0096
gaussian3 3	0,4278±0,0575	0,2870±0,0689	**0,0000±0,0000**	0,3944±0,0437	0,0111±0,0096
gaussian3 4	0,4704±0,0710	0,2537±0,0699	**0,0000±0,0000**	0,3963±0,0351	0,0074±0,0096
gaussian3 5	0,3556±0,0604	0,2852±0,0587	**0,0000±0,0000**	0,3463±0,0391	0,0074±0,0096
gaussian3 6	0,4463±0,0738	0,2167±0,0671	**0,0000±0,0000**	0,3630±0,0429	0,0074±0,0096
gaussian3 7	0,4444±0,0611	0,2981±0,0608	**0,0000±0,0000**	0,3944±0,0479	0,0000±0,0000
gaussian3 8	0,4685±0,0437	0,3130±0,0743	**0,0000±0,0000**	0,3852±0,0500	0,0000±0,0000
gaussian3 9	0,4852±0,0358	0,2481±0,0660	**0,0000±0,0000**	0,4019±0,0428	0,0111±0,0096
gaussian3 10	0,3889±0,0509	0,2907±0,0559	**0,0000±0,0000**	0,3185±0,0388	0,0148±0,0078
glass 1	0,3229±0,0211	**0,2781±0,0167**	0,4344±0,0223	0,4854±0,0175	0,3234±0,0162
glass 2	0,3104±0,0146	**0,2755±0,0148**	0,4458±0,0216	0,4953±0,0164	0,3036±0,0163
glass 3	0,2964±0,0171	**0,2609±0,0297**	0,4385±0,0180	0,4922±0,0156	0,3333±0,0095
glass 4	0,3161±0,0259	**0,2870±0,0148**	0,4260±0,0231	0,5094±0,0149	0,3135±0,0145
glass 5	0,3073±0,0171	**0,2896±0,0123**	0,4425±0,0139	0,5052±0,0202	0,3446±0,0115
glass 6	0,3005±0,0166	**0,2886±0,0220**	0,4238±0,0184	0,5067±0,0225	0,2953±0,0192
glass 7	0,2990±0,0153	**0,2845±0,0145**	0,4321±0,0118	0,5264±0,0206	0,3124±0,0188
glass 8	0,3140±0,0194	**0,2658±0,0173**	0,4663±0,0232	0,5378±0,0174	0,3197±0,0200
glass 9	0,2705±0,0124	**0,2477±0,0231**	0,4337±0,0206	0,5109±0,0164	0,2772±0,0164
glass 10	0,2907±0,0187	**0,2554±0,0158**	0,4202±0,0222	0,4912±0,0187	0,3109±0,0206

Tabela 39: Valores de erro médio e desvio padrão para *Bagging* quando aplicado às bases "haberman", "heart-c" e "hepatitis"

Base	RBF	J48	SMO	NB	IBk
haberman 1	0,2651±0,0074	0,2880±0,0140	0,2716±0,0049	**0,2513±0,0032**	0,3451±0,0096
haberman 2	0,2607±0,0067	0,2742±0,0118	0,2738±0,0064	**0,2542±0,0053**	0,3331±0,0121
haberman 3	0,2749±0,0106	0,3007±0,0137	0,2724±0,0063	**0,2673±0,0067**	0,3465±0,0138
haberman 4	0,2564±0,0057	0,2636±0,0209	0,2742±0,0043	**0,2491±0,0057**	0,3298±0,0146
haberman 5	0,2578±0,0110	0,2840±0,0121	0,2716±0,0049	**0,2505±0,0065**	0,3404±0,0091
haberman 6	0,2491±0,0089	0,2625±0,0089	0,2669±0,0093	**0,2411±0,0054**	0,3298±0,0149
haberman 7	0,2529±0,0083	0,2757±0,0089	0,2707±0,0112	**0,2417±0,0048**	0,3453±0,0132
haberman 8	0,2627±0,0108	0,2754±0,0136	0,2678±0,0027	**0,2500±0,0068**	0,3304±0,0112
haberman 9	0,2402±0,0071	0,2699±0,0123	0,2717±0,0076	**0,2380±0,0051**	0,3167±0,0142
haberman 10	0,2609±0,0054	0,2685±0,0088	0,2674±0,0056	**0,2467±0,0058**	0,3210±0,0104
heart-c 1	**0,1544±0,0081**	0,1996±0,0160	0,1548±0,0110	0,1562±0,0090	0,2316±0,0062
heart-c 2	0,1676±0,0055	0,2059±0,0123	**0,1562±0,0109**	0,1658±0,0059	0,2301±0,0092
heart-c 3	0,1673±0,0082	0,2070±0,0207	**0,1636±0,0100**	0,1680±0,0078	0,2449±0,0146
heart-c 4	0,1571±0,0074	0,1919±0,0098	0,1678±0,0079	**0,1546±0,0038**	0,2231±0,0133
heart-c 5	**0,1784±0,0067**	0,2260±0,0091	0,1868±0,0098	0,1824±0,0048	0,2575±0,0123
heart-c 6	0,1744±0,0063	0,2227±0,0097	0,1722±0,0115	**0,1714±0,0116**	0,2513±0,0098
heart-c 7	0,1784±0,0065	0,2077±0,0099	0,1788±0,0059	**0,1744±0,0035**	0,2505±0,0112
heart-c 8	0,1538±0,0081	0,2070±0,0081	**0,1520±0,0093**	0,1527±0,0035	0,2410±0,0103
heart-c 9	0,1615±0,0101	0,2121±0,0175	**0,1604±0,0059**	0,1718±0,0061	0,2249±0,0046
heart-c 10	**0,1571±0,0050**	0,1974±0,0142	0,1663±0,0105	0,1689±0,0047	0,2311±0,0084
hepatitis 1	0,1482±0,0167	0,1691±0,0114	**0,1439±0,0102**	0,1583±0,0136	0,1978±0,0164
hepatitis 2	**0,1403±0,0156**	0,1806±0,0124	0,1568±0,0106	0,1604±0,0083	0,1806±0,0086
hepatitis 3	**0,1403±0,0132**	0,1863±0,0142	0,1504±0,0153	0,1655±0,0127	0,1784±0,0172
hepatitis 4	**0,1504±0,0110**	0,1993±0,0204	0,1619±0,0193	0,1655±0,0048	0,2273±0,0170
hepatitis 5	**0,1338±0,0084**	0,2036±0,0198	0,1619±0,0183	0,1568±0,0116	0,2014±0,0206
hepatitis 6	**0,1371±0,0120**	0,1857±0,0126	0,1593±0,0143	0,1557±0,0120	0,2043±0,0176
hepatitis 7	**0,1457±0,0102**	0,1993±0,0132	0,1536±0,0091	0,1579±0,0071	0,1907±0,0122
hepatitis 8	**0,1200±0,0134**	0,1914±0,0120	0,1414±0,0088	0,1429±0,0112	0,1736±0,0172
hepatitis 9	**0,1479±0,0122**	0,1921±0,0148	0,1586±0,0120	0,1679±0,0131	0,2050±0,0107
hepatitis 10	**0,1450±0,0107**	0,1829±0,0221	0,1607±0,0131	0,1593±0,0131	0,1864±0,0209

Tabela 40: Valores de erro médio e desvio padrão para *Bagging* quando aplicado às bases "ionosphere", "iris" e "segment"

Base	RBF	J48	SMO	NB	IBk
ionosphere 1	0,0905±0,0043	**0,0902±0,0118**	0,1210±0,0061	0,1787±0,0067	0,1222±0,0062
ionosphere 2	0,0753±0,0036	**0,0715±0,0065**	0,1082±0,0068	0,1671±0,0029	0,1209±0,0071
ionosphere 3	0,0905±0,0045	**0,0842±0,0116**	0,1161±0,0078	0,1797±0,0039	0,1383±0,0076
ionosphere 4	0,0899±0,0058	**0,0778±0,0086**	0,1174±0,0067	0,1794±0,0045	0,1348±0,0062
ionosphere 5	**0,0778±0,0048**	0,0851±0,0096	0,1209±0,0083	0,1687±0,0050	0,1430±0,0070
ionosphere 6	0,0766±0,0049	**0,0753±0,0065**	0,1117±0,0070	0,1801±0,0075	0,1370±0,0054
ionosphere 7	**0,0826±0,0061**	0,0845±0,0076	0,1168±0,0080	0,1769±0,0055	0,1424±0,0049
ionosphere 8	0,0921±0,0071	**0,0731±0,0069**	0,1085±0,0062	0,1756±0,0031	0,1415±0,0075
ionosphere 9	0,0902±0,0090	**0,0854±0,0094**	0,1180±0,0054	0,1741±0,0049	0,1259±0,0063
ionosphere 10	0,0861±0,0059	**0,0804±0,0037**	0,1146±0,0076	0,1775±0,0038	0,1389±0,0062
iris 1	0,0393±0,0036	0,0504±0,0084	0,0400±0,0052	0,0489±0,0052	**0,0378±0,0023**
iris 2	0,0467±0,0061	0,0570±0,0070	**0,0422±0,0036**	0,0519±0,0049	0,0489±0,0038
iris 3	0,0459±0,0058	0,0541±0,0111	**0,0415±0,0038**	0,0430±0,0047	0,0452±0,0023
iris 4	**0,0319±0,0050**	0,0415±0,0106	0,0326±0,0062	0,0415±0,0052	0,0341±0,0062
iris 5	0,0459±0,0031	0,0563±0,0106	**0,0348±0,0036**	0,0407±0,0039	0,0400±0,0038
iris 6	0,0467±0,0050	0,0563±0,0100	**0,0422±0,0050**	0,0533±0,0047	0,0467±0,0070
iris 7	0,0400±0,0038	0,0615±0,0111	**0,0356±0,0058**	0,0437±0,0042	0,0393±0,0050
iris 8	0,0393±0,0050	0,0504±0,0143	**0,0348±0,0036**	0,0430±0,0077	0,0533±0,0047
iris 9	0,0467±0,0036	0,0674±0,0102	**0,0407±0,0039**	0,0467±0,0086	0,0459±0,0068
iris 10	0,0356±0,0031	0,0474±0,0072	**0,0348±0,0061**	0,0378±0,0023	0,0378±0,0055
segment 1	0,1160±0,0018	**0,0283±0,0016**	0,0712±0,0011	0,1993±0,0028	0,0353±0,0017
segment 2	0,1160±0,0016	**0,0275±0,0023**	0,0739±0,0018	0,2001±0,0022	0,0306±0,0024
segment 3	0,1142±0,0025	**0,0265±0,0013**	0,0733±0,0019	0,1963±0,0014	0,0320±0,0020
segment 4	0,1153±0,0027	**0,0282±0,0022**	0,0796±0,0016	0,1973±0,0009	0,0329±0,0016
segment 5	0,1155±0,0025	**0,0273±0,0026**	0,0779±0,0016	0,1995±0,0013	0,0298±0,0018
segment 6	0,1181±0,0026	**0,0259±0,0018**	0,0714±0,0025	0,2001±0,0026	0,0294±0,0021
segment 7	0,1101±0,0032	**0,0291±0,0015**	0,0705±0,0011	0,1937±0,0019	0,0295±0,0020
segment 8	0,1149±0,0024	**0,0271±0,0021**	0,0730±0,0019	0,1954±0,0016	0,0336±0,0017
segment 9	0,1166±0,0025	**0,0277±0,0024**	0,0702±0,0015	0,1926±0,0020	0,0323±0,0011
segment 10	0,1128±0,0022	**0,0265±0,0023**	0,0750±0,0015	0,1967±0,0023	0,0291±0,0014

Tabela 41: Valores de erro médio e desvio padrão para *Bagging* quando aplicado às bases "sick", "sonar" e "vehicle"

Base	RBF	J48	SMO	NB	IBk
sick 1	0,0367±0,0016	**0,0120±0,0010**	0,0613±0,0003	0,0725±0,0014	0,0414±0,0015
sick 2	0,0375±0,0012	**0,0131±0,0006**	0,0610±0,0003	0,0715±0,0015	0,0527±0,0016
sick 3	0,0363±0,0012	**0,0115±0,0008**	0,0613±0,0002	0,0722±0,0010	0,0398±0,0016
sick 4	0,0359±0,0015	**0,0124±0,0010**	0,0613±0,0002	0,0716±0,0017	0,0417±0,0015
sick 5	0,0353±0,0008	**0,0124±0,0012**	0,0612±0,0004	0,0675±0,0009	0,0406±0,0025
sick 6	0,0364±0,0011	**0,0120±0,0014**	0,0615±0,0004	0,0706±0,0016	0,0397±0,0017
sick 7	0,0368±0,0017	**0,0110±0,0008**	0,0610±0,0004	0,0916±0,0017	0,0382±0,0017
sick 8	0,0363±0,0013	**0,0115±0,0016**	0,0611±0,0002	0,0717±0,0021	0,0433±0,0010
sick 9	0,0364±0,0014	**0,0132±0,0011**	0,0610±0,0002	0,0727±0,0018	0,0411±0,0012
sick 10	0,0365±0,0014	**0,0115±0,0009**	0,0610±0,0003	0,0709±0,0017	0,0377±0,0020
sonar 1	0,2209±0,0275	0,2112±0,0210	0,2214±0,0168	0,3155±0,0163	**0,1342±0,0096**
sonar 2	0,2316±0,0149	0,2369±0,0234	0,2342±0,0090	0,3310±0,0154	**0,1524±0,0108**
sonar 3	0,2353±0,0178	0,2332±0,0268	0,2358±0,0089	0,3166±0,0142	**0,1353±0,0126**
sonar 4	0,2005±0,0132	0,2241±0,0152	0,2037±0,0134	0,3000±0,0144	**0,1422±0,0098**
sonar 5	0,2203±0,0128	0,2364±0,0270	0,2273±0,0170	0,3176±0,0129	**0,1519±0,0116**
sonar 6	0,2235±0,0234	0,2310±0,0142	0,2166±0,0211	0,3278±0,0147	**0,1594±0,0100**
sonar 7	0,2219±0,0180	0,2262±0,0134	0,2369±0,0098	0,3214±0,0164	**0,1326±0,0106**
sonar 8	0,2203±0,0257	0,2043±0,0249	0,2374±0,0173	0,3000±0,0102	**0,1310±0,0132**
sonar 9	0,2324±0,0159	0,2186±0,0264	0,2271±0,0148	0,3176±0,0130	**0,1527±0,0144**
sonar 10	0,2250±0,0159	0,2149±0,0246	0,2266±0,0181	0,3197±0,0134	**0,1426±0,0137**
vehicle 1	0,3385±0,0073	**0,2501±0,0088**	0,2625±0,0059	0,5518±0,0092	0,3074±0,0081
vehicle 2	0,3329±0,0081	**0,2574±0,0114**	0,2585±0,0046	0,5417±0,0074	0,3076±0,0078
vehicle 3	0,3353±0,0087	**0,2527±0,0059**	0,2652±0,0064	0,5505±0,0064	0,3041±0,0095
vehicle 4	0,3296±0,0098	**0,2568±0,0089**	0,2573±0,0060	0,5248±0,0068	0,3020±0,0133
vehicle 5	0,3311±0,0079	0,2632±0,0112	**0,2585±0,0054**	0,5510±0,0063	0,3068±0,0096
vehicle 6	0,3372±0,0067	**0,2576±0,0111**	0,2649±0,0053	0,5431±0,0130	0,3092±0,0060
vehicle 7	0,3270±0,0114	**0,2619±0,0085**	0,2630±0,0062	0,5483±0,0089	0,3073±0,0080
vehicle 8	0,3361±0,0072	**0,2643±0,0152**	0,2690±0,0044	0,5545±0,0089	0,3161±0,0091
vehicle 9	0,3364±0,0057	0,2634±0,0102	**0,2564±0,0055**	0,5496±0,0077	0,2966±0,0102
vehicle 10	0,3251±0,0042	**0,2560±0,0114**	0,2629±0,0074	0,5362±0,0066	0,2925±0,0066

Tabela 42: Valores de erro médio e desvio padrão para *Bagging* quando aplicado às bases "vote" e "waveform-5000"

Base	RBF	J48	SMO	NB	IBk
vote 1	0,0604±0,0062	**0,0427±0,0034**	0,0463±0,0043	0,1020±0,0028	0,0816±0,0056
vote 2	0,0537±0,0042	**0,0407±0,0053**	0,0437±0,0039	0,1059±0,0022	0,0875±0,0046
vote 3	0,0465±0,0051	**0,0376±0,0032**	0,0384±0,0032	0,1028±0,0024	0,0788±0,0034
vote 4	0,0596±0,0066	**0,0430±0,0052**	0,0448±0,0037	0,0985±0,0025	0,0821±0,0037
vote 5	0,0558±0,0055	0,0422±0,0057	**0,0358±0,0042**	0,1008±0,0022	0,0760±0,0042
vote 6	0,0622±0,0038	0,0436±0,0049	**0,0413±0,0048**	0,0936±0,0012	0,0758±0,0043
vote 7	0,0633±0,0031	0,0416±0,0042	**0,0390±0,0055**	0,1003±0,0017	0,0763±0,0042
vote 8	0,0543±0,0021	**0,0367±0,0032**	0,0398±0,0044	0,0941±0,0014	0,0702±0,0065
vote 9	0,0531±0,0048	**0,0301±0,0026**	0,0411±0,0041	0,0946±0,0019	0,0707±0,0038
vote 10	0,0533±0,0059	**0,0329±0,0037**	0,0365±0,0032	0,1015±0,0023	0,0750±0,0045
waveform-5000 1	0,1471±0,0010	0,1895±0,0043	**0,1361±0,0017**	0,1995±0,0009	0,2615±0,0034
waveform-5000 2	0,1494±0,0015	0,1868±0,0055	**0,1379±0,0019**	0,2019±0,0009	0,2647±0,0026
waveform-5000 3	0,1472±0,0012	0,1861±0,0031	**0,1356±0,0016**	0,1998±0,0013	0,2618±0,0044
waveform-5000 4	0,1470±0,0014	0,1866±0,0037	**0,1355±0,0015**	0,2016±0,0009	0,2624±0,0045
waveform-5000 5	0,1442±0,0011	0,1837±0,0044	**0,1316±0,0019**	0,1996±0,0010	0,2678±0,0039
waveform-5000 6	0,1455±0,0014	0,1841±0,0040	**0,1356±0,0012**	0,1999±0,0010	0,2575±0,0031
waveform-5000 7	0,1447±0,0019	0,1834±0,0055	**0,1358±0,0021**	0,1985±0,0009	0,2634±0,0039
waveform-5000 8	0,1492±0,0011	0,1879±0,0042	**0,1351±0,0024**	0,1981±0,0011	0,2638±0,0026
waveform-5000 9	0,1460±0,0015	0,1844±0,0042	**0,1358±0,0011**	0,1997±0,0008	0,2619±0,0048
waveform-5000 10	0,1474±0,0008	0,1828±0,0045	**0,1366±0,0014**	0,2014±0,0012	0,2667±0,0036

Tabela 43: Valores de erro médio e desvio padrão para *Boosting* quando aplicado às bases "anneal", "breast-cancer" e "bupa"

Base	RBF	J48	SMO	NB	IBk
anneal 1	0,0270±0,0053	**0,0062±0,0018**	0,0077±0,0017	0,0616±0,0047	0,0127±0,0031
anneal 2	0,0245±0,0027	**0,0077±0,0022**	0,0088±0,0016	0,0632±0,0047	0,0105±0,0028
anneal 3	0,0261±0,0053	**0,0054±0,0015**	0,0082±0,0012	0,0563±0,0070	0,0132±0,0023
anneal 4	0,0259±0,0045	**0,0066±0,0027**	0,0089±0,0021	0,0587±0,0045	0,0121±0,0028
anneal 5	0,0255±0,0072	**0,0047±0,0014**	0,0067±0,0021	0,0562±0,0037	0,0109±0,0040
anneal 6	0,0197±0,0040	0,0068±0,0017	**0,0054±0,0015**	0,0533±0,0058	0,0137±0,0026
anneal 7	0,0234±0,0035	**0,0067±0,0016**	0,0089±0,0015	0,0563±0,0063	0,0136±0,0042
anneal 8	0,0229±0,0034	**0,0061±0,0026**	0,0064±0,0017	0,0577±0,0058	0,0135±0,0027
anneal 9	0,0224±0,0026	**0,0042±0,0022**	0,0066±0,0018	0,0561±0,0053	0,0100±0,0024
anneal 10	0,0221±0,0050	**0,0062±0,0015**	0,0072±0,0010	0,0507±0,0054	0,0145±0,0018
breast-cancer 1	0,3237±0,0157	0,3284±0,0145	0,3191±0,0153	0,3191±0,0137	**0,2953±0,0091**
breast-cancer 2	0,3156±0,0195	0,3230±0,0157	0,3097±0,0097	0,3198±0,0150	**0,2615±0,0100**
breast-cancer 3	0,3016±0,0241	0,3179±0,0080	0,3086±0,0181	0,3093±0,0162	**0,2720±0,0093**
breast-cancer 4	0,3160±0,0218	0,3183±0,0259	0,3066±0,0217	0,3171±0,0199	**0,2805±0,0156**
breast-cancer 5	0,3089±0,0139	0,3362±0,0166	**0,3078±0,0167**	0,3300±0,0195	0,3249±0,0096
breast-cancer 6	**0,3222±0,0251**	0,3619±0,0184	0,3370±0,0155	0,3342±0,0199	0,3564±0,0107
breast-cancer 7	0,3229±0,0146	0,3349±0,0212	0,3163±0,0082	0,3275±0,0130	**0,3050±0,0095**
breast-cancer 8	0,3178±0,0235	0,3329±0,0147	0,3174±0,0175	0,3248±0,0128	**0,3019±0,0085**
breast-cancer 9	0,3248±0,0188	0,3442±0,0085	0,3109±0,0164	**0,2791±0,0140**	0,3124±0,0107
breast-cancer 10	0,3314±0,0246	0,3434±0,0254	0,3248±0,0239	0,3306±0,0189	**0,3027±0,0118**
bupa 1	0,3555±0,0204	**0,3045±0,0231**	0,3677±0,0142	0,3500±0,0200	0,3687±0,0068
bupa 2	0,3432±0,0174	**0,3216±0,0265**	0,3610±0,0123	0,3668±0,0165	0,3745±0,0099
bupa 3	0,3658±0,0188	**0,3116±0,0164**	0,3523±0,0169	0,3326±0,0125	0,3516±0,0154
bupa 4	0,3429±0,0120	**0,3306±0,0157**	0,3635±0,0230	0,3413±0,0169	0,3661±0,0112
bupa 5	0,3400±0,0161	**0,3126±0,0276**	0,3568±0,0249	0,3610±0,0253	0,4023±0,0140
bupa 6	0,3492±0,0213	**0,3151±0,0144**	0,3556±0,0174	0,3469±0,0187	0,4077±0,0142
bupa 7	0,3428±0,0135	**0,3051±0,0229**	0,3559±0,0179	0,3431±0,0191	0,3961±0,0187
bupa 8	0,3241±0,0150	**0,3228±0,0192**	0,3582±0,0176	0,3379±0,0245	0,3650±0,0170
bupa 9	0,3302±0,0208	**0,3090±0,0134**	0,3633±0,0147	0,3469±0,0200	0,3463±0,0133
bupa 10	0,3521±0,0131	**0,3045±0,0176**	0,3685±0,0186	0,3672±0,0217	0,3653±0,0154

Tabela 44: Valores de erro médio e desvio padrão para *Boosting* quando aplicado às bases “car”, “colic” e “credit-a”

Base	RBF	J48	SMO	NB	IBk
car 1	0,0717±0,0045	**0,0448±0,0045**	0,0612±0,0026	0,0939±0,0058	0,0753±0,0038
car 2	0,0698±0,0056	**0,0472±0,0048**	0,0558±0,0044	0,0949±0,0037	0,0743±0,0027
car 3	0,0689±0,0072	**0,0509±0,0042**	0,0588±0,0031	0,0970±0,0041	0,0767±0,0036
car 4	0,0731±0,0031	**0,0476±0,0031**	0,0604±0,0045	0,0979±0,0051	0,0765±0,0036
car 5	0,0676±0,0085	**0,0395±0,0031**	0,0541±0,0037	0,0898±0,0038	0,0720±0,0030
car 6	0,0704±0,0074	**0,0480±0,0029**	0,0605±0,0034	0,1015±0,0050	0,0794±0,0031
car 7	0,0756±0,0056	**0,0457±0,0033**	0,0570±0,0032	0,0990±0,0052	0,0793±0,0029
car 8	0,0725±0,0054	**0,0521±0,0039**	0,0621±0,0027	0,0950±0,0044	0,0774±0,0046
car 9	0,0721±0,0052	**0,0456±0,0030**	0,0546±0,0027	0,0987±0,0029	0,0730±0,0029
car 10	0,0725±0,0075	**0,0461±0,0062**	0,0646±0,0029	0,1016±0,0036	0,0752±0,0035
colic 1	0,2154±0,0216	**0,1852±0,0117**	0,2112±0,0168	0,2287±0,0166	0,2151±0,0072
colic 2	0,1906±0,0158	**0,1698±0,0140**	0,1804±0,0079	0,1967±0,0118	0,2024±0,0062
colic 3	0,2224±0,0221	**0,1961±0,0171**	0,2224±0,0107	0,2314±0,0180	0,2166±0,0059
colic 4	0,2091±0,0132	**0,1861±0,0176**	0,2163±0,0127	0,2193±0,0145	0,2148±0,0064
colic 5	0,2109±0,0151	**0,1906±0,0117**	0,2227±0,0120	0,2251±0,0163	0,2199±0,0091
colic 6	0,2073±0,0130	**0,1906±0,0044**	0,2036±0,0091	0,2057±0,0147	0,2112±0,0064
colic 7	0,2015±0,0116	**0,1770±0,0128**	0,2076±0,0118	0,2121±0,0089	0,2187±0,0066
colic 8	0,2184±0,0230	**0,1861±0,0119**	0,2127±0,0202	0,2193±0,0118	0,1991±0,0072
colic 9	0,1970±0,0121	**0,1771±0,0128**	0,2075±0,0078	0,2292±0,0209	0,2036±0,0079
colic 10	0,2190±0,0129	**0,1843±0,0161**	0,2199±0,0172	0,2298±0,0207	0,2111±0,0113
credit-a 1	0,1779±0,0093	**0,1533±0,0099**	0,1565±0,0062	0,1841±0,0100	0,1884±0,0094
credit-a 2	0,1775±0,0092	**0,1647±0,0077**	0,1731±0,0087	0,1876±0,0049	0,1926±0,0076
credit-a 3	0,1768±0,0053	**0,1670±0,0121**	0,1762±0,0100	0,1952±0,0058	0,1881±0,0071
credit-a 4	0,1757±0,0094	0,1583±0,0052	**0,1535±0,0069**	0,1874±0,0059	0,1709±0,0067
credit-a 5	0,1684±0,0121	**0,1585±0,0120**	0,1623±0,0124	0,1852±0,0077	0,1826±0,0070
credit-a 6	0,1715±0,0083	**0,1614±0,0094**	0,1641±0,0043	0,1878±0,0043	0,1903±0,0067
credit-a 7	0,1709±0,0108	0,1593±0,0099	**0,1591±0,0093**	0,1881±0,0066	0,1965±0,0055
credit-a 8	0,1792±0,0109	**0,1609±0,0104**	0,1694±0,0077	0,1899±0,0069	0,1910±0,0071
credit-a 9	0,1589±0,0047	**0,1452±0,0093**	0,1520±0,0081	0,1818±0,0074	0,1800±0,0041
credit-a 10	0,1717±0,0123	**0,1580±0,0077**	0,1631±0,0072	0,1895±0,0043	0,1833±0,0061

Tabela 45: Valores de erro médio e desvio padrão para *Boosting* quando aplicado às bases "diabetes", "gaussian3" e "glass"

Base	RBF	J48	SMO	NB	IBk
diabetes 1	0,2560±0,0065	0,2673±0,0131	**0,2294±0,0034**	0,2366±0,0083	0,2946±0,0058
diabetes 2	0,2670±0,0084	0,2816±0,0097	**0,2313±0,0037**	0,2493±0,0050	0,2928±0,0093
diabetes 3	0,2611±0,0088	0,2712±0,0080	**0,2302±0,0035**	0,2449±0,0065	0,3142±0,0066
diabetes 4	0,2648±0,0071	0,2705±0,0153	**0,2326±0,0036**	0,2440±0,0058	0,3217±0,0082
diabetes 5	0,2562±0,0113	0,2761±0,0132	**0,2323±0,0026**	0,2470±0,0063	0,2863±0,0084
diabetes 6	0,2508±0,0093	0,2648±0,0037	**0,2276±0,0033**	0,2346±0,0062	0,2951±0,0088
diabetes 7	0,2705±0,0069	0,2828±0,0143	**0,2343±0,0038**	0,2483±0,0076	0,3145±0,0075
diabetes 8	0,2630±0,0080	0,2810±0,0091	**0,2308±0,0032**	0,2476±0,0081	0,2987±0,0083
diabetes 9	0,2548±0,0059	0,2751±0,0059	**0,2269±0,0035**	0,2340±0,0052	0,2873±0,0064
diabetes 10	0,2464±0,0098	0,2718±0,0079	**0,2279±0,0039**	0,2374±0,0095	0,2845±0,0075
gaussian3 1	0,3574±0,0517	0,3389±0,0693	**0,0000±0,0000**	0,3278±0,0624	0,0111±0,0096
gaussian3 2	0,3722±0,0701	0,3056±0,0496	**0,0000±0,0000**	0,3444±0,0532	0,0111±0,0096
gaussian3 3	0,4019±0,0462	0,3204±0,0446	**0,0000±0,0000**	0,3815±0,0292	0,0185±0,0087
gaussian3 4	0,4407±0,0443	0,3222±0,0865	**0,0000±0,0000**	0,3926±0,0544	0,0130±0,0089
gaussian3 5	0,3333±0,0262	0,3278±0,0401	**0,0000±0,0000**	0,2907±0,0391	0,0111±0,0096
gaussian3 6	0,3759±0,0517	0,2537±0,0642	**0,0000±0,0000**	0,3278±0,0462	0,0111±0,0096
gaussian3 7	0,3852±0,0551	0,3278±0,0592	**0,0000±0,0000**	0,3556±0,0484	0,0000±0,0000
gaussian3 8	0,3667±0,0452	0,3519±0,0486	**0,0000±0,0000**	0,3204±0,0559	0,0000±0,0000
gaussian3 9	0,4667±0,0398	0,2722±0,0327	**0,0000±0,0000**	0,3870±0,0332	0,0167±0,0059
gaussian3 10	0,3333±0,0972	0,3519±0,0642	**0,0000±0,0000**	0,2704±0,0411	0,0167±0,0059
glass 1	0,3547±0,0126	**0,2776±0,0179**	0,3901±0,0278	0,5073±0,0130	0,3219±0,0136
glass 2	0,3104±0,0187	**0,2542±0,0188**	0,4219±0,0175	0,5094±0,0172	0,3042±0,0113
glass 3	0,3385±0,0184	**0,2505±0,0259**	0,4057±0,0166	0,5073±0,0148	0,3260±0,0121
glass 4	0,3568±0,0186	**0,2719±0,0201**	0,3906±0,0139	0,5141±0,0141	0,3130±0,0135
glass 5	0,3420±0,0260	**0,2611±0,0217**	0,4098±0,0206	0,5114±0,0127	0,3415±0,0133
glass 6	0,3176±0,0280	**0,2539±0,0238**	0,4073±0,0181	0,5078±0,0183	0,2896±0,0141
glass 7	0,3192±0,0161	**0,2674±0,0127**	0,4062±0,0107	0,5311±0,0132	0,3062±0,0135
glass 8	0,3306±0,0231	**0,2497±0,0106**	0,4368±0,0206	0,5705±0,0190	0,3212±0,0147
glass 9	0,2917±0,0160	**0,2145±0,0161**	0,4181±0,0120	0,5394±0,0182	0,2782±0,0127
glass 10	0,3212±0,0287	**0,2606±0,0158**	0,4067±0,0159	0,4922±0,0142	0,3067±0,0138

Tabela 46: Valores de erro médio e desvio padrão para *Boosting* quando aplicado às bases "haberman", "heart-c" e "hepatitis"

Base	RBF	J48	SMO	NB	IBk
haberman 1	0,2967±0,0188	0,3164±0,0173	0,2705±0,0102	**0,2633±0,0099**	0,3480±0,0111
haberman 2	0,3004±0,0150	0,3229±0,0193	**0,2600±0,0132**	0,2647±0,0103	0,3465±0,0115
haberman 3	0,2964±0,0162	0,3196±0,0129	**0,2775±0,0153**	0,2815±0,0132	0,3484±0,0098
haberman 4	0,2804±0,0131	0,2829±0,0176	0,2618±0,0152	**0,2571±0,0134**	0,3345±0,0104
haberman 5	0,2942±0,0135	0,3204±0,0132	0,2636±0,0099	**0,2629±0,0080**	0,3516±0,0060
haberman 6	0,2705±0,0121	0,2902±0,0164	0,2531±0,0089	**0,2469±0,0053**	0,3473±0,0142
haberman 7	0,2920±0,0199	0,3101±0,0215	0,2641±0,0089	**0,2565±0,0121**	0,3489±0,0128
haberman 8	0,3014±0,0123	0,3174±0,0144	0,2641±0,0121	**0,2565±0,0101**	0,3391±0,0097
haberman 9	0,2819±0,0161	0,3018±0,0064	0,2507±0,0074	**0,2442±0,0117**	0,3174±0,0103
haberman 10	0,2917±0,0160	0,3109±0,0133	0,2670±0,0139	**0,2558±0,0152**	0,3261±0,0085
heart-c 1	0,1879±0,0140	0,2022±0,0172	**0,1592±0,0098**	0,1621±0,0106	0,2349±0,0061
heart-c 2	0,1912±0,0168	0,2132±0,0144	0,1691±0,0110	**0,1676±0,0138**	0,2272±0,0064
heart-c 3	0,1967±0,0085	0,2070±0,0192	**0,1717±0,0126**	0,1743±0,0119	0,2489±0,0119
heart-c 4	0,1821±0,0121	0,2055±0,0185	0,1696±0,0113	**0,1626±0,0078**	0,2286±0,0112
heart-c 5	0,2026±0,0127	0,2286±0,0227	0,1879±0,0093	**0,1813±0,0103**	0,2630±0,0084
heart-c 6	0,2070±0,0202	0,2158±0,0190	**0,1740±0,0097**	0,1802±0,0149	0,2546±0,0074
heart-c 7	0,1960±0,0189	0,2209±0,0185	**0,1740±0,0117**	0,1744±0,0101	0,2502±0,0108
heart-c 8	0,1791±0,0190	0,2055±0,0158	**0,1538±0,0099**	0,1656±0,0094	0,2443±0,0107
heart-c 9	0,1813±0,0193	0,2044±0,0143	**0,1505±0,0082**	0,1641±0,0100	0,2300±0,0066
heart-c 10	**0,1747±0,0189**	0,2081±0,0138	0,1784±0,0065	0,1751±0,0117	0,2337±0,0096
hepatitis 1	**0,1647±0,0153**	0,1705±0,0230	0,1741±0,0194	0,1777±0,0201	0,1950±0,0115
hepatitis 2	0,1683±0,0152	0,1698±0,0215	0,1791±0,0110	**0,1647±0,0146**	0,1827±0,0103
hepatitis 3	**0,1698±0,0149**	0,1799±0,0179	0,1777±0,0225	0,1741±0,0165	0,1784±0,0147
hepatitis 4	0,1820±0,0198	0,1871±0,0269	0,1892±0,0212	**0,1691±0,0137**	0,2259±0,0152
hepatitis 5	0,1647±0,0208	0,1806±0,0138	0,1691±0,0167	**0,1626±0,0225**	0,1971±0,0170
hepatitis 6	**0,1436±0,0180**	0,1586±0,0199	0,1700±0,0146	0,1643±0,0158	0,2071±0,0158
hepatitis 7	0,1750±0,0162	0,1843±0,0125	0,1821±0,0176	**0,1664±0,0178**	0,1879±0,0126
hepatitis 8	**0,1414±0,0230**	0,1850±0,0152	0,1700±0,0223	0,1579±0,0086	0,1679±0,0113
hepatitis 9	**0,1714±0,0254**	0,1729±0,0223	0,1843±0,0157	0,1736±0,0162	0,2100±0,0108
hepatitis 10	**0,1586±0,0306**	0,1764±0,0221	0,1657±0,0125	0,1629±0,0120	0,1821±0,0179

Tabela 47: Valores de erro médio e desvio padrão para *Boosting* quando aplicado às bases "ionosphere", "iris" e "segment"

Base	RBF	J48	SMO	NB	IBk
ionosphere 1	0,0784±0,0092	**0,0749±0,0132**	0,1152±0,0091	0,0949±0,0093	0,1206±0,0067
ionosphere 2	**0,0592±0,0094**	0,0630±0,0099	0,1092±0,0122	0,0712±0,0050	0,1190±0,0065
ionosphere 3	0,0734±0,0087	**0,0731±0,0125**	0,1209±0,0140	0,1022±0,0092	0,1373±0,0062
ionosphere 4	0,0785±0,0074	**0,0712±0,0078**	0,1130±0,0094	0,0880±0,0080	0,1339±0,0047
ionosphere 5	0,0737±0,0076	**0,0728±0,0127**	0,1174±0,0103	0,0908±0,0113	0,1405±0,0062
ionosphere 6	0,0703±0,0139	**0,0677±0,0099**	0,1003±0,0086	0,0839±0,0072	0,1358±0,0055
ionosphere 7	0,0734±0,0055	**0,0706±0,0052**	0,1133±0,0108	0,0870±0,0110	0,1399±0,0068
ionosphere 8	0,0801±0,0080	**0,0722±0,0089**	0,1127±0,0065	0,0870±0,0098	0,1373±0,0064
ionosphere 9	0,0759±0,0117	**0,0734±0,0076**	0,1079±0,0080	0,0918±0,0070	0,1225±0,0068
ionosphere 10	0,0759±0,0096	**0,0655±0,0083**	0,1203±0,0090	0,0861±0,0104	0,1392±0,0039
iris 1	0,0444±0,0078	0,0526±0,0074	**0,0274±0,0086**	0,0400±0,0087	0,0378±0,0023
iris 2	0,0563±0,0106	0,0615±0,0116	**0,0363±0,0065**	0,0607±0,0109	0,0511±0,0023
iris 3	0,0563±0,0117	0,0607±0,0091	**0,0289±0,0065**	0,0467±0,0105	0,0452±0,0023
iris 4	0,0407±0,0117	0,0437±0,0113	**0,0274±0,0116**	0,0400±0,0117	0,0356±0,0058
iris 5	0,0607±0,0130	0,0733±0,0128	**0,0400±0,0100**	0,0489±0,0100	0,0444±0,0035
iris 6	0,0548±0,0094	0,0667±0,0135	0,0481±0,0112	0,0489±0,0100	**0,0474±0,0072**
iris 7	0,0593±0,0070	0,0630±0,0087	**0,0407±0,0072**	0,0474±0,0157	0,0444±0,0049
iris 8	0,0459±0,0058	0,0533±0,0155	**0,0341±0,0094**	0,0459±0,0098	0,0585±0,0042
iris 9	0,0533±0,0077	0,0756±0,0177	**0,0289±0,0123**	0,0556±0,0080	0,0504±0,0068
iris 10	0,0489±0,0106	0,0548±0,0062	**0,0385±0,0084**	0,0400±0,0072	0,0430±0,0047
segment 1	0,0924±0,0061	**0,0195±0,0017**	0,0701±0,0019	0,2012±0,0015	0,0347±0,0014
segment 2	0,0902±0,0077	**0,0174±0,0018**	0,0726±0,0013	0,2007±0,0025	0,0296±0,0020
segment 3	0,0865±0,0073	**0,0188±0,0021**	0,0744±0,0035	0,1971±0,0013	0,0311±0,0021
segment 4	0,0837±0,0089	**0,0197±0,0021**	0,0791±0,0022	0,1997±0,0016	0,0327±0,0019
segment 5	0,0911±0,0073	**0,0183±0,0029**	0,0776±0,0014	0,2002±0,0016	0,0288±0,0020
segment 6	0,0960±0,0066	**0,0187±0,0013**	0,0730±0,0021	0,2000±0,0011	0,0285±0,0024
segment 7	0,0930±0,0065	**0,0214±0,0015**	0,0715±0,0015	0,1962±0,0012	0,0287±0,0022
segment 8	0,0906±0,0048	**0,0202±0,0014**	0,0714±0,0027	0,1963±0,0011	0,0327±0,0018
segment 9	0,0848±0,0063	**0,0199±0,0015**	0,0695±0,0024	0,1935±0,0020	0,0316±0,0011
segment 10	0,0848±0,0059	**0,0169±0,0018**	0,0742±0,0028	0,1981±0,0015	0,0286±0,0010

Tabela 48: Valores de erro médio e desvio padrão para *Boosting* quando aplicado às bases "sick", "sonar" e "vehicle"

Base	RBF	J48	SMO	NB	IBk
sick 1	0,0425±0,0022	**0,0096±0,0011**	0,0573±0,0033	0,0669±0,0033	0,0406±0,0016
sick 2	0,0446±0,0032	**0,0114±0,0016**	0,0566±0,0033	0,0677±0,0031	0,0533±0,0011
sick 3	0,0439±0,0032	**0,0098±0,0011**	0,0565±0,0035	0,0668±0,0019	0,0396±0,0014
sick 4	0,0442±0,0040	**0,0106±0,0010**	0,0551±0,0042	0,0648±0,0033	0,0410±0,0013
sick 5	0,0436±0,0032	**0,0101±0,0009**	0,0594±0,0029	0,0664±0,0026	0,0403±0,0018
sick 6	0,0434±0,0032	**0,0101±0,0007**	0,0590±0,0049	0,0627±0,0035	0,0395±0,0014
sick 7	0,0411±0,0023	**0,0099±0,0010**	0,0566±0,0034	0,0700±0,0036	0,0370±0,0013
sick 8	0,0419±0,0026	**0,0110±0,0008**	0,0555±0,0044	0,0691±0,0019	0,0427±0,0010
sick 9	0,0430±0,0018	**0,0116±0,0010**	0,0528±0,0032	0,0666±0,0040	0,0394±0,0009
sick 10	0,0414±0,0022	**0,0105±0,0007**	0,0584±0,0048	0,0602±0,0034	0,0373±0,0008
sonar 1	0,1722±0,0238	0,2112±0,0189	0,2332±0,0189	0,2016±0,0212	**0,1337±0,0116**
sonar 2	0,1904±0,0330	0,2273±0,0282	0,2465±0,0245	0,2150±0,0203	**0,1540±0,0094**
sonar 3	0,1818±0,0333	0,2128±0,0157	0,2385±0,0156	0,2193±0,0240	**0,1283±0,0110**
sonar 4	0,1604±0,0235	0,1941±0,0220	0,2096±0,0188	0,1914±0,0201	**0,1433±0,0100**
sonar 5	0,1642±0,0238	0,2011±0,0269	0,2342±0,0204	0,1866±0,0154	**0,1513±0,0124**
sonar 6	0,1920±0,0327	0,2037±0,0232	0,2273±0,0058	0,2011±0,0192	**0,1561±0,0083**
sonar 7	0,1743±0,0200	0,2096±0,0174	0,2428±0,0217	0,1995±0,0194	**0,1267±0,0138**
sonar 8	0,1791±0,0194	0,1824±0,0226	0,1930±0,0225	0,1888±0,0260	**0,1358±0,0124**
sonar 9	0,1963±0,0254	0,1979±0,0280	0,2394±0,0177	0,1856±0,0203	**0,1527±0,0135**
sonar 10	0,1787±0,0157	0,2011±0,0293	0,2271±0,0112	0,2027±0,0229	**0,1394±0,0117**
vehicle 1	0,3302±0,0130	**0,2378±0,0103**	0,2620±0,0092	0,5572±0,0084	0,3097±0,0078
vehicle 2	0,3364±0,0107	**0,2371±0,0055**	0,2598±0,0048	0,5428±0,0081	0,3085±0,0081
vehicle 3	0,3252±0,0068	**0,2365±0,0158**	0,2678±0,0047	0,5552±0,0061	0,3042±0,0080
vehicle 4	0,3147±0,0109	**0,2344±0,0120**	0,2624±0,0064	0,5326±0,0055	0,3007±0,0115
vehicle 5	0,3269±0,0116	**0,2461±0,0098**	0,2606±0,0042	0,5568±0,0081	0,3080±0,0091
vehicle 6	0,3284±0,0123	**0,2373±0,0154**	0,2643±0,0076	0,5511±0,0081	0,3097±0,0063
vehicle 7	0,3176±0,0089	**0,2339±0,0132**	0,2665±0,0041	0,5512±0,0082	0,3073±0,0074
vehicle 8	0,3224±0,0066	**0,2453±0,0118**	0,2717±0,0061	0,5601±0,0070	0,3157±0,0091
vehicle 9	0,3192±0,0126	**0,2400±0,0089**	0,2583±0,0054	0,5551±0,0065	0,2953±0,0062
vehicle 10	0,3203±0,0092	**0,2270±0,0112**	0,2622±0,0061	0,5455±0,0050	0,2957±0,0049

Tabela 49: Valores de erro médio e desvio padrão para *Boosting* quando aplicado às bases "vote" e "waveform-5000"

Base	RBF	J48	SMO	NB	IBk
vote 1	0,0568±0,0056	0,0522±0,0053	**0,0473±0,0050**	0,0537±0,0071	0,0772±0,0041
vote 2	0,0527±0,0049	0,0460±0,0078	**0,0435±0,0051**	0,0460±0,0030	0,0890±0,0043
vote 3	0,0455±0,0070	0,0453±0,0042	**0,0412±0,0058**	0,0478±0,0040	0,0734±0,0021
vote 4	0,0545±0,0065	0,0519±0,0087	**0,0450±0,0053**	0,0476±0,0032	0,0777±0,0022
vote 5	0,0514±0,0067	0,0486±0,0060	**0,0458±0,0046**	0,0476±0,0027	0,0762±0,0041
vote 6	0,0536±0,0082	0,0577±0,0065	**0,0436±0,0080**	0,0490±0,0061	0,0786±0,0029
vote 7	0,0523±0,0086	0,0485±0,0058	**0,0390±0,0064**	0,0495±0,0058	0,0765±0,0050
vote 8	0,0490±0,0075	0,0490±0,0075	0,0421±0,0058	**0,0406±0,0035**	0,0681±0,0068
vote 9	0,0503±0,0052	0,0416±0,0073	**0,0406±0,0061**	0,0492±0,0043	0,0681±0,0042
vote 10	0,0564±0,0069	**0,0441±0,0051**	0,0454±0,0048	0,0452±0,0032	0,0699±0,0025
waveform-5000 1	0,1534±0,0017	0,1907±0,0041	**0,1369±0,0016**	0,1996±0,0009	0,2648±0,0020
waveform-5000 2	0,1550±0,0026	0,1889±0,0025	**0,1382±0,0019**	0,2020±0,0007	0,2695±0,0021
waveform-5000 3	0,1516±0,0020	0,1873±0,0044	**0,1364±0,0013**	0,1995±0,0009	0,2640±0,0026
waveform-5000 4	0,1554±0,0018	0,1864±0,0028	**0,1367±0,0018**	0,2017±0,0005	0,2654±0,0029
waveform-5000 5	0,1521±0,0021	0,1818±0,0048	**0,1321±0,0012**	0,1993±0,0010	0,2708±0,0025
waveform-5000 6	0,1512±0,0026	0,1853±0,0049	**0,1361±0,0013**	0,2001±0,0008	0,2590±0,0020
waveform-5000 7	0,1510±0,0020	0,1899±0,0045	**0,1347±0,0011**	0,1986±0,0007	0,2672±0,0024
waveform-5000 8	0,1542±0,0021	0,1893±0,0032	**0,1347±0,0021**	0,1979±0,0007	0,2658±0,0027
waveform-5000 9	0,1518±0,0027	0,1835±0,0048	**0,1358±0,0023**	0,1995±0,0007	0,2645±0,0034
waveform-5000 10	0,1517±0,0024	0,1856±0,0045	**0,1353±0,0014**	0,2015±0,0009	0,2695±0,0022

Tabela 50: Valores de erro médio e desvio padrão para *Multiboosting* quando aplicado às bases "anneal", "breast-cancer" e "bupa"

Base	RBF	J48	SMO	NB	IBk
anneal 1	0,0358±0,0041	**0,0079±0,0017**	0,0097±0,0021	0,1207±0,0055	0,0127±0,0031
anneal 2	0,0339±0,0041	**0,0087±0,0023**	0,0095±0,0023	0,1277±0,0060	0,0105±0,0028
anneal 3	0,0314±0,0032	**0,0072±0,0014**	0,0078±0,0024	0,1130±0,0066	0,0132±0,0023
anneal 4	0,0363±0,0040	**0,0078±0,0014**	0,0095±0,0030	0,1172±0,0072	0,0121±0,0028
anneal 5	0,0317±0,0037	**0,0074±0,0013**	0,0079±0,0020	0,1161±0,0085	0,0109±0,0040
anneal 6	0,0281±0,0050	0,0071±0,0013	**0,0067±0,0013**	0,0994±0,0059	0,0137±0,0026
anneal 7	0,0265±0,0041	**0,0053±0,0014**	0,0100±0,0026	0,1078±0,0058	0,0136±0,0042
anneal 8	0,0344±0,0041	**0,0079±0,0022**	0,0080±0,0029	0,1150±0,0083	0,0135±0,0027
anneal 9	0,0314±0,0036	**0,0058±0,0041**	0,0073±0,0020	0,1125±0,0088	0,0100±0,0024
anneal 10	0,0324±0,0034	**0,0074±0,0015**	0,0098±0,0033	0,1072±0,0057	0,0145±0,0018
breast-cancer 1	0,2809±0,0088	0,3043±0,0125	0,3070±0,0141	0,2821±0,0088	**0,2689±0,0085**
breast-cancer 2	0,2840±0,0193	0,3000±0,0152	0,3121±0,0105	0,2848±0,0111	**0,2615±0,0107**
breast-cancer 3	**0,2700±0,0129**	0,2868±0,0140	0,2965±0,0121	0,2735±0,0132	0,2743±0,0147
breast-cancer 4	0,2720±0,0109	0,3012±0,0102	0,3023±0,0168	0,2844±0,0100	**0,2677±0,0127**
breast-cancer 5	**0,2708±0,0134**	0,2965±0,0245	0,3054±0,0131	0,2856±0,0129	0,2864±0,0131
breast-cancer 6	**0,3082±0,0151**	0,3249±0,0149	0,3226±0,0143	0,3132±0,0172	0,3230±0,0153
breast-cancer 7	**0,2802±0,0182**	0,2969±0,0211	0,3019±0,0104	0,2833±0,0144	0,2853±0,0095
breast-cancer 8	0,2837±0,0140	0,2903±0,0168	0,3043±0,0123	0,2845±0,0132	**0,2829±0,0095**
breast-cancer 9	0,2771±0,0133	0,2880±0,0124	0,2930±0,0159	**0,2663±0,0095**	0,2767±0,0110
breast-cancer 10	0,2826±0,0150	0,2946±0,0139	0,3085±0,0201	0,2934±0,0164	**0,2822±0,0101**
bupa 1	0,3484±0,0202	**0,2952±0,0162**	0,4035±0,0115	0,3884±0,0116	0,3687±0,0068
bupa 2	0,3552±0,0216	**0,3019±0,0168**	0,4194±0,0104	0,3871±0,0196	0,3745±0,0099
bupa 3	0,3594±0,0157	**0,2932±0,0149**	0,4103±0,0123	0,3626±0,0280	0,3516±0,0154
bupa 4	0,3445±0,0146	**0,3190±0,0157**	0,4094±0,0157	0,3787±0,0119	0,3661±0,0112
bupa 5	0,3426±0,0150	**0,3113±0,0123**	0,3939±0,0171	0,3923±0,0170	0,4023±0,0140
bupa 6	0,3447±0,0138	**0,3016±0,0203**	0,4106±0,0080	0,3675±0,0079	0,4077±0,0142
bupa 7	0,3469±0,0136	**0,2881±0,0187**	0,4006±0,0094	0,3794±0,0165	0,3961±0,0187
bupa 8	0,3241±0,0128	**0,3087±0,0209**	0,3936±0,0092	0,3701±0,0126	0,3650±0,0170
bupa 9	0,3418±0,0170	**0,2900±0,0182**	0,3965±0,0154	0,3756±0,0172	0,3463±0,0133
bupa 10	0,3569±0,0205	**0,2968±0,0110**	0,4090±0,0116	0,3916±0,0266	0,3653±0,0154

Tabela 51: Valores de erro médio e desvio padrão para *Multiboosting* quando aplicado às bases "car", "colic" e "credit-a"

Base	RBF	J48	SMO	NB	IBk
car 1	0,0808±0,0033	**0,0535±0,0042**	0,0639±0,0044	0,1192±0,0033	0,0753±0,0038
car 2	0,0821±0,0062	**0,0567±0,0045**	0,0599±0,0032	0,1223±0,0049	0,0743±0,0027
car 3	0,0796±0,0049	0,0602±0,0031	**0,0588±0,0032**	0,1200±0,0048	0,0767±0,0036
car 4	0,0822±0,0079	**0,0583±0,0045**	0,0636±0,0023	0,1224±0,0059	0,0765±0,0036
car 5	0,0792±0,0048	**0,0549±0,0031**	0,0570±0,0036	0,1189±0,0052	0,0720±0,0030
car 6	0,0815±0,0038	**0,0595±0,0031**	0,0622±0,0018	0,1224±0,0038	0,0794±0,0031
car 7	0,0854±0,0055	**0,0554±0,0023**	0,0597±0,0030	0,1244±0,0055	0,0793±0,0029
car 8	0,0850±0,0056	**0,0609±0,0041**	0,0648±0,0028	0,1193±0,0059	0,0774±0,0046
car 9	0,0840±0,0076	**0,0539±0,0038**	0,0600±0,0028	0,1246±0,0045	0,0730±0,0029
car 10	0,0847±0,0056	**0,0564±0,0036**	0,0695±0,0031	0,1249±0,0044	0,0752±0,0035
colic 1	0,1912±0,0062	**0,1589±0,0134**	0,1912±0,0106	0,1991±0,0103	0,2151±0,0072
colic 2	0,1710±0,0114	**0,1402±0,0052**	0,1580±0,0097	0,1873±0,0096	0,2024±0,0062
colic 3	0,2015±0,0118	**0,1577±0,0068**	0,2009±0,0085	0,2085±0,0094	0,2166±0,0059
colic 4	0,1918±0,0077	**0,1637±0,0086**	0,1991±0,0118	0,2018±0,0062	0,2148±0,0064
colic 5	0,1973±0,0124	**0,1601±0,0092**	0,1976±0,0140	0,2024±0,0091	0,2199±0,0091
colic 6	0,1867±0,0131	**0,1619±0,0083**	0,1858±0,0085	0,1885±0,0112	0,2112±0,0064
colic 7	0,1749±0,0113	**0,1438±0,0118**	0,1912±0,0131	0,1918±0,0112	0,2187±0,0066
colic 8	0,1982±0,0130	**0,1529±0,0117**	0,1982±0,0128	0,1985±0,0081	0,1991±0,0072
colic 9	0,1928±0,0154	**0,1587±0,0093**	0,1895±0,0110	0,2000±0,0075	0,2036±0,0079
colic 10	0,2048±0,0074	**0,1684±0,0150**	0,1958±0,0095	0,2021±0,0102	0,2111±0,0113
credit-a 1	0,1852±0,0101	**0,1436±0,0063**	0,1490±0,0042	0,2055±0,0070	0,1884±0,0094
credit-a 2	0,1789±0,0061	**0,1448±0,0111**	0,1512±0,0082	0,2101±0,0048	0,1926±0,0076
credit-a 3	0,1829±0,0076	**0,1454±0,0069**	0,1552±0,0090	0,2190±0,0061	0,1881±0,0071
credit-a 4	0,1771±0,0076	**0,1406±0,0095**	0,1496±0,0077	0,2130±0,0054	0,1709±0,0067
credit-a 5	0,1789±0,0069	**0,1370±0,0075**	0,1464±0,0078	0,2069±0,0052	0,1826±0,0070
credit-a 6	0,1833±0,0079	0,1506±0,0104	**0,1491±0,0055**	0,2071±0,0055	0,1903±0,0067
credit-a 7	0,1794±0,0082	**0,1457±0,0073**	0,1514±0,0065	0,2110±0,0055	0,1965±0,0055
credit-a 8	0,1799±0,0079	**0,1409±0,0094**	0,1543±0,0079	0,2143±0,0054	0,1910±0,0071
credit-a 9	0,1726±0,0121	**0,1330±0,0084**	0,1440±0,0078	0,2048±0,0057	0,1800±0,0041
credit-a 10	0,1752±0,0087	**0,1403±0,0052**	0,1522±0,0040	0,2093±0,0038	0,1833±0,0061

Tabela 52: Valores de erro médio e desvio padrão para *Multiboosting* quando aplicado às bases "diabetes", "gaussian3" e "glass"

Base	RBF	J48	SMO	NB	IBk
diabetes 1	0,2478±0,0101	0,2482±0,0108	**0,2285±0,0054**	0,2359±0,0063	0,2946±0,0058
diabetes 2	0,2582±0,0105	0,2502±0,0097	**0,2357±0,0030**	0,2449±0,0046	0,2928±0,0093
diabetes 3	0,2538±0,0076	0,2534±0,0080	**0,2323±0,0049**	0,2415±0,0054	0,3142±0,0066
diabetes 4	0,2505±0,0079	0,2482±0,0072	**0,2311±0,0032**	0,2389±0,0028	0,3217±0,0082
diabetes 5	0,2554±0,0090	0,2577±0,0074	**0,2343±0,0047**	0,2437±0,0038	0,2863±0,0084
diabetes 6	0,2379±0,0092	0,2470±0,0097	**0,2249±0,0036**	0,2321±0,0045	0,2951±0,0088
diabetes 7	0,2637±0,0081	0,2664±0,0086	**0,2373±0,0040**	0,2495±0,0051	0,3145±0,0075
diabetes 8	0,2577±0,0070	0,2644±0,0155	**0,2340±0,0043**	0,2392±0,0059	0,2987±0,0083
diabetes 9	0,2480±0,0086	0,2506±0,0116	**0,2266±0,0033**	0,2303±0,0080	0,2873±0,0064
diabetes 10	0,2402±0,0081	0,2480±0,0083	**0,2285±0,0047**	0,2367±0,0039	0,2845±0,0075
gaussian3 1	0,3574±0,0517	0,3648±0,0791	**0,0000±0,0000**	0,3278±0,0624	0,0111±0,0096
gaussian3 2	0,3722±0,0701	0,3130±0,0673	**0,0000±0,0000**	0,3444±0,0532	0,0111±0,0096
gaussian3 3	0,4019±0,0462	0,3722±0,0432	**0,0000±0,0000**	0,3815±0,0292	0,0185±0,0087
gaussian3 4	0,4407±0,0443	0,3389±0,0824	**0,0000±0,0000**	0,3926±0,0544	0,0130±0,0089
gaussian3 5	0,3333±0,0262	0,3722±0,0375	**0,0000±0,0000**	0,2907±0,0391	0,0111±0,0096
gaussian3 6	0,3759±0,0517	0,2759±0,0662	**0,0000±0,0000**	0,3278±0,0462	0,0111±0,0096
gaussian3 7	0,3852±0,0551	0,3556±0,0681	**0,0000±0,0000**	0,3556±0,0484	0,0000±0,0000
gaussian3 8	0,3667±0,0452	0,3981±0,0711	**0,0000±0,0000**	0,3204±0,0559	0,0000±0,0000
gaussian3 9	0,4667±0,0398	0,2907±0,0471	**0,0000±0,0000**	0,3870±0,0332	0,0167±0,0059
gaussian3 10	0,3333±0,0972	0,3796±0,0503	**0,0000±0,0000**	0,2704±0,0411	0,0167±0,0059
glass 1	0,3349±0,0189	**0,2703±0,0135**	0,4068±0,0240	0,4958±0,0188	0,3219±0,0136
glass 2	0,3146±0,0232	**0,2630±0,0227**	0,4271±0,0179	0,5073±0,0165	0,3042±0,0113
glass 3	0,3229±0,0207	**0,2604±0,0214**	0,4125±0,0239	0,5073±0,0141	0,3260±0,0121
glass 4	0,3427±0,0240	**0,2599±0,0135**	0,3948±0,0106	0,5146±0,0149	0,3130±0,0135
glass 5	0,3264±0,0330	**0,2751±0,0179**	0,4067±0,0161	0,5031±0,0208	0,3415±0,0133
glass 6	0,3047±0,0237	**0,2617±0,0216**	0,4135±0,0213	0,5016±0,0240	0,2896±0,0141
glass 7	0,3109±0,0289	**0,2715±0,0190**	0,4114±0,0171	0,5285±0,0160	0,3062±0,0135
glass 8	0,3166±0,0288	**0,2554±0,0236**	0,4497±0,0287	0,5642±0,0173	0,3212±0,0147
glass 9	0,2808±0,0187	**0,2249±0,0157**	0,4285±0,0191	0,5311±0,0212	0,2782±0,0127
glass 10	0,3047±0,0204	**0,2487±0,0214**	0,4036±0,0159	0,4927±0,0137	0,3067±0,0138

Tabela 53: Valores de erro médio e desvio padrão para *Multiboosting* quando aplicado às bases "haberman", "heart-c" e "hepatitis"

Base	RBF	J48	SMO	NB	IBk
haberman 1	0,2844±0,0167	0,3047±0,0165	0,2742±0,0100	**0,2589±0,0100**	0,3436±0,0138
haberman 2	0,2775±0,0126	0,3007±0,0138	0,2727±0,0110	**0,2542±0,0070**	0,3356±0,0134
haberman 3	0,2880±0,0133	0,3135±0,0095	0,2789±0,0094	**0,2716±0,0177**	0,3476±0,0119
haberman 4	0,2749±0,0119	0,2829±0,0074	0,2727±0,0082	**0,2538±0,0072**	0,3327±0,0089
haberman 5	0,2760±0,0117	0,2967±0,0102	0,2713±0,0127	**0,2524±0,0075**	0,3480±0,0073
haberman 6	0,2516±0,0173	0,2764±0,0161	0,2713±0,0120	**0,2422±0,0073**	0,3396±0,0163
haberman 7	0,2699±0,0092	0,3000±0,0203	0,2736±0,0092	**0,2543±0,0063**	0,3438±0,0118
haberman 8	0,2801±0,0157	0,2895±0,0206	0,2743±0,0157	**0,2522±0,0108**	0,3312±0,0110
haberman 9	0,2627±0,0134	0,2888±0,0128	0,2638±0,0088	**0,2420±0,0087**	0,3120±0,0091
haberman 10	0,2746±0,0051	0,2931±0,0156	0,2743±0,0073	**0,2540±0,0116**	0,3236±0,0080
heart-c 1	0,1562±0,0090	0,1934±0,0115	0,1559±0,0128	**0,1540±0,0096**	0,2349±0,0061
heart-c 2	0,1647±0,0085	0,1934±0,0148	0,1614±0,0126	**0,1585±0,0074**	0,2272±0,0064
heart-c 3	**0,1588±0,0109**	0,2007±0,0106	0,1607±0,0141	0,1629±0,0104	0,2489±0,0119
heart-c 4	**0,1560±0,0105**	0,1978±0,0153	0,1674±0,0081	0,1571±0,0053	0,2286±0,0112
heart-c 5	**0,1769±0,0125**	0,2176±0,0124	0,1886±0,0085	0,1773±0,0060	0,2630±0,0084
heart-c 6	**0,1667±0,0092**	0,2187±0,0117	0,1769±0,0088	0,1685±0,0102	0,2546±0,0074
heart-c 7	0,1718±0,0092	0,2103±0,0136	0,1744±0,0076	**0,1674±0,0083**	0,2502±0,0108
heart-c 8	**0,1505±0,0080**	0,1930±0,0173	0,1641±0,0073	0,1509±0,0109	0,2443±0,0107
heart-c 9	**0,1549±0,0115**	0,2000±0,0138	0,1553±0,0063	0,1593±0,0076	0,2300±0,0066
heart-c 10	**0,1502±0,0085**	0,2000±0,0189	0,1685±0,0079	0,1623±0,0083	0,2337±0,0096
hepatitis 1	0,1525±0,0185	0,1698±0,0223	**0,1381±0,0151**	0,1489±0,0108	0,1950±0,0115
hepatitis 2	**0,1496±0,0162**	0,1683±0,0189	0,1626±0,0145	0,1532±0,0068	0,1827±0,0103
hepatitis 3	**0,1403±0,0149**	0,1755±0,0245	0,1647±0,0210	0,1540±0,0091	0,1784±0,0147
hepatitis 4	**0,1518±0,0241**	0,1885±0,0126	0,1655±0,0166	0,1612±0,0118	0,2259±0,0152
hepatitis 5	**0,1360±0,0153**	0,1813±0,0155	0,1633±0,0204	0,1403±0,0141	0,1971±0,0170
hepatitis 6	**0,1407±0,0126**	0,1771±0,0154	0,1621±0,0194	0,1464±0,0102	0,2071±0,0158
hepatitis 7	**0,1529±0,0151**	0,1857±0,0178	0,1636±0,0160	0,1571±0,0082	0,1879±0,0126
hepatitis 8	0,1300±0,0130	0,1679±0,0084	0,1514±0,0120	**0,1286±0,0112**	0,1679±0,0113
hepatitis 9	**0,1486±0,0142**	0,1814±0,0166	0,1671±0,0172	0,1557±0,0100	0,2100±0,0108
hepatitis 10	**0,1400±0,0084**	0,1829±0,0148	0,1571±0,0101	0,1536±0,0148	0,1821±0,0179

Tabela 54: Valores de erro médio e desvio padrão para *Multiboosting* quando aplicado às bases "ionosphere", "iris" e "segment"

Base	RBF	J48	SMO	NB	IBk
ionosphere 1	0,0797±0,0061	**0,0740±0,0127**	0,1200±0,0076	0,0895±0,0058	0,1206±0,0067
ionosphere 2	0,0649±0,0048	**0,0604±0,0057**	0,1114±0,0089	0,0722±0,0095	0,1190±0,0065
ionosphere 3	**0,0747±0,0037**	0,0772±0,0098	0,1168±0,0090	0,1006±0,0083	0,1373±0,0062
ionosphere 4	0,0807±0,0058	**0,0731±0,0084**	0,1209±0,0066	0,0902±0,0082	0,1339±0,0047
ionosphere 5	**0,0715±0,0050**	0,0794±0,0062	0,1180±0,0090	0,0911±0,0068	0,1405±0,0062
ionosphere 6	**0,0680±0,0081**	0,0687±0,0080	0,1095±0,0086	0,0816±0,0088	0,1358±0,0055
ionosphere 7	0,0750±0,0075	**0,0744±0,0078**	0,1180±0,0060	0,0911±0,0080	0,1399±0,0068
ionosphere 8	0,0778±0,0091	**0,0699±0,0109**	0,1101±0,0044	0,0924±0,0123	0,1373±0,0064
ionosphere 9	0,0769±0,0072	**0,0753±0,0055**	0,1139±0,0042	0,0845±0,0086	0,1225±0,0068
ionosphere 10	**0,0744±0,0090**	0,0753±0,0114	0,1174±0,0041	0,0902±0,0043	0,1392±0,0039
iris 1	**0,0378±0,0042**	0,0541±0,0078	0,0400±0,0052	0,0415±0,0052	0,0378±0,0023
iris 2	0,0496±0,0050	0,0607±0,0098	**0,0407±0,0072**	0,0474±0,0062	0,0511±0,0023
iris 3	0,0496±0,0050	0,0563±0,0087	**0,0430±0,0058**	0,0430±0,0047	0,0452±0,0023
iris 4	0,0400±0,0072	0,0459±0,0091	**0,0348±0,0111**	0,0422±0,0061	0,0356±0,0058
iris 5	0,0459±0,0077	0,0652±0,0104	**0,0363±0,0102**	0,0444±0,0049	0,0444±0,0035
iris 6	0,0504±0,0058	0,0667±0,0110	**0,0459±0,0098**	0,0519±0,0049	0,0474±0,0072
iris 7	0,0489±0,0087	0,0622±0,0094	**0,0378±0,0065**	0,0393±0,0050	0,0444±0,0049
iris 8	0,0422±0,0070	0,0519±0,0144	**0,0341±0,0100**	0,0415±0,0080	0,0585±0,0042
iris 9	0,0496±0,0061	0,0630±0,0112	**0,0452±0,0089**	0,0481±0,0072	0,0504±0,0068
iris 10	0,0467±0,0061	0,0496±0,0086	**0,0370±0,0070**	0,0407±0,0072	0,0430±0,0047
segment 1	0,1057±0,0055	**0,0239±0,0020**	0,0720±0,0020	0,2011±0,0016	0,0347±0,0014
segment 2	0,1015±0,0071	**0,0221±0,0013**	0,0726±0,0018	0,2015±0,0011	0,0296±0,0020
segment 3	0,1023±0,0053	**0,0232±0,0024**	0,0729±0,0027	0,1971±0,0013	0,0311±0,0021
segment 4	0,1022±0,0069	**0,0228±0,0020**	0,0761±0,0027	0,1996±0,0016	0,0327±0,0019
segment 5	0,1054±0,0083	**0,0235±0,0016**	0,0768±0,0020	0,1998±0,0014	0,0288±0,0020
segment 6	0,1099±0,0083	**0,0228±0,0026**	0,0725±0,0026	0,2000±0,0011	0,0285±0,0024
segment 7	0,1029±0,0079	**0,0245±0,0022**	0,0705±0,0012	0,1958±0,0011	0,0287±0,0022
segment 8	0,1043±0,0072	**0,0241±0,0022**	0,0716±0,0031	0,1961±0,0011	0,0327±0,0018
segment 9	0,1017±0,0070	**0,0244±0,0019**	0,0703±0,0018	0,1931±0,0022	0,0316±0,0011
segment 10	0,1020±0,0074	**0,0214±0,0021**	0,0755±0,0025	0,1978±0,0019	0,0286±0,0010

Tabela 55: Valores de erro médio e desvio padrão para *Multiboosting* quando aplicado às bases "sick", "sonar" e "vehicle"

Base	RBF	J48	SMO	NB	IBk
sick 1	0,0396±0,0018	**0,0110±0,0016**	0,0606±0,0006	0,0678±0,0022	0,0406±0,0016
sick 2	0,0405±0,0017	**0,0123±0,0007**	0,0604±0,0008	0,0673±0,0020	0,0533±0,0011
sick 3	0,0404±0,0023	**0,0103±0,0011**	0,0609±0,0007	0,0657±0,0015	0,0396±0,0014
sick 4	0,0406±0,0030	**0,0115±0,0009**	0,0607±0,0009	0,0680±0,0014	0,0410±0,0013
sick 5	0,0393±0,0014	**0,0104±0,0014**	0,0603±0,0009	0,0641±0,0015	0,0403±0,0018
sick 6	0,0402±0,0018	**0,0103±0,0011**	0,0576±0,0017	0,0662±0,0016	0,0395±0,0014
sick 7	0,0389±0,0018	**0,0105±0,0010**	0,0593±0,0011	0,0801±0,0029	0,0370±0,0013
sick 8	0,0401±0,0023	**0,0107±0,0011**	0,0600±0,0010	0,0666±0,0023	0,0427±0,0010
sick 9	0,0400±0,0013	**0,0123±0,0018**	0,0599±0,0009	0,0692±0,0019	0,0394±0,0009
sick 10	0,0401±0,0020	**0,0116±0,0009**	0,0600±0,0009	0,0671±0,0018	0,0373±0,0008
sonar 1	0,1995±0,0286	0,2187±0,0150	0,2176±0,0140	0,2406±0,0155	**0,1337±0,0116**
sonar 2	0,2091±0,0234	0,2070±0,0331	0,2294±0,0181	0,2529±0,0195	**0,1540±0,0094**
sonar 3	0,2027±0,0205	0,2118±0,0318	0,2428±0,0143	0,2406±0,0153	**0,1283±0,0110**
sonar 4	0,1979±0,0257	0,1936±0,0183	0,2032±0,0178	0,2390±0,0200	**0,1433±0,0100**
sonar 5	0,1984±0,0310	0,2118±0,0227	0,2262±0,0192	0,2182±0,0159	**0,1513±0,0124**
sonar 6	0,2080±0,0329	0,2182±0,0211	0,2193±0,0136	0,2326±0,0199	**0,1561±0,0083**
sonar 7	0,2027±0,0259	0,2187±0,0160	0,2444±0,0158	0,2492±0,0131	**0,1267±0,0138**
sonar 8	0,2037±0,0220	0,1861±0,0254	0,2128±0,0232	0,2422±0,0143	**0,1358±0,0124**
sonar 9	0,2133±0,0304	0,2170±0,0186	0,2351±0,0197	0,2617±0,0231	**0,1527±0,0135**
sonar 10	0,2229±0,0197	0,2043±0,0185	0,2250±0,0157	0,2532±0,0226	**0,1394±0,0117**
vehicle 1	0,3430±0,0146	**0,2468±0,0175**	0,2537±0,0080	0,5572±0,0084	0,3097±0,0078
vehicle 2	0,3300±0,0076	**0,2455±0,0119**	0,2543±0,0075	0,5428±0,0081	0,3085±0,0081
vehicle 3	0,3356±0,0052	**0,2417±0,0042**	0,2593±0,0063	0,5552±0,0061	0,3042±0,0080
vehicle 4	0,3192±0,0105	**0,2470±0,0105**	0,2607±0,0032	0,5317±0,0063	0,3007±0,0115
vehicle 5	0,3321±0,0070	**0,2566±0,0073**	0,2611±0,0070	0,5568±0,0081	0,3080±0,0091
vehicle 6	0,3306±0,0094	**0,2461±0,0091**	0,2612±0,0068	0,5511±0,0081	0,3097±0,0063
vehicle 7	0,3260±0,0110	**0,2467±0,0119**	0,2598±0,0070	0,5512±0,0082	0,3073±0,0074
vehicle 8	0,3243±0,0106	**0,2508±0,0083**	0,2705±0,0076	0,5601±0,0070	0,3157±0,0091
vehicle 9	0,3333±0,0102	**0,2486±0,0171**	0,2541±0,0082	0,5551±0,0065	0,2953±0,0062
vehicle 10	0,3231±0,0096	**0,2438±0,0168**	0,2579±0,0052	0,5455±0,0050	0,2957±0,0049

Tabela 56: Valores de erro médio e desvio padrão para *Multiboosting* quando aplicado às bases "vote" e "waveform-5000"

Base	RBF	J48	SMO	NB	IBk
vote 1	0,0532±0,0062	**0,0409±0,0036**	0,0448±0,0065	0,0908±0,0072	0,0772±0,0041
vote 2	0,0473±0,0049	**0,0402±0,0050**	0,0448±0,0039	0,0801±0,0121	0,0890±0,0043
vote 3	0,0425±0,0051	0,0402±0,0044	**0,0384±0,0058**	0,0783±0,0108	0,0734±0,0021
vote 4	0,0496±0,0079	0,0427±0,0048	**0,0422±0,0056**	0,0885±0,0050	0,0777±0,0022
vote 5	0,0471±0,0056	0,0425±0,0058	**0,0422±0,0042**	0,0744±0,0103	0,0762±0,0041
vote 6	0,0518±0,0048	0,0454±0,0057	**0,0418±0,0044**	0,0855±0,0065	0,0786±0,0029
vote 7	0,0477±0,0097	0,0449±0,0032	**0,0413±0,0060**	0,0895±0,0065	0,0765±0,0050
vote 8	0,0485±0,0093	0,0416±0,0048	**0,0390±0,0043**	0,0724±0,0084	0,0681±0,0068
vote 9	0,0444±0,0051	**0,0334±0,0037**	0,0385±0,0039	0,0684±0,0080	0,0681±0,0042
vote 10	0,0477±0,0048	**0,0347±0,0045**	0,0401±0,0047	0,0842±0,0081	0,0699±0,0025
waveform-5000 1	0,1453±0,0021	0,1819±0,0037	**0,1366±0,0014**	0,1984±0,0011	0,2648±0,0020
waveform-5000 2	0,1472±0,0016	0,1804±0,0052	**0,1384±0,0022**	0,2006±0,0011	0,2695±0,0021
waveform-5000 3	0,1446±0,0012	0,1808±0,0040	**0,1378±0,0022**	0,1982±0,0013	0,2640±0,0026
waveform-5000 4	0,1443±0,0021	0,1771±0,0031	**0,1372±0,0022**	0,2007±0,0009	0,2654±0,0029
waveform-5000 5	0,1417±0,0019	0,1759±0,0043	**0,1334±0,0023**	0,1978±0,0009	0,2708±0,0025
waveform-5000 6	0,1414±0,0018	0,1770±0,0042	**0,1358±0,0014**	0,1988±0,0007	0,2590±0,0020
waveform-5000 7	0,1425±0,0017	0,1795±0,0046	**0,1366±0,0016**	0,1967±0,0010	0,2672±0,0024
waveform-5000 8	0,1472±0,0019	0,1782±0,0038	**0,1368±0,0020**	0,1966±0,0009	0,2658±0,0027
waveform-5000 9	0,1422±0,0014	0,1783±0,0046	**0,1359±0,0010**	0,1977±0,0009	0,2645±0,0034
waveform-5000 10	0,1448±0,0017	0,1780±0,0041	**0,1352±0,0021**	0,2001±0,0015	0,2695±0,0022

APÊNDICE D – Extração de Meta-características

Das Tabelas 57 a 70, são apresentos os valores das meta-características extraídas das 200 bases utilizadas durante a realização dos experimentos computacionais, e estão organizadas em: número de atributo (NA), número de atributos simbólico (NAS), número de atributos numérico (NAN), número de exemplos (NE), número de classe (NC), valores faltosos total (VFT), valores faltosos relativo (VFR), linha com valores faltosos total (LVFT), linhas com valores faltosos relativo (LVFR), assimetria média absoluta dos atributos (AMAA), curtose média dos atributos (CMA), número de atributos com *outliers* (NAO), entropia de classe (EC), entropia dos atributos (EA), informação mútua (IM) e entropia conjunta (ECON).

Tabela 57: Extração de meta-características para as bases “anneal”, “breast-cancer” e “bupa”

Base	NA	NAS	NAN	NE	NC	VFT	VFR	LVFT
anneal 1	39	32	6	808	6	0	0	0
anneal 2	39	32	6	808	6	0	0	0
anneal 3	39	32	6	808	6	0	0	0
anneal 4	39	32	6	808	6	0	0	0
anneal 5	39	32	6	808	6	0	0	0
anneal 6	39	32	6	808	6	0	0	0
anneal 7	39	32	6	808	6	0	0	0
anneal 8	39	32	6	808	6	0	0	0
anneal 9	39	32	6	809	6	0	0	0
anneal 10	39	32	6	809	6	0	0	0
breast-cancer 1	10	9	0	257	2	9	0	9
breast-cancer 2	10	9	0	257	2	9	0	9
breast-cancer 3	10	9	0	257	2	8	0	8
breast-cancer 4	10	9	0	257	2	9	0	9
breast-cancer 5	10	9	0	257	2	6	0	6
breast-cancer 6	10	9	0	257	2	9	0	9
breast-cancer 7	10	9	0	258	2	7	0	7
breast-cancer 8	10	9	0	258	2	8	0	8
breast-cancer 9	10	9	0	258	2	9	0	9
breast-cancer 10	10	9	0	258	2	7	0	7
bupa 1	7	0	6	310	2	0	0	0
bupa 2	7	0	6	310	2	0	0	0
bupa 3	7	0	6	310	2	0	0	0
bupa 4	7	0	6	310	2	0	0	0
bupa 5	7	0	6	310	2	0	0	0
bupa 6	7	0	6	311	2	0	0	0
bupa 7	7	0	6	311	2	0	0	0
bupa 8	7	0	6	311	2	0	0	0
bupa 9	7	0	6	311	2	0	0	0
bupa 10	7	0	6	311	2	0	0	0

Tabela 58: Extração de meta-características para as bases “anneal”, “breast-cancer” e “bupa”

Base	LVFR	AMAA	CMA	NAO	EC	EA	IM	ECON
anneal 1	0,00	1,4226	5,6052	2	1,1947	0,4614	0,0819	1,5742
anneal 2	0,00	1,3838	5,2539	2	1,1910	0,4626	0,0815	1,5721
anneal 3	0,00	1,4169	5,1739	2	1,1910	0,4606	0,0819	1,5697
anneal 4	0,00	1,3946	5,2732	2	1,1910	0,4650	0,0825	1,5735
anneal 5	0,00	1,3726	5,2316	2	1,1868	0,4645	0,0815	1,5699
anneal 6	0,00	1,5130	6,7371	2	1,1868	0,4649	0,0822	1,5696
anneal 7	0,00	1,5093	6,6511	2	1,1868	0,4663	0,0815	1,5716
anneal 8	0,00	1,3772	5,2417	2	1,1868	0,4614	0,0815	1,5668
anneal 9	0,00	1,4095	5,2047	2	1,1893	0,4634	0,0828	1,5699
anneal 10	0,00	1,3931	5,3380	2	1,1937	0,4645	0,0826	1,5756
breast-cancer 1	0,04	–	–	–	0,8760	1,5012	0,0376	2,3396
breast-cancer 2	0,04	–	–	–	0,8760	1,5132	0,0391	2,3501
breast-cancer 3	0,03	–	–	–	0,8760	1,5127	0,0458	2,3429
breast-cancer 4	0,04	–	–	–	0,8760	1,5009	0,0386	2,3383
breast-cancer 5	0,02	–	–	–	0,8760	1,5035	0,0373	2,3422
breast-cancer 6	0,04	–	–	–	0,8808	1,5029	0,0320	2,3517
breast-cancer 7	0,03	–	–	–	0,8794	1,5157	0,0396	2,3555
breast-cancer 8	0,03	–	–	–	0,8794	1,5092	0,0395	2,3491
breast-cancer 9	0,03	–	–	–	0,8794	1,5091	0,0398	2,3486
breast-cancer 10	0,03	–	–	–	0,8794	1,5152	0,0401	2,3545
bupa 1	0,00	1,5861	7,5167	3	0,9812	–	–	–
bupa 2	0,00	1,5559	7,5193	3	0,9812	–	–	–
bupa 3	0,00	1,5619	7,7040	3	0,9812	–	–	–
bupa 4	0,00	1,5509	7,1846	3	0,9812	–	–	–
bupa 5	0,00	1,5640	7,5179	3	0,9812	–	–	–
bupa 6	0,00	1,5821	7,7426	3	0,9820	–	–	–
bupa 7	0,00	1,6478	7,9691	3	0,9820	–	–	–
bupa 8	0,00	1,5529	7,5686	3	0,9820	–	–	–
bupa 9	0,00	1,5727	7,7173	3	0,9820	–	–	–
bupa 10	0,00	1,5272	7,4359	3	0,9820	–	–	–

Tabela 59: Extração de meta-características para as bases “car”, “colic” e “credit-a”

Base	NA	NAS	NAN	NE	NC	VFT	VFR	LVFT
car 1	7	6	0	1555	4	0	0	0
car 2	7	6	0	1555	4	0	0	0
car 3	7	6	0	1555	4	0	0	0
car 4	7	6	0	1555	4	0	0	0
car 5	7	6	0	1555	4	0	0	0
car 6	7	6	0	1555	4	0	0	0
car 7	7	6	0	1555	4	0	0	0
car 8	7	6	0	1555	4	0	0	0
car 9	7	6	0	1556	4	0	0	0
car 10	7	6	0	1556	4	0	0	0
colic 1	23	15	7	331	2	1745	0,23	328
colic 2	23	15	7	331	2	1734	0,23	324
colic 3	23	15	7	331	2	1758	0,23	324
colic 4	23	15	7	331	2	1714	0,23	325
colic 5	23	15	7	331	2	1705	0,22	324
colic 6	23	15	7	331	2	1729	0,23	324
colic 7	23	15	7	331	2	1729	0,23	324
colic 8	23	15	7	331	2	1754	0,23	324
colic 9	23	15	7	332	2	1728	0,23	326
colic 10	23	15	7	332	2	1747	0,23	326
credit-a 1	16	9	6	621	2	60	0,01	34
credit-a 2	16	9	6	621	2	59	0,01	31
credit-a 3	16	9	6	621	2	66	0,01	36
credit-a 4	16	9	6	621	2	64	0,01	34
credit-a 5	16	9	6	621	2	66	0,01	36
credit-a 6	16	9	6	621	2	58	0,01	33
credit-a 7	16	9	6	621	2	52	0,01	30
credit-a 8	16	9	6	621	2	63	0,01	33
credit-a 9	16	9	6	621	2	66	0,01	36
credit-a 10	16	9	6	621	2	49	0	30

Tabela 60: Extração de meta-características para as bases “car”, “colic” e “credit-a”

Base	LVFR	AMAA	CMA	NAO	EC	EA	IM	ECON
car 1	0,00	–	–	–	1,2063	1,7924	0,1153	2,8834
car 2	0,00	–	–	–	1,2063	1,7924	0,1144	2,8843
car 3	0,00	–	–	–	1,2063	1,7924	0,1146	2,8840
car 4	0,00	–	–	–	1,2062	1,7924	0,1145	2,8841
car 5	0,00	–	–	–	1,2046	1,7925	0,1147	2,8824
car 6	0,00	–	–	–	1,2046	1,7924	0,1140	2,8831
car 7	0,00	–	–	–	1,2046	1,7924	0,1141	2,8829
car 8	0,00	–	–	–	1,2046	1,7924	0,1144	2,8827
car 9	0,00	–	–	–	1,2069	1,7924	0,1152	2,8842
car 10	0,00	–	–	–	1,2069	1,7925	0,1148	2,8846
colic 1	0,99	1,4237	5,0629	0	0,9496	1,4419	0,2775	2,1140
colic 2	0,98	1,3726	4,8413	1	0,9496	1,4369	0,2791	2,1074
colic 3	0,98	1,3635	4,7256	0	0,9496	1,4365	0,2751	2,1110
colic 4	0,98	1,3881	4,8033	0	0,9496	1,4461	0,2671	2,1286
colic 5	0,98	1,3666	4,7951	0	0,9496	1,4473	0,2722	2,1246
colic 6	0,98	1,3719	4,8979	0	0,9496	1,4476	0,2869	2,1103
colic 7	0,98	1,3649	4,8684	1	0,9519	1,4389	0,2845	2,1063
colic 8	0,98	1,3801	4,9054	0	0,9519	1,4380	0,2803	2,1095
colic 9	0,98	1,4282	5,2519	0	0,9510	1,4421	0,2786	2,1146
colic 10	0,98	1,4132	5,0751	0	0,9510	1,4407	0,2781	2,1137
credit-a 1	0,05	3,0938	21,1980	3	0,9911	1,2575	0,0971	2,1515
credit-a 2	0,05	3,1950	23,3284	4	0,9911	1,2609	0,0969	2,1551
credit-a 3	0,06	3,2455	24,0592	4	0,9911	1,2635	0,0961	2,1586
credit-a 4	0,05	3,1916	23,6347	3	0,9911	1,2584	0,0950	2,1545
credit-a 5	0,06	2,8935	20,4973	3	0,9911	1,2548	0,0989	2,1470
credit-a 6	0,05	3,3789	28,2542	4	0,9911	1,2521	0,0968	2,1464
credit-a 7	0,05	3,2837	24,9598	4	0,9911	1,2614	0,0979	2,1546
credit-a 8	0,05	3,0612	21,9057	3	0,9916	1,2541	0,0954	2,1503
credit-a 9	0,06	3,1872	23,2854	4	0,9916	1,2516	0,1008	2,1423
credit-a 10	0,05	3,0627	21,7307	4	0,9916	1,2545	0,0974	2,1486

Tabela 61: Extração de meta-características para as bases “diabetes”, “gaussian3” e “glass”

Base	NA	NAS	NAN	NE	NC	VFT	VFR	LVFT
diabetes 1	9	0	8	691	2	0	0	0
diabetes 2	9	0	8	691	2	0	0	0
diabetes 3	9	0	8	691	2	0	0	0
diabetes 4	9	0	8	691	2	0	0	0
diabetes 5	9	0	8	691	2	0	0	0
diabetes 6	9	0	8	691	2	0	0	0
diabetes 7	9	0	8	691	2	0	0	0
diabetes 8	9	0	8	691	2	0	0	0
diabetes 9	9	0	8	692	2	0	0	0
diabetes 10	9	0	8	692	2	0	0	0
gaussian3 1	601	0	600	54	3	0	0	0
gaussian3 2	601	0	600	54	3	0	0	0
gaussian3 3	601	0	600	54	3	0	0	0
gaussian3 4	601	0	600	54	3	0	0	0
gaussian3 5	601	0	600	54	3	0	0	0
gaussian3 6	601	0	600	54	3	0	0	0
gaussian3 7	601	0	600	54	3	0	0	0
gaussian3 8	601	0	600	54	3	0	0	0
gaussian3 9	601	0	600	54	3	0	0	0
gaussian3 10	601	0	600	54	3	0	0	0
glass 1	10	0	9	192	7	0	0	0
glass 2	10	0	9	192	7	0	0	0
glass 3	10	0	9	192	7	0	0	0
glass 4	10	0	9	192	7	0	0	0
glass 5	10	0	9	193	7	0	0	0
glass 6	10	0	9	193	7	0	0	0
glass 7	10	0	9	193	7	0	0	0
glass 8	10	0	9	193	7	0	0	0
glass 9	10	0	9	193	7	0	0	0
glass 10	10	0	9	193	7	0	0	0

Tabela 62: Extração de meta-características para as bases “diabetes”, “gaussian3” e “glass”

Base	LVFR	AMAA	CMA	NAO	EC	EA	IM	ECON
diabetes 1	0,00	1,0673	5,8494	2	0,9330	–	–	–
diabetes 2	0,00	1,0597	5,8124	2	0,9330	–	–	–
diabetes 3	0,00	1,0860	5,7824	2	0,9330	–	–	–
diabetes 4	0,00	1,0901	5,8111	2	0,9330	–	–	–
diabetes 5	0,00	1,0383	5,7203	3	0,9330	–	–	–
diabetes 6	0,00	1,1078	5,7906	2	0,9330	–	–	–
diabetes 7	0,00	1,0823	6,0078	3	0,9330	–	–	–
diabetes 8	0,00	1,0493	5,7805	3	0,9330	–	–	–
diabetes 9	0,00	1,0739	5,9549	4	0,9338	–	–	–
diabetes 10	0,00	1,0793	5,8669	4	0,9338	–	–	–
gaussian3 1	0,00	0,7151	4,2970	66	1,5850	–	–	–
gaussian3 2	0,00	0,7113	4,2724	64	1,5850	–	–	–
gaussian3 3	0,00	0,7184	4,2715	67	1,5850	–	–	–
gaussian3 4	0,00	0,7313	4,2998	62	1,5850	–	–	–
gaussian3 5	0,00	0,7245	4,2746	66	1,5850	–	–	–
gaussian3 6	0,00	0,7163	4,2783	61	1,5850	–	–	–
gaussian3 7	0,00	0,7211	4,2936	65	1,5850	–	–	–
gaussian3 8	0,00	0,7206	4,2608	66	1,5850	–	–	–
gaussian3 9	0,00	0,7192	4,2757	57	1,5850	–	–	–
gaussian3 10	0,00	0,7142	4,2522	63	1,5850	–	–	–
glass 1	0,00	1,3636	5,8429	5	2,1769	–	–	–
glass 2	0,00	1,4002	6,2450	5	2,1706	–	–	–
glass 3	0,00	1,4631	6,7963	5	2,1746	–	–	–
glass 4	0,00	1,4192	6,3835	5	2,1746	–	–	–
glass 5	0,00	1,4945	6,6307	6	2,1844	–	–	–
glass 6	0,00	1,4070	6,2637	5	2,1733	–	–	–
glass 7	0,00	1,4231	6,4930	6	2,1733	–	–	–
glass 8	0,00	1,5352	6,7067	5	2,1733	–	–	–
glass 9	0,00	1,4589	6,6457	6	2,1733	–	–	–
glass 10	0,00	1,3319	5,6234	5	2,1889	–	–	–

Tabela 63: Extração de meta-características para as bases “haberman”, “heart-c” e “hepatitis”

Base	NA	NAS	NAN	NE	NC	VFT	VFR	LVFT
haberman 1	4	1	2	275	2	0	0	0
haberman 2	4	1	2	275	2	0	0	0
haberman 3	4	1	2	275	2	0	0	0
haberman 4	4	1	2	275	2	0	0	0
haberman 5	4	1	2	275	2	0	0	0
haberman 6	4	1	2	275	2	0	0	0
haberman 7	4	1	2	276	2	0	0	0
haberman 8	4	1	2	276	2	0	0	0
haberman 9	4	1	2	276	2	0	0	0
haberman 10	4	1	2	276	2	0	0	0
heart-c 1	14	7	6	272	5	7	0	7
heart-c 2	14	7	6	272	5	6	0	6
heart-c 3	14	7	6	272	5	6	0	6
heart-c 4	14	7	6	273	5	6	0	6
heart-c 5	14	7	6	273	5	6	0	6
heart-c 6	14	7	6	273	5	6	0	6
heart-c 7	14	7	6	273	5	7	0	7
heart-c 8	14	7	6	273	5	7	0	7
heart-c 9	14	7	6	273	5	5	0	5
heart-c 10	14	7	6	273	5	7	0	7
hepatitis 1	20	13	6	139	2	154	0,06	67
hepatitis 2	20	13	6	139	2	148	0,05	65
hepatitis 3	20	13	6	139	2	142	0,05	66
hepatitis 4	20	13	6	139	2	145	0,05	66
hepatitis 5	20	13	6	139	2	161	0,06	71
hepatitis 6	20	13	6	140	2	157	0,06	66
hepatitis 7	20	13	6	140	2	136	0,05	65
hepatitis 8	20	13	6	140	2	144	0,05	67
hepatitis 9	20	13	6	140	2	151	0,05	69
hepatitis 10	20	13	6	140	2	165	0,06	73

Tabela 64: Extração de meta-características para as bases “haberman”, “heart-c” e “hepatitis”

Base	LVFR	AMAA	CMA	NAO	EC	EA	IM	ECON
haberman 1	0,00	1,7028	9,2216	1	0,8349	3,5194	0,0262	4,3281
haberman 2	0,00	1,5455	8,0932	1	0,8349	3,5274	0,0459	4,3163
haberman 3	0,00	1,2795	6,7229	1	0,8349	3,5239	0,0434	4,3154
haberman 4	0,00	1,5982	8,6777	1	0,8349	3,5296	0,0339	4,3306
haberman 5	0,00	1,5344	8,3674	1	0,8349	3,5276	0,0352	4,3272
haberman 6	0,00	1,3329	6,3546	1	0,8295	3,5060	0,0457	4,2898
haberman 7	0,00	1,5695	8,6152	1	0,8335	3,5254	0,0329	4,3260
haberman 8	0,00	1,6449	9,2949	1	0,8335	3,5220	0,0427	4,3127
haberman 9	0,00	1,6673	9,4509	1	0,8335	3,5264	0,0356	4,3242
haberman 10	0,00	1,5304	8,0506	1	0,8335	3,5265	0,0432	4,3167
heart-c 1	0,03	0,9259	4,7673	0	0,9944	1,1074	0,1122	1,9896
heart-c 2	0,02	0,8288	4,2304	0	0,9944	1,1066	0,1127	1,9883
heart-c 3	0,02	0,8676	4,5126	0	0,9944	1,1166	0,1121	1,9989
heart-c 4	0,02	0,8714	4,4879	0	0,9949	1,1104	0,1161	1,9892
heart-c 5	0,02	0,8735	4,4592	0	0,9949	1,1087	0,0989	2,0047
heart-c 6	0,02	0,7644	3,8714	0	0,9939	1,1039	0,1020	1,9958
heart-c 7	0,03	0,8734	4,5072	0	0,9939	1,1171	0,1067	2,0043
heart-c 8	0,03	0,8546	4,3877	0	0,9939	1,1146	0,1185	1,9900
heart-c 9	0,02	0,8369	4,4509	0	0,9939	1,1168	0,1117	1,9991
heart-c 10	0,03	0,8756	4,5263	0	0,9939	1,1189	0,1113	2,0016
hepatitis 1	0,48	1,3646	6,2746	2	0,7389	0,7923	0,0765	1,4547
hepatitis 2	0,47	1,2114	5,0522	2	0,7389	0,7938	0,0763	1,4564
hepatitis 3	0,47	1,3115	5,8902	2	0,7389	0,7933	0,0738	1,4583
hepatitis 4	0,47	1,3389	6,2733	2	0,7248	0,7864	0,0721	1,4391
hepatitis 5	0,51	1,2559	5,6770	2	0,7248	0,8000	0,0775	1,4473
hepatitis 6	0,47	1,2991	5,7981	2	0,7360	0,7864	0,0807	1,4417
hepatitis 7	0,46	1,3208	5,9897	2	0,7360	0,7878	0,0710	1,4528
hepatitis 8	0,48	1,1704	5,2628	2	0,7360	0,7964	0,0680	1,4643
hepatitis 9	0,49	1,3387	6,0808	2	0,7360	0,7998	0,0675	1,4683
hepatitis 10	0,52	1,3181	5,9710	2	0,7360	0,7861	0,0764	1,4457

Tabela 65: Extração de meta-características para as bases “ionosphere”, “iris” e “segment”

Base	NA	NAS	NAN	NE	NC	VFT	VFR	LVFT
ionosphere 1	35	0	34	315	2	0	0	0
ionosphere 2	35	0	34	316	2	0	0	0
ionosphere 3	35	0	34	316	2	0	0	0
ionosphere 4	35	0	34	316	2	0	0	0
ionosphere 5	35	0	34	316	2	0	0	0
ionosphere 6	35	0	34	316	2	0	0	0
ionosphere 7	35	0	34	316	2	0	0	0
ionosphere 8	35	0	34	316	2	0	0	0
ionosphere 9	35	0	34	316	2	0	0	0
ionosphere 10	35	0	34	316	2	0	0	0
iris 1	5	0	4	135	3	0	0	0
iris 2	5	0	4	135	3	0	0	0
iris 3	5	0	4	135	3	0	0	0
iris 4	5	0	4	135	3	0	0	0
iris 5	5	0	4	135	3	0	0	0
iris 6	5	0	4	135	3	0	0	0
iris 7	5	0	4	135	3	0	0	0
iris 8	5	0	4	135	3	0	0	0
iris 9	5	0	4	135	3	0	0	0
iris 10	5	0	4	135	3	0	0	0
segment 1	20	0	19	2079	7	0	0	0
segment 2	20	0	19	2079	7	0	0	0
segment 3	20	0	19	2079	7	0	0	0
segment 4	20	0	19	2079	7	0	0	0
segment 5	20	0	19	2079	7	0	0	0
segment 6	20	0	19	2079	7	0	0	0
segment 7	20	0	19	2079	7	0	0	0
segment 8	20	0	19	2079	7	0	0	0
segment 9	20	0	19	2079	7	0	0	0
segment 10	20	0	19	2079	7	0	0	0

Tabela 66: Extração de meta-características para as bases “ionosphere”, “iris” e “segment”

Base	LVFR	AMAA	CMA	NAO	EC	EA	IM	ECON
ionosphere 1	0,00	0,4930	2,7890	0	0,9416	–	–	–
ionosphere 2	0,00	0,4618	2,7043	0	0,9433	–	–	–
ionosphere 3	0,00	0,4874	2,8127	0	0,9433	–	–	–
ionosphere 4	0,00	0,4897	2,8085	0	0,9433	–	–	–
ionosphere 5	0,00	0,4825	2,7469	0	0,9433	–	–	–
ionosphere 6	0,00	0,4779	2,7886	0	0,9407	–	–	–
ionosphere 7	0,00	0,4792	2,7437	0	0,9407	–	–	–
ionosphere 8	0,00	0,4813	2,7871	0	0,9407	–	–	–
ionosphere 9	0,00	0,4915	2,7799	0	0,9407	–	–	–
ionosphere 10	0,00	0,4860	2,7824	0	0,9407	–	–	–
iris 1	0,00	0,3376	2,9946	0	1,5850	–	–	–
iris 2	0,00	0,3326	2,9975	0	1,5850	–	–	–
iris 3	0,00	0,3141	3,0176	0	1,5850	–	–	–
iris 4	0,00	0,2769	2,9565	0	1,5850	–	–	–
iris 5	0,00	0,3305	2,9700	0	1,5850	–	–	–
iris 6	0,00	0,3323	3,1098	0	1,5850	–	–	–
iris 7	0,00	0,3567	3,0599	0	1,5850	–	–	–
iris 8	0,00	0,3131	2,9789	0	1,5850	–	–	–
iris 9	0,00	0,2647	3,0082	0	1,5850	–	–	–
iris 10	0,00	0,3570	3,0173	0	1,5850	–	–	–
segment 1	0,00	2,0551	15,8217	5	2,8074	–	–	–
segment 2	0,00	2,0474	15,8291	5	2,8074	–	–	–
segment 3	0,00	2,0163	15,1006	5	2,8074	–	–	–
segment 4	0,00	2,0972	16,5665	5	2,8074	–	–	–
segment 5	0,00	1,9263	13,6272	5	2,8074	–	–	–
segment 6	0,00	2,0265	15,8389	5	2,8074	–	–	–
segment 7	0,00	2,0598	16,3393	5	2,8074	–	–	–
segment 8	0,00	2,0853	16,1185	5	2,8074	–	–	–
segment 9	0,00	2,0794	16,6842	5	2,8074	–	–	–
segment 10	0,00	2,0857	16,4928	5	2,8074	–	–	–

Tabela 67: Extração de meta-características para as bases “sick”, “sonar” e “vehicle”

Base	NA	NAS	NAN	NE	NC	VFT	VFR	LVFT
sick 1	30	22	7	3394	2	5448	0,05	3394
sick 2	30	22	7	3394	2	5495	0,05	3394
sick 3	30	22	7	3395	2	5433	0,05	3395
sick 4	30	22	7	3395	2	5451	0,05	3395
sick 5	30	22	7	3395	2	5443	0,05	3395
sick 6	30	22	7	3395	2	5454	0,05	3395
sick 7	30	22	7	3395	2	5490	0,05	3395
sick 8	30	22	7	3395	2	5433	0,05	3395
sick 9	30	22	7	3395	2	5470	0,05	3395
sick 10	30	22	7	3395	2	5459	0,05	3395
sonar 1	61	0	60	187	2	0	0	0
sonar 2	61	0	60	187	2	0	0	0
sonar 3	61	0	60	187	2	0	0	0
sonar 4	61	0	60	187	2	0	0	0
sonar 5	61	0	60	187	2	0	0	0
sonar 6	61	0	60	187	2	0	0	0
sonar 7	61	0	60	187	2	0	0	0
sonar 8	61	0	60	187	2	0	0	0
sonar 9	61	0	60	188	2	0	0	0
sonar 10	61	0	60	188	2	0	0	0
vehicle 1	19	0	18	761	4	0	0	0
vehicle 2	19	0	18	761	4	0	0	0
vehicle 3	19	0	18	761	4	0	0	0
vehicle 4	19	0	18	761	4	0	0	0
vehicle 5	19	0	18	761	4	0	0	0
vehicle 6	19	0	18	761	4	0	0	0
vehicle 7	19	0	18	762	4	0	0	0
vehicle 8	19	0	18	762	4	0	0	0
vehicle 9	19	0	18	762	4	0	0	0
vehicle 10	19	0	18	762	4	0	0	0

Tabela 68: Extração de meta-características para as bases “sick”, “sonar” e “vehicle”

Base	LVFR	AMAA	CMA	NAO	EC	EA	IM	ECON
sick 1	1,00	2,7354	38,3639	5	0,3325	0,3382	0,0057	0,6650
sick 2	1,00	2,4716	31,9670	5	0,3314	0,3402	0,0056	0,6660
sick 3	1,00	2,6963	36,2825	5	0,3325	0,3382	0,0054	0,6653
sick 4	1,00	2,6324	34,5804	5	0,3325	0,3389	0,0056	0,6658
sick 5	1,00	2,6573	36,7664	5	0,3325	0,3389	0,0056	0,6657
sick 6	1,00	2,6098	34,8084	5	0,3325	0,3401	0,0056	0,6669
sick 7	1,00	2,5277	35,3211	5	0,3325	0,3398	0,0059	0,6663
sick 8	1,00	2,6890	37,0400	5	0,3325	0,3382	0,0055	0,6652
sick 9	1,00	2,6395	35,7292	5	0,3325	0,3398	0,0056	0,6666
sick 10	1,00	2,6280	34,8179	5	0,3325	0,3391	0,0054	0,6661
sonar 1	0,00	1,0850	4,7438	12	0,9965	–	–	–
sonar 2	0,00	1,0858	4,7978	12	0,9965	–	–	–
sonar 3	0,00	1,0440	4,5014	13	0,9965	–	–	–
sonar 4	0,00	1,0806	4,6569	13	0,9965	–	–	–
sonar 5	0,00	1,0602	4,5672	13	0,9965	–	–	–
sonar 6	0,00	1,0872	4,6720	13	0,9965	–	–	–
sonar 7	0,00	1,0855	4,6481	9	0,9965	–	–	–
sonar 8	0,00	1,0583	4,5086	10	0,9975	–	–	–
sonar 9	0,00	1,0421	4,4043	7	0,9971	–	–	–
sonar 10	0,00	1,0591	4,5608	12	0,9971	–	–	–
vehicle 1	0,00	0,8243	5,0626	2	1,9991	–	–	–
vehicle 2	0,00	0,8438	5,2771	2	1,9991	–	–	–
vehicle 3	0,00	0,8304	5,2911	2	1,9991	–	–	–
vehicle 4	0,00	0,8161	5,3958	2	1,9991	–	–	–
vehicle 5	0,00	0,8417	5,2350	2	1,9990	–	–	–
vehicle 6	0,00	0,8369	5,0675	2	1,9990	–	–	–
vehicle 7	0,00	0,8367	5,2449	2	1,9990	–	–	–
vehicle 8	0,00	0,8397	5,0882	2	1,9990	–	–	–
vehicle 9	0,00	0,7925	4,9403	2	1,9990	–	–	–
vehicle 10	0,00	0,8510	5,6526	2	1,9992	–	–	–

Tabela 69: Extração de meta-características para as bases “vote” e “waveform-5000”

Base	NA	NAS	NAN	NE	NC	VFT	VFR	LVFT
vote 1	17	16	0	391	2	335	0,05	180
vote 2	17	16	0	391	2	348	0,05	181
vote 3	17	16	0	391	2	361	0,05	179
vote 4	17	16	0	391	2	339	0,05	182
vote 5	17	16	0	391	2	343	0,05	180
vote 6	17	16	0	392	2	363	0,05	187
vote 7	17	16	0	392	2	359	0,05	186
vote 8	17	16	0	392	2	369	0,06	183
vote 9	17	16	0	392	2	362	0,05	183
vote 10	17	16	0	392	2	349	0,05	186
waveform-5000 1	41	0	40	4500	3	0	0	0
waveform-5000 2	41	0	40	4500	3	0	0	0
waveform-5000 3	41	0	40	4500	3	0	0	0
waveform-5000 4	41	0	40	4500	3	0	0	0
waveform-5000 5	41	0	40	4500	3	0	0	0
waveform-5000 6	41	0	40	4500	3	0	0	0
waveform-5000 7	41	0	40	4500	3	0	0	0
waveform-5000 8	41	0	40	4500	3	0	0	0
waveform-5000 9	41	0	40	4500	3	0	0	0
waveform-5000 10	41	0	40	4500	3	0	0	0

Tabela 70: Extração de meta-características para as bases “vote” e “waveform-5000”

Base	LVFR	AMAA	CMA	NAO	EC	EA	IM	ECON
vote 1	0,46	–	–	–	0,9623	0,9892	0,2999	1,6516
vote 2	0,46	–	–	–	0,9623	0,9913	0,3030	1,6506
vote 3	0,46	–	–	–	0,9623	0,9900	0,3042	1,6482
vote 4	0,47	–	–	–	0,9623	0,9901	0,3048	1,6476
vote 5	0,46	–	–	–	0,9623	0,9888	0,2989	1,6522
vote 6	0,48	–	–	–	0,9616	0,9909	0,3109	1,6416
vote 7	0,47	–	–	–	0,9616	0,9918	0,3057	1,6478
vote 8	0,47	–	–	–	0,9616	0,9911	0,3078	1,6449
vote 9	0,47	–	–	–	0,9633	0,9900	0,3057	1,6476
vote 10	0,47	–	–	–	0,9633	0,9891	0,3018	1,6507
waveform-5000 1	0,00	0,0520	2,8873	0	1,5849	–	–	–
waveform-5000 2	0,00	0,0555	2,8919	0	1,5849	–	–	–
waveform-5000 3	0,00	0,0498	2,8933	0	1,5849	–	–	–
waveform-5000 4	0,00	0,0502	2,8912	0	1,5849	–	–	–
waveform-5000 5	0,00	0,0507	2,8898	0	1,5849	–	–	–
waveform-5000 6	0,00	0,0490	2,8879	0	1,5849	–	–	–
waveform-5000 7	0,00	0,0526	2,8855	0	1,5849	–	–	–
waveform-5000 8	0,00	0,0530	2,8946	0	1,5849	–	–	–
waveform-5000 9	0,00	0,0528	2,8912	0	1,5849	–	–	–
waveform-5000 10	0,00	0,0496	2,8898	0	1,5849	–	–	–