



UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE
UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO



WILAMIS KLEITON NUNES DA SILVA

CONSTRUÇÕES DE COMITÊS DE CLASSIFICADORES
MULTIRRÓTULOS NO APRENDIZADO
SEMISSUPERVISIONADO MULTIDESCRICÃO

MOSSORÓ - RN
2017

WILAMIS KLEITON NUNES DA SILVA

**CONSTRUÇÕES DE COMITÊS DE CLASSIFICADORES
MULTIRRÓTULOS NO APRENDIZADO
SEMISSUPERVISIONADO MULTIDESCRICÃO**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação - associação ampla entre a Universidade do Estado do Rio Grande do Norte e a Universidade Federal Rural do Semi-Árido, para a obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof^o ARAKEN DE MEDEIROS SANTOS,
D.Sc.

**MOSSORÓ - RN
2017**

© Todos os direitos estão reservados a Universidade Federal Rural do Semi-Árido. O conteúdo desta obra é de inteira responsabilidade do (a) autor (a), sendo o mesmo, passível de sanções administrativas ou penais, caso sejam infringidas as leis que regulamentam a Propriedade Intelectual, respectivamente, Patentes: Lei nº 9.279/1996 e Direitos Autorais: Lei nº 9.610/1998. O conteúdo desta obra tomar-se-á de domínio público após a data de defesa e homologação da sua respectiva ata. A mesma poderá servir de base literária para novas pesquisas, desde que a obra e seu (a) respectivo (a) autor (a) sejam devidamente citados e mencionados os seus créditos bibliográficos.

S Silva, Wilamis Kleiton Nunes da.
586 c CONSTRUÇÕES DE COMITÊS DE CLASSIFICADORES
MULTIRRÓTULOS NO APRENDIZADO SEMISSUPERVISIONADO
MULTIDESCRIÇÃO / Wilamis Kleiton Nunes da Silva. -
2017.
108 f. : il.

Orientador: Araken de Medeiros Santos Santos.
Dissertação (Mestrado) - Universidade Federal
Rural do Semi-árido, Programa de Pós-graduação em
, 2017.

1. Aprendizado de Máquina. 2. Aprendizado
Semissupervisionado Multidescrição. 3.
Classificação Multirrótulo. 4. Comitês de
Classificadores. I. Santos, Araken de Medeiros
Santos, orient. II. Título.

O serviço de Geração Automática de Ficha Catalográfica para Trabalhos de Conclusão de Curso (TCC's) foi desenvolvido pelo Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (USP) e gentilmente cedido para o Sistema de Bibliotecas da Universidade Federal Rural do Semi-Árido (SISBI-UFERSA), sendo customizado pela Superintendência de Tecnologia da Informação e Comunicação (SUTIC) sob orientação dos bibliotecários da instituição para ser adaptado às necessidades dos alunos dos Cursos de Graduação e Programas de Pós-Graduação da Universidade.

WILAMIS KLEITON NUNES DA SILVA

CONSTRUÇÕES DE COMITÊS DE CLASSIFICADORES MULTIRRÓTULOS NO
APRENDIZADO SEMISSUPERVISIONADO MULTIDISCRICÃO

Dissertação apresentada ao Programa de Pós-Graduação
em Ciência da Computação para a obtenção do título de
Mestre em Ciência da Computação.

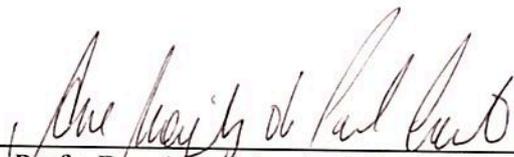
APROVADA EM: 18 / 08 / 2017



Prof. Dr. Araken de Medeiros Santos
Orientador e Presidente



Prof. Dr. Daniel Sabino Amorim de Araújo
Examinador Exteno - UFRN



Profa. Dra. Anne Magaly de Paula Canuto
Examinadora Extena - UFRN

AGRADECIMENTOS

Agradeço fundamentalmente a Deus pelo dom da vida, Deus que me capacita a todo momento para que eu possa enfrentar todos os desafios que aparecem em minha vida.

A minha avó (MÃE), Ozelita Nunes da Silva, pelo apoio e educação durante toda a minha vida sem o seu esforço e incentivo jamais chegaria aonde cheguei.

A minha esposa Lêdimila Nogueira da Rocha, pelo carinho, amizade, incentivo e compressão pelos meus momentos de ausência durante a realização deste trabalho. Você é o meu porto seguro.

As minhas filhas Sofia e Pétala pelo carinho, sorriso e abraços que me acalmam em momentos difíceis. Amo vocês.

A todos meus amigos do PPgCC. Sou grato em especial aos amigos Renê, Rafael, Hugo, Alex Cavaleira e Alex Trindade pelos incentivos e trocas de experiências durante o decorrer deste mestrado.

Ao meu orientador Prof^o. Araken de Medeiros Santos, pelo apoio, confiança, amizade e compreensão. Pelas várias mensagens trocadas via whatsapp, ligações e e-mail trocados sempre se dispondo a ajudar no desenvolvimento e nas orientações deste trabalho. O meu muito obrigado sem você jamais esse sonho de ser mestre chegaria a ser realizado.

Aos membros da banca examinadora, pela disponibilidade e contribuição para a evolução deste trabalho.

Aos professores do PPgCC, pelos ensinamentos, experiências e dedicação.

Muito obrigado a todos vocês!

RESUMO

São cada vez mais comuns problemas multirrótulos onde um rótulo pode ser atribuído a mais de uma instância, sendo chamados de problemas de classificação multirrótulo. Dentre os diferentes métodos de classificação multirrótulo, podemos citar os métodos BR (*Binary Relevance*), LP (*Label Powerset*) e RAKEL (*RAndom k-labELsets*). Tais métodos são ditos métodos de transformação do problema, pois consistem em transformar o problema multirrótulo em vários problemas de classificação tradicional (monorrótulo). A adoção de comitês de classificadores em problemas de classificação multirrótulo ainda é algo muito recente, com muito a ser explorado para a realização de pesquisas. O objetivo deste trabalho é realizar um estudo sobre a construção de comitês de classificadores multirrótulos construídos através da aplicação das técnicas de aprendizado semissupervisionado multidescrição, a fim de verificar se aplicação desse tipo de aprendizado na construção de comitês acarreta melhorias nos resultados. Os comitês de classificadores utilizados nos experimentos foram o *Bagging*, *Boosting* e *Stacking* como métodos de transformação do problemas foram utilizados os métodos BR, LP e Rakel e para a classificação multirrótulo semissupervisionada multidescrição foi utilizado o *Co-Training*. Ao fim das análises experimentais verificou-se que a utilização da abordagem semissupervisionado apresentou resultados satisfatórios, uma vez que as duas abordagens supervisionada e semissupervisionada utilizadas no trabalho apresentaram resultados semelhantes.

Palavras-chave: Aprendizado de Máquina, Aprendizado Semissupervisionado Multidescrição, Classificação Multirrótulo, Comitês de Classificadores.

ABSTRACT

Multi-label problems have become increasingly common, for a label can be attributed to more than one instance, being called multi-label classification problems. Among the different multilabel classification methods we can mention: BR (Binary Relevance), LP (Label Powerset) And RAKEL (RANDOM k-labELsets). Such methods have been recognized as methods for transforming the Problem, since they consist of turning the multi-label problem into several problems of traditional classification (mono label). However, the adoption of Classificatory committees in multi-label classification problems has still been new-found so far, With a great field to be explored for conducting researches as well. This work aims of doing a study on the construction of multi-label classifiers committees Built through the application of multi- description semi-supervised learning techniques, in order to verify if application of this type of learning in the construction of committees results in improvements linked to the results. The committees of classifiers used in the experiments were Bagging, Boosting and Stacking as methods of transformation of the problems used were the BR, LP and Rakel methods and for classification multi-label multi-label semi-supervised multi-description was used Co-Training. At the end of the experimental analyzes, it was verified that the use of the semi-supervised approach presented satisfactory results, since the two approaches presented similar results.

Key-words: Machine Learning, Multi-description Semi Supervised Learning, Multi-Label Classification, Classification Committees.

LISTA DE FIGURAS

Figura 1 – Hierarquia dos Tipos de ML	21
Figura 2 – Abordagem do <i>Self-Training</i>	23
Figura 3 – Abordagem do <i>Co-training</i>	25
Figura 4 – Processo de Classificação.	28
Figura 5 – Problema de Classificação Multirrótulo.	30
Figura 6 – Criação do Classificador Multirrótulo no BR.	31
Figura 7 – Predição do Classificador Multirrótulo BR.	31
Figura 8 – Construção do Classificador Multirrótulo.	33
Figura 9 – Construção do Classificador Multirrótulo RAKEL.	34
Figura 10 – Representação do Método <i>Bagging</i>	41
Figura 11 – Representação do Método <i>AdaBoosting</i>	42
Figura 12 – Representação do Método <i>Stacking</i>	44
Figura 13 – Média dos melhores resultados do método BR com os comitês de classificadores <i>Bagging</i> , <i>Boosting</i> e <i>Stacking</i>	59
Figura 14 – Média dos melhores resultados do método LP com os comitês de classificadores <i>Bagging</i> , <i>Boosting</i> e <i>Stacking</i>	61
Figura 15 – Média dos melhores resultados do método RAKEL com os comitês de classificadores <i>Bagging</i> , <i>Boosting</i> e <i>Stacking</i>	63
Figura 16 – Média dos melhores resultados dos métodos BR, LP e RAKEL com o comitê de classificadores <i>Bagging</i>	65
Figura 17 – Média dos melhores resultados dos métodos BR, LP e RAKEL com o comitê de classificadores <i>Boosting</i>	67
Figura 18 – Média dos melhores resultados dos métodos BR, LP e RAKEL com o comitê de classificadores <i>Stacking</i>	69
Figura 19 – Média dos melhores resultados do método BR com os comitês de classificadores <i>Bagging</i> , <i>Boosting</i> e <i>Stacking</i>	73
Figura 20 – Média dos melhores resultados do método LP com os comitês de classificadores <i>Bagging</i> , <i>Boosting</i> e <i>Stacking</i>	75
Figura 21 – Média dos melhores resultados do método RAKEL com os comitês de classificadores <i>Bagging</i> , <i>Boosting</i> e <i>Stacking</i>	77
Figura 22 – Média dos melhores resultados dos métodos BR, LP e RAKEL com o comitê de classificadores <i>Bagging</i>	79
Figura 23 – Média dos melhores resultados dos métodos BR, LP e RAKEL com o comitê de classificadores <i>Boosting</i>	81
Figura 24 – Média dos melhores resultados dos métodos BR, LP e RAKEL com o comitê de classificadores <i>Stacking</i>	83

Figura 25 – Média dos melhores resultados do método BR utilizando a abordagem do aprendizado supervisionado <i>versus</i> método BR utilizando a abordagem do aprendizado semissupervisionado.	85
Figura 26 – Média dos melhores resultados do método LP utilizando a abordagem do aprendizado supervisionado <i>versus</i> método LP utilizando a abordagem do aprendizado semissupervisionado.	87
Figura 27 – Média dos melhores resultados do método Rakel utilizando a abordagem do aprendizado supervisionado <i>versus</i> método Rakel utilizando a abordagem do aprendizado semissupervisionado.	89
Figura 28 – Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores <i>Bagging</i> utilizando a abordagem do aprendizado supervisionado <i>versus</i> métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado.	91
Figura 29 – Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores <i>Boosting</i> utilizando a abordagem do aprendizado supervisionado <i>versus</i> métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado.	93
Figura 30 – Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores <i>Stacking</i> utilizando a abordagem do aprendizado supervisionado <i>versus</i> métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado.	95

LISTA DE TABELAS

Tabela 1 – Conjunto de Dados Multirrótulo	29
Tabela 2 – Conjunto de Exemplos Multirrótulo	32
Tabela 3 – Conjunto de Dados Transformados Usando RAKEL	34
Tabela 4 – Detalhamentos das Bases de Dados Multirrótulo	50
Tabela 5 – Média e desvio padrão do métodos BR com os comitês de classifica- dores <i>Bagging</i> , <i>Boosting</i> e <i>Stacking</i>	60
Tabela 6 – Média e desvio padrão do método LP com os comitês de Classificado- res <i>Bagging</i> , <i>Boosting</i> e <i>Stacking</i>	62
Tabela 7 – Média e desvio padrão dos melhores resultados do método Rakel com os comitês de classificadores <i>Bagging</i> , <i>Boosting</i> e <i>Stacking</i>	64
Tabela 8 – Média e desvio padrão dos método BR, LP e Rakel com o comitê de classificadores <i>Bagging</i>	66
Tabela 9 – Média e desvio padrão dos métodos BR, LP e Rakel com o comitê de classificadores <i>Boosting</i>	68
Tabela 10 – Média e desvio padrão dos métodos BR, LP e Rakel com o comitê de classificadores <i>Stacking</i>	70
Tabela 11 – Média e desvio padrão do método BR com os comitês de classificadores <i>Bagging</i> , <i>Boosting</i> e <i>Stacking</i>	74
Tabela 12 – Média e desvio padrão do método LP com os comitês de classificadores <i>Bagging</i> , <i>Boosting</i> e <i>Stacking</i>	76
Tabela 13 – Média e desvio padrão do método Rakel com os comitês de classifica- dores <i>Bagging</i> , <i>Boosting</i> e <i>Stacking</i>	78
Tabela 14 – Média e desvio padrão dos métodos BR, LP e Rakel com o comitê de classificadores <i>Bagging</i>	80
Tabela 15 – Média e desvio padrão dos métodos BR, LP e Rakel com o comitê de classificadores <i>Boosting</i>	82
Tabela 16 – Média e desvio padrão dos métodos BR, LP e Rakel com o comitê de classificadores <i>Stacking</i>	84
Tabela 17 – Média dos melhores resultados do método BR utilizando a aborda- gem do aprendizado supervisionado <i>versus</i> método BR utilizando a abordagem do aprendizado semissupervisionado.	86
Tabela 18 – Abordagem do aprendizado supervisionado <i>versus</i> abordagem do aprendizado semissupervisionado.	88
Tabela 19 – Média dos melhores resultados do método Rakel utilizando a aborda- gem do aprendizado supervisionado <i>versus</i> método Rakel utilizando a abordagem do aprendizado semissupervisionado.	90

Tabela 20 – Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores <i>Bagging</i> utilizando a abordagem do aprendizado supervisionado <i>versus</i> métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado.	92
Tabela 21 – Média dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores <i>Boosting</i> utilizando a abordagem do aprendizado supervisionado <i>versus</i> métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado.	94
Tabela 22 – Média dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores <i>Stacking</i> utilizando a abordagem do aprendizado supervisionado <i>versus</i> métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado.	96

LISTA DE ABREVIATURAS E SIGLAS

AD	Árvore de Decisão
AM	Aprendizado de Máquina
AVPREC	<i>Average Precision</i>
BAGGING	<i>Bootstrap AGGregatING</i>
BPMLL	<i>BackPropagation for MultiLabel Learning</i>
BR	<i>Binary Relevance</i>
CBMLC	<i>Clustering-based MultiLabel Classification</i>
C4.5	Algoritmo de AM para construção de Árvore de Decisão
DAG	<i>Directed Acyclic Graph</i>
DT	<i>Decision Tree</i>
ERRSS	<i>Error Set Size</i>
FM	<i>F-Measure</i>
HLOSS	<i>Hamming Loss</i>
IA	Inteligência Artificial
K-NN	<i>k-Nearest-Neighbor</i>
LP	<i>Label Powerset</i>
MMAC	<i>Multi-Class Multilabel Associative Classification</i>
MIML-RBF	<i>Multi-Instance MultiLabel Radial Basis Function</i>
MLR	<i>Multi-response Linear Regression</i>
ML-kNN	<i>MultiLabel k Nearest Neighbors</i>
MMP	<i>Multiclass Multilabel Perceptron</i>
NB	<i>Naive-Bayes</i>
PREC	<i>Precision</i>
RAkEL	<i>RAandom k-LabELsets</i>

RAkELd	<i>RAndom K-LabELsets disjoint</i>
RBF	<i>Radial Basis Function</i>
RLOSS	<i>Ranking Loss</i>
RNA	<i>Redes Neurais Artificiais</i>
PNN	<i>Probabilistic Neural Networks</i>
STACKING	<i>Stacked Generalization</i>
SVM	<i>Support Vector Machines</i>
1-ERR	<i>One-Error</i>

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Motivação	17
1.2	Objetivos	17
1.2.1	Geral	17
1.2.2	Específicos	18
1.3	Estrutura do trabalho	18
2	APRENDIZADO DE MÁQUINA	19
2.1	Conceitos Básicos do Aprendizado de Máquina	19
2.2	Tipos de Aprendizado de Máquina	21
2.3	Aprendizado Supervisionado	21
2.4	Aprendizado Não Supervisionado	22
2.5	Aprendizado Semissupervisionado	22
2.5.1	Descrição Simples (<i>Self-Training</i>)	23
2.5.2	Multidescrição (<i>Co-training</i>)	24
2.6	Considerações finais	26
3	CLASSIFICAÇÃO DE DADOS	27
3.1	Classificação de Dados Multirrótulo	28
3.2	Métodos de Transformação do Problema	30
3.3	Adaptação de Algoritmos	34
3.4	Considerações finais	38
4	COMITÊ DE CLASSIFICADORES	39
4.1	Multiclassificadores Homogêneos	40
4.2	Multiclassificadores Heterogêneos	42
4.3	Considerações finais	45
5	TRABALHOS RELACIONADOS	46
5.1	Aprendizado semissupervisionado	46
5.2	Classificação multirrótulo	47
5.3	Comitês de classificadores com aprendizado semissupervisionado	48
6	METODOLOGIA DOS EXPERIMENTOS	49
6.1	Bases de Dados Multirrótulo	50
6.1.1	<i>Birds</i>	50
6.1.2	<i>Emotions</i>	50

6.1.3	<i>Enron</i>	51
6.1.4	<i>Yeast</i>	51
6.1.5	<i>CAL500</i>	51
6.1.6	<i>Flags</i>	51
6.2	Métricas de Avaliação	51
6.2.1	Métricas baseadas em Bipartição	52
6.2.2	<i>Hamming Loss</i>	52
6.2.3	<i>Recall</i>	52
6.2.4	<i>F-Measure</i>	52
6.2.5	<i>Precision</i>	53
6.3	Métricas baseadas em Ranking	53
6.3.1	<i>One-Error</i>	53
6.3.2	<i>Average Precision</i>	54
6.4	Testes Estatísticos	54
6.5	Configuração dos Experimentos	55
6.6	Considerações finais	56
7	RESULTADOS EXPERIMENTAIS: MÉTODOS DE CLASSIFICAÇÃO MULTIRRÓTULO SUPERVISIONADO	58
7.1	Melhores resultados do método BR com os comitês de classificadores <i>Bagging</i> , <i>Boosting</i> e <i>Stacking</i>	59
7.2	Melhores resultados do método LP com os comitês de classificadores <i>Bagging</i> , <i>Boosting</i> e <i>Stacking</i>	61
7.3	Melhores resultados do método Rakel com os comitês de classificadores <i>Bagging</i> , <i>Boosting</i> e <i>Stacking</i>	63
7.4	Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores <i>Bagging</i>	65
7.5	Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores <i>Boosting</i>	67
7.6	Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores <i>Stacking</i>	69
7.7	Considerações Finais	71
8	MÉTODOS DE CLASSIFICAÇÃO MULTIRRÓTULO SEMISUPERVISIONADO MULTIDESCRICÇÃO	72
8.1	Melhores resultados do método BR com os comitês de classificadores <i>Bagging</i> , <i>Boosting</i> e <i>Stacking</i>	73
8.2	Melhores resultados do método LP com os comitês de classificadores <i>Bagging</i> , <i>Boosting</i> e <i>Stacking</i>	75

8.3	Melhores resultados do método Rakel com os comitês de classificadores <i>Bagging</i> , <i>Boosting</i> e <i>Stacking</i>	77
8.4	Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores <i>Bagging</i>	79
8.5	Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores <i>Boosting</i>	81
8.6	Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores <i>Stacking</i>	83
8.7	Melhores resultados do método BR utilizando a abordagem do aprendizado supervisionado <i>versus</i> método BR utilizando a abordagem do aprendizado semissupervisionado	85
8.7.1	Melhores resultados do método LP utilizando a abordagem do aprendizado supervisionado <i>versus</i> método LP utilizando a abordagem do aprendizado semissupervisionado	87
8.7.2	Melhores resultados do método Rakel utilizando a abordagem do aprendizado supervisionado <i>versus</i> método Rakel utilizando a abordagem do aprendizado semissupervisionado	89
8.7.3	Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores <i>Bagging</i> utilizando a abordagem do aprendizado supervisionado <i>versus</i> métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado	91
8.7.4	Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores <i>Boosting</i> utilizando a abordagem do aprendizado supervisionado <i>versus</i> métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado	93
8.7.5	Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores <i>Stacking</i> utilizando a abordagem do aprendizado supervisionado <i>versus</i> métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado	95
8.8	Considerações Finais	97
9	CONCLUSÃO	99
9.0.1	Trabalhos Futuros	100
9.0.2	Produções Científicas	101
	REFERÊNCIAS	102

1 INTRODUÇÃO

Aprendizado de máquina é uma área de Inteligência Artificial cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática através de experiências passadas bem sucedidas (MONARD; BARANAUSKAS, 2003). Considerando o grau de supervisão aplicado durante o processo de aprendizagem, o aprendizado de máquina pode ser dividido em duas categorias principais: aprendizado supervisionado e aprendizado não supervisionado. A principal diferença entre o aprendizado supervisionado e o não supervisionado diz respeito à forma como é realizado o processo de generalização do conhecimento.

No aprendizado supervisionado, durante o treinamento, os métodos recebem como entrada exemplos juntamente com a informação de saída desejada, representando a classe a que aquele exemplo pertence. Por sua vez, no aprendizado não supervisionado, os métodos recebem como entrada exemplos sem a informação de saída, ou seja, não se conhece a priori a classe a que os exemplos de entrada do conjunto de treinamento pertencem. A comunidade científica vem estudando uma terceira categoria de aprendizado, surgida através da junção do aprendizado supervisionado com o aprendizado não supervisionado. A essa nova categoria de aprendizado dá-se o nome de aprendizado semissupervisionado (MATSUBARA; MONARD; BATISTA, 2005). O aprendizado semissupervisionado combina dados rotulados e dados não-rotulados, durante a fase de treinamento, reduzindo assim, a necessidade de uma grande quantidade de dados rotulados quando somente um pequeno conjunto de exemplos rotulados está disponível. Como exemplos de técnicas de aprendizado semissupervisionado temos o *Self-Training* (descrição simples) (YAROWSKY, 1995) e o *Co-training* (multidescrição) (BLUM; MITCHELL, 1998).

No *Self-Training*, um classificador é primeiramente gerado a partir de uma pequena quantidade de exemplos rotulados. Logo após ser gerado, o classificador é então utilizado para classificar exemplos não rotulados. Assim, os exemplos não rotulados, juntamente com os rótulos preditos pelo classificador, são adicionados ao conjunto de treinamento. Em seguida, o classificador é retreinado e todo o procedimento é repetido. É importante ressaltar que o classificador faz uso de suas próprias previsões para ensinar a si próprio, daí o nome *Self*.

O *Co-training* é uma poderosa abordagem do aprendizado semissupervisionado baseada na ideia da multidescrição. Nesta abordagem, os exemplos de treinamento são descritos por dois ou mais conjuntos de atributos disjuntos, ou seja, através de diferentes descrições (MATSUBARA; MONARD; BATISTA, 2005). Assim, dois ou mais classificadores são inicialmente gerados, cada um utilizando uma descrição diferente

dos exemplos rotulados. Desse modo, pode-se afirmar que *Co-training* baseia-se na ideia de que cada um dos classificadores rotulará exemplos que conterão informações desconhecidas pelo outro classificador.

Muitos dos algoritmos de aprendizado de máquina supervisionado propostos na literatura tem como objetivo a solução de problemas conhecidos como problemas de classificação de dados. Desse modo, um processo de classificação consiste na atribuição de rótulos aos dados, de tal forma que esses rótulos confirmem informações aos dados categorizados sob o mesmo rótulo.

Com o aumento do uso de sistemas de classificadores em diversas tarefas, surge a necessidade de sistemas cada vez mais eficazes. Por este motivo, verifica-se um crescente interesse da comunidade científica por pesquisas acerca da combinação ou integração de classificadores com o objetivo de obter melhorias de desempenho. Sistemas nos quais é feita a combinação de dois ou mais classificadores são chamados de comitês de classificadores que podem ser do mesmo tipo ou de diferentes tipos utilizando diferentes algoritmos de aprendizado de máquina. Dentre os comitês de classificadores mais utilizados na literatura destacam-se o *Bagging* (BREIMAN, 1996), *Boosting* (FREUND; SCHAPIRE, 1999) e *Stacking* (WOLPERT, 1992). Considerando as vantagens do aprendizado semissupervisionado e o dos comitês de classificadores, o objetivo do trabalho é realizar um estudo sobre as construções de comitês de classificadores multirrótulos através da aplicação das técnicas de aprendizado semissupervisionado multidescrição, a fim de verificar se aplicação desse tipo acarreta melhorias nos resultados.

1.1 MOTIVAÇÃO

Apesar de, tanto a classificação multirrótulo, quanto o aprendizado semissupervisionado e a utilização de comitês de classificadores estarem apresentando resultados favoráveis à sua utilização, poucos trabalhos foram realizados unindo as vantagens das três áreas. Diante desse fato, como não foram encontradas pesquisas que explorem as construções de comitês classificadores multirrótulo no aprendizado semissupervisionado multidescrição, esse passou a ser o foco desse trabalho.

1.2 OBJETIVOS

Nesse trabalho, os referidos objetivos a seguir deverão ser alcançados:

1.2.1 Geral

- Realizar um estudo sobre as construções de comitês de classificadores multirrótulos através da aplicação das técnicas de aprendizado semissupervisionado multidescrição, a fim de verificar se aplicação desse tipo de aprendizado nas construções de comitês acarreta melhorias nos resultados.

1.2.2 Específicos

- Pesquisar na literatura as aplicações das técnicas de classificação multirrótulo e dos principais métodos de construções de comitês de classificadores existentes em diferentes domínios de aplicações: esta fase inicia-se com uma vasta pesquisa bibliográfica, tendo como objetivo a identificação das técnicas de classificação multirrótulo, assim como dos métodos de construções de comitês de classificadores mais utilizados;
- Implementar as técnicas de construções de comitês de classificadores para problemas de classificação Multirrótulo: nesta fase, serão propostas e implementadas as técnicas de construções de comitês de classificadores para problemas de classificação Multirrótulo baseadas em algumas das abordagens investigadas, abrindo possibilidade para contribuições originais para o avanço do estado da arte na área;
- Avaliar os modelos propostos a diferentes problemas de classificação: esta fase inicia-se com a busca por diferentes domínios nos quais os modelos existentes possam ser aplicados. Em seguida, pode ser executado um estágio de pré-processamento dos dados, viabilizando sua aplicação nos modelos a serem investigados. Neste estágio de pré-processamento podem ser executados passos como: padronização do formato dos dados, seleção de atributos, seleção de rótulos, dentre outros;
- Analisar comparativamente os resultados obtidos por comitês de classificadores construídos através da abordagem supervisionada tradicional e da abordagem semissupervisionada multidescrição.

1.3 ESTRUTURA DO TRABALHO

O texto está organizado da seguinte forma: o Capítulo 2 apresenta os fundamentos teóricos associados à área de aprendizado de máquina. O Capítulo 3 apresenta os principais conceitos inerentes a classificação de dados multirrótulo e também os principais algoritmos encontrados na literatura para problemas de classificação de dados multirrótulo. O Capítulo 4 apresenta os principais conceitos sobre os comitês de classificadores homogêneos e heterogêneos. O Capítulo 5 apresenta os principais trabalhos relacionados com a pesquisa. O Capítulo 6 aborda os procedimentos adotados para a realização dos experimentos, assim como as bases de dados e as medidas de avaliação utilizadas nos experimentos. O Capítulo 7 apresenta os resultados experimentais obtidos pela classificação multirrótulo supervisionada. O Capítulo 8 apresenta os resultados experimentais obtidos pela classificação multirrótulo semissupervisionada multidescrição, e por fim, o Capítulo 9 traz as considerações finais do trabalho.

2 APRENDIZADO DE MÁQUINA

A inteligência é a chave do sucesso da espécie humana. Graças a ela o homem criou inúmeros artefatos que garantiram a sua supremacia sobre outras espécies, mesmo não sendo o de maior estatura, o mais rápido ou o mais forte entre os animais. Esta capacidade há milhares de anos intriga os estudiosos, que procuram entender como o ser humano pensa e aprende e quais são as leis que governam o processo de aprendizado. A Inteligência Artificial (IA) é uma área de pesquisa em ciência da computação que não só tenta compreender, mas também construir entidades inteligentes (NORVIG; RUSSELL, 2014). Desde sua criação, na década de 50, a IA tem evoluído muito, com essa evolução surgiu o aprendizado de máquina.

Aprendizado de máquina (AM) é um campo da Inteligência Artificial que tem como objetivo desenvolver técnicas computacionais de aprendizado e de aquisição de conhecimentos de forma automática. O sistema de aprendizado é um programa de computador que toma decisões baseadas em experiências armazenadas através da solução bem sucedida de problemas antecedentes.

Este capítulo está dividido nas seguintes seções: a primeira Seção trata dos conceitos básicos do aprendizado de máquina; a segunda Seção trata do tipos de aprendizado de máquina; a terceira Seção trata do aprendizado supervisionado; a quarta Seção trata do aprendizado não supervisionado; a quinta Seção trata do aprendizado semissupervisionado; e por fim, a sexta Seção traz as considerações finais do capítulo.

2.1 CONCEITOS BÁSICOS DO APRENDIZADO DE MÁQUINA

Na literatura encontra-se várias definições sobre conceitos básicos de aprendizado de máquina. Nesta Seção, é apresentados os principais termos referentes à área de aprendizado de máquina, bem como suas definições e características. Dentre os principais termos temos:

- **Conjunto de exemplos:** conhecido como conjunto de dados ou conjunto de instâncias, é composto por um número de exemplos com seus respectivos valores de atributos.
- **Conjunto de treinamento:** utilizado como entrada pelos algoritmos de aprendizado. É a partir do conjunto treinamento que são construídos classificadores, portanto, ele deve ser representativo na distribuição da população dos dados do domínio.
- **Conjunto de teste:** tem a função de avaliar o modelo construído. Esse conjunto não deve ser apresentado ao algoritmo de aprendizado durante a elaboração do

modelo. O ideal é que o conjunto de testes não tenha exemplos em comum com os exemplos do conjunto de treinamento.

- **Conjunto de validação:** utilizado na realização de ajustes no modelo construído pelo algoritmo de aprendizado. Os exemplos do conjunto de teste não são utilizados diretamente na construção do modelo, são utilizados para o seu ajuste. Dessa forma, esses exemplos são perceptíveis de maneira indireta durante o processo de aprendizado, o que obriga aos exemplos de validação que sejam distintos dos exemplos de teste.
- **Indutor:** é o algoritmo de aprendizado que faz uso de um processo indutivo para gerar a sua hipótese ou modelo. Essa hipótese é utilizada para classificar uma instância ou exemplo.
- **Classificador ou hipótese:** dado um conjunto de exemplos do conjunto de treinamento, um indutor, ou algoritmo de aprendizado, gera como saída um classificador, ou hipótese, de forma que, dado um novo padrão, ele possa identificar sua classe com a maior precisão possível (MONARD, 2003).
- **Atributo:** também chamados campos ou variáveis, são utilizados para descrever características ou aspectos de um exemplo. Geralmente, os atributos se dividem em dois modelos: discretos ou contínuos. Atributos discretos contêm um número finito ou infinito contável de valores. Um especial tipo de atributo discreto é o atributo binário, ou *booleano*, que representa apenas dois valores de saída, como 0/1, sim/não, ausência/presença, verdadeiro/falso, entre outros. Já os atributos contínuos, por sua vez, podem ser representados por números infinitos de valores. Geralmente esses atributos contínuos são resultados de medidas e representados por números reais, como peso, tamanho e distância (FACELLI et al., 2011).
- **Classe ou Rótulo:** também chamado de atributo-alvo, uma classe ou rótulo, descreve o conceito-meta, ou seja, o conceito que se deseja aprender para tornar viável a realização de previsões a seu respeito (FACELLI et al., 2011). No aprendizado supervisionado, todo exemplo possui pelo menos um atributo especial denominado rótulo ou classe.
- **Exemplo:** um objeto único do mundo a partir do qual um modelo será aprendido ou usado, é também conhecido como padrão ou instância. Um exemplo descreve um objeto de uma determinada situação na qual se deseja aprender através de um vetor de valores de características ou atributos (MONARD, 2003). Além do vetor de características, um exemplo pode ter um atributo especial, que indica a que classe ou rótulo esse exemplo pertence.

2.2 TIPOS DE APRENDIZADO DE MÁQUINA

As técnicas de aprendizado de máquina empregam um princípio de inferência denominado indução, no qual é possível obter conclusões genéricas a partir de um conjunto particular de exemplos. Pode-se dividir, o aprendizado de máquina em três principais categorias: aprendizado supervisionado, aprendizado não supervisionado e aprendizado semissupervisionado (LORENA; CARVALHO, 2007), ver figura 1.

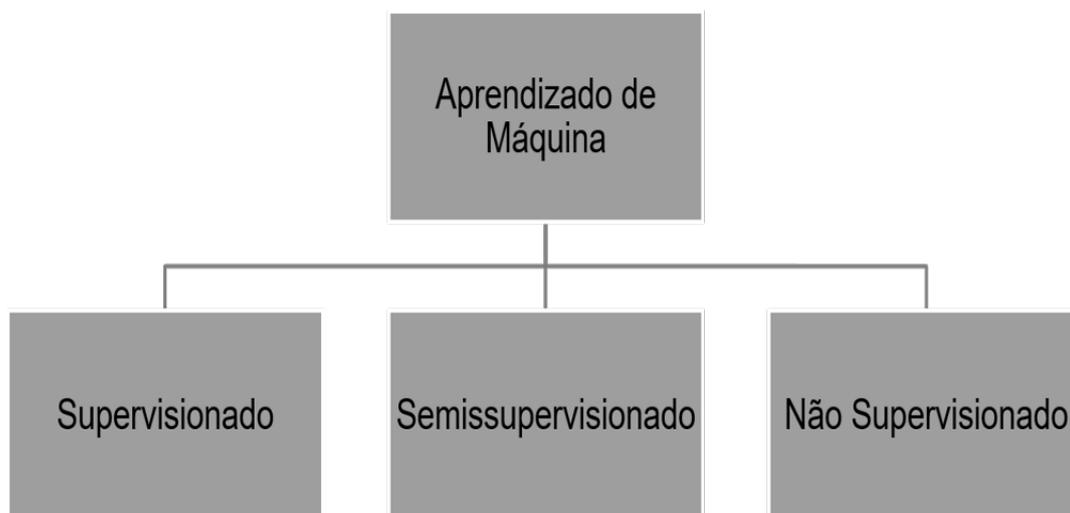


Figura 1 – Hierarquia dos Tipos de ML

Fonte: Autoria Própria

O aprendizado supervisionado é utilizado quando há uma grande quantidade de exemplos rotulados, enquanto no aprendizado não supervisionado é utilizado quando os exemplos não estão rotulados, já no aprendizado semissupervisionado é utilizado quando há poucos exemplos rotulados e um número expressivos de exemplos não rotulados.

2.3 APRENDIZADO SUPERVISIONADO

De acordo com Rezende (2003), no aprendizado supervisionado, existe um supervisor no qual fornece um conjunto de exemplos (instâncias) com seus respectivos atributos de entradas e de saídas (rótulos) sendo utilizados pelo algoritmo de aprendizado, com isso, o supervisor externo pode avaliar a capacidade da hipótese induzida de prever o valor de saída para novos exemplos do conjunto.

Quando um problema possui rótulo com um valor discreto, este é denominado classificação e, quando os rótulos são contínuos, é chamado de regressão.

2.4 APRENDIZADO NÃO SUPERVISIONADO

Diferentemente do aprendizado supervisionado, o não supervisionado não se utiliza referências, ou seja, não ocorre um treinamento com o conhecimento do ambiente. Lorena (2007) destaca que o algoritmo de aprendizado de máquina não supervisionado aprende a representar (ou agrupar) as entradas submetidas, segundo medidas de similaridade. Contudo, é importante enfatizar que os grupos no aprendizado não-supervisionado e as classes no aprendizado supervisionado não agrupam ou representam, necessariamente, os mesmos exemplos (MATSUBARA; MONARD; BATISTA, 2005).

O processo de agrupamento dos algoritmos de aprendizado não-supervisionado é realizado duas abordagens principais: por meio de alguma medida de distância ou de alguma medida de correlação, no agrupamento através de uma medida de distância, os exemplos podem ser agrupados com outros mais próximos, formando assim um *cluster*. Diversas medidas de distância podem ser utilizadas, dentre as quais podemos destacar a distância Euclidiana e a distância de Mahalanobis (MAHALANOBIS, 1936). No aprendizado não-supervisionado, a tarefa do algoritmo é agrupar os dados que não estão rotulados no conjunto de exemplos, o principal objetivo é agrupar exemplos não-rotulados através da identificação de padrões ou tendências, gerando agrupamentos ou *clusters* distintos, auxiliando assim no entendimento desses dados (COSTA, 2008).

2.5 APRENDIZADO SEMISSUPERVISIONADO

O aprendizado semissupervisionado emprega estratégia que estendem o aprendizado supervisionado ou não supervisionado para incluir informações adicionais do outro paradigma de aprendizagem. Por exemplo, a classificação semissupervisionado tem como objetivo treinar um classificador com dados rotulados e não rotulados, para obter um melhor classificador do que se fosse treinado só com dados rotulados (MANCHEGO, 2013).

De acordo com Manchego (2013), uma das vantagens do aprendizado semissupervisionado é o potencial de reduzir a necessidade de uma grande quantidade de dados rotulados, em domínios nos quais somente um pequeno conjunto de padrões rotulados está disponível. Outra vantagem desse tipo de aprendizado pode ser verificada quando o especialista não apresenta um total conhecimento sobre o conceito a ser aprendido, ou seja, esse especialista tem apenas o conhecimento de alguns padrões de um determinado conjunto de dados, apresentando assim, grande dificuldade de rotular exemplos para incrementar o conjunto de dados de treinamento.

Os dois algoritmos de aprendizado semissupervisionado mais utilizados na literatura são o *Self-Training* (descrição simples) e *Co-training* (multidescrição).

2.5.1 Descrição Simples (Self-Training)

Uma das primeiras formas de aprendizado semissupervisionado, denominado *Self-Training*, surgiu entre os anos de 60 e 70 (FOGELMAN-SOULIÉ et al., 2008) e foi provavelmente o primeiro algoritmo a utilizar dados não rotulados na classificação.

No *Self-Training* um classificador é primeiramente gerado a partir de uma pequena quantidade de exemplos rotulados. Logo após ser gerado, o classificador é então utilizado para classificar exemplos não-rotulados. Assim, os exemplos não-rotulados, juntamente com os rótulos preditos pelo classificador, são adicionados ao conjunto de treinamento. Em seguida, o classificador é retreinado e todo o procedimento é repetido. É importante ressaltar que o classificador faz uso de suas próprias previsões para ensinar a si próprio, daí o nome *self-training* (autotreinamento) (FOGELMAN-SOULIÉ et al., 2008).

De acordo com Ghinozzi (2012) foram os primeiros a usar o termo “Semissupervisionado” para a classificação com os dados rotulados e não rotulados no contexto do aprendizado de máquina. Com a expansão da *Internet* existe uma quantidade cada vez maior de dados e devido a impossibilidade de rotular todo esse volume de dados, o aprendizado semissupervisionado vem se popularizando cada vez mais. A Figura 2 apresenta a abordagem do *Self-Training* na forma original.

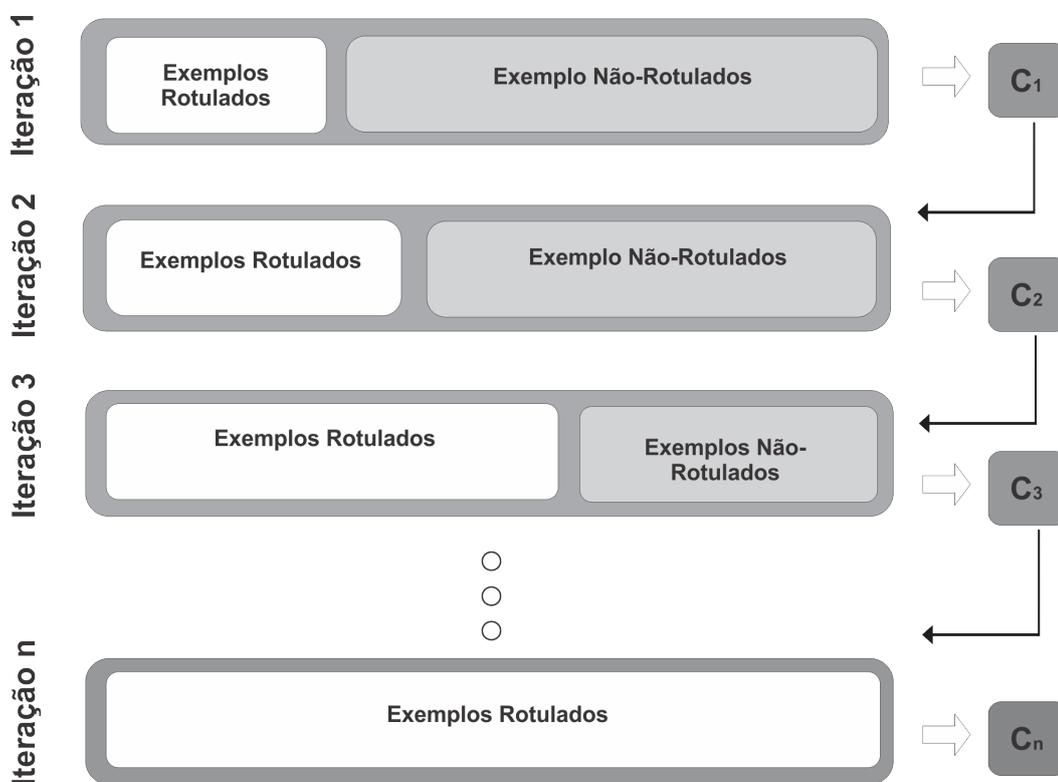


Figura 2 – Abordagem do *Self-Training*.

Fonte: (SANTOS, 2012).

Na sua forma original o *Self-Training* apresenta a seguinte forma básica:

1. Usar um conjunto de dados rotulados iniciais para treinar o classificador;
2. Aplicar este classificador a dados não rotulados e tomar as predições do classificador como certas para aquelas instâncias com maior confiança;
3. Expandir os dados rotulados, adicionando aqueles rotulados pelo classificador, e treinar novamente;
4. Repetir este processo – até satisfazer a condição de parada.

2.5.2 Multidescrição (Co-training)

O *Co-training* de Blum e Mitchell (1998), é uma poderosa abordagem ao aprendizado semissupervisionado baseada na ideia da multidescrição. Nesta abordagem, os exemplos de treinamento são descritos por dois ou mais conjuntos de atributos disjuntos, ou seja, através de diferentes descrições (MATSUBARA; MONARD; BATISTA, 2005). Nesse algoritmo, dois classificadores são inicialmente gerados, cada um utilizando uma descrição diferente dos exemplos rotulados. Uma vez que são utilizadas duas descrições, pode-se utilizar o mesmo algoritmo de aprendizado supervisionado para induzir as duas hipóteses (classificadores), assim como dois algoritmos de aprendizado supervisionado distintos. Assim, é gerado um classificador sobre os mesmos exemplos, mas cada um induzido segundo visões (descrições) diferentes. Desse modo, pode-se afirmar que o *Co-training* baseia-se na ideia de que cada um dos classificadores rotular exemplos que conterão informações desconhecidas pelo outro classificador (BLUM; MITCHELL, 1998), ver figura 3.

Blum e Mitchell (1998) mostram que o *Co-training* assume inicialmente que:

1. Os recursos podem ser divididos em dois grupos ou subconjuntos;
2. Cada subconjunto de recurso é suficiente para formar um bom classificador;
3. Os subconjuntos são condicionalmente independentes.

O *Co-training* gera as duas descrições a partir do conjunto de exemplos E . Dado que o conjunto A que representa o conjunto de atributos e o conjunto R que representa o conjunto de rótulos, as duas descrições A_{D1} e A_{D2} são geradas a partir da divisão do conjunto de atributos A , a qual descreve os exemplos e na qual, por definição, temos que $A = A_{D1} \cup A_{D2}$ e $A_{D1} \cap A_{D2} = \phi$, onde $A_{D1} = \{A_1 + A_2 + A_3, \dots, A_j\}$ e $A_{D2} = \{A_{j+1} + A_{j+2} + A_{j+3}, \dots, A_M\}$. Um rótulo com o valor '?' mostra que o exemplo não está rotulado (ASSIS, 2016).

Além da divisão em duas descrições, o conjunto de exemplos E deve ser dividido em subconjuntos de exemplos rotulados L e não rotulados U . Os exemplos do

subconjunto de exemplos rotulados L , por sua vez, devem ser subdivididos em L_{D_1} e L_{D_2} , nos quais representam o subconjunto de exemplos rotulados da descrição D_1 e D_2 , respectivamente, sendo $L = L_{D_1} \cup L_{D_2}$ e $L_{D_1} \cap L_{D_2} = \phi$. Da mesma forma, os exemplos do subconjunto de exemplos não rotulados U , por sua vez, devem ser subdivididos em U_{D_1} e U_{D_2} , sendo $U = U_{D_1} \cup U_{D_2}$ e $U_{D_1} \cap U_{D_2} = \phi$ (ASSIS, 2016). A Figura 3 mostra como o conjunto de exemplos E deve ser dividido em subconjuntos de exemplos rotulados L e não rotulados U .

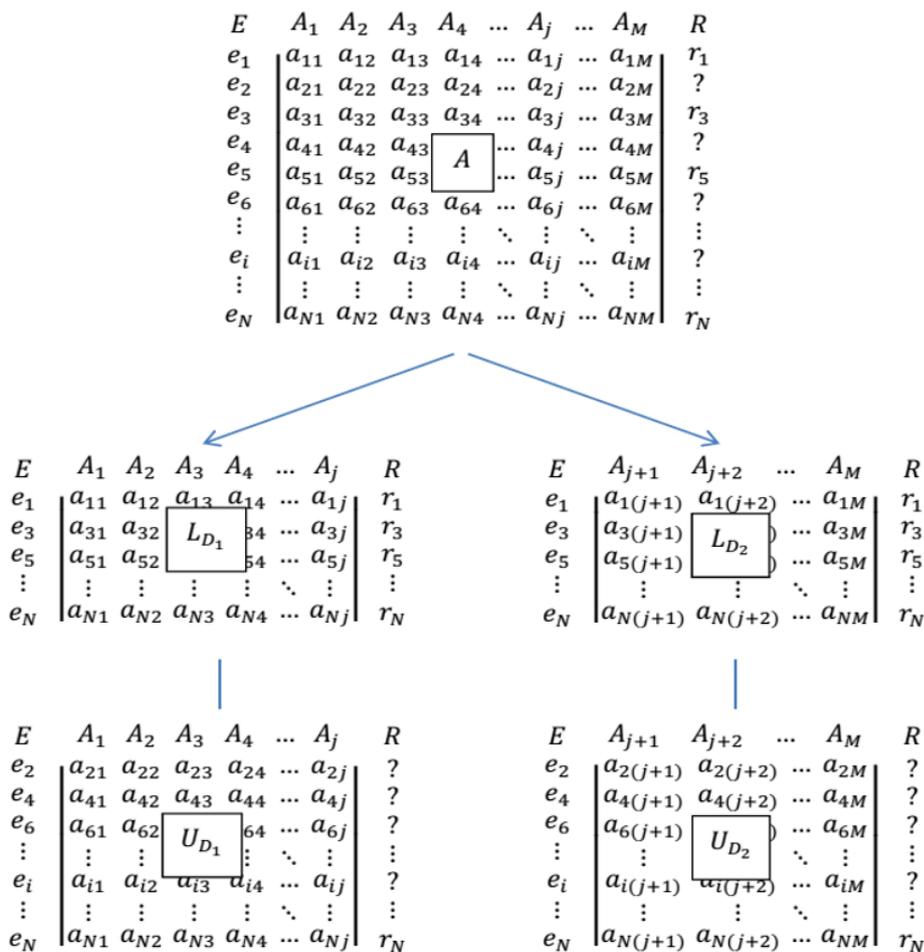


Figura 3 – Abordagem do Co-training.

Fonte: (ASSIS, 2016).

Krogl e Scheffer (2004) afirmam que o *Co-training* só é vantajoso se os conjuntos de dados utilizados na classificação são independentes. *Co-training* só pode funcionar se um dos classificadores rotula corretamente um pedaço de dados que o outro classificador previamente classificou erroneamente. Se ambos os classificadores concordar com todos os dados não rotulados, ou seja, eles não são independentes, rotulando os dados não cria novas informações.

2.6 CONSIDERAÇÕES FINAIS

Este capítulo aborda os principais conceitos básicos inerentes à área de aprendizado de máquina. Em seguida, foram apresentadas as principais características de cada uma das categorias de aprendizado: aprendizado supervisionado, aprendizado não supervisionado e aprendizado semissupervisionado.

No aprendizado supervisionado, os métodos tratam da construção de indutores com o objetivo de realizar inferências a partir de um conjunto de exemplos de treinamento rotulados, ou seja, para os quais se conhece a priori as classes às quais esses exemplos pertencem. O aprendizado supervisionado supõe a existência de um "Professor" que ensina que tipo de comportamento deve ser exibido em cada situação. Enquanto isso, no aprendizado não supervisionado, os métodos recebem como entrada exemplos sem a informação de saída, ou seja, não se conhece a priori a classe a que as instâncias do conjunto de exemplos pertencem.

O aprendizado não supervisionado ao contrário do aprendizado supervisionado não supõe a existência de um "Professor" que ensina que tipo de comportamento deve ser exibido em cada situação. O aprendizado semissupervisionado representa a junção do aprendizado supervisionado e não supervisionado. A junção dos dois primeiros tipos de aprendizado tem como objetivo reduzir a necessidade de uma grande quantidade de dados rotulados, quando somente um pequeno conjunto de exemplos rotulados está disponível, através da atribuição de rótulos a exemplos não rotulados por meio de um processo automático.

3 CLASSIFICAÇÃO DE DADOS

A classificação de dados é um processo no qual, a partir de um conjunto de dados brutos, são extraídas informações por meio de categorização. Desse modo, um processo de classificação consiste na atribuição de rótulos aos dados, de tal forma que esses rótulos confirmem informações aos dados categorizados sob o mesmo rótulo (COSTA, 2008).

Nas últimas décadas, intensas pesquisas vêm sendo realizadas na área de aprendizado de máquina, focando principalmente a classificação de dados, tais como em (CANUTO; SANTOS; VARGAS, 2011; GHINOZZI, 2012; CAMPOS et al., 2016; SILVA; CARVALHO et al., 2015; RODRIGUES, 2014). Problemas de classificação de dados são um subconjunto dos problemas de aprendizado supervisionado. Por tanto, o processo de classificação dá-se através de algoritmo de aprendizado, do qual a meta é gerar um classificador a partir de dados de entrada rotulados.

O processo de geração de um classificador se inicia com uma fase de pré-processamento dos dados. Esta fase tem como objetivo fazer com que os dados brutos pertencentes ao domínio sobre o qual será aplicado o algoritmo de classificação sejam representados de forma adequada, possibilitando a sua utilização na fase de treinamento.

Após a fase de pré-processamento, inicia-se a fase de treinamento. Nesta fase, cada exemplo E_i , representado por um par (x_i, y_i) , tal que $i = 1, 2, \dots, N$ é o número de exemplos de treinamento, x_i é vetor de atributos que descrevem cada exemplo e y_i é o valor do atributo-alvo Y , também conhecido como rótulo ou classe. O objetivo do processo de treinamento é obter uma função que mapeie cada E_i a seu Y_i correspondente. A função de mapeamento é encontrada com base nos ajustes dos parâmetros livres do modelo. Depois da fase de treinamento é gerado o classificador que é capaz de realizar previsões dos novos exemplos, por tanto a previsão é gerada pelo processo de generalização, ver Figura 4.

Pode-se observar através da Figura 4, que na fase de classificação os dados são organizados em conjuntos de N exemplos que fazem parte do domínio do problema, logo depois é aplicado algum algoritmo de classificação, após esta etapa vem a fase de processamento, logo depois é gerado o classificador para realizar as novas previsões.

Este capítulo está dividido nas seguintes seções: a primeira seção trata da classificação de dados multirrótulo; a segunda seção trata dos métodos de transformação do problema; a terceira seção trata da adaptação de algoritmos; e por fim, a quarta seção traz as considerações finais do capítulo.

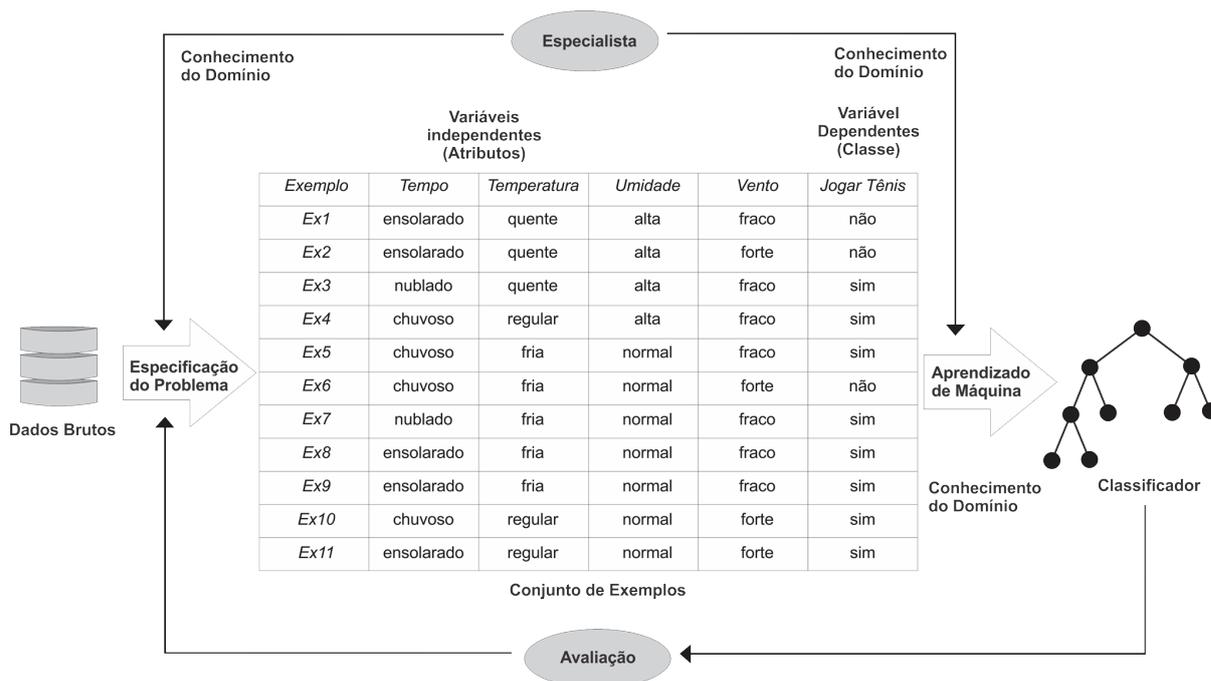


Figura 4 – Processo de Classificação.

Fonte: (RODRIGUES, 2014).

3.1 CLASSIFICAÇÃO DE DADOS MULTIRRÓTULO

A classificação de dados tradicional trata do aprendizado de um conjunto de exemplos (instâncias) em que cada um é associado a uma única classe (rótulo) de um conjunto de classes Y , de modo que $|Y| > 1$. Essa tarefa é chamada de classificação unirrótulo (*single-label classification*). Se $|Y| = 2$, então a tarefa de aprendizado é chamada de classificação unirrótulo binária. Por outro lado, se $|Y| > 2$, a tarefa é chamada de classificação unirrótulo multiclasse. Embora a classificação unirrótulo seja bastante utilizada, há vários domínios nos quais as instâncias estão associadas a um conjunto de classes L , tal que $L \in Y$, ou seja, cada exemplo pode ser associado a mais de uma classe simultaneamente.

Nesses casos, em que as classes não são disjuntas, à tarefa de classificação dá-se o nome de classificação multirrótulo (*Multi-label Classification*) (KATAKIS; TSOUMAKAS; VLAHAVAS, 2008). Em jornais, uma notícia pode estar associada a seções de esporte, músicas, ensino etc.

Para uma descrição formal, seja $L = \{j : j = 1, \dots, M\}$ um conjunto finito de classes (rótulos) em uma tarefa de aprendizado multirrótulo e $D = (x_i | L_i), i = 1, \dots, N$ o conjunto de exemplos de treinamento multirrótulo, onde x_i é o vetor de características (atributos) e $L_i \in Y$ o conjunto de classes da i -ésima instância (TSOUMAKAS; KATAKIS; VLAHAVAS, 2008). A Tabela 1 apresenta um conjunto de exemplos rotulados com mais de um rótulo representando uma notícia de um jornal, o rótulo 1 = esporte, 2 = música,

3= cinema, 4= educação.

Tabela 1 – Conjunto de Dados Multirrótulo

Exemplo	Rótulos
1	{Y1, Y2, Y3, Y4}
2	{Y1, Y3}
3	{Y2, Y3}
4	{Y2, Y3, Y4}
5	{Y3, Y4}

Fonte: (SANTOS, 2012).

Apesar de grande parte das pesquisas na área de classificação de dados se concentrarem na análise de dados com um único rótulo, nos últimos anos, a área de classificação de dados multirrótulo têm atraído a atenção de grande parte da comunidade científica, motivado principalmente pelo grande aumento no número de novas aplicações, tais como anotação semântica de imagens (BOUTELL et al., 2004); (ZHANG; ZHOU, 2007); (YANG; KIM; RO, 2007) e vídeo (QI et al., 2007); (SNOEK et al., 2006), função genômica (BLOCHEEL et al., 2006); (BARUTCUOGLU; SCHAPIRE; TROYANSKAYA, 2006); categorização de músicas através de emoções (LI; OGIHARA, 2006); (LI; OGIHARA, 2003); (WIECZORKOWSKA; SYNAK; RAŚ, 2006); (TROHIDIS et al., 2008) e marketing direcionado (ZHANG; BURER; STREET, 2006). Classificação de textos pode ser identificada como a área que tem maior aplicação de técnicas de classificação multirrótulo (TSOUMAKAS; KATAKIS; VLAHAVAS, 2008); (KATAKIS; TSOUMAKAS; VLAHAVAS, 2008); (PESTIAN et al., 2007). A Figura 5 apresenta um exemplo de classificação multirrótulo no qual está associada a música *Jazz n' Blues* da artista *Amy Winehouse*, pode-se perguntar: "a música *Jazz n' Blues* é *Jazz*"?. Nessa classificação, as possíveis classes são: "sim" ou "Não" tratando-se de uma problema da classificação tradicional monorrótulo. Entretanto, as aplicações multirrótulos estão associadas a mais de uma classe, quando se faz a seguinte pergunta: "Qual o gênero musical da música *Jazz n' Blues*"?. Porém, as possíveis respostas entram em conflito, pois a música é de fato *Jazz* e *Blues*, se ambas forem corretas tratando-se de uma classificação multirrótulo.

Na literatura, podem ser encontradas diferentes métodos e técnicas para tratar problemas de classificação multirrótulo. Em algumas destas técnicas, problemas multirrótulo são quebrados em vários problemas unirrótulo. Assim, classificadores unirrótulo podem ser combinados para viabilizar o tratamento de problemas de classificação multirrótulo. Por outro lado, há técnicas 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

Tipo	Pergunta	Rótulos possíveis	Resposta
Binária	A música é Jazz?	Sim, Não	Sim
Multiclasse	Qual o gênero musical?	Blues , Pop , Jazz	Jazz
Multirrótulo	Quais os gêneros musicais?	Blues , Pop , Jazz	Blues e Jazz

Figura 5 – Problema de Classificação Multirrótulo.

Fonte: Adaptado-(SILVA, 2014)

especificamente problemas de classificação multirrótulo (TSOUMAKAS; KATAKIS; VLAHAVAS, 2008). Os métodos de classificação multirrótulo podem ser divididos em grupos: Métodos de transformação do problema e Adaptação de algoritmos, são descritos a seguir.

3.2 MÉTODOS DE TRANSFORMAÇÃO DO PROBLEMA

No métodos de transformação do problema o problema multirrótulo existente é dividido em um ou vários problemas unirrótulos utilizando algoritmos de classificação unirrótulo para resolução de cada problema de maneira distinta. É importante considerar a independência de cada método sobre o algoritmo base. Assim, existe a possibilidade de um algoritmo de aprendizado unirrótulo construir os classificadores individualmente.

Existem na literatura diversos métodos capazes de transformar problemas de classificação multirrótulos em problemas de classificação unirrótulo, dentre eles os métodos *Binary Relevance*, *Label Powerset* e *RAkEL (RAndom k-LabELsets)* (TSOUMAKAS; KATAKIS; VLAHAVAS, 2011).

Binary Relevance

Em problemas multirrótulo, o método BR (*Binary Relevance*) é um dos métodos de transformação mais popular. É com o BR que é tida a visão de dividir os problemas em L problemas de classificação unirrótulo binário, sendo um para cada rótulo contido em L . Dessa maneira, o método considera que, de maneira independente, cada rótulo tem sua predição (TSOUMAKAS; KATAKIS, 2007). O conjunto de exemplos unirrótulos gerados a partir da divisão do problema multirrótulo em L problemas de classificação binária serão replicados e seus atuais rótulos serão modificados. Cada exemplo do conjunto replicado que contém o unirrótulo Y_j será identificado como positivo; para os exemplos que não contêm esse unirrótulo, estes serão classificados como negativos. A partir desses classificadores unirrótulos, uma nova instância tem sua predição multirrótulo formada pela união dos rótulos associados preditos positivamente pelos classificadores binários (TSOUMAKAS; KATAKIS, 2007). Ver Figura 6 e 7.

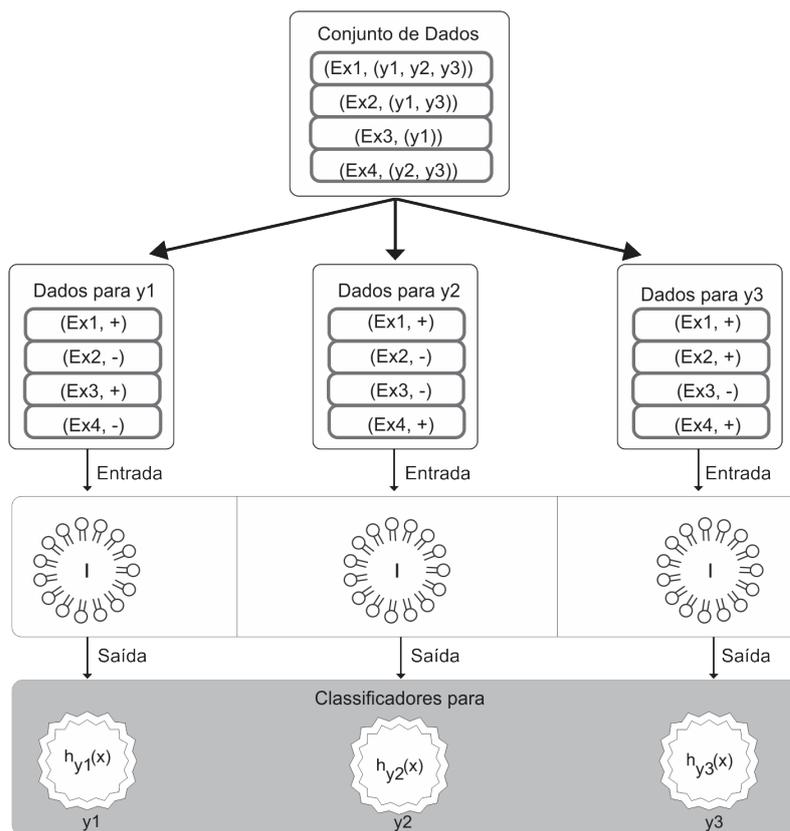


Figura 6 – Criação do Classificador Multirrótulo no BR.

Fonte: (RODRIGUES, 2014).

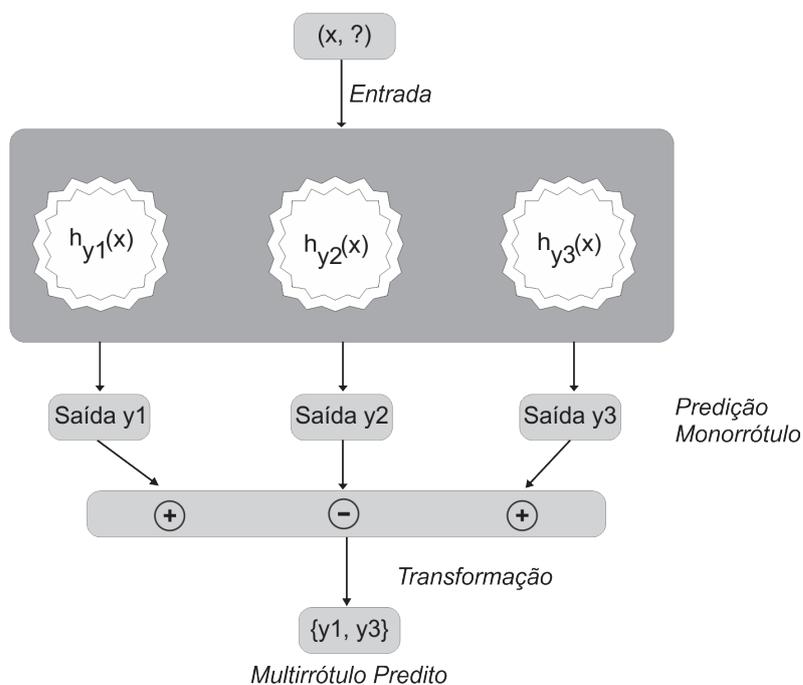


Figura 7 – Predição do Classificador Multirrótulo BR.

Fonte: Adaptado-(SANTOS, 2012).

Uma das deficiências do método *Binary Relevance* é por não serem consideradas correlações entre instâncias (rótulos) durante a construção do modelo de classificação multirrótulo, proporcionando, de certa maneira, uma perda nas informações relevantes à tarefa de classificação. Além disso, esse método propicia um modelo de classificação com pouca generalização e legibilidade. Isso se deve à construção de um único modelo para cada rótulo correlacionado a um exemplo de teste.

Label Powerset

O método LP (*Label Powerset*) procura transformar um problema multirrótulo em distintos subconjuntos do problema. Para tal feito, é criado, através dessa transformação, um conjunto unirrótulo multiclasse, na qual os subconjuntos, ou conjuntos únicos de rótulos, estão presentes no conjunto de treinamento e representados por valores unitários do atributo classe, a partir dos quais um classificador multiclasse é construído (METZ, 2011). A Tabela 2 ilustra um conjunto multirrótulo, no qual é composto por quatro exemplos e $L = y1, y2, y3$ como conjunto de unirrótulos.

Tabela 2 – Conjunto de Exemplos Multirrótulo

Exemplo	Yi
1	Y1={Y1, Y2, Y3, Y4}
2	Y2={Y1, Y3}
3	Y3={Y2, Y3}
4	Y4={Y2, Y3, Y4}
5	Y5={Y3, Y4}

Fonte: (METZ, 2011).

No método *Label Powerset*, os exemplos multirrótulos são transformados em novos tipos de dados, agora unirrótulos. Essa transformação ocorre por meio da concatenação dos múltiplos rótulos para a criação de um novo unirrótulo. Essa transformação pode ser realizada através da criação de um índice único para cada agrupamento de dados multirrótulos no conjunto de treinamento. Por exemplo, o dado multirrótulo $y1, y2, y3$ poderia estar associado a um índice apresentado pela letra "A". A partir de então, cada vez que o unirrótulo "A" fosse utilizado, automaticamente seriam recuperados os múltiplos rótulos originais. Por fim, os exemplos multirrótulos transformados são apresentados a um algoritmo de aprendizado para a construção de um classificador unirrótulo multiclasse (METZ, 2011). A Figura 8 ilustra a construção do classificador multirrótulo.

Ao contrário do método de classificação *Binary Relevance* (BR), o *Label Powerset* tem a vantagem de considerar as correlações entre os rótulos, ver Figura 8. Entretanto, pode haver um crescimento desproporcional do número de rótulos de uma classe caso haja um grande número de subconjuntos de rótulos, resultando, dessa maneira, em muitas classes com poucos exemplos associados. Isso porventura ocasionará em um

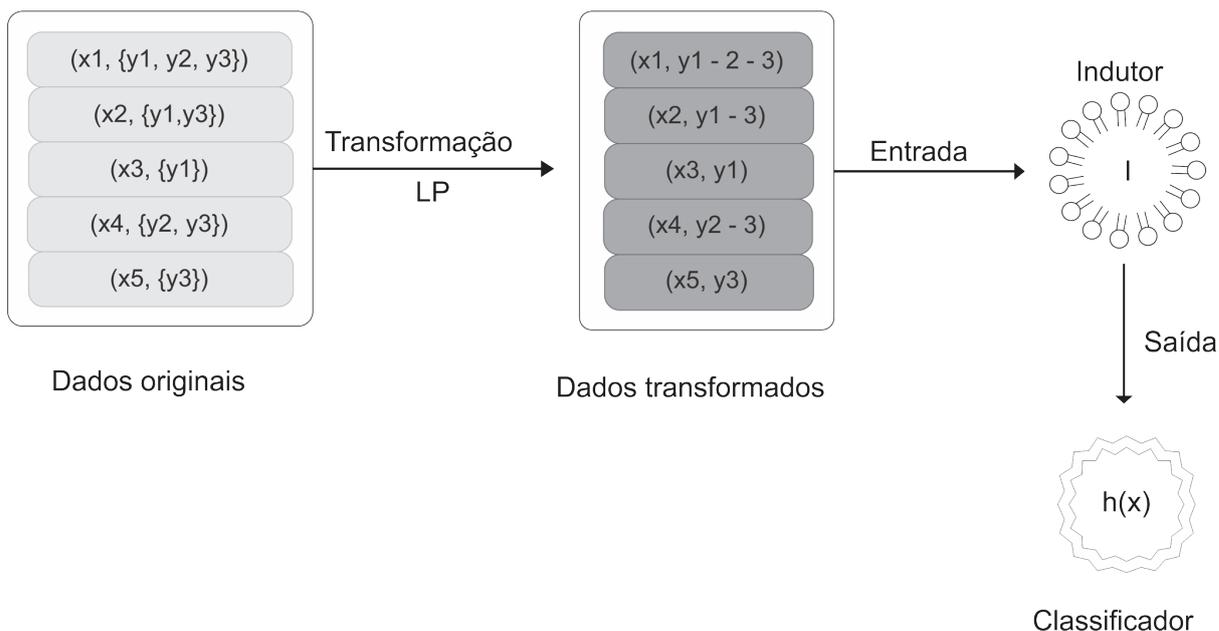


Figura 8 – Construção do Classificador Multirrótulo.

Fonte: (METZ, 2011).

alto custo computacional do método *Label Powerset* e a diminuição da acurácia dos classificadores. Por outro lado, o método de classificação LP apresenta-se por ser um método simples, menos comum, mas bastante utilizado (METZ, 2011).

Rakel e Rakeld

O método de classificação RAKEL (*R*ANdom *k*-*Lab*ELsets) é construído por meio de comitês de classificadores LP. Um dos seus objetivos é minimizar os problemas oriundos do LP, pois mesmo considerando a correlação entre os rótulos, busca-se coibir o problema da suscetibilidade ao surgimento de muitas classes com uma quantidade mínima de exemplos. Foi proposta uma abordagem em (TSOUMAKAS; KATAKIS; VLAHAVAS, 2011) onde um parâmetro definido k especifica o tamanho do subconjunto de rótulos que devem ser considerados (RODRIGUES, 2014).

No RAKEL, cada classificador do comitê de classificadores LP é treinado usando um pequeno subconjunto aleatório de *labelsets*. Portanto, é possível afirmar que nesse método, os classificadores unirrótulos, mesmo considerando o correlacionamento dos rótulos, são aplicados em subtarefas cujo número de rótulos e exemplos por rótulos sejam adequados (NASIERDING; TSOUMAKAS; KOUZANI, 2009); (TSOUMAKAS; KATAKIS; VLAHAVAS, 2008). A Figura 9 apresenta o resultado da criação do classificador através do RAKEL. Além disso, Tabela 3 expõe o resultado da transformação do conjunto de dados originais usando o método RAKEL.

Enquanto no *Random k-Labelsets* é permitida a sobreposição dos *labelsets*, em (TSOUMAKAS; KATAKIS; VLAHAVAS, 2011), uma extensão do RAKEL, chamada RAKELd (*R*ANdom *K*-*Lab*ELsets *dis*joint), considera que os *labelsets* de cada classificador

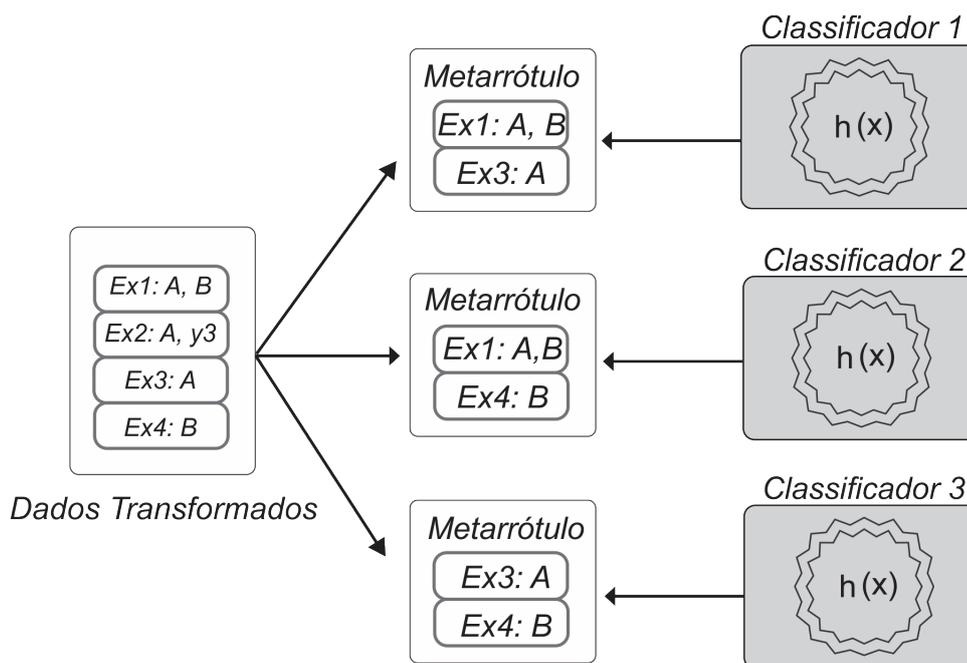


Figura 9 – Construção do Classificador Multirrótulo RAKEL.

Fonte: Adaptado-(SANTOS, 2012).

Tabela 3 – Conjunto de Dados Transformados Usando RAKEL

Exemplo	Y_i	Metarrótulo
1	$Y_1 = \{Y_1, Y_2, Y_3, Y_4\}$	=M1
2	$Y_2 = \{Y_1, Y_3\}$	=M2
3	$Y_3 = \{Y_2, Y_3\}$	=M3
4	$Y_4 = \{Y_2, Y_3, Y_4\}$	=M4
5	$Y_5 = \{Y_3, Y_4\}$	=M5

Fonte: (SANTOS, 2012).

do comitê sejam disjuntos.

3.3 ADAPTAÇÃO DE ALGORITMOS

Os métodos de adaptação de algoritmos são gerados como resultado de uma extensão de um determinado algoritmo de aprendizado unirrótulo, possibilitando assim a manipulação de dados multirrótulo diretamente. Na literatura, há uma variedade de trabalhos tratando da adaptação das mais variadas técnicas de aprendizado de máquina, como por exemplo: Algoritmos de Indução de Árvores de Decisão (CLARE; KING, 2001), Redes Neurais Artificiais (ZHANG; BURER; STREET, 2006), k vizinhos mais próximos (ZHANG; BURER; STREET, 2006), *AdaBoost* (SCHAPIRE; SINGER, 2000), dentre outras. A seguir são apresentados os principais métodos de adaptação de algoritmos encontrados na Literatura:

Árvore de Decisão

Em Clare e King (2001), um dos algoritmos mais utilizados para construção de árvores de decisão, chamado C4.5, foi adaptado de modo a viabilizar a manipulação de dados multirrótulo. No algoritmo C4.5, os nós da árvore de decisão são definidos através de uma medida de entropia. Em Clare e King (2001), a fórmula dessa medida, originalmente elaborada para problemas unirrótulo, foi modificada, de modo a permitir seu uso em problemas multirrótulo.

Redes Neurais Artificiais

Na literatura podem ser encontrados vários trabalhos propondo adaptações de modelos de redes neurais artificiais para o tratamento de dados multirrótulo. Em Zhang, Burer e Street (2006) foi proposta uma adaptação do popular algoritmo *backpropagation* para viabilizar a utilização no processo de aprendizagem multirrótulo, chamada BP-MLL (*BackPropagation for MultiLabel Learning*). A principal modificação do algoritmo é a introdução de uma nova função de erro capaz de capturar as características do aprendizado multirrótulo, ou seja, considerar múltiplos rótulos associados a uma mesma instância. Já em Crammer e Singer (2003), é proposta uma adaptação da arquitetura do modelo de rede neural perceptron para tratar dados multirrótulo. O MMP (*Multiclass Multilabel Perceptron*) (CRAMMER; SINGER, 2003) é uma família de algoritmos online para a geração de *ranking* dos rótulos de dados multirrótulo baseado no algoritmo *backpropagation*, onde as atualizações nos pesos de cada um destes perceptrons são realizadas de modo a conseguir um *ranking* perfeito de todos os rótulos.

Em Zhang (2009) são propostas adaptações no modelo de rede neural artificial RBF (*Radial Basis Function*) para aplicação em problemas multirrótulo. A adaptação proposta, chamado MIML-RBF (*Multi-Instance MultiLabel Radial Basis Function*), trata do aprendizado multi-instância multirrótulo, ou seja, dados nos quais cada exemplo é não só representado por mais de uma instância, mas também associado a múltiplos rótulos de classes. No MIML-RBF, conexões entre instâncias e rótulos são diretamente exploradas no processo de agrupamento (*clustering*) da primeira camada e otimização da segunda camada. Enquanto a primeira camada consiste de conjuntos de instâncias formadas pela execução do agrupamento para cada possível classe, na segunda camada os pesos são otimizados pela diminuição das somas dos erros quadráticos médios. Já em (OLIVEIRA et al., 2008), é proposta uma modificação do algoritmo PNN (*Probabilistic Neural Networks*) (SPECHT, 1990) para problemas multirrótulo. No algoritmo proposto, podem ser destacados dois passos principais: [A] tentativa de aproximação das fronteiras de decisões utilizando o teorema de *Bayes* e; [B] estimativa da função de densidade das instâncias.

Máquinas de Vetores Suporte

Em Elisseff e Weston (2001), é proposta uma variação do SVM para manipular dados multirrótulo. Neste algoritmo, busca-se minimizar o número de vezes que rótulos

irrelevantes estão posicionados acima dos rótulos relevantes no *ranking*. Godbole e Sarawagi (2004) propôs três métodos para instanciação do método BR com classificadores SVM. Os dois primeiros métodos podem facilmente ser abstraídos para serem usados com qualquer algoritmo de classificação, podendo assim serem considerados uma extensão do próprio BR, enquanto o terceiro é específico aos SVMs.

A principal ideia do primeiro método é estender o conjunto de dados com q características adicionais contendo as predições de cada classificador binário. Então, uma segunda rodada de treinamento de q classificadores binários é realizada, desta vez usando os novos conjuntos de dados estendidos. Para a classificação de um novo exemplo, os classificadores binários da primeira rodada são inicialmente usados e suas saídas são adicionadas às características do exemplo, para formar um meta-exemplo. Este meta-exemplo é então classificado pelos classificadores binários da segunda rodada. Através dessa extensão, são consideradas as potenciais dependências entre os diferentes rótulos. O segundo método, chamado *ConfMat*, consiste em remover as instâncias de treinamento negativas de um rótulo completo, se elas forem muito semelhante ao rótulo positivo. A decisão toma como base a matriz de confusão estimada usando qualquer classificador rápido e moderadamente acurado sobre um conjunto de validação. Por sua vez, o terceiro método, chamado *BandSVM*, consiste na remoção de instâncias de treinamento negativas muito similares que estão dentro de um limiar de distância do hiperplano aprendido.

Ademais, em Aioli e Sperduti (2004) é proposto um método em que o problema multirrótulo é dividido entre vários SVMs em paralelo, combinando todos os classificadores em um único processo de otimização, por meio do compartilhamento da matriz *kernel* entre os diferentes classificadores.

k Vizinhos Mais Próximos

Na literatura, há vários trabalhos propondo métodos baseados no popular algoritmo de aprendizado k vizinhos mais próximos (*k Nearest Neighbors*) (COVER; HART, 1967), tais como: (LUO; ZINCIR-HEYWOOD, 2005; WIECZORKOWSKA; SYNAK; RAŚ, 2006; BRINKER; HÜLLERMEIER, 2007; ZHANG; ZHOU, 2007; SPYROMITROS; TSOUMAKAS; VLAHAVAS, 2008). O primeiro passo em todas estas abordagens é o mesmo que no KNN, isto é, recuperar os k exemplos mais próximos. Assim, o que os diferencia é o modo de agregação dos conjuntos de rótulos destes exemplos. O ML-kNN (*MultiLabel k Nearest Neighbors*) (ZHANG; ZHOU, 2007), adaptação do algoritmo kNN para dados multirrótulo, usa 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-kNN usa o algoritmo kNN independentemente para cada rótulo. Desse modo, o ML-kNN busca os vizinhos mais próximos para a instância de teste, considerando aquelas instâncias que são rotuladas ao menos com como positiva e as demais instâncias

como negativas. Assim, o que diferencia este método do KNN original é justamente o uso de probabilidades a priori. Adicionalmente, o ML-kNN tem a capacidade de produzir um *ranking* dos rótulos como saída (COSTA et al., 2007).

Em Luo e Zincir-Heywood (2005), são apresentados dois sistemas de classificação de documentos multirrótulo, baseado no algoritmo kNN. A principal contribuição deste trabalho é no estágio de pré-processamento para a efetiva representação dos documentos. Para a classificação de uma nova instância, o sistema inicialmente busca os exemplos mais próximos. Então, a cada manifestação de um rótulo em cada um destes exemplos, um contador correspondente a este rótulo é incrementado. Finalmente, os rótulos com os maiores contadores são retornados (TSOUMAKAS; KATAKIS, 2007).

Algoritmos de Agrupamento

Em Nasierding, Tsoumakas e Kouzani (2009) é proposto um sistema para classificação multirrótulo baseado em agrupamento (*clustering*), chamado CBMLC (*Clustering-based MultiLabel Classification*). O treinamento do CBMLC consiste em dois passos principais: no primeiro passo, o CBMLC agrupa os dados de treinamento em um número k de grupos (*clusters*) especificado pelo usuário, usando um algoritmo de agrupamento. Os rótulos do conjunto de dados de treinamento não são considerados durante este passo. No segundo passo, CBMLC usa o algoritmo multirrótulo nos dados de cada grupo (*cluster*) e produz k modelos de classificação multirrótulo. Para a classificação de uma nova instância, CBMLC primeiro busca o grupo mais fechado (mais próximo) desta instância, e então usa o modelo multirrótulo correspondente para classificá-lo.

O processo de agrupamento divide o conjunto de dados de treinamento em partes menores de objetos similares. A expectativa é que objetos similares sejam anotados com rótulos similares. Com isso, hipoteticamente, cada parte será anotada com muito menos rótulos comparados àqueles anotados ao conjunto de dados de treinamento completo. Assim, espera-se melhorar o desempenho preditivo dos modelos multirrótulos correspondentes e reduzir o tempo de treinamento e classificação comparado a quando é utilizado o mesmo algoritmo de classificação multirrótulo no conjunto de dados original (NASIERDING; TSOUMAKAS; KOUZANI, 2009).

Métodos Probabilísticos

Em McCallum (1999), é proposto um modelo probabilístico generativo, aplicado à tarefa de classificação de documentos. Neste modelo, é utilizada uma abordagem de classificação *bayesiana*, na qual cada rótulo gera diferentes palavras. Baseado neste modelo, um documento multirrótulo é produzido por uma mistura das distribuições das palavras de seus rótulos. Uma similar mistura de palavras para classificação de texto multirrótulo é apresentado em (UEDA; SAITO, 2003). Em Streich e Buhmann (2008), uma abordagem de deconvolução é proposta para estimar a contribuição individual de cada rótulo a um dado item.

Em Ghamrawi e McCallum (2005), é explorado o uso de campos aleatórios condicionais, no qual dois modelos gráficos que parametrizam co-ocorrências de rótulos são apresentados. O primeiro, multirrótulo coletivo, captura padrões de co-ocorrências entre rótulos, enquanto o segundo, multirrótulo coletivo com características, tenta capturar o impacto que uma característica individual tem sobre a probabilidade de co-ocorrência de um par de rótulos (TSOUMAKAS; KATAKIS; VLAHAVAS, 2010)

Métodos Associativos

O MMAC (*Multi-Class Multilabel Associative Classification*) (THABTAH et al., 2004) é um algoritmo que segue o paradigma de classificação associativa, lidando com a construção de conjuntos de regras de classificação usando mineração de regras de associação. O MMAC aprende um conjunto inicial de regras de classificação através da mineração de regras de associação, removendo os exemplos associados com este conjunto de regras e gerando recursivamente um novo conjunto de regras dos exemplos remanescentes até que nenhum item frequente esteja à esquerda. Estes múltiplos conjuntos de regras forçam regras controladas com pré-condições similares, mas diferentes rótulos no lado direito. Tais regras são fundidas em uma regra multirrótulo única. Os rótulos são ordenados de acordo com o suporte das regras individuais correspondentes.

3.4 CONSIDERAÇÕES FINAIS

Neste capítulo, foram apresentadas as tarefas a classificação multirrótulo e os principais métodos de transformação do problemas juntamente com os métodos de adaptação de algoritmos. Ao contrário da classificação tradicional que trata do aprendizado de um conjunto de exemplos em que cada um é associado a uma única classe a classificação multirrótulo ocorre em casos, em que as classes não são disjuntas, e estão associadas a mais de uma classe.

No método de transformação do problema o problema multirrótulo existente é dividido em um ou vários problemas unirrótulos utilizando algoritmos de classificação unirrótulos para resolução de cada problema de maneira distinta. É importante considerar a independência de cada método sobre o algoritmo base. Assim, existe a possibilidade de um algoritmo de aprendizado unirrótulos construir os classificadores individualmente.

Os métodos de adaptação de algoritmos são gerados como resultado de uma extensão de um determinado algoritmo de aprendizado unirrótulos, possibilitando assim a manipulação de dados multirrótulo diretamente. Na literatura, há uma variedade de trabalhos tratando da adaptação das mais variadas técnicas de aprendizado de máquina, como por exemplo: algoritmos de indução de árvores de decisão (CLARE; KING, 2001), redes neurais artificiais (ZHANG; ZHOU, 2006), *K* vizinhos mais próximos (ZHANG; ZHOU, 2006), *AdaBoost* (SCHAPIRE; SINGER, 2000), dentre outros.

4 COMITÊ DE CLASSIFICADORES

Comitês de classificadores são sistemas de classificação compostos por um conjunto de sistemas de classificação e por um método de combinação dos resultados. O conjunto de sistemas de classificação é chamado de classificadores base. O objetivo desse sistema é melhorar o desempenho de classificação combinando as decisões dos classificadores para tomar uma decisão mais acurada. A melhora na acurácia de classificação dos comitês de classificadores é diretamente relacionada com a acurácia dos classificadores. Se todos os classificadores são iguais então o sistema pode ser substituído por apenas um classificador. Ou seja, para que os comitês apresentem um melhora na acurácia de classificação é necessário que os classificadores cometam erros em diferentes instâncias. Assim, o método de combinação dos resultados ou não é ou é pouco influenciado por esse erro, fornecendo a saída correta no caso ideal (KUNCHEVA, 2004).

Comitês são construídos de diferentes maneiras. De um modo geral, é fornecida uma base de dados para o treinamento. Essa base de dados é distribuída entre os classificadores e os mesmos são treinados. Ao classificar uma instância, ela é distribuída para cada classificador base, os quais avaliam e informam o seu resultado para o método de combinação que então combina os resultados em uma resposta, a saída do comitê (HANSEN; SALAMON, 1990).

Comitês são construídos tomando diferentes decisões sobre sua arquitetura. São construídos escolhendo uma técnica de combinação dos resultados, usando ou mesmo tipo ou diferentes tipos de classificadores (sendo classificado como homogêneo ou heterogêneo respectivamente). São construídos distribuindo diferentes bases de dados aos classificadores ou a mesma base com diferentes conjuntos de atributos, ou construídos por uma combinação de todos os métodos (KUNCHEVA, 2004).

Todas as técnicas de construção citadas são formas implícitas de geração de diversidade, isto é, não otimizam uma medida de diversidade durante o processo de construção dos comitês de classificação. Essas técnicas apresentam bons resultados na geração de diversidade em comitês. No entanto, tentar medir a diversidade e usá-la explicitamente no processo de construção do comitê não apresenta o mesmo resultado que o uso das metodologias implícitas (BROWN et al., 2005).

As técnicas de combinação dos resultados dos classificadores podem ser categorizadas em técnicas baseadas em seleção ou fusão dos valores. As técnicas de fusão dos valores assumem que cada classificador base possui conhecimento sobre todo o espaço de características enquanto que as técnicas de seleção dos valores assumem que cada classificador é especializado em determinadas características da base de dados. Assim, técnicas baseadas em seleção costumam escolher um classificador para tomar

a decisão baseado nas características apresentadas pela instância sendo classificada (KUNCHEVA, 2004).

Este capítulo está dividido nas seguintes seções: a primeira Seção trata dos multiclassificadores Homogêneos; a segunda Seção trata dos multiclassificadores heterogêneos; e por fim, a terceira Seção traz as considerações finais do capítulo.

4.1 MULTICLASSIFICADORES HOMOGÊNEOS

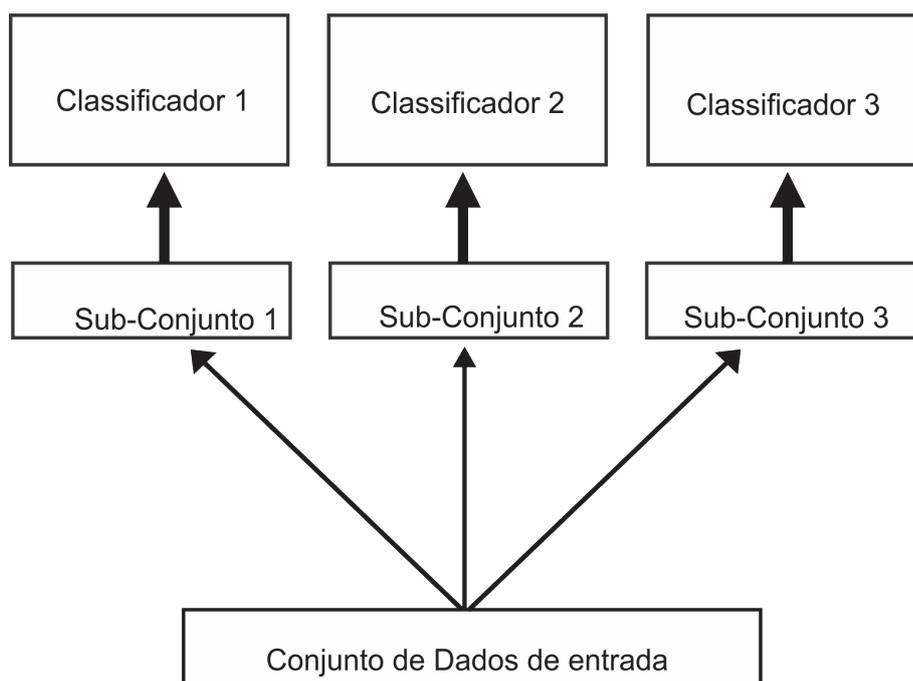
Como multi-classificadores homogêneos, destacam-se *Bagging* e *Boosting* (SCHAPIRE, 1990); (BREIMAN, 1996). Ambos os métodos trabalham com um único algoritmo de AM, utilizam diferentes amostragens do conjunto de dados original como dados de treinamento e combinam os resultados obtidos com os diferentes classificadores base através de algum esquema de votação.

Bagging

Bagging (*Bootstrap AGGregatING*) foi proposto por Breiman (1996) (BREIMAN, 1996), baseado na amostragem *bootstrap* (EFRON; TIBSHIRANI, 1993). Neste método, são gerados vários subconjuntos diferentes de treinamento a partir de uma amostragem aleatória do conjunto original de dados, com reposição. Então, são induzidos diferentes classificadores a partir de cada um destes subconjuntos de treinamento, ver Figura 10 baseada em (CANUTO, 2001). Os subconjuntos de treinamento têm o mesmo número de instâncias do conjunto original. Porém, algumas instâncias do conjunto original podem aparecer mais de uma vez, enquanto que outras instâncias podem não ser selecionadas. Esta distinção aleatória entre os vários subconjuntos de treinamento confere diversidade aos modelos de classificação que são obtidos a partir de cada um desses subconjuntos (KUNCHEVA, 2004). Para combinar os diferentes classificadores base um esquema de votação simples é normalmente empregado.

A amostragem com reposição representa uma técnica eficaz para a produção de conjuntos de treinamento representativos do conjunto de treinamento original e, ao mesmo tempo, distintos entre si (LIMA, 2005). No entanto, a existência de subconjuntos de dados distintos não implica que todos os classificadores induzidos irão generalizar de forma distinta. Para que isso ocorra, os classificadores devem ser instáveis, no sentido de produzirem comportamentos distintos sempre que submetidos a conjuntos de treinamento distintos, como os métodos AD e RNA.

O método *Bagging* busca, portanto, classificadores que sejam instáveis, capazes de generalizar de forma distinta, não necessariamente de forma ótima. Isso porque, mesmo que os componentes não apresentem uma boa capacidade de generalização, a combinação deles tendem a generalizar bem. Na verdade, a capacidade de generalização dos classificadores base não pode ser muito ruim, pois a melhora obtida com a combinação deles pode não ser suficiente para garantir um bom desempenho do sistema classificador. Por outro lado, também não precisa ser muito boa, pois neste

Figura 10 – Representação do Método *Bagging*.

Fonte: Adaptado-(SANTOS, 2009).

caso a combinação dos diferentes classificadores base pode não produzir ganho no desempenho de classificação (LIMA, 2005).

Boosting

O método *Boosting* abrange, na verdade, uma família de diferentes técnicas (por exemplo, *Arcing*, *Adaboosting*, *MultiBoosting* e *AveBoost*). Em geral, no *Boosting*, o conjunto de treinamento usado por cada classificador base, membro de uma sequência de classificadores, é escolhido baseado no desempenho dos primeiros classificadores. É necessário, assim, que os vários classificadores base do *Boosting* sejam treinados sequencialmente, visando definir as instâncias que irão constituir os próximos conjuntos de treinamento (LIMA, 2005).

Uma das formas de *Boosting* muito difundida e aplicada é o *AdaBoosting*. Neste caso, a probabilidade de escolha de uma instância para compor um conjunto de treinamento depende da contribuição dela para o erro de treinamento dos classificadores já treinados. Isto é, caso uma instância não tenha sido corretamente classificada pelos classificadores já gerados, a probabilidade de escolha desta instância aumenta em relação às demais, quando do treinamento de novos classificadores (LIMA, 2005).

Conseqüentemente, esta instância terá uma chance maior de ser escolhida para compor o conjunto de treinamento do próximo classificador a ser gerado. Assim, pode-se observar que, ao contrário do que ocorre no *Bagging*, apenas o primeiro classificador base da sequência do método é treinado a partir de uma amostragem aleatória do conjunto de dados original.

O *AdaBoosting* também pode seguir um outro modelo para a construção de multi-classificadores, no qual todas as instâncias, e seus pesos associados, fazem parte do conjunto de treinamento dos classificadores base. Esta abordagem tem a vantagem de que todas as instâncias, com pouca ou muita representatividade (de acordo com o peso correspondente), são incorporadas aos conjuntos de treinamento, ver Figura 11.

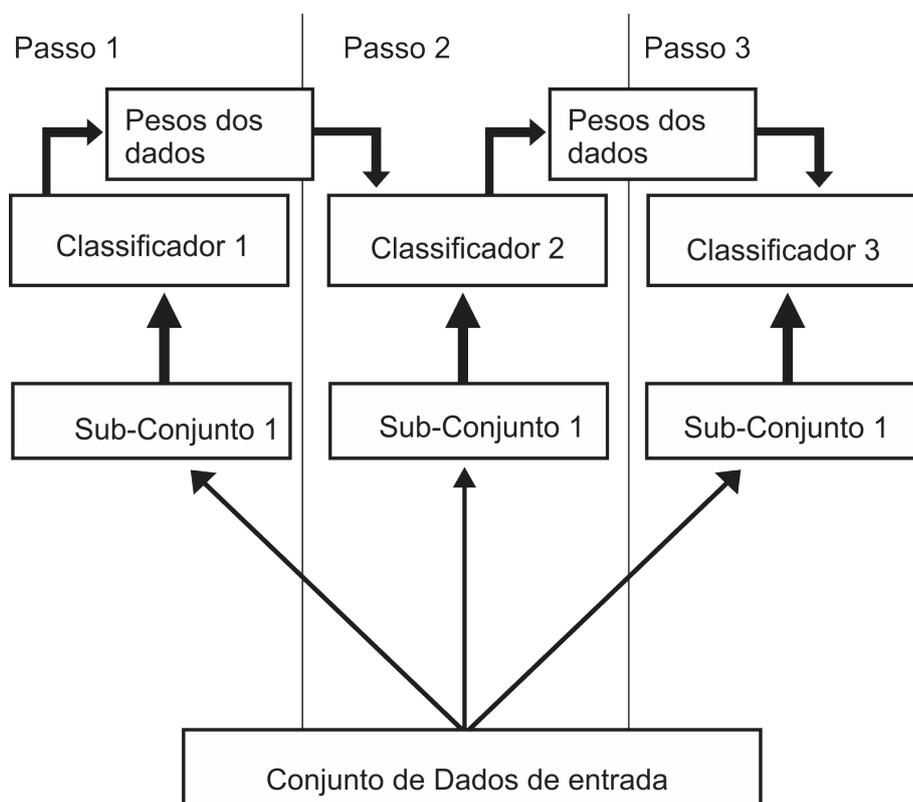


Figura 11 – Representação do Método *AdaBoosting*.

Fonte: (SANTOS, 2009).

Normalmente, *Bagging* é mais consistente do que *Boosting*, por raramente aumentar a taxa de erro dos classificadores base. Entretanto, *Boosting* conduz a reduções substancialmente maiores da taxa de erro do que o *Bagging*, apesar de, às vezes, poder ter um desempenho inferior ao de alguns dos seus classificadores base (OPITZ; MACLIN, 1999).

4.2 MULTICLASSIFICADORES HETEROGÊNEOS

Menos conhecido do que as técnicas de multiclassificação homogênea, os multiclassificadores heterogêneos utilizam diferentes algoritmos de aprendizado de máquina e utilizam como dados de treinamento o mesmo conjunto de dados original. A seguir, são apresentados os métodos de Multiclassificação Heterogênea: *Voting*, *Stacking* e *StackingC*.

Voting

O método *Voting* exige que cada classificador base apresente como saída um voto à classe que considera ser a mais provável para uma instância X . Então, é realizada uma contagem do número de votos por classe para todos os classificadores. A classe com o maior número de votos é escolhida como a predição final para a instância X . Este esquema de votação, o mais fácil de se executar, é denominado votação majoritária ou por maioria simples (JI; MA, 1997).

Algumas variações dessa ideia foram propostas, originando os vários esquemas de votação que existem, como por exemplo, votação ponderada, votação com confiança máxima e votação com predição de probabilidade (KUNCHEVA, 2004). Nesta proposta de trabalho, o termo *Voting* é usado referindo-se à aplicação do esquema de votação majoritária.

Vale ressaltar que os sistemas de multiclassificação homogênea usam o esquema de votação como este para combinar seus diferentes classificadores obtidos com o mesmo algoritmo de AM e diferentes amostragens do conjunto de dados original.

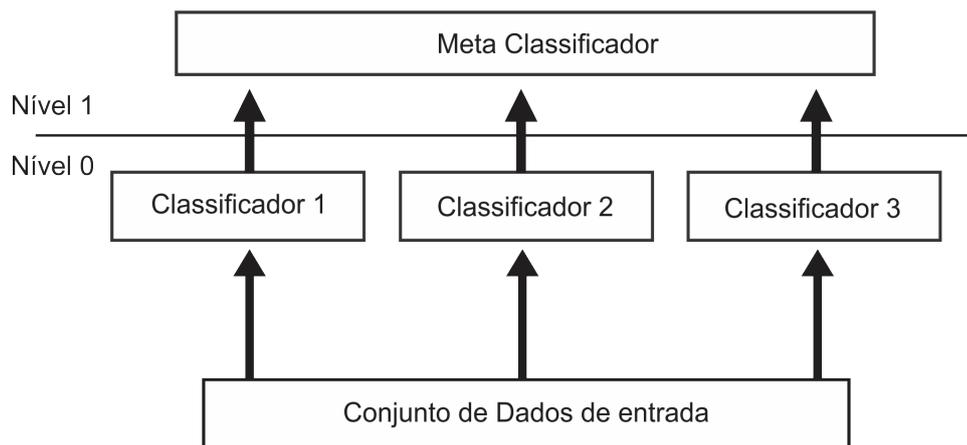
O *Voting* é uma das formas mais simples para a construção de um sistema multiclassificador (WITTEN; FRANK, 2005) e tem a vantagem de não requerer recursos computacionais ou espaço de memória extra. No entanto, o *Voting* tem a possibilidade de não conseguir aproveitar ao máximo o que cada classificador base tem de melhor. Ele pode piorar o desempenho de alguns classificadores base, caso existam entre eles classificadores com desempenho global ruim. Isto ocorre porque as saídas de todos os classificadores têm o mesmo peso, independente do seu desempenho global (LIMA, 2005).

Stacking

Ao contrário de um simples método de votação, o *Stacking* (*Stacked Generalization*) leva em consideração o fato de um certo classificador ser mais seguro que outro (KUNCHEVA, 2004). De fato, o *Stacking* tenta aprender quais dos classificadores base são os mais confiáveis e descobrir a melhor maneira de combinar as suas saídas (WITTEN; FRANK, 2005).

Stacking (WOLPERT, 1992) combina as saídas de classificadores (classificadores base, também chamados de classificadores de nível zero), induzidos por uma variedade de algoritmos de aprendizado através do mesmo conjunto de dados, por meio de um outro classificador (meta classificador, também conhecido como classificador de nível um), do qual se espera um melhor desempenho (MERZ, 1999).

Mais especificamente, o *Stacking* é construído em duas fases, ver Figura 12. Na primeira fase, no nível zero, um conjunto de classificadores base (C_1, C_2, \dots, C_k) é gerado por diferentes algoritmos de aprendizado e um único conjunto de dados T . Na segunda fase, nível um, o meta classificador combina as saídas dos classificadores base em uma predição final. Os dados do meta classificador tem, pois, tantos atributos

Figura 12 – Representação do Método *Stacking*.

Fonte: (SANTOS, 2009).

quantos forem os classificadores base.

StackingC

StackingC (SEEWALD, 2002) é uma variação do método *Stacking* baseada na remoção prévia de características irrelevantes e na redução da dimensionalidade do conjunto de dados usado como entrada para o meta classificador.

A ideia principal do *Stacking*, como visto anteriormente, é usar as saídas de um conjunto de classificadores base no treinamento do meta classificador, para fornecer a predição final. Ting and Witten (1999) (TING; WITTEN, 1999) propuseram um tipo de *Stacking* que usa como meta classificador a técnica *Multi-response Linear Regression* (MLR) (GUERMEUR; PAUGAM-MOISY; GALLINARI, 1998). Neste caso, durante a predição, os classificadores base fornecem suas distribuições de probabilidade das classes para serem usadas como entrada para as funções de regressão (uma para cada classe) do MLR. Com isso, Ting and Witten (1999) realizaram uma inovação, pois outros meta classificadores não aprendem geralmente um modelo distinto para cada classe, mas sim, um único modelo para todas as classes de uma vez.

O *StackingC* foi proposto por Seewald (2002) motivado pelo fraco desempenho que se obtém com o *Stacking MLR* ao trabalhar com bases de dados com mais de duas classes. O *StackingC* difere do *Stacking MLR* nos seguintes pontos: para cada modelo linear associado a uma classe específica, somente a distribuição parcial de probabilidade desta classe é usada durante o treinamento e teste. Enquanto o *Stacking* usa as distribuições de probabilidade de todas as classes e de todos os classificadores base para cada modelo linear, *StackingC* usa somente as probabilidades da classe associadas com a classe que se quer que o modelo linear faça a predição.

O *StackingC* não deve mudar significativamente o desempenho do *Stacking* para bases de dados com duas classes. Uma vez que a soma da distribuição de probabilidade de cada classe tem que ser um, a probabilidade de uma classe é um

menos a probabilidade da outra classe. Assim, apenas um destes valores é suficiente para codificar a informação completa em problemas com duas classes. No entanto, espera-se que o tempo de treinamento seja bem menor, por causa da redução da dimensionalidade dos dados do meta classificador (SEEWALD, 2003).

Com base em Seewald (2003), o *StackingC* deve se mostrar, teoricamente, melhor que o *Stacking* em termos de precisão e de tempo de treinamento. Estas melhorias são mais evidentes para bases de dados com mais de duas classes e têm tendência de se tornarem mais pronunciadas quando o número de classes aumenta. No entanto, o *StackingC*, assim como o *Stacking*, também apresenta a dificuldade de se definir quais e quantos devem ser os classificadores base usados e qual o meta classificador mais apropriado, uma vez que não existem regras que determinem tais possibilidades (SEEWALD, 2003).

4.3 CONSIDERAÇÕES FINAIS

Comitês de Classificadores são sistemas de classificação compostos por um conjunto de sistemas de classificação e por um método de combinação dos resultados. O conjunto de sistemas de classificação é chamado de Classificadores Base. O objetivo desse sistema é melhorar o desempenho de classificação combinando as decisões dos classificadores para tomar uma decisão mais acurada (KUNCHEVA, 2004).

Comitês são construídos de diferentes maneiras. De um modo geral, é fornecida uma base de dados para o treinamento. Essa base de dados é distribuída entre os classificadores e os mesmos são treinados. Ao classificar uma instância, ela é distribuída para cada classificador base, os quais avaliam e informam o seu resultado para o método de combinação que então combina os resultados em uma resposta, a saída do comitê (KUNCHEVA, 2004).

Os comitês de classificadores podem ser homogêneos, heterogêneos ou híbridos. Como multiclassificadores homogêneos, destacam-se *Bagging* e *Boosting* (SCHAPIRE, 1990); (BREIMAN, 1996). Ambos os métodos trabalham com um único algoritmo de AM, utilizam diferentes amostragens do conjunto de dados original como dados de treinamento e combinam os resultados obtidos com os diferentes classificadores base através de algum esquema de votação. Os classificadores heterogêneos são menos conhecidos do que as técnicas de multi-classificação homogênea, os multiclassificadores heterogêneos utilizam diferentes algoritmos de aprendizado de máquina e utilizam como dados de treinamento o mesmo conjunto de dados original, sendo o comitê *Stacking* o mais conhecidos na literatura.

5 TRABALHOS RELACIONADOS

Neste Capítulo são apresentados trabalhos relacionados ao aprendizado de máquina, principalmente trabalhos relacionados com a classificação multirrótulo ao aprendizado semissupervisionado e os comitês de classificadores com aprendizado semissupervisionado.

5.1 APRENDIZADO SEMISSUPERVISIONADO

Como exemplos de técnicas de aprendizado semissupervisionado bastante utilizados podemos citar o *Co-Training* (BLUM; MITCHELL, 1998) e o *Self-Training* (ROSENBERG; HEBERT; SCHNEIDERMAN, 2005). Além destes, podem ser citar:

Braga (2010) aplica o aprendizado semissupervisionado multidescrição em classificação de textos, o objetivo do trabalho é analisar o desempenho de algoritmos semissupervisionados multidescrição na classificação de textos, usando unigramas e bigramas para compor duas descrições distintas de documentos textuais. Assim, é considerado inicialmente o difundido algoritmo multidescrição *Co-training*, para o qual são propostas modificações a fim de se tratar o problema dos pontos de contenção. É também proposto o algoritmo COAL, o qual pode melhorar ainda mais o algoritmo *Co-training* pela incorporação de aprendizado ativo como uma maneira de tratar pontos de contenção. Uma ampla avaliação experimental desses algoritmos foi conduzida em bases de textos reais. Os resultados mostram que o algoritmo COAL, usando unigramas como uma descrição das bases textuais e bigramas como uma outra descrição, atinge um desempenho significativamente melhor que um algoritmo semissupervisionado monodescrição. Neste trabalho, foram obtidas coleções textuais provenientes de fontes variadas, como páginas da web, mensagens eletrônicas em grupos de discussão e artigos científicos, que podem ser encontradas em <http://www.labic.icmc.usp.br/extradata/igorab/msc/datasets/>. Para a realização dos testes estáticos foram utilizados o teste de *Nemenyi*. Levando em consideração os bons resultados obtidos por COAL, conclui-se que o uso de unigramas e bigramas como duas descrições distintas de bases de textos pode ser bastante compensador.

Rodrigues (2014) apresenta uma análise experimental dos resultados obtidos por meio da utilização do aprendizado semissupervisionado em problemas de classificação multirrótulo usando um parâmetro de confiabilidade como auxílio na classificação dos dados. Dessa maneira, a utilização de técnicas de aprendizado semissupervisionado, bem como de métodos de classificação multirrótulos, foram imprescindíveis na apresentação dos resultados. Neste trabalho foram utilizadas 10 dados multirrótulo: *emotions*, *bibtex*, *CAL500*, *corel5k*, *Enron*, *medical*, *genbase*, *mediamill*, *scene* e *yeast*. Como

métricas de avaliação foram utilizadas as métricas baseados em Bipartição: *Hamming Loss*, *F-Measure* e *Accuracy* e as métricas baseadas em *Ranking*: *One-Error*, *Average Precision* e *Ranking Loss*. O *T-test* foi utilizado para a realização das estáticas.

Assis (2016) propõe novos métodos para classificação multirrótulo semissupervisionado em uma abordagem para dados de visão múltiplas, mostra os resultados de alguns experimentos com esses novos métodos e compara alguns desses resultados com resultados já existentes na literatura. Para este trabalho foram propostos os métodos C_oXBR e BR_{CoT} . Esses métodos tratam de novas técnicas para classificação multirrótulo semissupervisionado DVM, aplicando a ideia do *Co-training* em sua forma original. Além desses métodos, também foi proposto o método C_pXBR , método DVM que utiliza a abordagem *co-perspective* em sua composição. Por último, também foi proposto o método S_TXBR . Nesse trabalho foram utilizadas três diferentes bases de dados multirrótulo retiradas do repositório de bases de dados MULAN: *emotions*, *flags* e *genbase*. Como teste estáticos foram utilizados os teste de *Friedman* e *Wilcoxon*.

5.2 CLASSIFICAÇÃO MULTIRRÓTULO

Na literatura encontramos diferentes trabalhos que tratam dos problemas de classificação multirrótulo, como por exemplo: (VALLIM; FREITAS, 2009; TSOUMAKAS et al., 2008; TSOUMAKAS; KATAKIS, 2007). Além desse trabalhos pode citar os trabalhos de:

Vallim (2009) apresenta uma proposta de LCS (do inglês, *Learning Classifier System*) para problemas de classificação multirrótulo, na qual se pretende induzir um conjunto de regras de classificação que possa gerar um resultado eficaz e que, também, seja capaz de ser comparável com outras técnicas de classificação. Os LCS, *Learning Classifier System*, são representado por técnicas de indução de regras de classificação, cujo principal mecanismo de busca é um algoritmo genético. Essa técnica almeja encontrar um conjunto de regras que disponha de uma alta precisão de classificação. Na literatura, os estudos sobre os LCS são temas muito bem fundamentados e vêm se mostrando capazes de atingir resultados satisfatórios com classes mutuamente exclusivas.

Silva (2014) considera que a utilização de uma cadeia mais adequada para cada instância possa aumentar a capacidade preditiva do método *Classifier Chains* (CC). Para isso, duas novas técnicas foram propostas, denominadas: *Lazy Classifier Chains* (LCC) e *One-To-One Classifier Chains* (OOCC). Os experimentos mostram que os dois métodos propostos LCC e OOCC são capazes de obter um desempenho preditivo superior aos seguintes métodos do estado da arte: *Binary Relevance*, CC e *Ensemble of Classifier Chains*. No trabalho foi utilizando as medidas Acurácia, *F-Measure*, Casamento Exato e *Hamming Loss*. Para determinar se há diferença de desempenho entre os algoritmos em cada medida, foram utilizados os testes de *Friedman* e pós-teste de *Nemenyi*.

Silva, Carvalho et al. (2015) investigam a hipótese de que utilizar o classificador adequado para cada conjunto binário melhora a classificação multirrótulo. Utilizando o método de transformação Binary Relevance, foi adotada uma estratégia de meta-aprendizado para recomendar o classificador adequado para cada subproblema. Os resultados experimentais validam a hipótese investigada e mostram o potencial da abordagem utilizada. Além disso, a estratégia proposta é genérica, de modo que, pode ser aplicada em outros problemas de transformação multirrótulo. No trabalho foram utilizados quatro classificadores distintos: *Support Vector Machine* (SVM); *Random Forest* (RF), *Naive Bayes* (NB); e o *k-Nearest-Neighbors* (kNN). Neste trabalho foram utilizados as bases *bibtex*, *birds*, *CAL500*, *corel5k*, *emotions*, *enron*, *flags*, *genbase*, *mediamill*, *medical*, *scene*, *tmc2007-500* e *yeast*. No trabalho não foram utilizados testes estatísticos.

5.3 COMITÊS DE CLASSIFICADORES COM APRENDIZADO SEMISSUPERVISADO

Na literatura encontramos somente o trabalho (SANTOS, 2012) que trata do uso de comitês de classificadores com aprendizado semissupervisionado:

Santos (2012) utiliza a aplicação de aprendizado semissupervisionado em tarefas de classificação hierárquica multirrótulo, de modo a se atender eficientemente as principais necessidades das duas áreas. Uma análise experimental dos métodos propostos verificou que a utilização do aprendizado semissupervisionado em métodos de classificação hierárquica multirrótulo apresentou resultados satisfatórios, uma vez que as duas abordagens apresentaram resultados estatisticamente semelhantes. Neste trabalho foram utilizadas as seguintes bases de dados multirrótulo: *emotions*, *genbase*, *medical*, *scene* e *yeast*. Como bases de dados Hierárquica Multirrótulo foram utilizadas: *Cellcycle*, *Church*, *Derisi*, *Eisen* e *Pheno*. Como métricas de avaliação foram utilizadas as métricas baseadas em Bipartição: *Hamming Loss*, *Subset Accuracy*, *Precision*, *Recall*, *F-Measure* e *Accuracy* e as métricas baseadas em Ranking: *One-Error*, *Average Precision*, *Is-Error*, *Error Set Size*, *Coverage* e *Ranking Loss*. Para a realização dos testes estáticos foram utilizados os *T-test* e o teste de *Friedman*. No trabalho surgiu a proposta da utilização de comitês (*ensembles*) no processo de atribuição de rótulos aos exemplos do conjunto de dados não-rotulados. Desse modo, com o objetivo minimizar a perda de informação das variações dos métodos SSBR e SSBRc, chamadas, respectivamente, de ESSBR (do inglês, *Ensemble of Semi-Supervised Binary Relevance*) e ESSBRc (do inglês, *Ensemble of Semi-Supervised Binary Relevance with confidence*).

6 METODOLOGIA DOS EXPERIMENTOS

Este capítulo apresenta os procedimentos adotados para a realização dos experimentos e está organizado da seguinte forma:

- A primeira Seção trata das bases de dados utilizadas: para a realização dos experimentos foram utilizadas seis bases de dados. A Tabela 4 apresenta algumas características dessas bases as quais podem ser encontradas no repositório do MULAN. Com o objetivo de melhor descrever a origem dos dados das bases de dados, a Tabela 4 faz uma breve descrição apresentando tais informações.
- A segunda Seção trata das métricas de avaliação: é possível encontrar várias medidas de classificação multirrótulo na literatura. Segundo Tsoumakas et al. (2009), estas medidas propostas podem ser desmembradas em dois grupos: As medidas baseadas em bipartição: que avaliam as bipartições sobre todos os exemplos da base de dados de avaliação, além de, em outros casos, decompor o processo de avaliação em avaliações separadas para cada rótulo. O outro grupo refere-se às medidas baseadas em *Ranking*: que avaliam os *rankings* com respeito à precisão da base de dados multirrótulo. No trabalho foram utilizados as métricas de avaliação baseadas em Bipartição: *Hamming Loss (HLoss)*, *Precision (Prec)*, *Recall (Rec)* e *FMeasure (FM)* e as medidas baseadas em *Ranking*: *One-Error (1-Err)* e *Average Precision (AvPrec)*.
- A terceira Seção trata dos testes estatísticos: os testes estatísticos fornecem um respaldo científico às pesquisas para que estas tenham validade e credibilidade. Testes estatísticos são utilizados para analisar a significância estatística das diferenças apresentadas pelos resultados experimentais dos métodos de classificação multirrótulo utilizados neste trabalho. Foram utilizados os teste estatístico de *Friedman* e *Wilcoxon*, a terceira Seção apresenta com detalhes o funcionamentos dos testes de *Friedman* e *Wilcoxon*.
- A quarta Seção trata das configurações dos experimentos: para cada um dos métodos de classificação multirrótulo supervisionada e semissupervisionada foram aplicados cinco diferentes algoritmos como classificadores base: k Vizinhos Mais Próximos (k-NN), Algoritmo de Indução de Árvores de Decisão (J48), Máquinas de Vetores Suporte (SVM), *Naive Bayes* (NB) e *Extended Repeated Incremental Pruning* (JRip). Os comitês de classificadores *Bagging*, *Boosting* e *Stacking* foram utilizados tanto na classificação multirrótulo supervisionado quanto na classificação semissupervisionada. Como tamanho dos comitês foram adotados os valores 3, 5, 7 e 9. Na classificação multirrótulo semissupervisionada foi utilizado o *Co-training*.

Como método de transformação do problema foram utilizados o BR, LP e Rakel. Todos os experimentos foram conduzidos usando a metodologia de validação cruzada com 10 grupos, $K=10$ (*10-fold Cross-Validation*) e o *framework* MULAN, o qual é implementado utilizando o WEKA. Para a realização dos testes estatísticos de *Friedman* e *Wilcoxon* foi utilizando o IBM SPSS (*Statistical Package for the Social Sciences*).

6.1 BASES DE DADOS MULTIRRÓTULO

O trabalho propõe o uso de diferentes domínios de aplicação. A seguir, são descritas as características e especificações das bases de dados que foram utilizadas durante os experimentos. A Tabela 4 apresenta algumas características das bases de dados multirrótulo, tais como o número de exemplos, o número de atributos numéricos, o número de atributos discretos e, por fim, o número de rótulos. Todas as bases foram obtidas através do repositório *Multi-Label* do MULAN (<http://mulan.sourceforge.net/datasets-mlc.html>).

Tabela 4 – Detalhamentos das Bases de Dados Multirrótulo

Base	Num. Exemplos	Num. Atrib. Numérico	Num. Atrib. Discreto	Rótulos
<i>Birds</i>	645	258	0	19
<i>Emotions</i>	593	78	0	6
<i>Enron</i>	1702	0	1054	53
<i>Yeast</i>	2417	117	0	14
<i>CAL500</i>	502	68	0	174
<i>Flags</i>	194	09	1186	7

Fonte: Autória Própria

6.1.1 Birds

A base de dados *Birds* apresenta novos métodos para a classificação acústica de várias espécies simultâneas de aves em um ambiente ruidoso. A base de dados contém 645 instâncias, sendo 258 atributos contínuos, com possibilidade de associação a 19 rótulos.

6.1.2 Emotions

A base de dados *Emotions* (WIECZORKOWSKA, 2006) está relacionada com a classificação de músicas de acordo com as emoções provocadas por elas. A base de dados contém 593 instâncias, sendo 78 atributos contínuos, com possibilidade de associação a 6 rótulos.

6.1.3 Enron

A base *enron* representa um conjunto de dados que foi coletado e preparado pelo projeto CALO – assistência cognitiva que aprende e organiza –. Essa base contém os dados de 150 usuários organizados em pastas. Esse *dataset* é formada por 1702 instâncias, 1054 atributos discretos e 53 rótulos.

6.1.4 Yeast

A base de dados biológicos *Yeast* (CLARE; KING, 2001) está relacionada à classificação de funções de proteínas. Em sua composição encontramos microvetores de 60 expressões e perfis filogenéticos de 2417 genes *yeast*, descritos por 117 atributos contínuos. Cada é associado a um subconjunto de 14 categorias com 2417 atributos e 14 rótulos.

6.1.5 CAL500

Essa base de dados é a composição em um conjunto de 1.700 anotações musicais do ser humano que, gerados, descrevem 500 faixas musicais populares do ocidente. A *CAL500* é definida por suas 502 instâncias, 68 atributos numéricos, bem como seus 174 rótulos. Ela está disponível em <http://cosmal.ucsd.edu/cal/projects/AnnRet/>.

6.1.6 Flags

A base de dados multirrótulo *Flags*, formada principalmente por dados obtidos do guia "*Guide to flags*" (SHAW, 1986), possui 194 instâncias e 30 atributos e 7 rótulos. Os atributos trazem informações dos países dos quais descrevem ou associam às cores utilizadas nas bandeiras daqueles países, para tal, sete são o número de classes.

6.2 MÉTRICAS DE AVALIAÇÃO

É possível encontrar várias medidas de classificação multirrótulo na literatura. Segundo (TSOUMAKAS et al., 2009), estas medidas propostas podem ser desmembradas em dois grupos: As medidas baseadas em bipartição: que avaliam as bipartições sobre todos os exemplos da base de dados de avaliação, além de, em outros casos, decompor o processo de avaliação em avaliações separadas para cada rótulo. O outro grupo refere-se às medidas baseadas em *ranking*: que avaliam os *rankings* com respeito à precisão da base de dados multirrótulo. A seguir, é apresentadas as medidas de avaliação baseadas em Bipartição e as medidas baseadas em *Ranking* às mesmas foram utilizadas durante todos os experimentos.

6.2.1 Métricas baseadas em Bipartição

Durante todos os experimentos foram utilizadas as seguintes medidas de Bipartição: *Hamming Loss (HLoss)*, *Precision (Prec)*, *Recall (Rec)* e *FMeasure (FM)*.

6.2.2 Hamming Loss

O *Hamming Loss* avalia a porcentagem correta de rótulos não preditos e rótulos preditos incorretamente. Dessa maneira, esta medida considera a frequência com que um par exemplo/rótulo é classificado de maneira incorreta, ou seja, cada exemplo é associado a um rótulo errado ou um rótulo não predito. Esta medida define que o melhor desempenho é alcançado quando o resultado é igual a zero, isto é, quanto mais o valor se aproxima a zero, melhor será seu desempenho. A seguir é apresentado a formulação matemática.

$$H_{loss} = \frac{1}{N} \sum_{i=1}^N \left| \frac{Y_i \Delta Z_i}{M} \right| \quad (6.1)$$

Formulação matemática da métrica *Hamming Loss*.

Onde: N é o número de instâncias de teste, M o número de rótulos, Y_i e Z_i representam, respectivamente, o conjunto de rótulos reais e preditos da instância de teste i e Δ é a diferença simétrica entre os dois conjuntos.

6.2.3 Recall

Esta métrica representa a porcentagem de exemplos classificados como positivos por um modelo de classificação que são verdadeiros positivos. A seguir é apresentado a formulação matemática.

$$Rec = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i \cap z_i|}{|Y_i|} \quad (6.2)$$

Formulação matemática da métrica *Recall*.

Onde: N é o número de instâncias de teste, Y_i e Z_i representam, respectivamente, o conjunto de rótulos reais e preditos da instância de teste i . Assim, quanto maior o valor alcançado, melhor é o desempenho.

6.2.4 F-Measure

A *F-Measure* é uma métrica que representa a combinação entre Precisão e Reconhecimento. Sua representação é definida pela média hamônica das duas métricas,

Precision e *Recall*, e é usada como um *score* de desempenho agregado (SANTOS, 2012). A seguir é apresentado a formulação matemática.

$$FM = \frac{1}{N} \sum_{i=1}^N \frac{2 |Y_i \cap Z_i|}{|Z_i| + |Y_i|} \quad (6.3)$$

Formulação matemática da métrica *F-Measure*.

Onde: N é o número de instâncias de teste, Y_i e Z_i representam, respectivamente, o conjunto de rótulos reais e preditos da instância de teste i . Assim, quanto maior o valor alcançado, melhor é o desempenho.

6.2.5 Precision

Esta métrica pode ser definida como a porcentagem de exemplos verdadeiros positivos de todos os exemplos classificados como positivos pelo modelo de classificação. A seguir é apresentado a formulação matemática.

$$Prec = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i \cap Z_i|}{|Z_i|} \quad (6.4)$$

Formulação matemática da métrica *Precision*.

Onde: N é o número de instâncias de teste, Y_i e Z_i representam, respectivamente, o conjunto de rótulos reais e preditos da instância de teste i . Assim, quanto maior o valor alcançado, melhor é o desempenho.

6.3 MÉTRICAS BASEADAS EM RANKING

Esta proposta de trabalho pretende utilizar as seguintes medidas baseadas em *Ranking*: *One-Error* (*1-Err*) e *Average Precision* (*AvPrec*).

6.3.1 One-Error

A *One-Error*, também conhecida por *1-Err*, avalia a frequência de rótulos no topo do *ranking* que não estavam no conjunto de rótulos (relevantes) corretos. Nessa métrica, o melhor desempenho é alcançado quando essa medida vai se aproximando a zero. A seguir é apresentado a formulação matemática.

$$1 - Err = \frac{1}{N} \sum_{i=1}^{|N|} \delta(\text{arg}_{y \in L} \min r_i(\lambda)) \quad (6.5)$$

Formulação matemática da métrica *One-Error*.

onde:

$$\text{delta}(\lambda) = \begin{cases} 1 & \text{se } \lambda \in Y_i \\ 0 & \text{caso contrario} \end{cases} \quad (6.6)$$

Por tanto, N são os números de rótulos, Y_i representa o conjunto de rótulos reais da instância de teste i .

6.3.2 Average Precision

A métrica *Average Precision*, conhecida por *AvPrec*, avalia a fração média de rótulos *rankeados* acima de um rótulo particular $l \in Y_i$ que está em Y_i . O melhor resultado é alcançado quando a precisão média é igual a um. Ou seja, essa medida de avaliação representa a precisão média tomada para todos os possíveis rótulos. A seguir é apresentado a formulação matemática.

$$\text{AvPrec}(H, |D|) = \frac{1}{|D|} * \sum_{i=1}^{|D|} \frac{1}{Z_i} * \sum_{y \in Z_i} \frac{|y' \in Z_i : r(y') \leq r(y)|}{r(y)} \quad (6.7)$$

Formulação matemática da métrica *Average Precision*.

Onde: D é o número de instâncias de teste, Z_i representa o número de interações, respectivamente, o conjunto de rótulos reais e preditos da instância de teste $r(y)$. Assim, quanto maior o valor alcançado, melhor é o desempenho.

6.4 TESTES ESTATÍSTICOS

Testes estatísticos são utilizados para analisar a significância estatística das diferenças apresentadas pelos resultados experimentais dos métodos de classificação utilizados neste trabalho. Os testes estatísticos fornecem um respaldo científico às pesquisas para que estas tenham validade e credibilidade. Em estatística, diz-se que um resultado é significativo se for improvável que tenha ocorrido por acaso. Dessa forma, busca-se com a utilização de um teste estatístico, determinar se a diferença encontrada entre os sistemas indica superioridade de um sistema em relação ao outro. Para isso formulam-se duas hipóteses, a primeira, hipótese nula, é tida como verdadeira até que as provas estatísticas indiquem o contrário. A segunda, hipótese alternativa, deve ser oposta à primeira. Dessa forma, as hipóteses sendo contraditórias não poderão ser simultaneamente verdadeiras. Para o problema proposto neste trabalho, as hipóteses levantadas são:

- Hipótese nula (H_0): Não existe diferença estatística entre o desempenho dos sistemas, ou seja, não pode-se dizer que um sistema seja estatisticamente superior ao outro;

- Hipótese alternativa (H_1): Os sistemas possuem desempenhos estatisticamente diferentes, podendo-se dizer que um sistema é estatisticamente superior ao outro.

Deve-se levar em consideração que para a execução do teste estatístico é preciso definir previamente o nível de significância ou nível de confiança, que representa a probabilidade de que o resultado estatístico encontrado não seja verdadeiro. Em geral, são aceitos como significativos os níveis $P = 0,05$ e $P = 0,01$, ou seja, no caso de rejeitar-se a hipótese nula, o faz com 95% ou 99% de confiança sobre sua correção. Neste trabalho é adotado o nível de confiança de 95% ($\alpha = 0,05$). Existe uma série de testes estatísticos que podem auxiliar as pesquisas. Neste trabalho, foram utilizados dois testes estatísticos bastante utilizados pela comunidade científica: o teste de *Wilcoxon* e Teste de *Friedman*. O teste de *Wilcoxon* é um método não-paramétrico para comparação de duas amostras pareadas. A princípio são calculados os valores numéricos da diferença entre cada par, sendo possível três condições: aumento (+), diminuição (-) ou igualdade (=). Uma vez calculadas todas as diferenças entre os valores obtidos para cada par de dados, essas diferenças são ordenadas pelo seu valor absoluto (sem considerar o sinal), substituindo-se então os valores originais pelo posto que ocupam na escala ordenada. O teste da hipótese de igualdade entre os grupos é baseado na soma dos postos das diferenças negativas e positivas. Este teste para dados pareados, ao invés de considerar apenas o sinal das diferenças entre os pares, considera o valor dessas diferenças, sendo assim um teste não-paramétrico dos mais poderosos e populares.

Já o Teste de *Friedman* é uma alternativa não-paramétrica para o teste de experimentos em blocos ao acaso (RBD - *Randon Blocks Design*) na ANOVA regular. Ele substitui o RBD quando os pressupostos de normalidade não estão assegurados, ou quando as variações são possivelmente diferentes de população para população. Este teste utiliza os *osranks* dos dados ao invés de seus valores brutos para o cálculo da estatística de teste. O teste de *Friedman* é amplamente utilizado para comparar três ou mais amostras. O algoritmo que obtiver o melhor desempenho recebe a posição 1 do *ranking*, o segundo melhor recebe a posição 2 e assim em diante.

6.5 CONFIGURAÇÃO DOS EXPERIMENTOS

Para cada um dos métodos de classificação multirrótulo supervisionada e semisupervisionada foram aplicados cinco diferentes algoritmos como classificadores base: k Vizinhos Mais Próximos (k-NN), Algoritmo de Indução de Árvores de Decisão (J48), Máquinas de Vetores Suporte (SVM), *Naive Bayes* (NB) e *Extended Repeated Incremental Pruning* (JRip). Estes classificadores foram escolhidos por serem bastante utilizados na literatura. Todos os resultados apresentados neste trabalho referem-se às médias dos melhores resultados obtidos pelos 5 classificadores base. Os comitês de classificadores *Bagging*, *Boosting* e *Stacking* foram utilizados tanto na classificação multir-

rótulo supervisionado quanto na classificação semissupervisionada. Como tamanho dos comitês foram adotados os valores 3, 5, 7 e 9. De acordo com Kuncheva (2004) mais de 25 membros em um comitê de classificadores não geram ganhos significativos. A ideia é categorizar os tamanhos dos componentes dos comitês, de maneira que seja possível o estabelecimento de parâmetros de referências mais adequados para as construções dos comitês. Na classificação multirrótulo semissupervisionada foi utilizado o *Co-training*. Como método de transformação do problema foram utilizados o BR, LP e Rakel por serem bem difundido e os mais utilizados na literatura.

Todos os experimentos foram conduzidos usando a metodologia de validação cruzada com 10 grupos, $K=10$ (*10-fold Cross-Validation*) e o *framework MULAN*, o qual é implementado utilizando o WEKA (MULAN, 2017). O WEKA tem como objetivo agregar algoritmos provenientes de diferentes abordagens/paradigmas na sub-área da inteligência artificial dedicada ao estudo de aprendizagem de máquina (WEKA, 2017). Vale ressaltar, que em relação as configurações internas dos classificadores durante os experimentos, foram utilizados o padrão *default* definidas pelo WEKA. Assim, todos os resultados apresentados a referem-se à média sobre 6 diferentes conjuntos de dados. Os classificadores foram avaliados utilizando as medidas de avaliação definidas na Seção Métricas de Avaliação.

Para a análise dos resultados obtidos pelos diferentes métodos de aprendizado foram utilizados dois testes bastante conhecidos: o Teste de *Wilcoxon* e o Teste de *Friedman*. O primeiro foi utilizado para as comparações pareadas, ou seja, entre duas amostras, enquanto o segundo foi utilizado nas comparações entre três ou mais amostras. Foi utilizado o nível de confiança de 95% ($\alpha = 0,05$) durante todos os testes estatísticos. Para a realização dos testes estatísticos de *Friedman* e *Wilcoxon* foi utilizado o IBM SPSS (*Statistical Package for the Social Sciences*) que é uma família integrada de produtos usado para a análise estatística que inclui ferramentas para a criação de gráficos e relatórios. O IBM SPSS é uma ferramenta modular e integrada que cobre todo o processo analítico ajudando a validar as hipóteses e a escolher a técnica mais adequada para responder ao problema colocado.

6.6 CONSIDERAÇÕES FINAIS

Neste capítulo são descritas as características e especificações das seis bases de dados que foram utilizadas durante os experimentos, tais como o número de exemplos, o número de atributos numéricos, o número de atributos discretos e, por fim, o número de rótulos. Todas as bases foram obtidas através do repositório *Multi-Label* do MULAN (<http://mulan.sourceforge.net/datasets-mlc.html>).

O capítulo também aborda as medidas de classificação multirrótulo baseadas em bipartição e às medidas baseadas em *ranking* que avaliam os *rankings* com respeito à precisão da base de dados multirrótulo.

Em seguida, foram apresentadas os testes estatísticos e a hipótese do trabalho. Em estatística, diz-se que um resultado é significativo se for improvável que tenha ocorrido por acaso. Dessa forma, busca-se com a utilização de um teste estatístico, determinar se a diferença encontrada entre os sistemas indica superioridade de um sistema em relação ao outro. Para isso formulam-se duas hipóteses, a primeira, hipótese nula, é tida como verdadeira até que as provas estatísticas indiquem o contrário. A segunda, hipótese alternativa, deve ser oposta à primeira. Dessa forma, as hipóteses sendo contraditórias não poderão ser simultaneamente verdadeiras. No trabalho são levantadas as hipóteses:

- Hipótese nula (H_0): Não existe diferença estatística entre o desempenho dos sistemas, ou seja, não pode-se dizer que um sistema seja estatisticamente superior ao outro;
- Hipótese alternativa (H_1): Os sistemas possuem desempenhos estatisticamente diferentes, podendo-se dizer que um sistema é estatisticamente superior ao outro.

Para a análise dos resultados obtidos pelos diferentes métodos de aprendizado foram utilizados dois testes bastante conhecidos: o Teste de *Wilcoxon* e o Teste de *Friedman* com o nível de confiança de 95% ($\alpha = 0,05$). O primeiro foi utilizado para as comparações pareadas, ou seja, entre duas amostras, enquanto o segundo foi utilizado nas comparações entre três ou mais amostras.

Para cada um dos métodos de classificação multirrótulo supervisionado e semissupervisionada foram aplicados cinco diferentes algoritmos como classificadores base: k Vizinhos Mais Próximos (k-NN), Algoritmo de Indução de Árvores de Decisão (J48), Máquinas de Vetores Suporte (SVM), *Naive Bayes* (NB) e *Extended Repeated Incremental Pruning* (JRip). Todos os resultados apresentados referem-se às médias dos melhores resultados obtidos pelos 5 classificadores base. Os comitês de classificadores *Bagging*, *Boosting* e *Stacking* foram utilizados tanto na classificação multirrótulo supervisionado quanto na classificação semissupervisionada. Para o tamanho dos comitês foram adotados os valores 3, 5, 7 e 9. Na classificação Multirrótulo Semissupervisionada foi utilizado o *Co-training*. Para os métodos de transformação do problema foram utilizados o BR, LP e Raket por serem bem difundido e os mais utilizados na literatura.

Todos os experimentos foram conduzidos usando a metodologia de validação cruzada com 10 grupos, $K=10$ (*10-fold Cross-Validation*) e o *framework* MULAN. Em relação as configurações internas dos classificadores durante os experimentos, foram utilizados o padrão *default* definidas pelo WEKA. Para a realização dos testes de *Friedman* e *Wilcoxon* foi utilizado o IBM SPSS (*Statistical Package for the Social Sciences*).

7 RESULTADOS EXPERIMENTAIS: MÉTODOS DE CLASSIFICAÇÃO MULTIRRÓTULO SUPERVISIONADO

Neste sétimo capítulo, é realizada uma investigação dos comportamentos dos métodos de classificação multirrótulo supervisionados. Os experimentos utilizaram diferentes valores para dois parâmetros: número de classificadores e tamanho de comitê. Desse modo, em virtude do grande número de métodos e parâmetros variáveis utilizados, a análise experimental foi realizada de uma perspectiva mais geral, de acordo com a abordagem de aprendizado utilizada no processo de construção dos comitês de classificadores.

Neste contexto, todos os resultados apresentam a média e o desvio padrão das melhores métricas, independente do tamanho do comitê e dos classificadores supervisionado, tendo sido sombreados os melhores resultados em cada uma das medidas. Além disso, para análise da relevância estatística foram utilizados os testes de *Friedman* e *Wilcoxon*. Os resultados nos quais os testes estatísticos mostraram relevância do ponto de vista estatístico tiveram a coluna **FRI** marcada em negrito.

No Teste de *Friedman*, se a hipótese nula for rejeitada, ou seja, se for detectada relevância estatística nas diferenças apresentadas pelas amostras, são realizadas comparações para detectar entre quais amostras essas diferenças ocorrem utilizando o teste de *Wilcoxon*. A coluna **WIL** nas tabelas exibe os resultados das comparações nas quais foi verificada significância estatística nas diferenças apresentadas. Vale ressaltar que, para algumas métricas de avaliação, os melhores resultados foram obtidos com valores próximos a zero, enquanto que, para outras métricas, os melhores resultados foram obtidos com valores próximos a um. Portanto, ao lado de cada métrica, foi adicionado o símbolo ↓ (o menor valor significa o melhor) e ↑ (o maior valor significa o melhor) para representar o comportamento das métricas de avaliação.

As Seções que apresentam os resultados experimentais relacionados às estratégias métodos BR, LP e Rakel com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*, a coluna **WIL** nas tabelas exibe os resultados das comparações nas quais foi verificada significância estatística nas diferenças apresentadas, utilizando a seguinte nomenclatura:

- A: Método BR vs LP
- B: Método BR vs Rakel
- C: Método LP vs Rakel

E por fim, as Seções que apresentam os melhores resultados experimentais dos métodos BR, LP e Rakel com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*,

a coluna **WIL** nas tabelas exibe os resultados das comparações nas quais foi verificada significância estatística nas diferenças apresentadas, utilizando a seguinte nomenclatura:

- A: *Bagging* vs *Boosting*
- B: *Bagging* vs *Stacking*
- C: *Boosting* vs *Stacking*

7.1 MELHORES RESULTADOS DO MÉTODO BR COM OS COMITÊS DE CLASSIFICADORES BAGGING, BOOSTING E STACKING

A Tabela 5 apresenta a média e o desvio padrão dos melhores resultados do métodos BR com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*. A Figura 13 ilustra a média dos melhores resultados do método BR com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*.

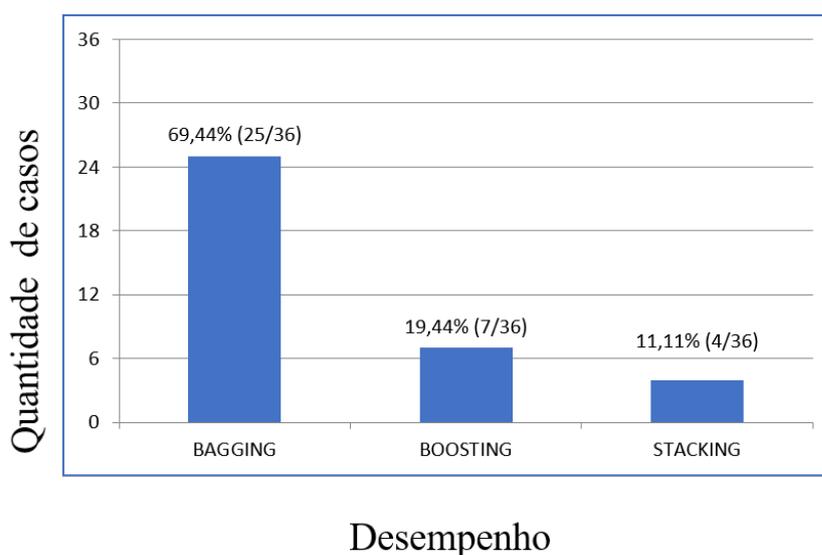


Figura 13 – Média dos melhores resultados do método BR com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*.

Tabela 5 – Média e desvio padrão do métodos BR com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*.

Comitê	BAGGING		BOOSTING		STACKING			
Base	BIRDS							
Medida	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	FRI	WIL
HLoss↓	0.0428	0.0111	0.0452	0.0108	0.0522	0.0106	0.015	BC
PPRec↑	0.7426	0.1363	0.6149	0.1315	0.4797	0.0961	0.000	ABC
PRec↑	0.4994	0.0875	0.5375	0.1281	0.3946	0.0731	0.007	BC
F-M↑	0.4641	0.0861	0.4909	0.1015	0.4028	0.0748	0.114	
1-Err↓	0.6836	0.1082	0.7278	0.0527	0.7830	0.0597	0.017	B
AvgPPRec↑	0.6137	0.0772	0.5537	0.1003	0.4757	0.0522	0.009	BC
Base	EMOTIONS							
HLoss↓	0.2058	0.0009	0.2097	0.0132	0.2139	0.0000	0.000	B
PPRec↑	0.7120	0.0024	0.7034	0.0272	0.6914	0.0000	0.000	BC
PRec↑	0.7544	0.0000	0.7254	0.0083	0.7632	0.0000	0.000	ABC
F-M↑	0.6815	0.0034	0.6805	0.0104	0.6932	0.0000	0.000	BC
1-Err↓	0.2083	0.0264	0.2600	0.0000	0.3167	0.0000	0.000	ABC
AvgPPRec↑	0.8091	0.0018	0.8017	0.0004	0.7768	0.0000	0.000	BC
Base	ENRON							
HLoss↓	0.0491	0.0016	0.0552	0.0033	0.0504	0.0019	0.000	ABC
PPRec↑	0.6607	0.0475	0.5827	0.0595	0.6793	0.0435	0.001	AC
PRec↑	0.6012	0.0408	0.5740	0.0470	0.5089	0.0284	0.000	BC
F-M↑	0.5438	0.0442	0.5107	0.0361	0.5227	0.0284	0.007	AB
1-Err↓	0.2642	0.0648	0.3213	0.0621	0.3075	0.0799	0.067	
AvgPPRec↑	0.6617	0.0414	0.5938	0.0284	0.6116	0.0345	0.001	AB
Base	YEAST							
HLoss↓	0.1994	0.0065	0.1331	0.0344	0.2586	0.0096	0.000	ABC
PPRec↑	0.7139	0.0176	0.6986	0.0224	0.5725	0.0210	0.000	BC
PRec↑	0.7265	0.0162	0.6062	0.0235	0.5751	0.0169	0.009	BC
F-M↑	0.6962	0.0143	0.6303	0.0128	0.5737	0.0167	0.000	ABC
1-Err↓	0.2420	0.0202	0.1648	0.1010	0.2726	0.0547	0.046	AC
AvgPPRec↑	0.7534	0.0140	0.7626	0.0819	0.6844	0.0263	0.026	AC
Base	FLAGS							
HLoss↓	0.2415	0.0328	0.6619	0.0274	0.3006	0.0329	0.000	ABC
PPRec↑	0.7388	0.0600	0.7668	0.0247	0.7048	0.0522	0.092	
PRec↑	0.7793	0.0437	0.7525	0.0494	0.6847	0.0661	0.004	BC
F-M↑	0.7573	0.0361	0.3929	0.0240	0.6838	0.0507	0.000	ABC
1-Err↓	0.1955	0.0750	0.7225	0.1089	0.2937	0.0781	0.000	ABC
AvgPPRec↑	0.8163	0.0268	0.8148	0.0320	0.7819	0.0247	0.024	BC
Base	CAL500							
HLoss↓	0.1375	0.0052	0.1507	0.0046	0.1345	0.0372	0.044	A
PPRec↑	0.6059	0.0258	0.4944	0.0190	0.3532	0.0155	0.000	ABC
PRec↑	0.5098	0.0389	0.4503	0.0271	0.3454	0.0177	0.000	ABC
F-M↑	0.3605	0.0174	0.3620	0.0185	0.3469	0.0139	0.193	
1-Err↓	0.1656	0.0713	0.2530	0.0618	0.8666	0.0363	0.000	ABC
AvgPPRec↑	0.4692	0.0169	0.4558	0.0179	0.2868	0.0504	0.000	ABC

Analisando a Tabela 5, é possível observar que os experimentos utilizando o método BR com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*, o comitê *Bagging* apresentou melhor desempenho em 69,44% (25/36) dos casos. O Teste de *Friedman* foi executado para verificar se há significância estatística nas diferenças apresentadas, a qual foi verificada em 88,88% dos casos (32/36). Nas comparações realizadas para detectar entre quais amostras as diferenças estatística são relevantes foi utilizado o teste de *Wilcoxon*, o qual obteve 57,81% (37/64) dos casos.

Neste contexto, pode-se afirmar que, em geral, nos experimentos realizados, o comitê *Bagging* obteve os melhores resultados. Nas análises realizadas, em todos os casos foi verificada que existe relevância estatística significativa nas diferenças apresentadas.

7.2 MELHORES RESULTADOS DO MÉTODO LP COM OS COMITÊS DE CLASSIFICADORES BAGGING, BOOSTING E STACKING

A Tabela 6 apresenta a média e o desvio padrão dos melhores resultados do método LP com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*. A Figura 14 ilustra a média dos melhores resultados do método LP com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*.

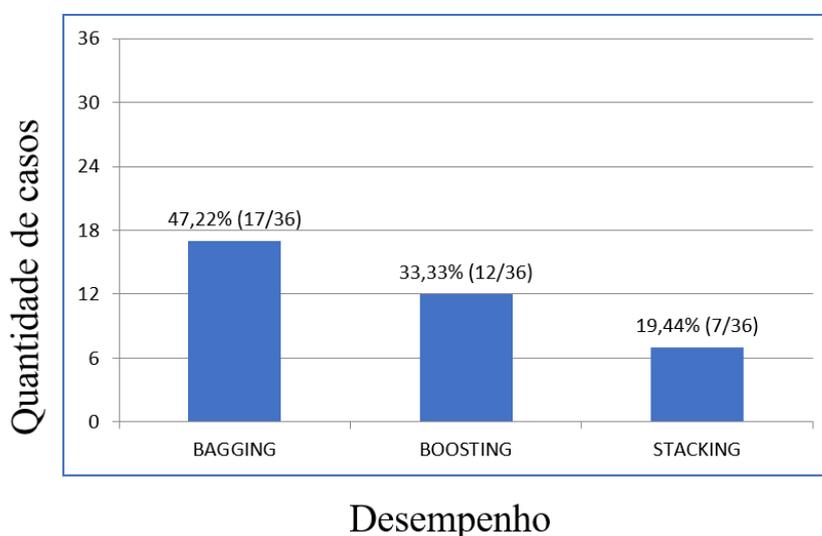


Figura 14 – Média dos melhores resultados do método LP com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*.

Tabela 6 – Média e desvio padrão do método LP com os comitês de Classificadores *Bagging*, *Boosting* e *Stacking*.

Comitê	BAGGING		BOOSTING		STACKING			
Base	BIRDS							
Medida	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	FRI	WIL
HLoss↓	0.0483	0.0112	0.0488	0.0105	0.0536	0.0126	0.319	
PPRec↑	0.5902	0.1681	0.6638	0.1145	0.5653	0.2145	0.182	
PRec↑	0.4340	0.1216	0.6069	0.1204	0.6256	0.1320	0.006	BC
F-M↑	0.4585	0.1061	0.6333	0.1168	0.5867	0.1901	0.044	BC
1-Err↓	0.7958	0.0795	0.6812	0.0332	0.5631	0.2283	0.001	AB
AvgPPRec↑	0.4456	0.1201	0.5243	0.0232	0.5906	0.1898	0.003	AB
Base	EMOTIONS							
HLoss↓	0.1922	0.0220	0.1964	0.0163	0.2425	0.0173	0.000	BC
PPRec↑	0.6858	0.0292	0.6741	0.0236	0.6104	0.0293	0.000	BC
PRec↑	0.7061	0.0482	0.7151	0.0381	0.6081	0.0489	0.001	BC
F-M↑	0.6955	0.0363	0.6937	0.0266	0.6088	0.0365	0.000	BC
1-Err↓	0.4081	0.0663	0.3777	0.0492	0.3792	0.0429	0.524	
AvgPPRec↑	0.6923	0.0355	0.7177	0.0345	0.7145	0.0251	0.222	
Base	ENRON							
HLoss↓	0.0595	0.0046	0.0576	0.0039	0.0683	0.0040	0.00	BC
PPRec↑	0.5449	0.0537	0.5573	0.0452	0.4275	0.0693	0.00	BC
PRec↑	0.4112	0.0374	0.4349	0.0492	0.3033	0.0683	0.00	BC
F-M↑	0.4442	0.0459	0.4896	0.0442	0.3309	0.0668	0.00	ABC
1-Err↓	0.5595	0.0407	0.4939	0.0575	0.6303	0.0494	0.00	ABC
AvgPPRec↑	0.3752	0.0369	0.4300	0.0351	0.3176	0.0428	0.00	ABC
Base	YEAST							
HLoss↓	0.2130	0.0122	0.2203	0.0118	0.2564	0.0138	0.00	BC
PPRec↑	0.6638	0.0236	0.6515	0.0321	0.5802	0.0272	0.00	BC
PRec↑	0.6069	0.0327	0.6012	0.0221	0.5590	0.0396	0.00	BC
F-M↑	0.6333	0.0237	0.6191	0.0240	0.5691	0.0307	0.00	BC
1-Err↓	0.3849	0.0563	0.2955	0.0440	0.3032	0.0417	0.00	AB
AvgPPRec↑	0.6340	0.0243	0.6734	0.0306	0.6441	0.0264	0.00	AC
Base	FLAGS							
HLoss↓	0.2571	0.0863	0.2714	0.0552	0.3077	0.0475	0.602	
PPRec↑	0.7309	0.0707	0.7236	0.0502	0.6953	0.0672	0.316	
PRec↑	0.7248	0.0582	0.7245	0.0659	0.6538	0.0811	0.041	BC
F-M↑	0.7215	0.0569	0.7201	0.0642	0.6717	0.0545	0.001	BC
1-Err↓	0.4553	0.1101	0.3929	0.1086	0.3442	0.0918	0.388	
AvgPPRec↑	0.6984	0.1095	0.7225	0.0364	0.7172	0.0489	0.704	
Base	CAL500							
HLoss↓	0.1931	0.0053	0.1970	0.0048	0.1411	0.1042	0.429	
PPRec↑	0.3507	0.0196	0.3403	0.0134	0.4511	0.2117	0.317	
PRec↑	0.3561	0.0227	0.3488	0.0173	0.3418	0.0272	0.367	
F-M↑	0.3477	0.0228	0.3390	0.0133	0.1031	0.0507	1.000	
1-Err↓	0.9284	0.0361	0.8845	0.0638	0.8000	0.2422	0.205	
AvgPPRec↑	0.1883	0.0149	0.2452	0.0143	0.2440	0.0666	0.000	AB

Analisando a Tabela 6, é possível observar que os experimentos utilizando o método LP com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*, o comitê *Bagging* apresentou melhor desempenho em 47,22% (17/36) dos casos. O Teste de *Friedman* foi executado para verificar se há significância estatística nas diferenças apresentadas, a qual foi verificada em 63,88% dos casos (23/36). Nas comparações realizadas para detectar entre quais amostras as diferenças estatística são relevantes foi utilizado o teste de *Wilcoxon*, o qual obteve 23,91% (11/46) dos casos.

Neste contexto, pode-se afirmar que, em geral, nos experimentos realizados, o comitê *Bagging* obteve os melhores resultados. Nas análises realizadas, em todos os casos foi verificada que existe relevância estatística significativa nas diferenças apresentadas.

7.3 MELHORES RESULTADOS DO MÉTODO RAKEL COM OS COMITÊS DE CLASSIFICADORES BAGGING, BOOSTING E STACKING

A Tabela 7 apresenta a média e o desvio padrão dos melhores resultados do método Rakel com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*. A Figura 15 ilustra a média dos melhores resultados do método Rakel com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*.

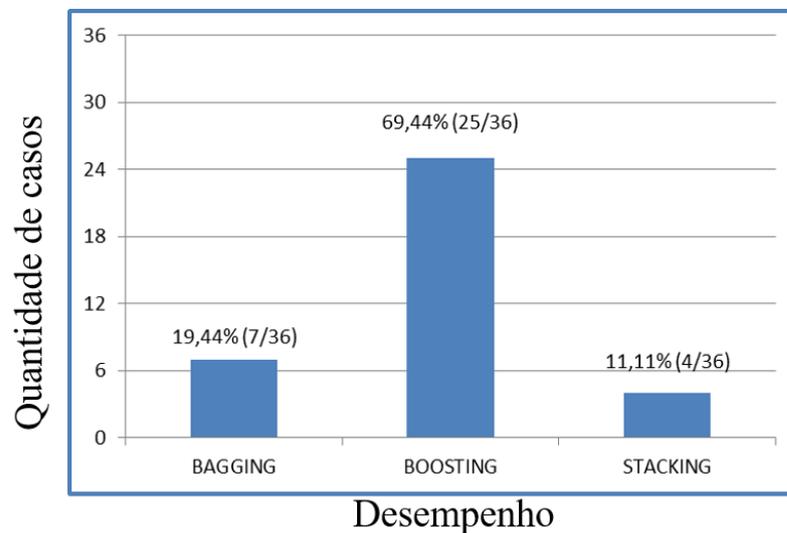


Figura 15 – Média dos melhores resultados do método Rakel com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*.

Tabela 7 – Média e desvio padrão dos melhores resultados do método Rakel com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*.

Comitê	BAGGING		BOOSTING		STACKING			
Base	BIRDS							
Medida	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	FRI	WIL
HLoss↓	0.0439	0.0109	0.0413	0.0079	0.0433	0.0114	0.030	BC
PPRec↑	0.7404	0.1442	0.7880	0.1523	0.6914	0.0696	0.001	BC
PRec↑	0.4669	0.1044	0.5033	0.1317	0.6556	0.0796	0.001	BC
F-M↑	0.4681	0.0939	0.4909	0.1035	0.4380	0.0760	0.114	
1-Err↓	0.7245	0.0780	0.6869	0.0712	0.7025	0.0953	0.042	B
AvgPPRec↑	0.5370	0.0682	0.5835	0.0830	0.5999	0.0961	0.082	
Base	EMOTIONS							
HLoss↓	0.1823	0.0170	0.1815	0.0165	0.1821	0.0100	0.000	AB
PPRec↑	0.7255	0.0483	0.7293	0.0457	0.7108	0.0192	0.071	
PRec↑	0.6894	0.0413	0.7209	0.0513	0.7007	0.0301	0.000	BC
F-M↑	0.6980	0.0270	0.7021	0.0240	0.7053	0.0186	0.002	AC
1-Err↓	0.2512	0.0247	0.2477	0.0381	0.2460	0.0324	0.001	BC
AvgPPRec↑	0.7993	0.0239	0.8021	0.0234	0.8068	0.0177	0.007	BC
Base	ENRON							
HLoss↓	0.0486	0.0030	0.0516	0.0034	0.0504	0.0019	0.000	AC
PPRec↑	0.6568	0.0505	0.6266	0.0425	0.6647	0.0481	0.000	AC
PRec↑	0.5875	0.0407	0.6007	0.0414	0.5094	0.0309	0.001	BC
F-M↑	0.5596	0.0403	0.5328	0.0342	0.5227	0.0284	0.016	AB
1-Err↓	0.2989	0.0721	0.2917	0.0737	0.3248	0.0603	0.736	
AvgPPRec↑	0.5891	0.0474	0.6032	0.0470	0.6116	0.0345	0.067	
Base	YEAST							
HLoss↓	0.2065	0.0108	0.2073	0.0154	0.2298	0.0168	0.000	ABC
PPRec↑	0.6962	0.0205	0.7063	0.0388	0.6225	0.0304	0.000	BC
PRec↑	0.6089	0.0331	0.6191	0.0274	0.6277	0.0185	0.007	BC
F-M↑	0.6292	0.0232	0.6324	0.0276	0.6201	0.0270	0.00	BC
1-Err↓	0.2562	0.0577	0.2497	0.0505	0.2584	0.0497	0.047	C
AvgPPRec↑	0.7274	0.0243	0.7318	0.0286	0.7243	0.0177	0.006	BC
Base	FLAGS							
HLoss↓	0.2344	0.0356	0.2416	0.0304	0.2470	0.0419	0.000	ABC
PPRec↑	0.7376	0.0498	0.7385	0.0442	0.7322	0.0599	0.078	
PRec↑	0.8070	0.0470	0.7786	0.0367	0.7895	0.0297	0.000	ABC
F-M↑	0.7684	0.0410	0.7572	0.0311	0.7515	0.0467	0.000	ABC
1-Err↓	0.2171	0.0985	0.2061	0.0476	0.2155	0.0682	0.000	ABC
AvgPPRec↑	0.8128	0.0424	0.8184	0.0345	0.8156	0.0321	0.048	C
Base	CAL500							
HLoss↓	0.1823	0.0170	0.1815	0.0165	0.1411	0.1042	0.001	AB
PPRec↑	0.7255	0.0483	0.7293	0.0457	0.4511	0.2117	0.000	ABC
PRec↑	0.6894	0.0413	0.7209	0.0513	0.3418	0.0272	0.000	ABC
F-M↑	0.6980	0.0270	0.7021	0.0239	0.3931	0.1031	0.000	AB
1-Err↓	0.2512	0.0247	0.2477	0.0381	0.8000	0.2422	0.000	BC
AvgPPRec↑	0.7993	0.0239	0.8021	0.0234	0.2440	0.0666	0.000	ABC

Analisando a Tabela 7, é possível observar que os experimentos utilizando o método Rakel com o comitê de classificadores *Bagging*, *Boosting* e *Stacking*, o comitê *Boosting* apresentou melhor desempenho em 69,44% (25/36) dos casos. O Teste de *Friedman* foi executado para verificar se há significância estatística nas diferenças apresentadas, a qual foi verificada em 83,33% dos casos (30/36). Nas comparações realizadas para detectar entre quais amostras as diferenças estatística são relevantes foi utilizado o teste de *Wilcoxon*, o qual obteve 40% (24/60) dos casos. Vale ressaltar que, as

bases *Emotions* e *CAL500* obtiveram os mesmos resultados das métricas de avaliação nos comitês de classificadores *Boosting*.

Neste contexto, pode-se afirmar que, em geral, nos experimentos realizados, o comitê *Boosting* obteve os melhores resultados. Nas análises realizadas, em todos os casos foi verificada que existe relevância estatística significativa nas diferenças apresentadas.

7.4 MELHORES RESULTADOS DOS MÉTODOS BR, LP E RAKEL COM O COMITÊ DE CLASSIFICADORES BAGGING

A Tabela 8 apresenta a média e o desvio padrão dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Bagging*. A Figura 16 ilustra a média dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Bagging*.

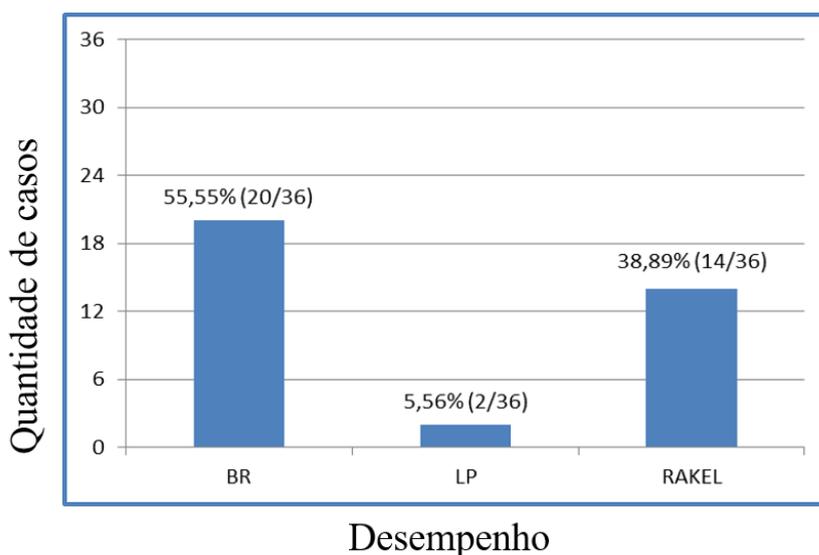


Figura 16 – Média dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Bagging*

Tabela 8 – Média e desvio padrão dos métodos BR, LP e Rakel com o comitê de classificadores *Bagging*.

Método	BR		LP		RAKEL			
Base	BIRDS							
Medida	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	FRI	WIL
HLoss↓	0.0428	0.0111	0.0483	0.0112	0.0439	0.0109	0.000	AC
PPRec↑	0.7426	0.1363	0.5902	0.1681	0.7404	0.1442	0.026	AC
PRec↑	0.4994	0.0875	0.4340	0.1216	0.4669	0.1044	0.145	
F-M↑	0.4641	0.0861	0.4585	0.1061	0.4681	0.0939	0.767	
1-Err↓	0.6836	0.1082	0.7958	0.0795	0.7245	0.0780	0.000	AC
AvgPPRec↑	0.6137	0.0772	0.4456	0.1201	0.5370	0.0682	0.003	ABC
Base	EMOTIONS							
HLoss↓	0.2058	0.0009	0.1922	0.0220	0.1823	0.0170	0.006	B
PPRec↑	0.7120	0.0024	0.6858	0.0292	0.7255	0.0483	0.007	AC
PRec↑	0.7544	0.0000	0.7061	0.0482	0.6894	0.0413	0.024	AB
F-M↑	0.6815	0.0034	0.6955	0.0363	0.6980	0.0270	0.430	
1-Err↓	0.2083	0.0264	0.4081	0.0663	0.2512	0.0247	0.000	ABC
AvgPPRec↑	0.8091	0.0018	0.6923	0.0355	0.7993	0.0239	0.000	AC
Base	ENRON							
HLoss↓	0.0491	0.0016	0.0595	0.0046	0.0486	0.0030	0.000	AC
PPRec↑	0.6607	0.0475	0.5449	0.0537	0.6568	0.0505	0.000	AC
PRec↑	0.6012	0.0408	0.4112	0.0374	0.5875	0.0407	0.000	AC
F-M↑	0.5438	0.0442	0.4442	0.0459	0.5596	0.0403	0.000	AC
1-Err↓	0.2642	0.0648	0.5595	0.0407	0.2989	0.0721	0.000	ABC
AvgPPRec↑	0.6617	0.0414	0.3752	0.0369	0.5891	0.0474	0.000	ABC
Base	YEAST							
HLoss↓	0.1994	0.0065	0.2130	0.0122	0.2065	0.0108	0.137	
PPRec↑	0.7139	0.0176	0.6638	0.0236	0.6962	0.0205	0.000	AC
PRec↑	0.7265	0.0162	0.6069	0.0327	0.6089	0.0331	0.606	
F-M↑	0.6962	0.0143	0.6333	0.0237	0.6292	0.0232	0.000	AB
1-Err↓	0.2420	0.0202	0.3849	0.0563	0.2562	0.0577	0.000	AC
AvgPPRec↑	0.7534	0.0140	0.6340	0.0243	0.7274	0.0243	0.000	ABC
Base	FLAGS							
HLoss↓	0.2415	0.0328	0.2571	0.0863	0.2344	0.0356	0.517	
PPRec↑	0.7388	0.0600	0.7309	0.0707	0.7376	0.0498	1.000	
PRec↑	0.7793	0.0437	0.7248	0.0582	0.8070	0.0470	0.000	ABC
F-M↑	0.7573	0.0361	0.7215	0.0569	0.7684	0.0410	0.003	AC
1-Err↓	0.1955	0.0750	0.4553	0.1101	0.2171	0.0985	0.000	AC
AvgPPRec↑	0.8163	0.0268	0.6984	0.1095	0.8128	0.0424	0.000	AC
Base	CAL500							
HLoss↓	0.1375	0.0052	0.1931	0.0053	0.1823	0.0170	0.000	AB
PPRec↑	0.6059	0.0258	0.3507	0.0196	0.7255	0.0483	0.000	ABC
PRec↑	0.5098	0.0389	0.3561	0.0227	0.6894	0.0413	0.000	ABC
F-M↑	0.3605	0.0174	0.3477	0.0228	0.6980	0.0270	0.000	CB
1-Err↓	0.1656	0.0713	0.9284	0.0361	0.2512	0.0247	0.000	ABC
AvgPPRec↑	0.4692	0.0169	0.1883	0.0149	0.7993	0.0239	0.000	ABC

Analisando a Tabela 8, é possível observar que os experimentos utilizando os métodos BR, LP e Rakel com os comitê de classificadores *Bagging*, o método BR apresentou melhor desempenho em 55,55% (20/36) dos casos. O Teste de *Friedman* foi executado para verificar se há significância estatística nas diferenças apresentadas, a qual foi verificada em 80,55% dos casos (29/36). Nas comparações realizadas para detectar entre quais amostras as diferenças estatística são relevantes foi utilizado o teste de *Wilcoxon*, o qual obteve 44,82% (26/58) dos casos. Neste contexto, pode-se afirmar que,

em geral, nos experimentos realizados, o método BR apresentou os melhores resultados. Nas análises realizadas, em todos os casos foi verificada que existe relevância estatística significativa nas diferenças apresentadas.

7.5 MELHORES RESULTADOS DOS MÉTODOS BR, LP E RAKEL COM O COMITÊ DE CLASSIFICADORES BOOSTING

A Tabela 9 apresenta a média e o desvio padrão dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Boosting*. A Figura 17 ilustra a média dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Boosting*.

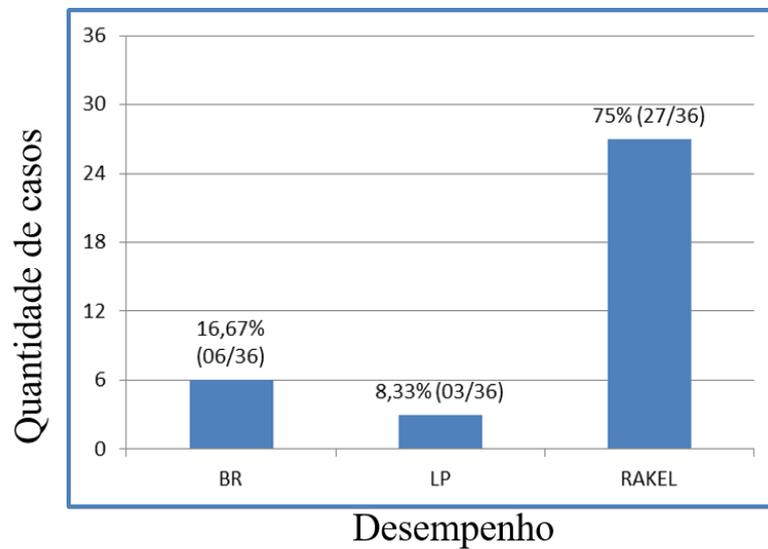


Figura 17 – Média dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Boosting*

Tabela 9 – Média e desvio padrão dos métodos BR, LP e Rakel com o comitê de classificadores *Boosting*.

Método	BR		LP		RAKEL			
Base	BIRDS							
Medida	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	FRI	WIL
HLoss↓	0.0452	0.0108	0.0488	0.0105	0.0413	0.0079	0.113	
PPRec↑	0.6149	0.1315	0.6638	0.1145	0.7880	0.1523	0.000	ABC
PRec↑	0.5375	0.1281	0.6069	0.1204	0.5033	0.1317	0.009	AC
F-M↑	0.4909	0.1015	0.6333	0.1168	0.4909	0.1035	0.083	
1-Err↓	0.7278	0.0527	0.6812	0.0332	0.6869	0.0712	0.442	
AvgPPRec↑	0.5537	0.1003	0.5243	0.0232	0.5835	0.0830	0.140	
Base	EMOTIONS							
HLoss↓	0.2097	0.0132	0.1964	0.0163	0.1815	0.0165	0.140	
PPRec↑	0.7034	0.0272	0.6741	0.0236	0.7293	0.0457	0.004	AC
PRec↑	0.7254	0.0083	0.7151	0.0381	0.7209	0.0513	0.684	
F-M↑	0.6805	0.0104	0.6937	0.0266	0.7021	0.0240	0.028	B
1-Err↓	0.2600	0.0000	0.3777	0.0492	0.2477	0.0381	0.000	AC
AvgPPRec↑	0.8017	0.0004	0.7177	0.0345	0.8021	0.0234	0.000	AC
Base	ENRON							
HLoss↓	0.0552	0.0033	0.0576	0.0039	0.0516	0.0034	0.007	BC
PPRec↑	0.5827	0.0595	0.5573	0.0452	0.6266	0.0425	0.007	ABC
PRec↑	0.5740	0.0470	0.4349	0.0492	0.6007	0.0414	0.000	AC
F-M↑	0.5107	0.0361	0.4896	0.0442	0.5328	0.0342	0.048	BC
1-Err↓	0.3213	0.0621	0.4939	0.0575	0.2917	0.0737	0.000	AC
AvgPPRec↑	0.5938	0.0284	0.4300	0.0351	0.6032	0.0470	0.000	AC
Base	YEAST							
HLoss↓	0.1331	0.0344	0.2203	0.0118	0.2073	0.0154	0.000	ABC
PPRec↑	0.6986	0.0224	0.6515	0.0321	0.7063	0.0388	0.001	AC
PRec↑	0.6062	0.0235	0.6012	0.0221	0.6191	0.0274	0.235	
F-M↑	0.6303	0.0128	0.6191	0.0240	0.6324	0.0276	0.143	
1-Err↓	0.1648	0.1010	0.2955	0.0440	0.2497	0.0505	0.003	AC
AvgPPRec↑	0.7626	0.0819	0.6734	0.0306	0.7318	0.0286	0.007	AC
Base	FLAGS							
HLoss↓	0.6619	0.0274	0.2714	0.0552	0.2416	0.0304	0.000	AB
PPRec↑	0.7668	0.0247	0.7236	0.0502	0.7385	0.0442	0.135	
PRec↑	0.7525	0.0494	0.7245	0.0659	0.7786	0.0367	0.105	
F-M↑	0.3929	0.0240	0.7201	0.0642	0.7572	0.0311	0.000	AB
1-Err↓	0.7225	0.1089	0.3929	0.1086	0.2061	0.0476	0.000	ABC
AvgPPRec↑	0.8148	0.0320	0.7225	0.0364	0.8184	0.0345	0.000	AC
Base	CAL500							
HLoss↓	0.1507	0.0046	0.1970	0.0048	0.1815	0.0165	0.00	ABC
PPRec↑	0.4944	0.0190	0.3403	0.0134	0.7293	0.0457	0.00	ABC
PRec↑	0.4503	0.0271	0.3488	0.0173	0.7209	0.0513	0.00	ABC
F-M↑	0.3620	0.0185	0.3390	0.0133	0.7021	0.0239	0.00	ABC
1-Err↓	0.2530	0.0618	0.8845	0.0638	0.2477	0.0381	0.00	AC
AvgPPRec↑	0.4558	0.0179	0.2452	0.0143	0.8021	0.0234	0.00	ABC

Analisando a Tabela 9, é possível observar que os experimentos utilizando os métodos BR, LP e Rakel com o comitê de classificadores *Boosting*, o método Rakel apresentou melhor desempenho em 75% (27/36) dos casos. O Teste de *Friedman* foi executado para verificar se há significância estatística nas diferenças apresentadas, a qual foi verificada em 72,22% (26/36) dos casos. Nas comparações realizadas para detectar entre quais amostras as diferenças estatística são relevantes foi utilizado o teste de *Wilcoxon*, o qual obteve 57,69% (30/52) dos casos.

Neste contexto, pode-se afirmar que, em geral, nos experimentos realizados, o método Rakel apresentou os melhores resultados. Nas análises realizadas, em todos os casos foi verificada que existe relevância estatística significativa nas diferenças apresentadas.

7.6 MELHORES RESULTADOS DOS MÉTODOS BR, LP E RAKEL COM O COMITÊ DE CLASSIFICADORES STACKING

A Tabela 10 apresenta a média e o desvio padrão dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Stacking*. A Figura 18 ilustra a média dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Stacking*.

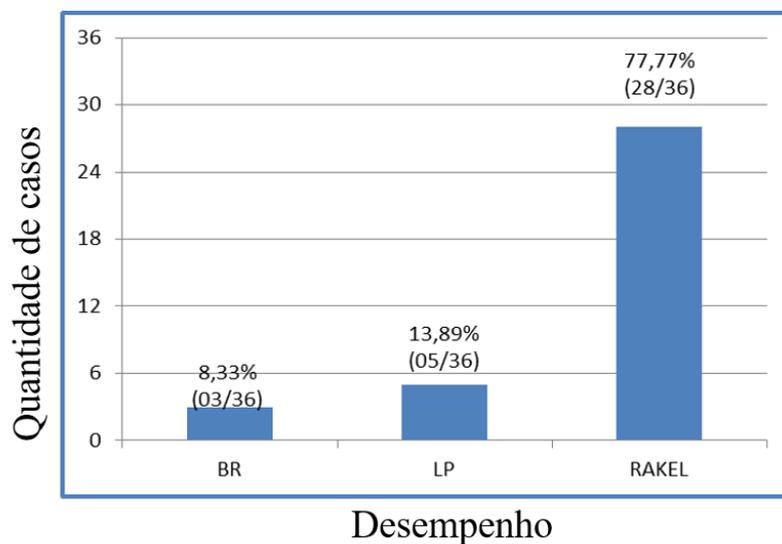


Figura 18 – Média dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Stacking*

Tabela 10 – Média e desvio padrão dos métodos BR, LP e Rakel com o comitê de classificadores *Stacking*.

Método	BR		LP		RAKEL			
Base	BIRDS							
Medida	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	FRI	WIL
HLoss↓	0.0522	0.0106	0.0536	0.0126	0.0433	0.0114	0.007	BC
PPRec↑	0.4797	0.0961	0.5653	0.2145	0.6914	0.0696	0.007	B
PRec↑	0.3946	0.0731	0.6256	0.1320	0.6556	0.0796	0.000	AB
F-M↑	0.4028	0.0748	0.5867	0.1901	0.4380	0.0760	0.078	
1-Err↓	0.7830	0.0597	0.5631	0.2283	0.7025	0.0953	0.018	AB
AvgPPRec↑	0.4757	0.0522	0.5906	0.1898	0.5999	0.0961	0.027	B
Base	EMOTIONS							
HLoss↓	0.2139	0.0000	0.2425	0.0173	0.1821	0.0100	0.00	ABC
PPRec↑	0.6914	0.0000	0.6104	0.0293	0.7108	0.0192	0.00	ABC
PRec↑	0.7632	0.0000	0.6081	0.0489	0.7007	0.0301	0.00	ABC
F-M↑	0.6932	0.0000	0.6088	0.0365	0.7053	0.0186	0.00	AC
1-Err↓	0.3167	0.0000	0.3792	0.0429	0.2460	0.0324	0.00	ABC
AvgPPRec↑	0.7768	0.0000	0.7145	0.0251	0.8068	0.0177	0.00	ABC
Base	ENRON							
HLoss↓	0.0504	0.0019	0.0683	0.0040	0.0504	0.0019	0.00	AC
PPRec↑	0.6793	0.0435	0.4275	0.0693	0.6647	0.0481	0.00	ABC
PRec↑	0.5089	0.0284	0.3033	0.0683	0.5094	0.0309	0.00	AC
F-M↑	0.5227	0.0284	0.3309	0.0668	0.5227	0.0284	0.00	ABC
1-Err↓	0.3075	0.0799	0.6303	0.0494	0.3248	0.0603	0.00	AC
AvgPPRec↑	0.6116	0.0345	0.3176	0.0428	0.6116	0.0345	0.00	ABC
Base	YEAST							
HLoss↓	0.2586	0.0096	0.2564	0.0138	0.2298	0.0168	0.001	BC
PPRec↑	0.5725	0.0210	0.5802	0.0272	0.6225	0.0304	0.001	BC
PRec↑	0.5751	0.0169	0.5590	0.0396	0.6277	0.0185	0.000	BC
F-M↑	0.5737	0.0167	0.5691	0.0307	0.6201	0.0270	0.001	BC
1-Err↓	0.2726	0.0547	0.3032	0.0417	0.2584	0.0497	0.005	BC
AvgPPRec↑	0.6844	0.0263	0.6441	0.0264	0.7243	0.0177	0.000	ABC
Base	FLAGS							
HLoss↓	0.3006	0.0329	0.3077	0.0475	0.2470	0.0419	0.046	BC
PPRec↑	0.7048	0.0522	0.6953	0.0672	0.7322	0.0599	0.614	
PRec↑	0.6847	0.0661	0.6538	0.0811	0.7895	0.0297	0.002	BC
F-M↑	0.6838	0.0507	0.6717	0.0545	0.7515	0.0467	0.002	BC
1-Err↓	0.2937	0.0781	0.3442	0.0918	0.2155	0.0682	0.000	ABC
AvgPPRec↑	0.7819	0.0247	0.7172	0.0489	0.8156	0.0321	0.000	AC
Base	CAL500							
HLoss↓	0.1345	0.0372	0.1411	0.1042	0.1411	0.1042	1.000	
PPRec↑	0.3532	0.0155	0.4511	0.2117	0.4511	0.2117	0.238	
PRec↑	0.3454	0.0177	0.3418	0.0272	0.3418	0.0272	0.784	
F-M↑	0.3469	0.0139	0.3931	0.1031	0.3931	0.1031	0.788	
1-Err↓	0.8666	0.0363	0.8000	0.2422	0.8000	0.2422	0.787	
AvgPPRec↑	0.2868	0.0504	0.2440	0.0666	0.2440	0.0666	0.037	C

Analisando a Tabela 10, é possível observar que os experimentos utilizando os métodos BR, LP e Rakel com o comitê de classificadores *Stacking*, o método Rakel apresentou melhor desempenho em 77,77% (28/36) dos casos. O Teste de *Friedman* foi executado para verificar se há significância estatística nas diferenças apresentadas, a qual foi verificada em 80,55% (26/36) dos casos. Nas comparações realizadas para detectar entre quais amostras as diferenças estatística são relevantes foi utilizado o teste de *Wilcoxon*, o qual obteve 63,46% (33/52) dos casos. Vale ressaltar que, os métodos LP e

Rakel obtiveram os mesmos resultados das métricas de avaliação com a base CAL500.

Neste contexto, pode-se afirmar que, em geral, nos experimentos realizados, o método Rakel apresentou os melhores resultados. Nas análises realizadas, em todos os casos foi verificada que existe relevância estatística significativa nas diferenças apresentadas.

7.7 CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentados os resultados experimentais obtidos pelos métodos de classificação multirrótulo utilizando as abordagens de aprendizado supervisionado. Os experimentos utilizando os métodos BR, LP e Rakel com os comitês de classificadores *Bagging*, *Boosting* e *Stacking* apresentam as médias dos melhores resultados dos vários métodos supervisionados utilizados na realização dos experimentos.

Em relação aos resultados dos métodos BR, LP e Rakel, observou-se que, na maioria dos casos, o método Rakel apresentou melhor resultado. Por sua vez, ao analisar segundo a perspectiva dos comitês de classificadores *Bagging*, *Boosting* e *Stacking* observou-se que, na maioria dos casos, o comitê de classificadores *Bagging* apresentou os melhores resultados em todos os experimentos realizados no aprendizado supervisionado. Foi verificada relevância estatística na maioria dos casos analisados.

8 MÉTODOS DE CLASSIFICAÇÃO MULTIRRÓTULO SEMISSUPERVISIONADO MULTIDESCRICÃO

Neste oitavo capítulo, será realizada uma investigação comparativa dos métodos de classificação multirrótulo utilizando a abordagem supervisionada e semisupervisionada.

Neste contexto, todos os resultados apresentam a média e o desvio padrão das melhores métricas, independente do tamanho do comitê e dos classificadores supervisionado, tendo sido sombreados os melhores resultados em cada uma das medidas. Assim como no capítulo sétimo, para análise da relevância estatística foram utilizados os testes de *Friedman* e *Wilcoxon*. Os resultados nos quais os testes estatísticos mostraram relevância do ponto de vista estatístico tiveram os valores **FRI** da coluna marcados em negrito.

A coluna **WIL** nas tabelas exibe os resultados das comparações nas quais foi verificada significância estatística nas diferenças apresentadas. Vale ressaltar que, para algumas métricas de avaliação, os melhores resultados foram obtidos com valores próximos a zero, enquanto que, para outras métricas, os melhores resultados foram obtidos com valores próximos a um. Portanto, ao lado de cada métrica, foi adicionado o símbolo ↓ (o menor valor significa o melhor) e ↑ (o maior valor significa o melhor) para representar o comportamento das métricas de avaliação.

As Seções que apresentam os resultados experimentais relacionados às estratégias métodos BR, LP e Rakel com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*, a coluna **WIL** nas tabelas exibe os resultados das comparações nas quais foi verificada significância estatística nas diferenças apresentadas utilizando o teste de *Wilcoxon*, com a seguinte nomenclatura:

- A: Método BR vs LP
- B: Método BR vs Rakel
- C: Método LP vs Rakel

Enquanto que, as Seções que apresentam os melhores resultados experimentais dos métodos BR, LP e Rakel com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*, a coluna **WIL** nas tabelas exibe os resultados das comparações nas quais foi verificada significância estatística nas diferenças apresentadas utilizando o teste de *Wilcoxon*, com a seguinte nomenclatura:

- A: *Bagging* vs *Boosting*
- B: *Bagging* vs *Stacking*

- C: *Boosting vs Stacking*

Por fim, as Seções que apresentam os melhores resultados experimentais do aprendizado supervisionado vs aprendizado semissupervisionado, a coluna **WIL** nas tabelas exibe os resultados das comparações nas quais foi verificada significância estatística nas diferenças apresentadas utilizando o teste de *Wilcoxon*. Os resultados nos quais os testes estatísticos mostraram relevância do ponto de vista estatístico tiveram os valores **WIL** marcados em negrito.

8.1 MELHORES RESULTADOS DO MÉTODO BR COM OS COMITÊS DE CLASSIFICADORES BAGGING, BOOSTING E STACKING

A Tabela 11 apresenta a média e o desvio padrão dos melhores resultados do método BR com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*. A Figura 19 ilustra a média dos melhores resultados do método BR com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*.

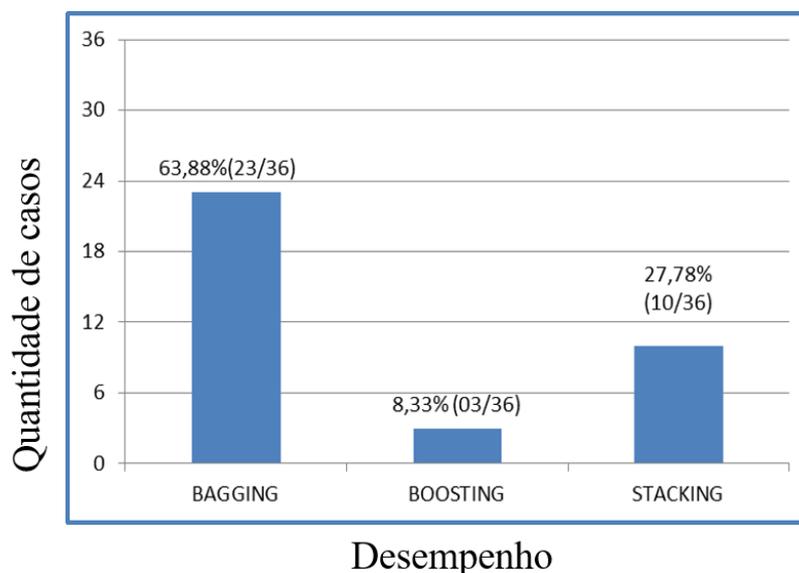


Figura 19 – Média dos melhores resultados do método BR com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*.

Tabela 11 – Média e desvio padrão do método BR com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*.

Comitê	BAGGING		BOOSTING		STACKING			
Base	BIRDS							
Medida	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	FRI	WIL
HLoss↓	0.0436	0.0083	0.0490	0.0121	0.0457	0.0098	0.191	
PPRec↑	0.6630	0.1282	0.6158	0.2110	0.6723	0.2247	0.766	
PRec↑	0.3463	0.0827	0.3914	0.0464	0.3792	0.0946	0.000	AC
F-M↑	0.4275	0.0696	0.4253	0.0992	0.3189	0.0651	0.000	BC
1-Err↓	0.6808	0.0892	0.7246	0.1129	0.4170	0.0610	0.010	BC
AvgPPRec↑	0.6293	0.0947	0.5591	0.1123	0.5457	0.0808	0.006	AB
Base	EMOTIONS							
HLoss↓	0.2089	0.0142	0.2161	0.0207	0.2222	0.0204	0.163	
PPRec↑	0.7424	0.0248	0.7479	0.0248	0.6744	0.0418	0.000	BC
PRec↑	0.6684	0.0261	0.6219	0.0497	0.6219	0.0357	0.032	AB
F-M↑	0.6545	0.0229	0.6453	0.0354	0.6233	0.0320	0.182	
1-Err↓	0.2517	0.0183	0.2667	0.0236	0.3167	0.0550	0.048	B
AvgPPRec↑	0.7876	0.0137	0.7672	0.0256	0.7503	0.0255	0.006	ABC
Base	ENRON							
HLoss↓	0.0588	0.0016	0.0436	0.0083	0.0457	0.0098	0.0	AB
PPRec↑	0.7339	0.0114	0.6544	0.1337	0.6723	0.2247	0.0	AB
PRec↑	0.6482	0.0342	0.3439	0.0685	0.3230	0.1224	0.0	AB
F-M↑	0.6463	0.0301	0.4315	0.0633	0.3981	0.0231	0.0	ABC
1-Err↓	0.2617	0.0352	0.5915	0.1027	0.5883	0.0827	0.0	AB
AvgPPRec↑	0.7621	0.0142	0.6241	0.1004	0.5457	0.0808	0.0	ABC
Base	YEAST							
HLoss↓	0.0082	0.0000	0.0465	0.0078	0.2236	0.0119	0.0	AB
PPRec↑	0.8942	0.0000	0.7478	0.0198	0.7011	0.0450	0.0	ABC
PRec↑	0.8571	0.0000	0.5975	0.0088	0.5768	0.0175	0.0	ABC
F-M↑	0.8500	0.0000	0.6156	0.0143	0.6025	0.0274	0.0	AB
1-Err↓	0.1031	0.0000	0.2545	0.0215	0.2519	0.0436	0.0	AB
AvgPPRec↑	0.9145	0.0000	0.7183	0.0150	0.7116	0.0303	0.0	AB
Base	FLAGS							
HLoss↓	0.2403	0.0416	0.2763	0.0490	0.2491	0.0399	0.006	BC
PPRec↑	0.7513	0.0615	0.7204	0.0570	0.7307	0.0544	0.223	
PRec↑	0.7533	0.0510	0.7379	0.0520	0.7648	0.0657	0.165	
F-M↑	0.7513	0.0489	0.7172	0.0546	0.7464	0.0532	0.026	AC
1-Err↓	0.1787	0.1075	0.2158	0.1244	0.2611	0.1089	0.209	
AvgPPRec↑	0.8319	0.0421	0.8092	0.0285	0.8101	0.0550	0.077	
Base	CAL500							
HLoss↓	0.1751	0.0052	0.1755	0.0037	0.0567	0.0055	0.000	BC
PPRec↑	0.3983	0.0226	0.3939	0.0241	0.6165	0.3089	0.313	
PRec↑	0.3353	0.0155	0.3213	0.0163	0.3470	0.0370	0.018	AC
F-M↑	0.3630	0.0136	0.3538	0.0182	0.3779	0.0537	0.608	
1-Err↓	0.3982	0.1119	0.4023	0.0548	0.1935	0.1033	0.006	BC
AvgPPRec↑	0.3489	0.0204	0.3879	0.0088	0.4252	0.0585	0.001	AB

Analisando a Tabela 11, é possível observar que os experimentos utilizando o método BR, o comitê de classificadores *Bagging* apresentou o melhor desempenho em 63,88% dos casos (23/36). O Teste de *Friedman* foi executado para verificar se há significância estatística nas diferenças apresentadas, a qual foi verificada em 72,22%

dos casos (26/36). Nas comparações realizadas para detectar entre quais amostras as diferenças estatística são relevantes foi utilizado o teste de *Wilcoxon*, o qual foi verificado diferenças estatística em 53,84% dos casos (28/52).

Neste contexto, pode-se afirmar que, em geral, nos experimentos realizados utilizando o método BR, o comitê de classificadores *Bagging* obteve os melhores resultados. Nas análises realizadas, em todos os casos foi verificada que existe relevância estatística significativa nas diferenças apresentadas.

8.2 MELHORES RESULTADOS DO MÉTODO LP COM OS COMITÊS DE CLASSIFICADORES BAGGING, BOOSTING E STACKING

A Tabela 12 apresenta a média e o desvio padrão dos melhores resultados do método LP com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*. A Figura 20 ilustra a média dos melhores resultados do método LP com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*.

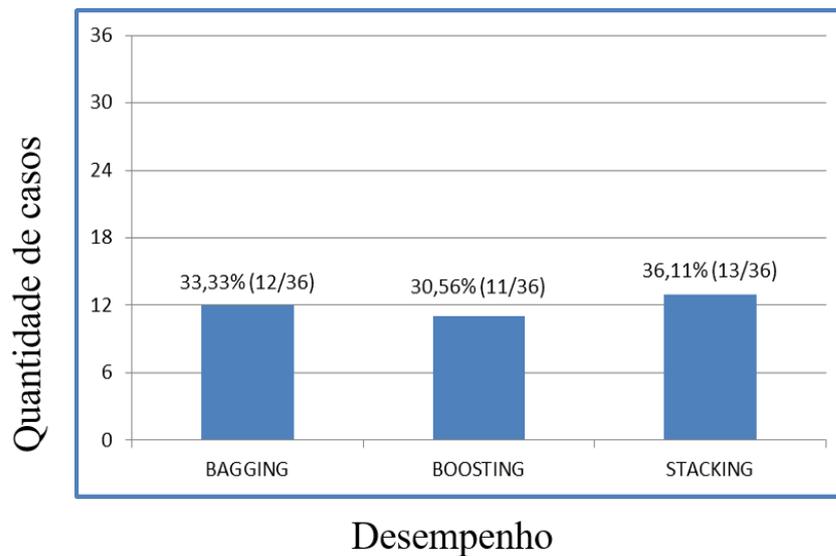


Figura 20 – Média dos melhores resultados do método LP com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*.

Tabela 12 – Média e desvio padrão do método LP com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*.

Comitê	BAGGING		BOOSTING		STACKING			
Base	BIRDS							
Medida	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	FRI	WIL
HLoss↓	0.0485	0.0141	0.0542	0.0135	0.0560	0.0134	0.026	BC
PPRec↑	0.5366	0.1426	0.4656	0.1377	0.2855	0.1168	0.012	AC
PRec	0.3802	0.0827	0.1299	0.0751	0.2488	0.1149	0.682	
F-M↑	0.9057	0.2982	0.4111	0.0738	0.2644	0.1144	0.002	AB
1-Err↓	0.8733	0.0529	0.8238	0.0738	0.8670	0.0665	0.000	ABC
AvgPPRec↑	0.3334	0.0959	0.4167	0.0880	0.3376	0.1154	0.000	ABC
Base	EMOTIONS							
HLoss↓	0.2212	0.0292	0.2298	0.0205	0.2541	0.0227	0.235	
PPRec↑	0.6413	0.0455	0.6325	0.0295	0.5925	0.0473	0.004	BC
PRec	0.6572	0.0531	0.6310	0.0807	0.5878	0.0499	0.002	BC
F-M↑	0.6488	0.0477	0.6304	0.0248	0.5899	0.0284	0.433	
1-Err↓	0.5059	0.0596	0.3912	0.0450	0.4133	0.0411	0.000	ABC
AvgPPRec↑	0.6269	0.0258	0.7091	0.0387	0.6969	0.0288	0.000	ABC
Base	ENRON							
HLoss↓	0.0681	0.0045	0.0690	0.0050	0.0207	0.0030	0.00	BC
PPRec↑	0.4584	0.0285	0.4375	0.0577	0.6287	0.0459	0.00	BC
PRec	0.3608	0.0456	0.3171	0.0304	0.5941	0.0643	0.00	ABC
F-M↑	0.4031	0.0379	0.3611	0.0275	0.6099	0.0516	0.00	ABC
1-Err↓	0.7564	0.0530	0.6460	0.0441	0.3754	0.0517	0.00	ABC
AvgPPRec↑	0.2611	0.0269	0.2653	0.0320	0.6298	0.0450	0.00	BC
Base	YEAST							
HLoss↓	0.2314	0.0102	0.2424	0.0128	0.2436	0.0000	0.00	ABC
PPRec↑	0.6236	0.0168	0.6004	0.0210	0.6646	0.0000	0.00	ABC
PRec	0.6102	0.0183	0.5969	0.0260	0.7063	0.0000	0.00	BC
F-M↑	0.6167	0.0159	0.5980	0.0210	0.6837	0.0000	0.00	ABC
1-Err↓	0.3711	0.0460	0.3641	0.0346	0.3767	0.0000	0.00	BC
AvgPPRec↑	0.5938	0.0243	0.5748	0.0136	0.7434	0.0000	0.00	ABC
Base	FLAGS							
HLoss↓	0.2976	0.0400	0.2911	0.0376	0.3257	0.0229	0.901	
PPRec↑	0.7010	0.0707	0.7069	0.0455	0.6538	0.1129	0.189	
PRec	0.6789	0.0642	0.6899	0.0415	0.6282	0.1929	0.0427	
F-M↑	0.6867	0.0576	0.6962	0.0328	0.6372	0.0416	0.316	
1-Err↓	0.5342	0.1907	0.4189	0.1969	0.3500	0.0000	0.001	BC
AvgPPRec↑	0.6264	0.0637	0.6663	0.0803	0.6906	0.0246	0.000	BC
Base	CAL500							
HLoss↓	0.1932	0.0082	0.0717	0.0047	0.1979	0.006	0.000	BC
PPRec↑	0.3548	0.0285	0.3473	0.0173	0.3430	0.0058	0.017	BC
PRec	0.3536	0.0222	0.3507	0.0639	0.3369	0.0357	0.000	BC
F-M↑	0.3539	0.0230	0.3484	0.0258	0.3261	0.0208	0.136	
1-Err↓	0.9542	0.0250	0.9318	0.1305	0.9421	0.0251	0.000	ABC
AvgPPRec↑	0.1452	0.0097	0.2373	0.0135	0.1745	0.0256	0.000	ABC

Analisando a Tabela 12, é possível observar que os experimentos utilizando o método LP, o comitê de classificadores *Stacking* apresentou o melhor desempenho em 36,11% dos casos (13/36). O Teste de *Friedman* foi executado para verificar se há significância estatística nas diferenças apresentadas, a qual foi verificada em 77,77%

dos casos (28/36). Nas comparações realizadas para detectar entre quais amostras as diferenças estatística são relevantes foi utilizado o teste de *Wilcoxon*, o qual foi verificado diferenças estatística em 35,71% dos casos (20/56).

Neste contexto, pode-se afirmar que, em geral, nos experimentos realizados utilizando o método LP, o comitê de classificadores *Stacking* obteve os melhores resultados. Nas análises realizadas, em todos os casos foi verificada que existe relevância estatística significativa nas diferenças apresentadas.

8.3 MELHORES RESULTADOS DO MÉTODO RAKEL COM OS COMITÊS DE CLASSIFICADORES BAGGING, BOOSTING E STACKING

A Tabela 13 apresenta a média e o desvio padrão dos melhores resultados do método Rakel com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*. A Figura 21 ilustra a média dos melhores resultados do método Rakel com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*.

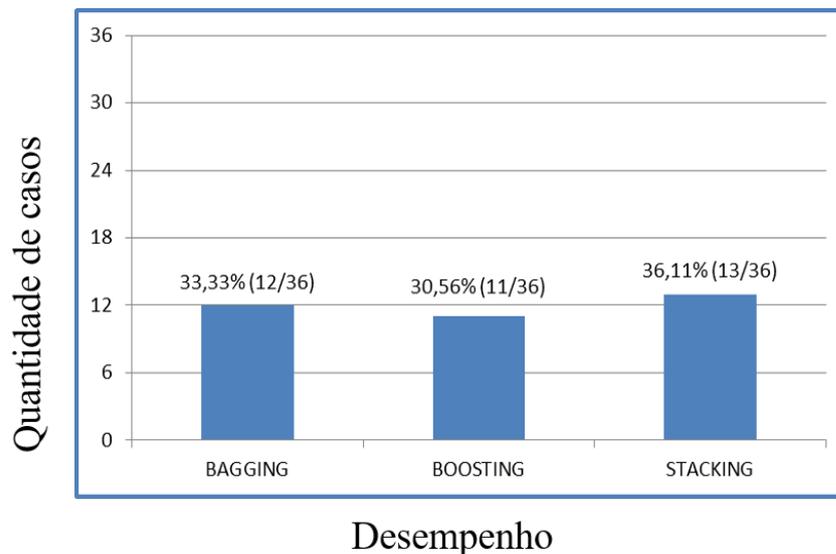


Figura 21 – Média dos melhores resultados do método Rakel com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*.

Tabela 13 – Média e desvio padrão do método Rakel com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*.

Comitê	BAGGING		BOOSTING		STACKING			
Base	BIRDS							
Medida	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	FRI	WIL
HLoss↓	0.0421	0.0092	0.0422	0.0114	0.0391	0.0014	1.000	
PPRec↑	0.6454	0.1147	0.6391	0.1435	0.5934	0.1409	0.261	
PRec↑	0.3742	0.1032	0.3791	0.1317	0.3867	0.0973	0.527	
F-M↑	0.4620	0.0975	0.4655	0.1226	0.4625	0.0979	0.827	
1-Err↓	0.1922	0.0274	0.6840	0.0712	0.7243	0.0715	0.000	ABC
AvgPPRec↑	0.6044	0.0797	0.6102	0.1099	0.5837	0.0906	0.825	
Base	EMOTIONS							
HLoss↓	0.0444	0.0269	0.2136	0.0114	0.2071	0.0227	0.804	
PPRec↑	0.6707	0.0513	0.7397	0.0000	0.6834	0.0473	0.017	AC
PRec↑	0.6757	0.0383	0.7561	0.0770	0.6938	0.0499	0.266	
F-M↑	0.6695	0.0397	0.6463	0.0301	0.6545	0.0284	0.317	
1-Err↓	0.2580	0.0468	0.2167	0.0352	0.2749	0.0411	0.650	
AvgPPRec↑	0.7971	0.0274	0.8111	0.0000	0.7880	0.0288	0.311	
Base	ENRON							
HLoss↓	0.0080	0.0029	0.0105	0.0017	0.0102	0.0014	0.832	
PPRec↑	0.9679	0.1417	0.8351	0.0519	0.8544	0.0432	0.372	
PRec↑	0.8335	0.0100	0.7921	0.0246	0.7963	0.0481	0.003	AB
F-M↑	0.8468	0.0532	0.8049	0.0266	0.8121	0.0290	0.609	
1-Err↓	0.0981	0.0640	0.1647	0.0359	0.2045	0.0362	0.007	ABC
AvgPPRec↑	0.9002	0.0488	0.8687	0.0221	0.8191	0.0324	0.00	ABC
Base	YEAST							
HLoss↓	0.2045	0.0077	0.0105	0.0017	0.0102	0.0014	0.00	AB
PPRec↑	0.6816	0.0179	0.8351	0.0519	0.8544	0.0432	0.00	ABC
PRec↑	0.6094	0.0171	0.7921	0.0246	0.7963	0.0481	0.00	AB
F-M↑	0.6432	0.0143	0.8049	0.0266	0.8121	0.0290	0.00	AB
1-Err↓	0.2396	0.0217	0.1647	0.0359	0.2045	0.0362	0.00	ABC
AvgPPRec↑	0.7422	0.0165	0.8687	0.0221	0.8191	0.0324	0.00	ABC
Base	FLAGS							
HLoss↓	0.2686	0.0274	0.1564	0.0250	0.2571	0.1466	0.977	
PPRec↑	0.6637	0.0274	0.4646	0.0494	1.1575	0.4596	0.013	ABC
PRec↑	0.7733	0.0439	0.3551	0.0295	0.8509	0.1929	0.715	
F-M↑	0.7070	0.0309	0.3851	0.0322	0.9996	0.3228	0.048	ABC
1-Err↓	0.2850	0.0669	0.2230	0.0985	0.2353	0.1035	0.263	
AvgPPRec↑	0.7944	0.0161	0.4150	0.0370	1.1896	0.3961	0.0	ABC
Base	CAL500							
HLoss↓	0.1564	0.0048	0.1515	0.0026	0.1697	0.0050	0.0	ABC
PPRec↑	0.4646	0.0187	0.4595	0.0082	0.4004	0.0235	0.0	BC
PRec↑	0.3551	0.0148	0.4501	0.0372	0.4990	0.0244	0.0	ABC
F-M↑	0.3851	0.0185	0.4566	0.0264	0.3665	0.0071	0.0	ABC
1-Err↓	0.2230	0.0599	0.0567	0.0492	0.2227	0.0979	0.0	AC
AvgPPRec↑	0.4150	0.0136	0.4281	0.0089	0.3893	0.0053	0.0	ABC

Analisando a Tabela 13, é possível observar que os experimentos utilizando o método Rakel, o comitê de classificadores *Bagging* apresentou o melhor desempenho em 36,11% dos casos (13/36). O Teste de *Friedman* foi executado para verificar se há significância estatística nas diferenças apresentadas, a qual foi verificada em 55,55%

dos casos (20/36). Nas comparações realizadas para detectar entre quais amostras as diferenças estatística são relevantes foi utilizado o teste de *Wilcoxon*, o qual foi verificado diferenças estatística em 27,50% dos casos (11/40). Vale ressaltar que, as bases *Enron* e *Yeast* obtiveram os mesmos resultados das métricas de avaliação com os comitê de classificadores *Boosting* e *Stacking*.

Neste contexto, pode-se afirmar que, em geral, nos experimentos realizados utilizando o método *Rakel*, o comitê de classificadores *Bagging* obteve os melhores resultados. Nas análises realizadas, em todos os casos foi verificada que existe relevância estatística significativa nas diferenças apresentadas.

8.4 MELHORES RESULTADOS DOS MÉTODOS BR, LP E RAKEL COM O COMITÊ DE CLASSIFICADORES BAGGING

A Tabela 14 apresenta a média e o desvio padrão dos resultados do melhores resultados dos métodos BR, LP e *Rakel* com o comitê de classificadores *Bagging*. A Figura 22 ilustra a média dos melhores resultados dos métodos BR, LP e *Rakel* com o comitê de classificadores *Bagging*.

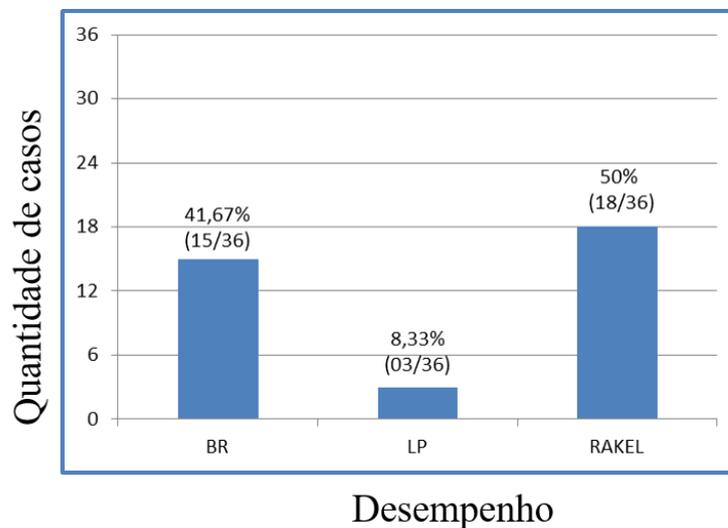


Figura 22 – Média dos melhores resultados dos métodos BR, LP e *Rakel* com o comitê de classificadores *Bagging*.

Tabela 14 – Média e desvio padrão dos métodos BR, LP e Rakel com o comitê de classificadores *Bagging*.

Método	BR		LP		RAKEL			
Base	BIRDS							
Medida	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	FRI	WIL
HLoss ↓	0.0436	0.0083	0.0485	0.0141	0.0421	0.0092	0.217	
PPRec↑	0.6630	0.1282	0.5366	0.1426	0.6454	0.1147	0.001	AC
PRec↑	0.3463	0.0827	0.3802	0.1299	0.3742	0.1032	0.653	
F-M↑	0.4275	0.0696	0.9057	0.2982	0.4620	0.0975	0.003	AC
1-Err↓	0.6808	0.0892	0.8733	0.0529	0.1922	0.0274	0.000	ABC
AvgPPRec↑	0.6293	0.0947	0.3334	0.0959	0.6044	0.0797	0.000	AC
Base	EMOTIONS							
HLoss ↓	0.2089	0.0142	0.2212	0.0292	0.0444	0.0269	0.115	
PPRec↑	0.7424	0.0248	0.6413	0.0455	0.6707	0.0513	0.001	ABC
PRec↑	0.6684	0.0261	0.6572	0.0531	0.6757	0.0383	0.312	
F-M↑	0.6545	0.0229	0.6488	0.0477	0.6695	0.0397	0.136	
1-Err↓	0.2517	0.0183	0.5059	0.0596	0.2580	0.0468	0.000	AC
AvgPPRec↑	0.7876	0.0137	0.6269	0.0258	0.7971	0.0274	0.000	AC
Base	ENRON							
HLoss ↓	0.0588	0.0016	0.0681	0.0045	0.0080	0.0029	0.0	ABC
PPRec↑	0.7339	0.0277	0.4584	0.0285	0.9679	0.1417	0.0	ABC
PRec↑	0.6482	0.0342	0.3608	0.0456	0.8335	0.0100	0.0	ABC
F-M↑	0.6463	0.0301	0.4031	0.0379	0.8468	0.0532	0.0	ABC
1-Err↓	0.2617	0.0352	0.7564	0.0348	0.0981	0.0640	0.0	ABC
AvgPPRec↑	0.7621	0.0142	0.2611	0.0269	0.9002	0.0488	0.0	ABC
Base	YEAST							
HLoss ↓	0.0082	0.0000	0.2314	0.0102	0.2045	0.0077	0.0	ABC
PPRec↑	0.8942	0.0000	0.6236	0.0168	0.6816	0.0179	0.0	ABC
PRec↑	0.8571	0.0000	0.6102	0.0183	0.6094	0.0171	0.0	AB
F-M↑	0.8500	0.0000	0.6167	0.0159	0.6432	0.0143	0.0	ABC
1-Err↓	0.1031	0.0000	0.3711	0.0460	0.2396	0.0217	0.0	ABC
AvgPPRec↑	0.9145	0.0000	0.5938	0.0243	0.7422	0.0165	0.0	ABC
Base	FLAGS							
HLoss ↓	0.2403	0.0416	0.2976	0.0400	0.2686	0.0274	0.001	ABC
PPRec↑	0.7513	0.0615	0.7010	0.0707	0.6637	0.0274	0.010	AB
PRec↑	0.7533	0.0510	0.6789	0.0642	0.7733	0.0439	0.011	AC
F-M↑	0.7513	0.0489	0.6867	0.0576	0.7070	0.0309	0.006	AB
1-Err↓	0.1787	0.1075	0.5342	0.1907	0.2850	0.0669	0.000	ABC
AvgPPRec↑	0.8319	0.0421	0.6264	0.0637	0.7944	0.0161	0.000	ABC
Base	CAL500							
HLoss ↓	0.1751	0.0052	0.1932	0.0082	0.1564	0.0048	0.000	ABC
PPRec↑	0.3983	0.0226	0.3548	0.0285	0.4646	0.0187	0.000	ABC
PRec↑	0.3353	0.0155	0.3536	0.0222	0.3551	0.0128	0.007	ABC
F-M↑	0.3630	0.0136	0.3539	0.0230	0.3851	0.0185	0.000	ABC
1-Err↓	0.3982	0.1119	0.9542	0.0250	0.2230	0.0599	0.000	ABC
AvgPPRec↑	0.3489	0.0204	0.1452	0.0097	0.4150	0.0136	0.000	ABC

Analisando a Tabela 14, é possível observar que os experimentos utilizando os métodos BR, LP e Rakel com o comitê de classificadores *Bagging*, o método Rakel

apresentou melhor desempenho em 50% (18/36) dos casos. O Teste de *Friedman* foi executado para verificar se há significância estatística nas diferenças apresentadas, a qual foi verificada em 86,11% dos casos (31/36). Nas comparações realizadas para detectar entre quais amostras as diferenças estatística são relevantes foi utilizado o teste de *Wilcoxon*, o qual foi verificado diferenças estatística em 41,93% dos casos (26/62).

Neste contexto, pode-se afirmar que, em geral, nos experimentos realizados utilizando os métodos BR, LP e Rakel com o comitê de classificadores *Stacking*, o método Rakel apresentou os melhores resultados. Nas análises realizadas, em todos os casos foi verificada que existe relevância estatística significativa nas diferenças apresentadas.

8.5 MELHORES RESULTADOS DOS MÉTODOS BR, LP E RAKEL COM O COMITÊ DE CLASSIFICADORES BOOSTING

A Tabela 15 apresenta a média e o desvio padrão dos resultados do melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Boosting*. A Figura 23 ilustra a média dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Boosting*.

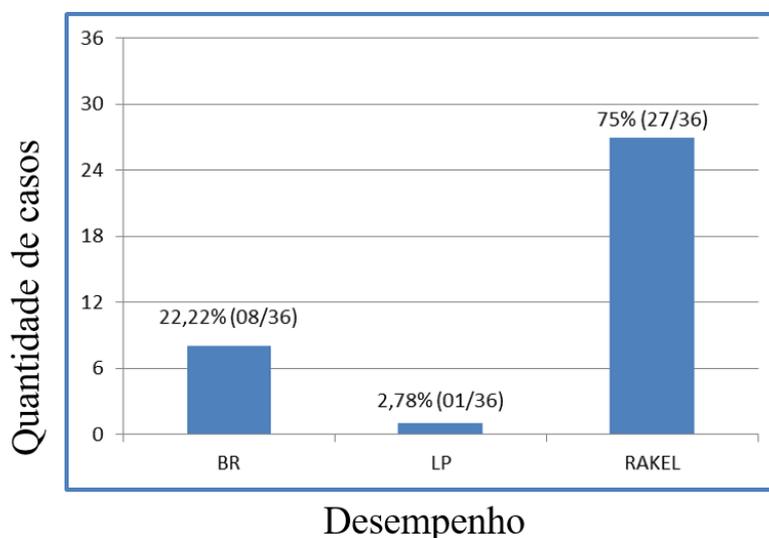


Figura 23 – Média dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Boosting*.

Tabela 15 – Média e desvio padrão dos métodos BR, LP e Rakel com o comitê de classificadores *Boosting*.

Método	BR		LP		RAKEL			
Base	BIRDS							
Medida	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	FRI	WIL
HLoss ↓	0.0490	0.0121	0.0542	0.0135	0.0422	0.0114	0.130	
PPRec↑	0.6158	0.2110	0.4656	0.1377	0.6391	0.1435	0.270	
PRec↑	0.3914	0.1157	0.3737	0.0751	0.3791	0.1317	0.564	
F-M↑	0.4253	0.0992	0.4111	0.0738	0.4655	0.1226	0.190	
1-Err↓	0.7246	0.1129	0.8238	0.0738	0.6840	0.0712	0.002	AC
AvgPPRec↑	0.5591	0.1123	0.4167	0.0880	0.6102	0.1099	0.00	ABC
Base	EMOTIONS							
HLoss ↓	0.2161	0.0207	0.2298	0.0205	0.2136	0.0114	0.180	
PPRec↑	0.7479	0.0248	0.6325	0.0295	0.7397	0.0000	0.000	AC
PRec↑	0.6219	0.0497	0.6310	0.0807	0.7561	0.0770	0.019	BC
F-M↑	0.6453	0.0354	0.6304	0.0248	0.6463	0.0301	0.833	
1-Err↓	0.2667	0.0236	0.3912	0.0450	0.2167	0.0352	0.000	ABC
AvgPPRec↑	0.7672	0.0256	0.7091	0.0387	0.8111	0.0000	0.000	ABC
Base	ENRON							
HLoss ↓	0.0436	0.0083	0.0690	0.0050	0.0105	0.0017	0.0	ABC
PPRec↑	0.6544	0.1337	0.4375	0.0577	0.8351	0.0519	0.0	ABC
PRec↑	0.3439	0.0685	0.3171	0.0304	0.7921	0.0246	0.0	ABC
F-M↑	0.4315	0.0633	0.3611	0.0275	0.8049	0.0266	0.0	ABC
1-Err↓	0.5915	0.1027	0.6460	0.0441	0.1647	0.0359	0.0	BC
AvgPPRec↑	0.6241	0.1004	0.2653	0.0320	0.8687	0.0221	0.0	ABC
Base	YEAST							
HLoss ↓	0.0465	0.0078	0.2424	0.0128	0.0105	0.0017	0.0	ABC
PPRec↑	0.7478	0.0198	0.6004	0.0210	0.8351	0.0519	0.0	ABC
PRec↑	0.5975	0.0088	0.5969	0.0260	0.7921	0.0246	0.0	BC
F-M↑	0.6156	0.0143	0.5980	0.0210	0.8049	0.0266	0.0	ABC
1-Err↓	0.2545	0.0215	0.3641	0.0346	0.1647	0.0359	0.0	ABC
AvgPPRec↑	0.7183	0.0150	0.5748	0.0136	0.8687	0.0221	0.0	ABC
Base	FLAGS							
HLoss ↓	0.2763	0.0490	0.2911	0.0376	0.1564	0.0250	0.593	
PPRec↑	0.7204	0.0570	0.7069	0.0455	0.4646	0.0494	0.445	
PRec↑	0.7379	0.0520	0.6899	0.0415	0.3551	0.0539	0.000	ABC
F-M↑	0.7172	0.0546	0.6962	0.0328	0.3851	0.0322	0.007	BC
1-Err↓	0.2158	0.1244	0.4189	0.1969	0.2230	0.0985	0.029	AC
AvgPPRec↑	0.8092	0.0285	0.6663	0.0803	0.4150	0.0370	0.000	AC
Base	CAL500							
HLoss ↓	0.1755	0.0037	0.0717	0.0047	0.1515	0.0026	0.0	ABC
PPRec↑	0.3939	0.0241	0.3473	0.0173	0.4595	0.0082	0.0	ABC
PRec↑	0.3213	0.0163	0.3507	0.0639	0.4501	0.0372	0.0	BC
F-M↑	0.3538	0.0182	0.3484	0.0258	0.4566	0.0264	0.0	BC
1-Err↓	0.4023	0.0548	0.9318	0.1305	0.0567	0.0492	0.0	ABC
AvgPPRec↑	0.3879	0.0088	0.2373	0.0135	0.4281	0.0089	0.0	ABC

Analisando a Tabela 15, é possível observar que os experimentos utilizando os métodos BR, LP e Rakel com o comitê de classificadores *Boosting*, o método Rakel

apresentou melhor desempenho em 75% (27/36) dos casos. O Teste de *Friedman* foi executado para verificar se há significância estatística nas diferenças apresentadas, a qual foi verificada em 77,77% (28/36) dos casos. Nas comparações realizadas para detectar entre quais amostras as diferenças estatística são relevantes foi utilizado o teste de *Wilcoxon*, o qual foi verificado diferenças estatística em 67,85% dos casos (38/56).

Neste contexto, pode-se afirmar que, em geral, nos experimentos realizados utilizando os métodos BR, LP e Rakel com o comitê de classificadores *Stacking*, o método Rakel apresentou os melhores resultados. Nas análises realizadas, em todos os casos foi verificada que existe relevância estatística significativa nas diferenças apresentadas.

8.6 MELHORES RESULTADOS DOS MÉTODOS BR, LP E RAKEL COM O COMITÊ DE CLASSIFICADORES STACKING

A Tabela 16 apresenta a média e o desvio padrão dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Stacking*. A Figura 24 ilustra a média dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Stacking*.

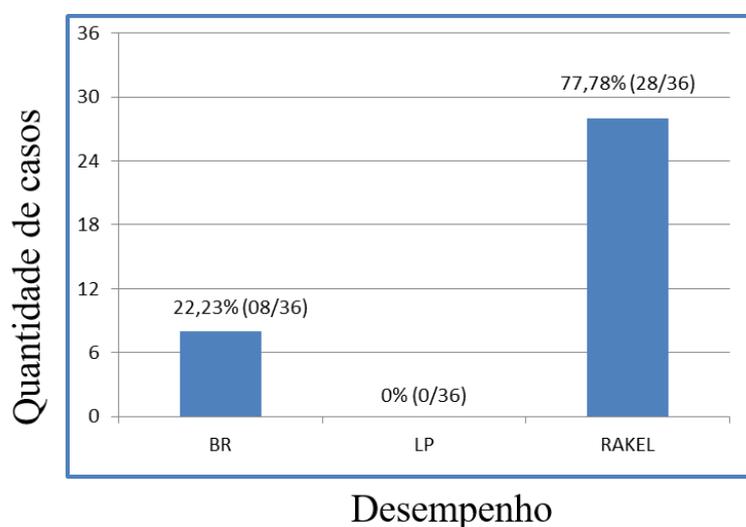


Figura 24 – Média dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Stacking*.

Tabela 16 – Média e desvio padrão dos métodos BR, LP e Rakel com o comitê de classificadores *Stacking*.

Método	BR		LP		RAKEL			
Base	BIRDS							
Medida	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	FRI	WIL
HLoss ↓	0.0457	0.0098	0.0560	0.0134	0.0391	0.0014	0.000	ABC
PPRec↑	0.6723	0.2247	0.2855	0.1168	0.5934	0.1409	0.000	AC
PRec↑	0.3792	0.0946	0.2488	0.1149	0.3867	0.0973	0.014	C
F-M↑	0.3189	0.0651	0.2644	0.1144	0.4625	0.0979	0.000	ABC
1-Err↓	0.4170	0.0610	0.8670	0.0665	0.7243	0.0715	0.000	BC
AvgPPRec↑	0.5457	0.0808	0.3376	0.1154	0.5837	0.0906	0.000	ABC
Base	EMOTIONS							
HLoss ↓	0.2222	0.0204	0.2541	0.0227	0.2071	0.0227	0.030	AC
PPRec↑	0.6744	0.0418	0.5925	0.0473	0.6834	0.0473	0.026	AC
PRec↑	0.6219	0.0357	0.5878	0.0499	0.6938	0.0499	0.000	ABC
F-M↑	0.6233	0.0320	0.5899	0.0284	0.6545	0.0284	0.120	BC
1-Err↓	0.3167	0.0550	0.4133	0.0629	0.2749	0.0411	0.002	ABC
AvgPPRec↑	0.7503	0.0255	0.6969	0.0288	0.7880	0.0288	0.000	ABC
Base	ENRON							
HLoss ↓	0.0457	0.0098	0.0207	0.0030	0.0102	0.0014	0.000	ABC
PPRec↑	0.6723	0.2247	0.6287	0.0459	0.8544	0.0432	0.006	BC
PRec↑	0.3230	0.1224	0.5941	0.0643	0.7963	0.0481	0.000	ABC
F-M↑	0.3981	0.0231	0.6099	0.0516	0.8121	0.0290	0.000	ABC
1-Err↓	0.5883	0.0827	0.3754	0.0517	0.2045	0.0362	0.000	ABC
AvgPPRec↑	0.5457	0.0808	0.6298	0.0450	0.8191	0.0324	0.000	ABC
Base	YEAST							
HLoss ↓	0.2236	0.0119	0.2436	0.000	0.0102	0.0014	0.000	ABC
PPRec↑	0.7011	0.0450	0.6646	0.000	0.8544	0.0432	0.000	ABC
PRec↑	0.5768	0.0175	0.7063	0.000	0.7963	0.0481	0.000	ABC
F-M↑	0.6025	0.0274	0.6837	0.000	0.8121	0.0290	0.000	ABC
1-Err↓	0.2519	0.0436	0.3767	0.000	0.2045	0.0362	0.000	ABC
AvgPPRec↑	0.7116	0.0303	0.7434	0.000	0.8191	0.0324	0.000	ABC
Base	FLAGS							
HLoss ↓	0.2491	0.0399	0.3257	0.0229	0.2571	0.1466	0.038	A
PPRec↑	0.7307	0.0544	0.6538	0.1129	1.1575	0.4596	0.003	ABC
PRec↑	0.7648	0.0657	0.6282	0.0427	0.8509	0.1929	0.009	AC
F-M↑	0.7464	0.0532	0.6372	0.0416	0.9996	0.3228	0.001	ABC
1-Err↓	0.2611	0.1089	0.3500	0.0000	0.2353	0.1035	0.044	AC
AvgPPRec↑	0.8101	0.0550	0.6906	0.0246	1.1896	0.3961	0.010	ABC
Base	CAL500							
HLoss ↓	0.0567	0.0055	0.1979	0.0060	0.1697	0.0050	0.000	ABC
PPRec↑	0.6165	0.3089	0.3430	0.0058	0.4004	0.0235	0.025	ABC
PRec↑	0.3470	0.0370	0.3369	0.0357	0.4990	0.0244	0.000	BC
F-M↑	0.3779	0.0537	0.3261	0.0208	0.3665	0.0071	0.000	AC
1-Err↓	0.1935	0.1033	0.9421	0.0251	0.2227	0.0979	0.000	AC
AvgPPRec↑	0.4252	0.0585	0.1745	0.0256	0.3893	0.0053	0.000	ABC

Analisando a Tabela 16, é possível observar que os experimentos utilizando os métodos BR, LP e Rakel com o comitê de classificadores *Stacking*, o método Rakel

apresentou melhor desempenho em 77,78% (28/36) dos casos. O Teste de *Friedman* foi executado para verificar se há significância estatística nas diferenças apresentadas, a qual foi verificada em 100% (36/36) dos casos. Nas comparações realizadas para detectar entre quais amostras as diferenças estatística são relevantes foi utilizado o teste de *Wilcoxon*, o qual foi verificado diferenças estatística em 65,27% dos casos (47/72).

Neste contexto, pode-se afirmar que, em geral, nos experimentos realizados utilizando os métodos BR, LP e Rakel com o comitê de classificadores *Stacking*, o método Rakel apresentou os melhores resultados. Nas análises realizadas, em todos os casos foi verificada que existe relevância estatística significativa nas diferenças apresentadas.

8.7 MELHORES RESULTADOS DO MÉTODO BR UTILIZANDO A ABORDAGEM DO APRENDIZADO SUPERVISIONADO VERSUS MÉTODO BR UTILIZANDO A ABORDAGEM DO APRENDIZADO SEMISSUPERVISIONADO

A Tabela 17 apresenta a média e o desvio padrão dos melhores resultados do método BR utilizando a abordagem do aprendizado supervisionado *versus* método BR utilizando a abordagem do aprendizado semissupervisionado. A Figura 25 ilustra a média dos melhores resultados do método BR utilizando a abordagem do aprendizado supervisionado *versus* método BR utilizando a abordagem do aprendizado semissupervisionado.

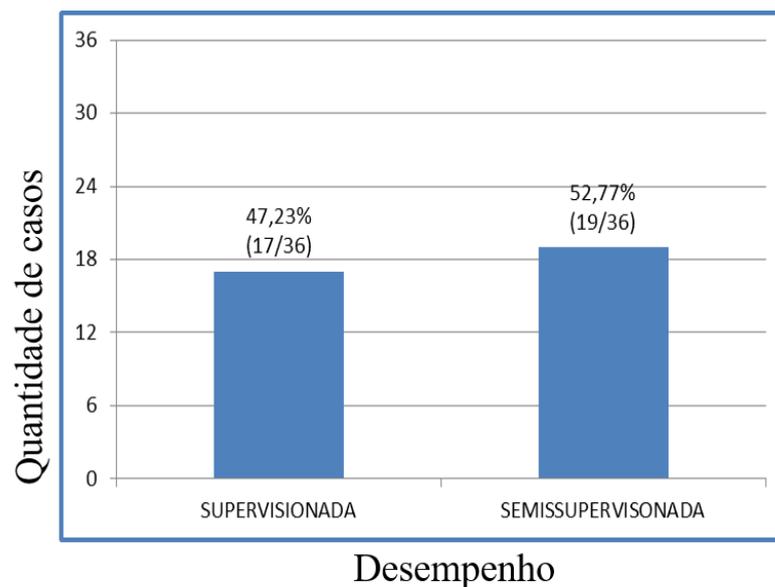


Figura 25 – Média dos melhores resultados do método BR utilizando a abordagem do aprendizado supervisionado *versus* método BR utilizando a abordagem do aprendizado semissupervisionado.

Tabela 17 – Média dos melhores resultados do método BR utilizando a abordagem do aprendizado supervisionado *versus* método BR utilizando a abordagem do aprendizado semissupervisionado.

Abordagem	SUPERVISIONADA		SEMISSUPERVISIONADA		
Comitê	BAGGING		BAGGING		
Base	BIRDS				
Medida	Média	Desvio Padrão	Média	Desvio Padrão	WIL
HLoss ↓	0.0428	0.0111	0.0436	0.0083	0.220
PPRec↑	0.7426	0.1363	0.6630	0.1282	0.241
PRec↑	0.4994	0.0875	0.3463	0.0827	0.020
F-M	0.4641	0.0861	0.4275	0.0696	0.342
1-Err↓	0.6836	0.1082	0.6808	0.0892	0.452
AvgPPRec↑	0.6137	0.0772	0.6293	0.0947	1.000
Base	EMOTIONS				
HLoss ↓	0.2058	0.0009	0.2089	0.0142	0.220
PPRec↑	0.7120	0.0024	0.7424	0.0248	0.111
PRec↑	0.7544	0.0000	0.6684	0.0261	0.020
F-M	0.6815	0.0034	0.6545	0.0229	0.210
1-Err↓	0.2083	0.0264	0.2517	0.0183	0.220
AvgPPRec↑	0.8091	0.0018	0.7876	0.0137	0.020
Base	ENRON				
HLoss ↓	0.0491	0.0016	0.0588	0.0016	0.020
PPRec↑	0.6607	0.0475	0.7339	0.0114	0.024
PRec↑	0.6012	0.0408	0.6482	0.0342	0.110
F-M	0.5438	0.0442	0.6463	0.0301	0.002
1-Err↓	0.2642	0.0648	0.2617	0.0352	0.751
AvgPPRec↑	0.6617	0.0414	0.7621	0.0142	0.003
Base	YEAST				
HLoss ↓	0.1994	0.0065	0.0082	0.0000	0.003
PPRec↑	0.7139	0.0176	0.8942	0.0000	0.002
PRec↑	0.7265	0.0162	0.8571	0.0000	0.002
F-M	0.6962	0.0143	0.8500	0.0000	0.005
1-Err↓	0.2420	0.0202	0.1031	0.0000	0.001
AvgPPRec↑	0.7534	0.0140	0.9145	0.0000	0.003
Base	FLAGS				
HLoss ↓	0.2415	0.0328	0.2403	0.0416	1.000
PPRec↑	0.7388	0.0600	0.7513	0.0615	0.348
PRec↑	0.7793	0.0437	0.7533	0.0510	0.342
F-M	0.7573	0.0361	0.7513	0.0489	0.340
1-Err↓	0.1955	0.0750	0.1787	0.1075	1.000
AvgPPRec↑	0.8163	0.0268	0.8319	0.0421	0.109
Base	CAL500				
HLoss ↓	0.1375	0.0052	0.1751	0.0052	0.001
PPRec↑	0.6059	0.0258	0.3983	0.0226	0.002
PRec↑	0.5098	0.0389	0.3353	0.0155	0.002
F-M	0.3605	0.0174	0.3630	0.0136	0.758
1-Err↓	0.1656	0.0713	0.3982	0.1119	0.002
AvgPPRec↑	0.4692	0.0169	0.3489	0.0204	0.001

Analisando a Tabela 17, é possível observar que os experimentos utilizando o

método BR utilizando a abordagem do aprendizado supervisionado *versus* BR utilizando a abordagem semissupervisionado, a abordagem semissupervisionada apresentou melhor desempenho em 52,77% dos casos (19/36). O Teste de *Wilcoxon* foi executado para verificar se há significância estatística nas diferenças apresentadas, o qual obteve 50% (18/36) dos casos.

Neste contexto, pode-se afirmar que, em geral, nos experimentos realizados a abordagem semissupervisionada apresentou os melhores resultados. Nas análises realizadas, em todos os casos foi verificada que existe relevância estatística significativa nas diferenças apresentadas.

8.7.1 Melhores resultados do método LP utilizando a abordagem do aprendizado supervisionado *versus* método LP utilizando a abordagem do aprendizado semissupervisionado

A Tabela 18 apresenta a média e o desvio padrão dos melhores resultados do método LP utilizando a abordagem do aprendizado supervisionado *versus* método LP utilizando a abordagem do aprendizado semissupervisionado. A Figura 26 ilustra a média dos melhores resultados do método LP utilizando a abordagem do aprendizado supervisionado *versus* método BR utilizando a abordagem do aprendizado semissupervisionado.

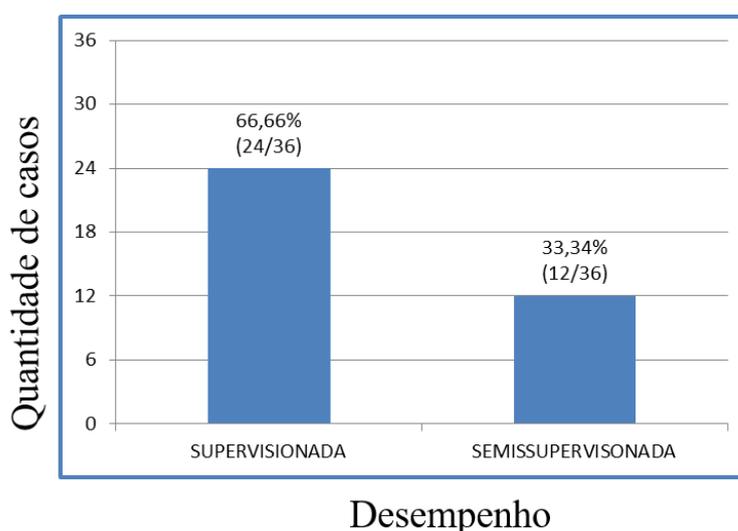


Figura 26 – Média dos melhores resultados do método LP utilizando a abordagem do aprendizado supervisionado *versus* método LP utilizando a abordagem do aprendizado semissupervisionado.

Tabela 18 – Abordagem do aprendizado supervisionado *versus* abordagem do aprendizado semissupervisionado.

Abordagem	SUPERVISIONADA		SEMI-SUPERVISIONADA		
Comitê	BAGGING		STACKING		
Base	BIRDS				
Medida	Média	Desvio Padrão	Média	Desvio Padrão	WIL
HLoss↓	0.0483	0.0112	0.0560	0.0134	0.508
PPRec↑	0.5902	0.1681	0.2855	0.1168	0.506
PRec↑	0.4340	0.1216	0.2488	0.1149	0.170
F-M↑	0.4585	0.1061	0.2644	0.1144	0.200
1-Err↓	0.7958	0.0795	0.8670	0.0665	0.003
AvgPPRec↑	0.4456	0.1201	0.3376	0.1154	0.021
Base	EMOTIONS				
HLoss↓	0.1922	0.0220	0.2541	0.0227	0.022
PPRec↑	0.6858	0.0292	0.5925	0.0473	0.021
PRec↑	0.7061	0.0482	0.5878	0.0499	0.020
F-M↑	0.6955	0.0363	0.5899	0.0284	0.024
1-Err↓	0.4081	0.0663	0.4133	0.0411	0.030
AvgPPRec↑	0.6923	0.0355	0.6969	0.0288	0.025
Base	ENRON				
HLoss↓	0.0595	0.0046	0.0207	0.0030	0.002
PPRec↑	0.5449	0.0537	0.6287	0.0459	0.002
PRec↑	0.4112	0.0374	0.5941	0.0643	0.109
F-M↑	0.4442	0.0459	0.6099	0.0516	0.025
1-Err↓	0.5595	0.0407	0.3754	0.0517	0.002
AvgPPRec↑	0.3752	0.0369	0.6298	0.0450	0.002
Base	YEAST				
HLoss↓	0.2130	0.0122	0.2436	0.0000	0.110
PPRec↑	0.6638	0.0236	0.6646	0.0000	0.002
PRec↑	0.6069	0.0327	0.7063	0.0000	0.751
F-M↑	0.6333	0.0237	0.6837	0.0000	0.338
1-Err↓	0.3849	0.0563	0.3767	0.0000	1.000
AvgPPRec↑	0.6340	0.0243	0.7434	0.0000	0.021
Base	FLAGS				
HLoss↓	0.2571	0.0863	0.3257	0.0229	0.184
PPRec↑	0.7309	0.0707	0.6538	0.1129	0.021
PRec↑	0.7248	0.0582	0.6282	0.1929	0.337
F-M↑	0.7215	0.0569	0.6372	0.0416	0.104
1-Err↓	0.4553	0.1101	0.3500	0.0000	0.495
AvgPPRec↑	0.6984	0.1095	0.6906	0.0246	0.349
Base	CAL500				
HLoss↓	0.1931	0.0053	0.1979	0.0060	1.000
PPRec↑	0.3507	0.0196	0.3430	0.0058	1.000
PRec↑	0.3561	0.3488	0.3369	0.0357	0.351
F-M↑	0.3477	0.0228	0.3261	0.0208	1.000
1-Err↓	0.9284	0.0361	0.9421	0.0251	0.108
AvgPPRec↑	0.1883	0.0149	0.1745	0.0256	0.002

Analisando a Tabela 18, é possível observar que os experimentos utilizando o método LP utilizando a abordagem do aprendizado supervisionado *versus* LP utilizando a abordagem semissupervisionado, a abordagem supervisionada apresentou melhor

desempenho em 66,66% dos casos (24/36). O Teste de *Wilcoxon* foi executado para verificar se há significância estatística nas diferenças apresentadas, o qual obteve 47,22% (17/36) dos casos.

Neste contexto, pode-se afirmar que, em geral, nos experimentos realizados a abordagem supervisionada apresentou os melhores resultados. Nas análises realizadas, em todos os casos não foi verificada que existe relevância estatística significativa nas diferenças apresentadas.

8.7.2 Melhores resultados do método Rakel utilizando a abordagem do aprendizado supervisionado versus método Rakel utilizando a abordagem do aprendizado semissupervisionado

A Tabela 19 apresenta a média e o desvio padrão dos melhores resultados do método Rakel utilizando a abordagem do aprendizado supervisionado *versus* método Rakel utilizando a abordagem do aprendizado semissupervisionado. A Figura 27 ilustra a média dos melhores resultados do método Rakel utilizando a abordagem do aprendizado supervisionado *versus* método Rakel utilizando a abordagem do aprendizado semissupervisionado.

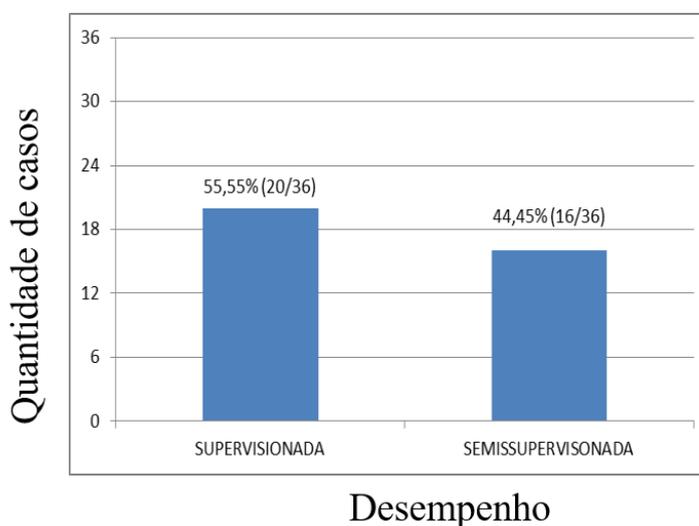


Figura 27 – Média dos melhores resultados do método Rakel utilizando a abordagem do aprendizado supervisionado *versus* método Rakel utilizando a abordagem do aprendizado semissupervisionado.

Tabela 19 – Média dos melhores resultados do método Rakel utilizando a abordagem do aprendizado supervisionado *versus* método Rakel utilizando a abordagem do aprendizado semissupervisionado.

Abordagem	SUPERVISIONADA		SEMISSUPERVISIONADA		
Comitê	BOOSTING		BAGGING		
Base	BIRDS				
Medida	Média	Desvio Padrão	Média	Desvio Padrão	WIL
HLoss↓	0.0413	0.0079	0.0421	0.0092	0.751
PPRec↑	0.7880	0.1523	0.6454	0.1147	0.200
PRec↑	0.5033	0.1317	0.3742	0.1032	0.108
F-M↑	0.4909	0.1035	0.4620	0.0975	0.105
1-Err↓	0.6869	0.0712	0.1922	0.0274	1.000
AvgPPRec↑	0.5835	0.0830	0.6044	0.0797	1.000
Base	EMOTIONS				
HLoss↓	0.1815	0.0165	0.0444	0.0269	0.024
PPRec↑	0.7293	0.0457	0.6707	0.0513	0.103
PRec↑	0.7209	0.0513	0.6757	0.0383	0.342
F-M↑	0.7021	0.0240	0.6695	0.0397	0.002
1-Err↓	0.2477	0.0381	0.2580	0.0468	0.730
AvgPPRec↑	0.8021	0.0234	0.7971	0.0274	0.348
Base	ENRON				
HLoss↓	0.0516	0.0034	0.0080	0.0029	0.002
PPRec↑	0.6266	0.0425	0.9679	0.1417	0.002
PRec↑	0.6007	0.0414	0.8335	0.0100	0.002
F-M↑	0.5328	0.0342	0.8468	0.0532	0.002
1-Err↓	0.2917	0.0737	0.0981	0.0640	0.019
AvgPPRec↑	0.6032	0.0470	0.9002	0.0488	0.021
Base	YEAST				
HLoss↓	0.2073	0.0154	0.2045	0.0077	0.003
PPRec↑	0.7063	0.0388	0.6816	0.0179	0.002
PRec↑	0.6191	0.0274	0.6094	0.0171	0.001
F-M↑	0.6324	0.0276	0.6432	0.0143	0.002
1-Err↓	0.2497	0.0505	0.2396	0.0217	0.107
AvgPPRec↑	0.7318	0.0286	0.7422	0.0165	0.003
Base	FLAGS				
HLoss↓	0.2416	0.0304	0.2686	0.0274	0.022
PPRec↑	0.7385	0.0442	0.6637	0.0274	0.022
PRec↑	0.7786	0.0367	0.7733	0.0439	1.000
F-M↑	0.7572	0.0311	0.7070	0.0309	0.755
1-Err↓	0.2061	0.0476	0.2850	0.0669	0.730
AvgPPRec↑	0.8184	0.0345	0.7944	0.0161	1.000
Base	CAL500				
HLoss↓	0.1815	0.0165	0.1564	0.0048	0.003
PPRec↑	0.7293	0.0457	0.4646	0.0187	0.002
PRec↑	0.7209	0.0513	0.3551	0.0148	0.002
F-M↑	0.7021	0.0239	0.3851	0.0185	0.002
1-Err↓	0.2477	0.0381	0.2230	0.0599	0.002
AvgPPRec↑	0.8021	0.0234	0.4150	0.0136	0.002

Analisando a Tabela 19, é possível observar que os experimentos utilizando o método Rakel utilizando a abordagem do aprendizado supervisionado *versus* Rakel utilizando a abordagem semissupervisionado, a abordagem supervisionada apresentou

melhor desempenho em 55,55% dos casos (20/36). O Teste de *Wilcoxon* foi executado para verificar se há significância estatística nas diferenças apresentadas, o qual obteve 58,33% (21/36) dos casos.

Neste contexto, pode-se afirmar que, em geral, nos experimentos realizados a abordagem supervisionada apresentou os melhores resultados. Nas análises realizadas, em todos os casos foi verificada que existe relevância estatística significativa nas diferenças apresentadas.

8.7.3 Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores Bagging utilizando a abordagem do aprendizado supervisionado versus métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado

A Tabela 20 apresenta a média e o desvio padrão dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Bagging* utilizando a abordagem do aprendizado supervisionado *versus* métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado. A Figura 28 ilustra a média dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Bagging* utilizando a abordagem do aprendizado supervisionado *versus* métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado

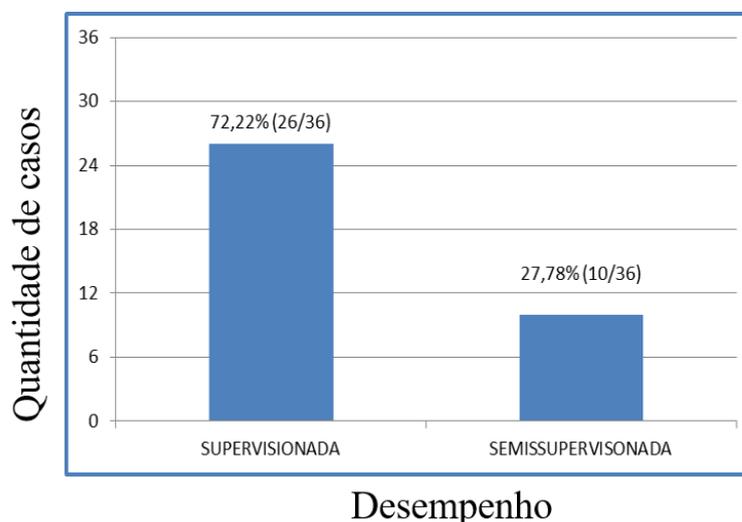


Figura 28 – Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Bagging* utilizando a abordagem do aprendizado supervisionado *versus* métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado.

Tabela 20 – Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Bagging* utilizando a abordagem do aprendizado supervisionado *versus* métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado.

Método	BR		RAKEL		
Base	BIRDS				
Medida	Média	Desvio Padrão	Média	Desvio Padrão	WIL
HLoss↓	0.0428	0.0111	0.0421	0.0092	1.000
PPRec↑	0.7426	0.1363	0.6454	0.1147	0.111
PRec↑	0.4994	0.0875	0.3742	0.1032	0.039
F-M↑	0.4641	0.0861	0.4620	0.0975	1.000
1-Err↓	0.6836	0.1082	0.1922	0.0274	0.002
AvgPPRec↑	0.6137	0.0772	0.6044	0.0797	0.757
Base	EMOTIONS				
HLoss↓	0.2058	0.0009	0.0444	0.0269	0.758
PPRec↑	0.7120	0.0024	0.6707	0.0513	0.108
PRec↑	0.7544	0.000	0.6757	0.0383	0.002
F-M↑	0.6815	0.0034	0.6695	0.0397	0.334
1-Err↓	0.2083	0.0264	0.258	0.0468	0.040
AvgPPRec↑	0.8091	0.0018	0.7971	0.0274	0.754
Base	ENRON				
HLoss↓	0.0491	0.0016	0.0080	0.0028	0.002
PPRec↑	0.6607	0.0475	0.9679	0.1417	0.002
PRec↑	0.6012	0.0408	0.8335	0.0100	0.002
F-M↑	0.5438	0.0442	0.8468	0.0532	0.003
1-Err↓	0.2642	0.0648	0.0981	0.0640	0.002
AvgPPRec↑	0.6617	0.0414	0.9002	0.0488	0.002
Base	YEAST				
HLoss↓	0.1994	0.0065	0.2045	0.0077	0.338
PPRec↑	0.7139	0.0176	0.6816	0.0179	0.003
PRec↑	0.7265	0.0162	0.6094	0.0171	0.348
F-M↑	0.6962	0.0143	0.6432	0.0143	0.002
1-Err↓	0.2420	0.0202	0.2396	0.0217	0.752
AvgPPRec↑	0.7534	0.0140	0.7422	0.0165	0.002
Base	FLAGS				
HLoss↓	0.2415	0.0328	0.2686	0.0274	0.003
PPRec↑	0.7388	0.0600	0.6637	0.0274	0.023
PRec↑	0.7793	0.0437	0.7733	0.0439	1.000
F-M↑	0.7573	0.0361	0.7070	0.0309	0.109
1-Err↓	0.1955	0.0750	0.2850	0.0669	0.020
AvgPPRec↑	0.8163	0.0268	0.7944	0.0161	0.022
Base	CAL500				
HLoss↓	0.1375	0.0052	0.1564	0.0048	0.002
PPRec↑	0.6059	0.0258	0.4646	0.0187	0.002
PRec↑	0.5098	0.0389	0.3551	0.0128	0.002
F-M↑	0.3605	0.0174	0.3851	0.0185	0.002
1-Err↓	0.1656	0.0713	0.2230	0.0599	0.019
AvgPPRec↑	0.4692	0.0169	0.4150	0.0136	0.002

Analisando a Tabela 20, é possível observar que os experimentos utilizando os métodos BR, LP e Rakel com o comitê de classificadores *Bagging* utilizando a abordagem do aprendizado supervisionado *versus* método BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado, a abordagem supervisionada apresentou melhor

desempenho em 72,22% dos casos (26/36). O Teste de *Wilcoxon* foi executado para verificar se há significância estatística nas diferenças apresentadas, o qual obteve 66,66% (24/36) dos casos.

Neste contexto, pode-se afirmar que, em geral, nos experimentos realizados a abordagem supervisionada apresentou os melhores resultados. Nas análises realizadas, em todos os casos foi verificada que existe relevância estatística significativa nas diferenças apresentadas.

8.7.4 Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores Boosting utilizando a abordagem do aprendizado supervisionado versus métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado

A Tabela 21 apresenta a média e o desvio padrão dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Boosting* utilizando a abordagem do aprendizado supervisionado *versus* método BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado. A Figura 29 ilustra a média dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Boosting* utilizando a abordagem do aprendizado supervisionado *versus* métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado.

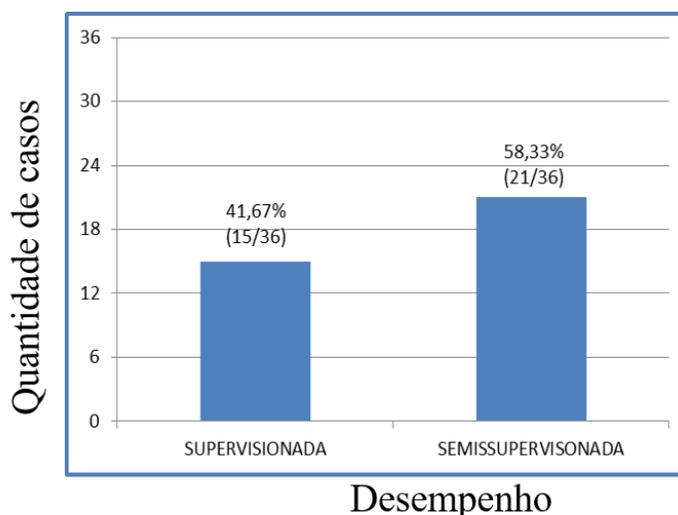


Figura 29 – Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Boosting* utilizando a abordagem do aprendizado supervisionado *versus* métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado.

Tabela 21 – Média dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Boosting* utilizando a abordagem do aprendizado supervisionado *versus* métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado.

Método	RAKEL		RAKEL		
Base	BIRDS				
Medida	Média	Desvio Padrão	Média	Desvio Padrão	WIL
HLoss↓	0.0413	0.0079	0.0422	0.0114	0.753
PPRec↑	0.7880	0.1523	0.6391	0.1435	0.022
PRec↑	0.5033	0.1317	0.3791	0.1317	0.111
F-M↑	0.4909	0.1035	0.4655	0.1226	0.757
1-Err↓	0.6869	0.0712	0.6840	0.0712	1.000
AvgPPRec↑	0.5835	0.0830	0.6102	0.1099	0.756
Base	EMOTIONS				
HLoss↓	0.1815	0.0165	0.2136	0.0114	0.023
PPRec↑	0.7293	0.0457	0.7397	0.0000	0.103
PRec↑	0.7209	0.0513	0.7561	0.0770	0.342
F-M↑	0.7021	0.0240	0.6463	0.0301	0.021
1-Err↓	0.2477	0.0381	0.2167	0.0352	0.179
AvgPPRec↑	0.8021	0.0234	0.8111	0.0000	0.343
Base	ENRON				
HLoss↓	0.0516	0.0034	0.0105	0.0017	0.002
PPRec↑	0.6266	0.0425	0.8351	0.0519	0.002
PRec↑	0.6007	0.0414	0.7921	0.0246	0.002
F-M↑	0.5328	0.0342	0.8049	0.0266	0.002
1-Err↓	0.2917	0.0737	0.1647	0.0359	0.021
AvgPPRec↑	0.6032	0.0470	0.8687	0.0221	0.002
Base	YEAST				
HLoss↓	0.2073	0.0154	0.0105	0.0017	0.002
PPRec↑	0.7063	0.0388	0.8351	0.0519	0.001
PRec↑	0.6191	0.0274	0.7921	0.0246	0.002
F-M↑	0.6324	0.0276	0.8049	0.0266	0.002
1-Err↓	0.2497	0.0505	0.1647	0.0359	0.116
AvgPPRec↑	0.7318	0.0286	0.8687	0.0221	0.002
Base	FLAGS				
HLoss↓	0.2416	0.0304	0.1564	0.0250	0.012
PPRec↑	0.7385	0.0442	0.4646	0.0494	0.013
PRec↑	0.7786	0.0367	0.3551	0.0539	0.315
F-M↑	0.7572	0.0311	0.3851	0.0322	0.501
1-Err↓	0.2061	0.0476	0.2230	0.0985	0.267
AvgPPRec↑	0.8184	0.0345	0.4150	0.0370	0.312
Base	CAL500				
HLoss↓	0.1815	0.0165	0.1515	0.0026	0.001
PPRec↑	0.7293	0.0457	0.4595	0.0082	0.002
PRec↑	0.7209	0.0513	0.4501	0.0372	0.001
F-M↑	0.7021	0.0239	0.4566	0.0264	0.001
1-Err↓	0.2477	0.0381	0.0567	0.0492	0.001
AvgPPRec↑	0.8021	0.0234	0.4281	0.0089	0.001

Analisando a Tabela 21, é possível observar que os experimentos utilizando os métodos BR, LP e Rakel com o comitê de classificadores *Boosting* utilizando a abordagem do aprendizado supervisionado *versus* método BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado, a abordagem semissupervisionada apresentou melhor desempenho em 58,33% dos casos (21/36). O Teste de *Wilcoxon* foi executado

para verificar se há significância estatística nas diferenças apresentadas, o qual obteve 63,88% (23/36) dos casos.

Neste contexto, pode-se afirmar que, em geral, nos experimentos realizados a abordagem semissupervisionada apresentou os melhores resultados. Nas análises realizadas, em todos os casos foi verificada que existe relevância estatística significativa nas diferenças apresentadas.

8.7.5 Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Stacking* utilizando a abordagem do aprendizado supervisionado versus métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado

A Tabela 22 apresenta a média e o desvio padrão dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Stacking* utilizando a abordagem do aprendizado supervisionado *versus* método BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado. A Figura 30 ilustra a média dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Stacking* utilizando a abordagem do aprendizado supervisionado *versus* métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado.

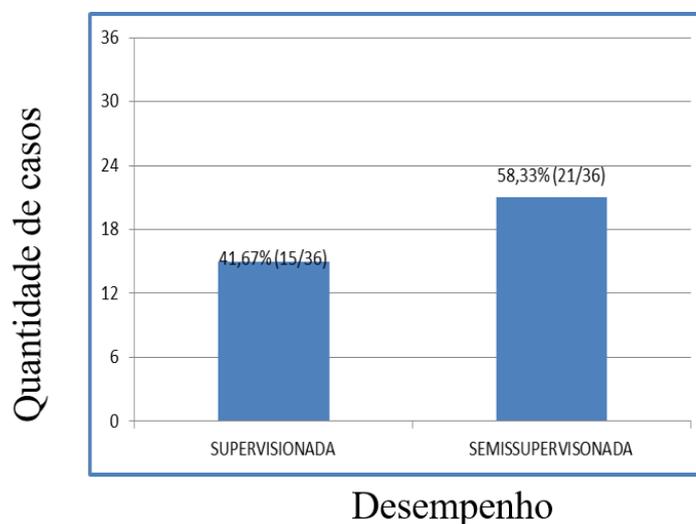


Figura 30 – Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Stacking* utilizando a abordagem do aprendizado supervisionado *versus* métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado.

Tabela 22 – Média dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Stacking* utilizando a abordagem do aprendizado supervisionado *versus* métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado.

Método	RAKEL		RAKEL		
Base	BIRDS				
Medida	Média	Desvio Padrão	Média	Desvio Padrão	WIL
HLoss↓	0.0433	0.0114	0.0391	0.0014	0.751
PPRec↑	0.6914	0.0696	0.5934	0.1409	0.013
PRec↑	0.6556	0.0796	0.3867	0.0973	0.001
F-M↑	0.4380	0.0760	0.4625	0.0979	0.092
1-Err↓	0.7025	0.0953	0.7243	0.0715	0.241
AvgPPRec↑	0.5999	0.0961	0.5837	0.0906	0.245
Base	EMOTIONS				
HLoss↓	0.1821	0.0100	0.2071	0.0227	0.003
PPRec↑	0.7108	0.0192	0.6834	0.0473	0.051
PRec↑	0.7007	0.0301	0.6938	0.0499	0.408
F-M↑	0.7053	0.0186	0.6545	0.0284	0.003
1-Err↓	0.2460	0.0324	0.2749	0.0411	0.048
AvgPPRec↑	0.8068	0.0177	0.7880	0.0288	0.022
Base	ENRON				
HLoss↓	0.0504	0.0019	0.0102	0.0014	0.001
PPRec↑	0.6647	0.0435	0.8544	0.0432	0.001
PRec↑	0.5094	0.0309	0.7963	0.0481	0.001
F-M↑	0.5227	0.0284	0.8121	0.0290	0.001
1-Err↓	0.3248	0.0603	0.2045	0.0362	0.001
AvgPPRec↑	0.6116	0.0345	0.8191	0.0324	0.001
Base	YEAST				
HLoss↓	0.2298	0.0168	0.0102	0.0014	0.001
PPRec↑	0.6225	0.0304	0.8544	0.0432	0.001
PRec↑	0.6277	0.0185	0.7963	0.0481	0.001
F-M↑	0.6201	0.0270	0.8121	0.0290	0.001
1-Err↓	0.2584	0.0497	0.2045	0.0362	0.005
AvgPPRec↑	0.7243	0.0177	0.8191	0.0324	0.001
Base	FLAGS				
HLoss↓	0.2470	0.0419	0.2571	0.1466	0.496
PPRec↑	0.7322	0.0599	1.1575	0.4596	0.014
PRec↑	0.7895	0.0297	0.8509	0.1929	0.217
F-M↑	0.7515	0.0467	0.9996	0.3228	0.017
1-Err↓	0.2155	0.0682	0.2353	0.1035	0.220
AvgPPRec↑	0.8156	0.0321	1.1896	0.3961	0.012
Base	CAL500				
HLoss↓	0.1411	0.1042	0.1697	0.0050	0.250
PPRec↑	0.4511	0.2117	0.4004	0.0235	0.242
PRec↑	0.3418	0.0272	0.4990	0.0244	0.001
F-M↑	0.3931	0.1031	0.3665	0.0071	0.003
1-Err↓	0.8000	0.2422	0.2227	0.0979	0.001
AvgPPRec↑	0.2440	0.0666	0.3893	0.0053	0.001

Analisando a Tabela 22, é possível observar que os experimentos utilizando os métodos BR, LP e Rakel com o comitê de classificadores *Stacking* utilizando a abordagem do aprendizado supervisionado *versus* método BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado, a abordagem semissupervisionada apresentou melhor desempenho em 58,33% dos casos (21/36). O Teste de *Wilcoxon* foi executado

para verificar se há significância estatística nas diferenças apresentadas, o qual obteve 69,44% (25/36) dos casos.

Neste contexto, pode-se afirmar que, em geral, nos experimentos realizados a abordagem semissupervisionada apresentou os melhores resultados. Nas análises realizadas, em todos os casos foi verificada que existe relevância estatística significativa nas diferenças apresentadas.

8.8 CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentados os resultados experimentais obtidos pelos métodos de classificação multirrótulo utilizando as abordagens de aprendizado supervisionado e semissupervisionado. Nos experimentos foram utilizando os métodos BR, LP e Raket com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*.

Ao analisar os resultados segundo a perspectiva dos comitês de classificadores *Bagging*, *Boosting* e *Stacking* observou-se que, na maioria dos casos, o comitê de classificadores *Bagging* apresentou os melhores resultados em todos os experimentos utilizando a abordagem semissupervisionada. Foi verificada relevância estatística na maioria dos casos analisados. Em seguida, foi realizada uma análise de uma perspectiva geral, possibilitando uma comparação das abordagens supervisionada e semissupervisionada dos métodos BR, LP e Raket com os comitês de classificadores *Bagging*, *Boosting* e *Stacking*. Ao analisar separadamente os métodos BR com a abordagem supervisionada *versus* BR com a abordagem semissupervisionada utilizando o comitê de classificadores *Bagging*, verificou-se que a abordagem semissupervisionada apresentou melhor desempenho em 52,77% dos casos (19/36). Ademais, ao analisar separadamente os métodos LP com a abordagem supervisionada *versus* LP com a abordagem semissupervisionada utilizando o comitê de classificadores *Bagging*, verificou-se que a abordagem supervisionada apresentou melhor desempenho em 66,66 % dos casos (24/36). Ao analisar o método métodos Raket com a abordagem supervisionada utilizando o comitê de classificadores *Boosting versus* Raket com a abordagem semissupervisionada utilizando o comitê de classificadores *Bagging*, verificou-se que a abordagem supervisionada apresentou melhor desempenho em 55,55% dos casos (20/36).

Ao analisar os resultados segundo a perspectiva dos métodos BR, LP e Raket, observou-se que, na maioria dos casos, o método Raket apresentou os melhores resultados. Observando os melhores resultados dos métodos BR, LP e Raket com o comitê de classificadores *Bagging* utilizando a abordagem do aprendizado supervisionado *versus* método BR, LP e Raket utilizando a abordagem do aprendizado semissupervisionado, pode verificar que, a abordagem supervisionada apresentou melhor desempenho em 72,22% dos casos (26/36). Ao analisar os melhores resultados dos métodos BR, LP e Raket com o comitê de classificadores *Boosting* utilizando a abordagem do aprendizado supervisionado *versus* métodos BR, LP e Raket utilizando a abordagem do aprendizado

semissupervisionado, verificou-se que a abordagem semissupervisionada apresentou melhor desempenho em 58,33% dos casos (21/36). Em relação aos melhores resultados do método BR, LP e Rakel com o comitê de classificadores *Stacking* utilizando a abordagem do aprendizado supervisionado *versus* método BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado, verificou-se que a abordagem semissupervisionada apresentou melhor desempenho em 58,33% dos casos (21/36).

9 CONCLUSÃO

Neste trabalho foram propostas várias estratégias de construções de comitês de classificadores multirrótulo utilizando os métodos de transformação do problema BR, LP e Rakel com a abordagem do aprendizado supervisionado e semissupervisionado, conforme descritos a seguir:

Abordagem supervisionada:

- Método de transformação do problema BR com o comitê de classificadores *Bagging*;
- Método de transformação do problema BR com o comitê de classificador *Boosting*;
- Método de transformação do problema BR com o comitê de classificadores *Stacking*;
- Método de transformação do problema LP com o comitê de classificadores *Bagging*;
- Método de transformação do problema LP com o comitê de classificadores *Bagging*;
- Método de transformação do problema LP com o comitê de classificadores *Stacking*;
- Método de transformação do problema Rakel com o comitê de classificadores *Bagging* ;
- Método de transformação do problema Rakel com o comitê de classificadores *Boosting*;
- Método de transformação do problema Rakel com o comitê de classificadores *Stacking*.

Abordagem semissupervisionada:

- Método de transformação do problema BR com o comitê de classificadores *Bagging*;
- Método de transformação do problema BR com o comitê de classificadores *Boosting*;
- Método de transformação do problema BR com o comitê de classificadores *Stacking*;
- Método de transformação do problema LP com o comitê de classificadores *Bagging*;
- Método de transformação do problema LP com o comitê de classificadores *Boosting*;
- Método de transformação do problema LP com o comitê de classificadores *Stacking*;
- Método de transformação do problema Rakel com o comitê de classificadores *Bagging*;

- Método de transformação do problema Rakel com o comitê de classificadores *Boosting*
- Método de transformação do problema Rakel com o comitê de classificadores *Stacking*.

Em cada uma das fases foi realizada uma análise comparativa dos métodos de classificação supervisionada e semissupervisionada. Ao fim das análises utilizando a abordagem supervisionada verificou-se que, na maioria das estratégias o comitê de classificadores *Bagging* apresentou os melhores resultados em 66,66% (02/03). Em relação ao método de transformação de problema o método Rakel apresentou os melhores resultados em 66,66% das estratégias (02/03).

Nas análises utilizando o abordagem semissupervisionada verificou-se que, na maioria dos casos o comitê de classificadores *Bagging* apresentou os melhores resultados em 66,66% das estratégias (02/03). Em relação ao método de transformação de problema o método Rakel apresentou os melhores resultados em 100% das estratégias de construções de comitês (03/03).

Analisar comparativamente os resultados obtidos por comitês de classificadores multirrótulo construídos através da abordagem supervisionada e da abordagem semissupervisionada multidescrição, verificou-se que, a abordagem semissupervisionada apresentou resultados semelhantes (50% das estratégias (03/06)) aos obtidos pela abordagem supervisionada (50% das estratégias (03/06)).

Assim, é possível afirmar que a utilização da abordagem semissupervisionado multidescrição mostrou-se bastante interessante, uma vez que, embora parte dos exemplos utilizados para o treinamento dos classificadores tenham sido rotulados através de um processo automático, no geral, o desempenho dos classificadores não foi deteriorado.

No entanto, como a área de aprendizado de máquina semissupervisionado é relativamente nova, muitos estudos ainda precisam ser realizados a fim de compreender melhor o uso de comitês classificadores e as suas estratégias de construções. Desse modo, este trabalho traz contribuições às pesquisas nas áreas de classificação de dados multirrótulo e aprendizado semissupervisionado multidescrição com a utilização de comitês classificadores.

9.0.1 Trabalhos Futuros

A seguir são apresentados alguns pontos que podem ser desenvolvidos em futuras pesquisas envolvendo as estratégias de construções de comitês de classificadores multirrótulo:

- Utilizar bases de dados maiores, com mais atributos, rótulos e instâncias, uma vez que no estudo realizado neste trabalhos foram utilizadas bases de dados pequenas;

- Combinar a classificação multirrótulo e hierárquica multirrótulo e o aprendizado semissupervisionado com comitês de classificadores utilizando outros métodos de classificação existentes na literatura;
- Realizar um estudo comparativo entre o aprendizado semissupervisionado multidescrição utilizando comitês de classificadores *versus* aprendizado semissupervisionado descrição simples utilizando comitês de classificadores.

9.0.2 Produções Científicas

Os resultados obtidos ao longo deste trabalho foram submetidos e/ou aprovados nos seguintes eventos científicos:

- **Título:** COMITÊS DE CLASSIFICADORES MULTIRRÓTULO EM DIFERENTES DOMÍNIOS DE APLICAÇÕES.

Veículo: LIVRO DE RESUMOS DA III JORNADA ACADÊMICA E II MOSTRA DE INICIAÇÃO CIENTÍFICA JÚNIOR E I SEMANA DE EXTENSÃO, 1ª ed. ISBN 978-85-66530-70-4.

- **Título:** COMITÊS DE CLASSIFICADORES MULTIRRÓTULO EM DIFERENTES DOMÍNIOS DE APLICAÇÕES. III JORNADA ACADÊMICA E II MOSTRA DE INICIAÇÃO CIENTÍFICA JÚNIOR E I SEMANA DE EXTENSÃO.

Veículo: Revista de Educação Profissional e Tecnológica - ‡ eduproftec

Situação: Aprovado: Aguardando *Qualis* da Capes para a 1ª edição da revista.

- **Título:** ESTRATÉGIAS DE CONSTRUÇÃO DE COMITÊS DE CLASSIFICADORES NO APRENDIZADO SEMISSUPERVISIONADO MULTIDESCRIÇÃO.

Veículo: Revista de Informática Teórica e Aplicada - RITA

Situação: Submetido

REFERÊNCIAS

- AIOLLI, F. P. F.; SPERDUTI, A. Speeding up the solution of multilabel problems with support vector machines. In: *Proceedings of the Joint 13th International Conference on Artificial Neural Networks and 10th International Conference on Neural Information Processing*. [S.l.: s.n.], 2004. p. 118–121. Citado na página 36.
- ASSIS, M. S. d. *Classificação multirrótulo com aprendizado semissupervisionado: uma análise multivisão de dados*. Dissertação (Mestrado) — Brasil, 2016. Citado 3 vezes nas páginas 24, 25 e 47.
- BARUTCUOGLU, Z.; SCHAPIRE, R. E.; TROYANSKAYA, O. G. Hierarchical multi-label prediction of gene function. *Bioinformatics*, Oxford Univ Press, v. 22, n. 7, p. 830–836, 2006. Citado na página 29.
- BLOCKEEL, H. et al. *Decision trees for hierarchical multilabel classification: A case study in functional genomics*. [S.l.]: Springer, 2006. Citado na página 29.
- BLUM, A.; MITCHELL, T. Combining labeled and unlabeled data with co-training. In: ACM. *Proceedings of the eleventh annual conference on Computational learning theory*. [S.l.], 1998. p. 92–100. Citado na página 24.
- BOUTELL, M. R. et al. Learning multi-label scene classification. *Pattern recognition*, Elsevier, v. 37, n. 9, p. 1757–1771, 2004. Citado na página 29.
- BRAGA, Í. A. *Aprendizado semissupervisionado multidescrição em classificação de textos*. Tese (Doutorado) — Universidade de São Paulo, 2010. Citado na página 46.
- BREIMAN, L. Bagging predictors. *Machine learning*, Springer, v. 24, n. 2, p. 123–140, 1996. Citado na página 40.
- BRINKER, K.; HÜLLERMEIER, E. Case-based multilabel ranking. In: *IJCAI*. [S.l.: s.n.], 2007. p. 702–707. Citado na página 36.
- BROWN, G. et al. Diversity creation methods: a survey and categorisation. *Information Fusion*, Elsevier, v. 6, n. 1, p. 5–20, 2005. Citado na página 39.
- CAMPOS, C. M. et al. Comitê de classificadores em bases de dados transacionais desbalanceadas com seleção de características baseada em padrões minerados. Universidade Federal de Juiz de Fora (UFJF), 2016. Citado na página 27.
- CANUTO, A. M.; SANTOS, A. M.; VARGAS, R. R. Ensembles of artmap-based neural networks: an experimental study. *Applied Intelligence*, Springer, v. 35, n. 1, p. 1–17, 2011. Citado na página 27.
- CANUTO, A. M. d. P. *Combining neural networks and fuzzy logic for applications in character recognition*. Tese (Doutorado) — University of Kent at Canterbury, 2001. Citado na página 40.
- CLARE, A.; KING, R. D. Knowledge discovery in multi-label phenotype data. In: *Principles of data mining and knowledge discovery*. [S.l.]: Springer, 2001. p. 42–53. Citado 3 vezes nas páginas 34, 35 e 51.

- COSTA, E. d. P. *Investigação de técnicas de classificação hierárquica para problemas de bioinformática*. Tese (Doutorado) — Universidade de São Paulo, 2008. Citado 2 vezes nas páginas 22 e 27.
- COSTA, E. P. et al. Comparing several approaches for hierarchical classification of proteins with decision trees. In: *Advances in Bioinformatics and Computational Biology*. [S.l.]: Springer, 2007. p. 126–137. Citado na página 37.
- COVER, T. M.; HART, P. E. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, IEEE, v. 13, n. 1, p. 21–27, 1967. Citado na página 36.
- CRAMMER, K.; SINGER, Y. A family of additive online algorithms for category ranking. *The Journal of Machine Learning Research*, JMLR. org, v. 3, p. 1025–1058, 2003. Citado na página 35.
- EFRON, B.; TIBSHIRANI, R.-J. *Introduction to the bootstrap*. [an]. Chapman and Hall, 1993. Citado na página 40.
- ELISSEEFF, A.; WESTON, J. A kernel method for multi-labelled classification. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2001. p. 681–687. Citado na página 35.
- FOGELMAN-SOULIÉ, F. et al. Large-scale semi-supervised learning. *Mining Massive Data Sets for Security: Advances in Data Mining, Search, Social Networks and Text Mining, and Their Applications to Security*, IOS Press, v. 19, p. 62, 2008. Citado na página 23.
- GHAMRAWI, N.; MCCALLUM, A. Collective multi-label classification. In: *ACM. Proceedings of the 14th ACM international conference on Information and knowledge management*. [S.l.], 2005. p. 195–200. Citado na página 38.
- GHINOZZI, G. G. Aprendizado semissupervisionado aplicado ao problema de valores ausentes. 2012. Citado 2 vezes nas páginas 23 e 27.
- GODBOLE, S.; SARAWAGI, S. Discriminative methods for multi-labeled classification. In: *Advances in Knowledge Discovery and Data Mining*. [S.l.]: Springer, 2004. p. 22–30. Citado na página 36.
- GUERMEUR, Y.; PAUGAM-MOISY, H.; GALLINARI, P. Multivariate linear regression on classifier outputs: a capacity study. In: *ICANN 98*. [S.l.]: Springer, 1998. p. 693–698. Citado na página 44.
- JL, C.; MA, S. Combinations of weak classifiers. *Neural Networks, IEEE Transactions on*, IEEE, v. 8, n. 1, p. 32–42, 1997. Citado na página 43.
- KATAKIS, I.; TSOUMAKAS, G.; VLAHAVAS, I. Multilabel text classification for automated tag suggestion. *ECML PKDD discovery challenge*, v. 75, 2008. Citado 2 vezes nas páginas 28 e 29.
- KROGEL, M.-A.; SCHEFFER, T. Multi-relational learning, text mining, and semi-supervised learning for functional genomics. *Machine Learning*, Springer, v. 57, n. 1-2, p. 61–81, 2004. Citado na página 25.
- KUNCHEVA, L. I. *Combining pattern classifiers: methods and algorithms*. [S.l.]: John Wiley & Sons, 2004. Citado 3 vezes nas páginas 39, 40 e 43.

- LI, T.; OGIHARA, M. Detecting emotion in music. In: *ISMIR*. [S.l.: s.n.], 2003. v. 3, p. 239–240. Citado na página 29.
- LI, T.; OGIHARA, M. Toward intelligent music information retrieval. *Multimedia, IEEE Transactions on*, IEEE, v. 8, n. 3, p. 564–574, 2006. Citado na página 29.
- LIMA, C. A. de M. *Comitê de Máquinas: uma abordagem unificada empregando máquinas de vetores-suporte*. Tese (Doutorado) — Universidade Estadual de Campinas, 2005. Citado 3 vezes nas páginas 40, 41 e 43.
- LORENA, A. C.; CARVALHO, A. C. de. Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 43–67, 2007. Citado na página 21.
- LUO, X.; ZINCIR-HEYWOOD, A. N. Evaluation of two systems on multi-class multi-label document classification. In: *Foundations of Intelligent Systems*. [S.l.]: Springer, 2005. p. 161–169. Citado 2 vezes nas páginas 36 e 37.
- MAHALANOBIS, P. C. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta)*, v. 2, p. 49–55, 1936. Citado na página 22.
- MANCHEGO, F. E. A. *Anotação automática semissupervisionada de papéis semânticos para o português do Brasil*. Tese (Doutorado) — Universidade de São Paulo, 2013. Citado na página 22.
- MATSUBARA, E. T.; MONARD, M. C.; BATISTA, G. E. Multi-view semi-supervised learning: An approach to obtain different views from text datasets. In: *LAPTEC*. [S.l.: s.n.], 2005. p. 97–104. Citado 3 vezes nas páginas 16, 22 e 24.
- MCCALLUM, A. Multi-label text classification with a mixture model trained by em. In: *AAAI'99 workshop on text learning*. [S.l.: s.n.], 1999. p. 1–7. Citado na página 37.
- MERZ, C. J. Using correspondence analysis to combine classifiers. *Machine Learning*, Springer, v. 36, n. 1-2, p. 33–58, 1999. Citado na página 43.
- METZ, J. *Abordagens para aprendizado semissupervisionado multirrotulo e hierárquico*. Tese (Doutorado) — Universidade de São Paulo, 2011. Citado 2 vezes nas páginas 32 e 33.
- MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. *Sistemas Inteligentes-Fundamentos e Aplicações*, v. 1, p. 1, 2003. Citado na página 16.
- MULAN. *Mulan: A java library for multi-label learning*. <http://mulan.sourceforge.net/>, 2017. Citado na página 56.
- NASIERDING, G.; TSOUMAKAS, G.; KOUZANI, A. Z. Clustering based multi-label classification for image annotation and retrieval. In: *IEEE. Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*. [S.l.], 2009. p. 4514–4519. Citado 2 vezes nas páginas 33 e 37.
- NORVIG, P.; RUSSELL, S. *Inteligência Artificial, 3ª Edição*. [S.l.]: Elsevier Brasil, 2014. Citado na página 19.
- OLIVEIRA, E. et al. Using a probabilistic neural network for a large multi-label problem. In: *IEEE. Neural Networks, 2008. SBRN'08. 10th Brazilian Symposium on*. [S.l.], 2008. p. 195–200. Citado na página 35.

- OPITZ, D.; MACLIN, R. Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, p. 169–198, 1999. Citado na página 42.
- PESTIAN, J. P. et al. A shared task involving multi-label classification of clinical free text. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. *Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing*. [S.l.], 2007. p. 97–104. Citado na página 29.
- QI, G.-J. et al. Correlative multi-label video annotation. In: ACM. *Proceedings of the 15th international conference on Multimedia*. [S.l.], 2007. p. 17–26. Citado na página 29.
- REZENDE, S. O. *Sistemas inteligentes: fundamentos e aplicações*. [S.l.]: Editora Manole Ltda, 2003. Citado na página 21.
- RODRIGUES, F. M. Uso de confiabilidade na rotulação de exemplos em problemas de classificação multirrótulo com aprendizado semissupervisionado. Universidade Federal do Rio Grande do Norte, 2014. Citado 5 vezes nas páginas 27, 28, 31, 33 e 46.
- SANTOS, A. d. M. Investigando a combinação de técnicas de aprendizado semissupervisionado e classificação hierárquica multirrótulo. Universidade Federal do Rio Grande do Norte, 2012. Citado 6 vezes nas páginas 23, 29, 31, 34, 48 e 53.
- SANTOS, S. P. d. Aplicação de sistemas multi-classificadores no diagnóstico de falhas em motores de indução trifásicos. Universidade Federal do Rio Grande do Norte, 2009. Citado 3 vezes nas páginas 41, 42 e 44.
- SCHAPIRE, R. E. The strength of weak learnability. *Machine learning*, Springer, v. 5, n. 2, p. 197–227, 1990. Citado na página 40.
- SCHAPIRE, R. E.; SINGER, Y. Boostexter: A boosting-based system for text categorization. *Machine learning*, Springer, v. 39, n. 2, p. 135–168, 2000. Citado na página 34.
- SEEWALD, A. K. How to make stacking better and faster while also taking care of an unknown weakness. In: MORGAN KAUFMANN PUBLISHERS INC. *Proceedings of the nineteenth international conference on machine learning*. [S.l.], 2002. p. 554–561. Citado na página 44.
- SEEWALD, A. K. *Towards understanding stacking: studies of a general ensemble learning scheme*. [S.l.]: na, 2003. Citado na página 45.
- SILVA, A. R. d.; CARVALHO, A. C. P. d. L. et al. O uso seletivo de classificadores binários na solução de problemas multirrótulos. *Encontro Nacional de Inteligência Artificial e Computacional, XII*, Universidade Federal do Rio Grande do Norte–UFRN, 2015. Citado 2 vezes nas páginas 27 e 48.
- SILVA, P. N. D. Classificação multirrótulo em cadeia: Novas abordagens. 2014. Citado 2 vezes nas páginas 30 e 47.
- SNOEK, C. G. et al. The challenge problem for automated detection of 101 semantic concepts in multimedia. In: ACM. *Proceedings of the 14th annual ACM international conference on Multimedia*. [S.l.], 2006. p. 421–430. Citado na página 29.

- SPECHT, D. F. Probabilistic neural networks. *Neural networks*, Elsevier, v. 3, n. 1, p. 109–118, 1990. Citado na página 35.
- SPYROMITROS, E.; TSOUMAKAS, G.; VLAHAVAS, I. An empirical study of lazy multilabel classification algorithms. In: *Artificial Intelligence: Theories, Models and Applications*. [S.l.]: Springer, 2008. p. 401–406. Citado na página 36.
- STREICH, A. P.; BUHMANN, J. M. Classification of multi-labeled data: A generative approach. In: *Machine learning and knowledge discovery in databases*. [S.l.]: Springer, 2008. p. 390–405. Citado na página 37.
- THABTAH, F. et al. Mmac: A new multi-class, multi-label associative classification approach. In: *IEEE. Data Mining, 2004. ICDM'04. Fourth IEEE International Conference on*. [S.l.], 2004. p. 217–224. Citado na página 38.
- TING, K. M.; WITTEN, I. H. Issues in stacked generalization. *J. Artif. Intell. Res.(JAIR)*, v. 10, p. 271–289, 1999. Citado na página 44.
- TROHIDIS, K. et al. Multi-label classification of music into emotions. In: *ISMIR*. [S.l.: s.n.], 2008. v. 8, p. 325–330. Citado na página 29.
- TSOUMAKAS, G.; KATAKIS, I. Multi-label classification: An overview. *Dept. of Informatics, Aristotle University of Thessaloniki, Greece*, 2007. Citado 2 vezes nas páginas 30 e 37.
- TSOUMAKAS, G.; KATAKIS, I.; VLAHAVAS, I. Effective and efficient multilabel classification in domains with large number of labels. In: *Proc. ECML/PKDD 2008 Workshop on Mining Multidimensional Data (MMD'08)*. [S.l.: s.n.], 2008. p. 30–44. Citado 4 vezes nas páginas 28, 29, 30 e 33.
- TSOUMAKAS, G.; KATAKIS, I.; VLAHAVAS, I. Mining multi-label data. In: *Data mining and knowledge discovery handbook*. [S.l.]: Springer, 2010. p. 667–685. Citado na página 38.
- TSOUMAKAS, G.; KATAKIS, I.; VLAHAVAS, I. Random k-labelsets for multilabel classification. *Knowledge and Data Engineering, IEEE Transactions on*, IEEE, v. 23, n. 7, p. 1079–1089, 2011. Citado 2 vezes nas páginas 30 e 33.
- UEDA, N.; SAITO, K. Parametric mixture models for multi-labeled text. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2003. p. 721–728. Citado na página 37.
- WEKA. Data mining software in java. 2017. Citado na página 56.
- WIECZORKOWSKA, A.; SYNAK, P.; RAŚ, Z. W. Multi-label classification of emotions in music. In: *Intelligent Information Processing and Web Mining*. [S.l.]: Springer, 2006. p. 307–315. Citado 2 vezes nas páginas 29 e 36.
- WITTEN, I. H.; FRANK, E. *Data Mining: Practical machine learning tools and techniques*. [S.l.]: Morgan Kaufmann, 2005. Citado na página 43.
- WOLPERT, D. H. Stacked generalization. *Neural networks*, Elsevier, v. 5, n. 2, p. 241–259, 1992. Citado na página 43.
- YANG, S.; KIM, S.-K.; RO, Y. M. Semantic home photo categorization. *Circuits and Systems for Video Technology, IEEE Transactions on*, IEEE, v. 17, n. 3, p. 324–335, 2007. Citado na página 29.

YAROWSKY, D. Unsupervised word sense disambiguation rivaling supervised methods. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*. [S.l.], 1995. p. 189–196. Citado na página 16.

ZHANG, M.-L. MI-rbf: Rbf neural networks for multi-label learning. *Neural Processing Letters*, Springer, v. 29, n. 2, p. 61–74, 2009. Citado na página 35.

ZHANG, M.-L.; ZHOU, Z.-H. MI-knn: A lazy learning approach to multi-label learning. *Pattern recognition*, Elsevier, v. 40, n. 7, p. 2038–2048, 2007. Citado 2 vezes nas páginas 29 e 36.

ZHANG, Y.; BURER, S.; STREET, W. N. Ensemble pruning via semi-definite programming. *The Journal of Machine Learning Research*, JMLR. org, v. 7, p. 1315–1338, 2006. Citado 3 vezes nas páginas 29, 34 e 35.