

# Estratégias de Construções de Comitês de Classificadores Multirrótulos no Aprendizado Semissupervisionado Multidescrição

## Strategies of Construction of Committees of Multi-Label Classifiers in the Semi-Supervised Learning Multi-description

Wilamis Kleiton Nunes da Silva <sup>1</sup>  
Araken de Medeiros Santos <sup>1 2</sup>

*Data de submissão: 26/07/2017, Data de aceite: 25/10/2017*

**Resumo:** Classificação multirrótulo é um problema de aprendizado supervisionado no qual um objeto pode estar associado a múltiplas classes. Dentre os diferentes métodos de classificação multirrótulo destacam-se os métodos BR (*Binary Relevance*), LP (*Label Powerset*) e RAKEL (*RANdom k-labELsets*). O trabalho realizou um estudo sobre as construções de comitês de classificadores multirrótulos construídos através da aplicação de técnicas de aprendizado semissupervisionado multidescrição. Os comitês de classificadores utilizados nos experimentos foram o *Bagging*, *Boosting* e *Stacking*; como métodos de transformação do problema utilizamos os métodos BR, LP e RAKEL; na classificação multirrótulo semissupervisionada multidescrição foi utilizado o *Co-Training*; foram aplicados cinco diferentes algoritmos como classificadores base: k-NN (k Vizinhos Mais Próximos), J48 (Algoritmo de Indução de Árvores de Decisão), SVM (Máquinas de Vetores Suporte), NB (*Naive Bayes*) e o JRip (*Extended Repeated Incremental Pruning*). Todos os experimentos utilizaram a metodologia de validação cruzada com 10 grupos (*10-fold Cross-Validation*) e o *framework* MULAN, o qual é implementado utilizando o WEKA. Para os tamanhos dos comitês de classificadores adotamos os valores 3, 5, 7 e 9. Para a análise dos resultados foi utilizado o teste estatístico de *Wilcoxon*. Ao final das análises experimentais, verificou-se que a abordagem semissupervisionada apresentou resultados competitivos em relação ao aprendizado supervisionado, uma vez que as duas abordagens utilizadas apresentaram resultados estatisticamente semelhantes.

---

<sup>1</sup>Programa de Pós-graduação em Ciência da Computação, UERN/UFERSA, Av. Francisco Mota, 572 - Bairro Costa e Silva, Mossoró - RN, Brasil. Caixa Postal 137

{[wilamiskleiton@yahoo.com.br](mailto:wilamiskleiton@yahoo.com.br)}

<sup>2</sup>[araken@ufersa.edu.br](mailto:araken@ufersa.edu.br)

**Abstract:** Multi-label classification is a supervised learning problem in which an object can be associated with multiple classes. Among the different methods of multi-label classification are the BR (Binary Relevance), LP (Label Powerset) and RAKEL (RANdom k-labELsets). The work conducted a study on the construction of multi-label classifier committees built through the application of multi-description semi-supervised learning techniques. The classifier committees used in the experiments were Bagging, Boosting and Stacking; BR, LP and Rakel were used as Problem Transformation Methods, and in the multi-description semi-supervised multi-label classification we used Co-Training; five different algorithms were applied as base classifiers: k-NN (k Nearest Neighbors), J48 (Decision Trees Induction Algorithm), SVM (Supporting Vector Machines), NB (Naive Bayes) and JRip (Extended Repeated Incremental Pruning). All experiments were performed through the cross validation methodology with 10 groups (10-fold Cross-Validation) and the MULAN framework, which is implemented using WEKA. For the sizes of the classifier committees, the following values were adopted: 3, 5, 7 and 9. For the analysis of the results it was used the statistical test of Wilcoxon. At the end of experimental analyzes it was verified that the semi-supervised approach had competitive results in relation to supervised learning, since these two approaches used in the experiments presented statistically similar results.

**Palavras-chave:** Aprendizado de máquina, aprendizado semissupervisionado multidescrição, classificação multirrotulo, comitês de classificadores.

**Keywords:** Machine learning, semi-supervised learning multi-description, multi-label classification, classifier committees.

## 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 [1]. 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 primeiro e o segundo diz respeito à forma como é realizado o processo de generalização do conhecimento [2].

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. Nos últimos anos, a comunidade científica vem estudando uma terceira categoria de aprendizado, surgida através da junção do aprendizado supervisionado com o aprendizado não supervisionado. A essa categoria de aprendizado dá-se o nome de aprendizado semissupervisionado [3].

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 destaca-se o *Co-training* que é uma poderosa abordagem ao aprendizado semissupervisionado baseada na ideia da multidescrição [4].

Nesta abordagem, os exemplos de treinamento são descritos por dois ou mais conjuntos de atributos disjuntos, ou seja, através de diferentes descrições [3]. 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 [5].

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, tais como em [6–10]. Todavia, o aprendizado semissupervisionado tem tomado papel de destaque quando se trata de classificação multirrótulo, por se tratar do método que está tendo resultados positivos quando o conjunto de exemplos de treinamento possui poucos exemplos rotulados e muitos exemplos não rotulados. Com o intuito de obter sistemas cada vez mais eficazes, têm-se utilizado cada vez mais comitês de classificadores em diversas tarefas, tais como na classificação para detecção de intrusão em redes de computadores [11]; predição de rendimento escolar [12]; identificação da dislalia em crianças [13]; detecção de opinião em mensagens curtas [14]; diagnóstico de tuberculose por meio de imagens radiológicas [15].

Os classificadores que compõem um comitê de classificadores podem ser do mesmo tipo, com variações no conjunto de treinamento, parâmetros iniciais e/ou configuração ou de diferentes tipos, utilizando diferentes algoritmos de aprendizado de máquina. 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. Apesar de tanto a classificação multirrótulo, quanto o aprendizado semissupervisionado e a utilização de comitês de classificadores estarem apresentando resultados favoráveis à sua utilização, até o momento nenhum trabalho foi realizado na literatura unindo as vantagens das três áreas. Diante desse fato, como não foram encontradas pesquisas que exploram as construções de comitês classificadores multirrótulo no aprendizado semissupervisionado multidescrição, esse passou a ser o foco do trabalho.

## 2 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, cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas de aprendizado capazes de adquirir conhecimento de forma automática [1]. 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 [16].

1. **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].
2. **No 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. [16] destaca que o algoritmo de aprendizado de máquina não supervisionado aprende a representar (ou agrupar) as entradas submetidas, segundo medidas de similaridade.

3. **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 [17].

De acordo com [5], 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. Dentre os algoritmos de aprendizado semissupervisionado mais utilizados na literatura está o *Co-training*.

O *Co-training* [4], 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 [3]. 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 [5]. [4] mostram que o *Co-training* assume inicialmente que:

- (a) Os recursos podem ser divididos em dois grupos ou subconjuntos;
- (b) Cada subconjunto de recurso é suficiente para formar um bom classificador;
- (c) 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 [18].

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  $L_{D1}$  e  $L_{D2}$ , nos quais representam o subconjunto de exemplos rotulados da descrição  $D_1$  e  $D_2$ , respectivamente, sendo  $L = L_{D1} \cup L_{D2}$  e  $L_{D1} \cap L_{D2} = \phi$ . Da mesma forma, os exemplos do subconjunto de exemplos não rotulados  $U$ , por sua vez, devem ser subdivididos em  $U_{D1}$  e  $U_{D2}$ , sendo  $U = U_{D1} \cup U_{D2}$  e  $U_{D1} \cap U_{D2} = \phi$  [18]. [19]. A Figura 1 mostra como o conjunto de exemplos  $E$  deve ser dividido em subconjuntos de exemplos rotulados  $L$  e não rotulados  $U$ .

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 [5].

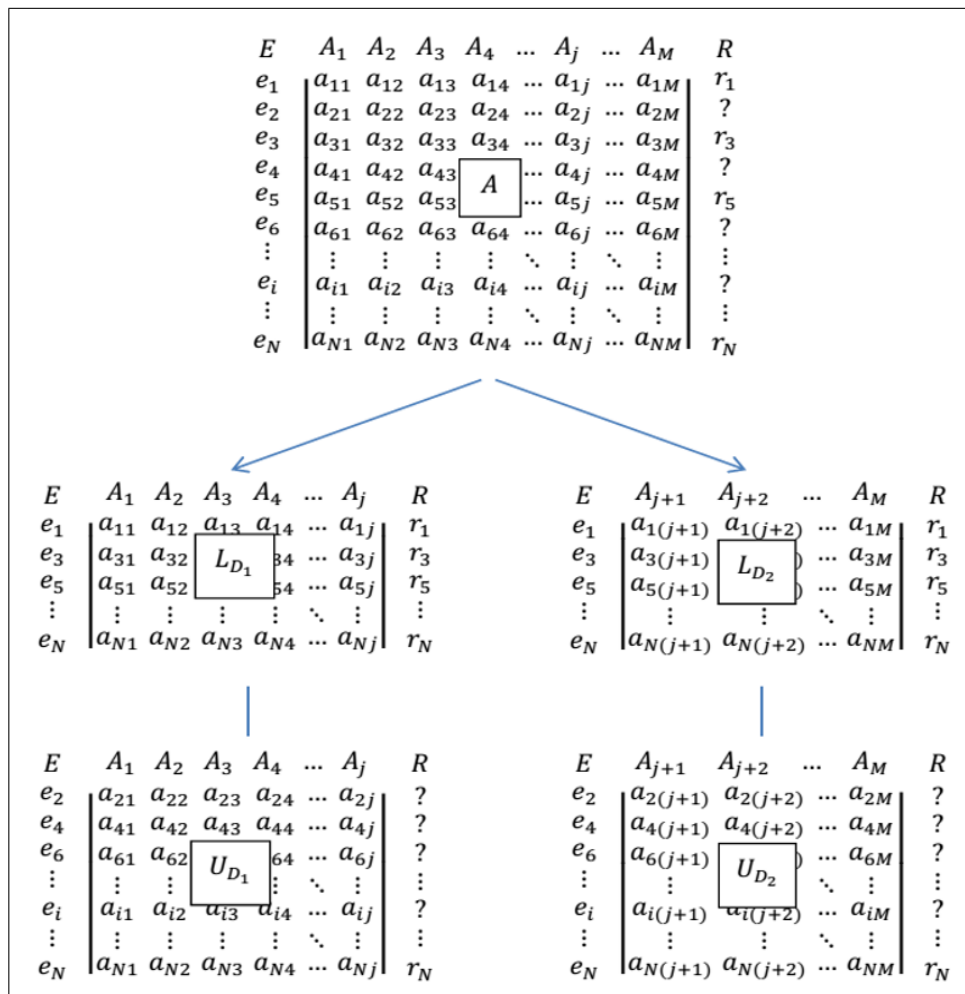


Figura 1. Abordagem do *Co-training*.  
[18]

### 3 Classificação de Dados Multirrótulo

A classificação de dados multirrótulo tem recebido atenção da comunidade de pesquisa, devido ao fato de não se enquadrarem em um modelo de classificação tradicional. Essa classe de problemas, que permite que múltiplos rótulos sejam associados ao mesmo objeto, é conhecida como classificação multirrótulo [20].

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 tem atraído a atenção de grande parte da comunidade científica, motivado principalmente pelo grande aumento no número de aplicações, tais como para classificação de proteínas [21]; mineração de mídias sociais

[22]; avaliação semântica da integração da gestão de riscos de segurança em documentos de software da administração pública [23]; função genômica [24, 25]; categorização de músicas através de emoções [26–29] e marketing direcionado [30].

A classificação multirrótulo pode ser formalmente definida como a seguir: 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 [31]. A Tabela 1 apresenta um exemplo de classificação multirrótulo no qual está associada à 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 um 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 [20].

**Tabela 1.** Tipos de classificação, utilizando como exemplo a música *Jazz n' Blues* da artista *Amy Winehouse*.

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

Fonte:[20]

A Tabela 2 apresenta um conjunto de exemplos rotulados com mais de um rótulo representando uma notícia de um jornal, o rótulo  $\lambda_1$  = esporte,  $\lambda_2$ = música,  $\lambda_3$ = cinema e  $\lambda_4$ = educação.

**Tabela 2.** Conjunto de Dados Multirrótulo

Exemplo	Rótulos
1	$\{\lambda_{1,4}\}$
2	$\{\lambda_{3,4}\}$
3	$\{\lambda_1\}$
4	$\{\lambda_{2,3,4}\}$

Fonte: [5]

### 3.1 Métodos de Transformação do Problema

Nos métodos de transformação do problema, o 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 BR (*Binary Relevance*), LP (*Label Powerset*) e RAKEL (*RAndom k-LabELsets*) [32].

### 3.2 Binary Relevance

Em problemas multirrótulo, o método BR (*Binary Relevance*) é um dos métodos de transformação mais popular [33]. 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 [34].

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 [34].

Uma das deficiências do método BR 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 [5]. A Figura 2 ilustra o processo utilizado pelo BR após ser aplicado sobre o conjunto de dados da Tabela 2.

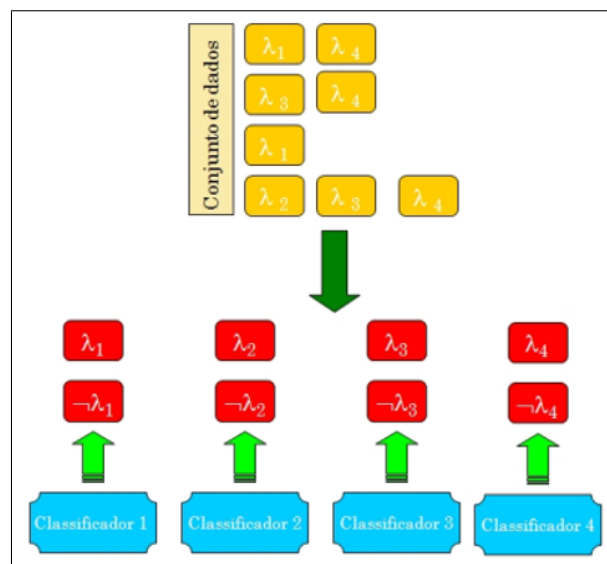
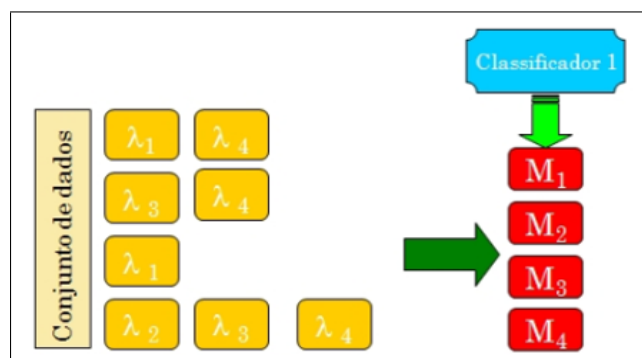


Figura 2. Processo utilizado pelo método BR [5].

### 3.3 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 multiclases, 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 [35]. Neste método, cada subconjunto diferente de rótulos de  $L$  é considerado como uma única classe da nova tarefa de classificação unirrótulo. Desse modo, um classificador unirrótulo

$C : X \rightarrow P(L)$  é treinado, onde  $P(L)$  é o *Powerset* de  $L$ , contendo todos os subconjuntos de rótulos possíveis. Assim, dada uma nova instância, o classificador unirrotulo  $C$  retorna como saída a classe mais provável, que neste caso é um conjunto de rótulos [31]. A Figura 3 ilustra o processo utilizado pelo LP após ser aplicado sobre o conjunto de dados da Tabela 2.



**Figura 3.** Processo utilizado pelo método LP [5].

Um das vantagens deste método é que, ao contrário do BR, no LP, as correlações entre os rótulos são consideradas. Contudo, é suscetível ao fato de, no caso de haver um número muito grande de subconjuntos de rótulos, o número de rótulos de uma classe pode crescer exponencialmente, resultando em muitas classes com poucos exemplos associados, aumentando o custo computacional do LP e diminuindo a acurácia dos classificadores. Além disso, o LP só pode prever confiavelmente conjuntos de rótulos (*labelsets*) observados no conjunto de treinamento. Esta é uma importante limitação, uma vez que novos *labelsets* tipicamente aparecerão em instâncias de teste [5].

### 3.4 RakeL

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 [32] onde um parâmetro definido  $k$  especifica o tamanho do subconjunto de rótulos que devem ser considerados [34].

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 unirrotulos, mesmo considerando o correlacionamento dos rótulos, são aplicados em subtarefas cujo número de rótulos e exemplos por rótulos sejam adequados [31, 36]. A Figura 4 ilustra o processo utilizado pelo RAKEL após ser aplicado sobre o conjunto de dados da Tabela 2.



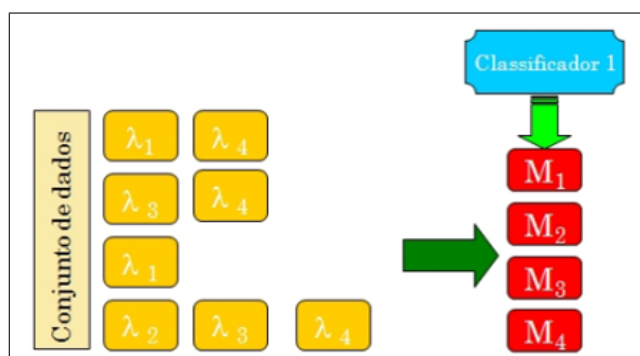


Figura 4. Processo utilizado pelo método RAKEL [5].

## 4 Comitês 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 [37].

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 [37].

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 [37].

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. 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 [37]. Dentre os principais comitês de classificadores encontrados na literatura destaca-se o *Bagging*, *Boosting* e *Stacking*.

### 4.1 Bagging

*Bagging (Bootstrap AGGregatING)* foi proposto por [38] baseado na amostragem *bootstrap*. Este consiste em gerar N subconjuntos de treinamento de forma aleatória sendo que os exemplos selecionados são devolvidos ao conjunto base para geração dos demais subconjuntos. 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 [37].

Para combinar os diferentes classificadores base um esquema de votação simples é normalmente empregado [39]. Além disso, o total de classificadores obtidos equivale ao número de subconjuntos gerados, ou seja,  $N$ . Outra característica do *Bagging* é que o mesmo algoritmo é aplicado para todos os  $N$  subconjuntos, podendo ser gerados em paralelo [37]. A Figura 5 representa o processo realizado pelo *Bagging*, no qual divide o conjunto de dados original em  $N$  subconjuntos ( $S_1, S_2, \dots, S_n$ ) ao passo que cada subconjunto obtido é designado para um determinado classificador ( $C_1, C_2, \dots, C_n$ ).

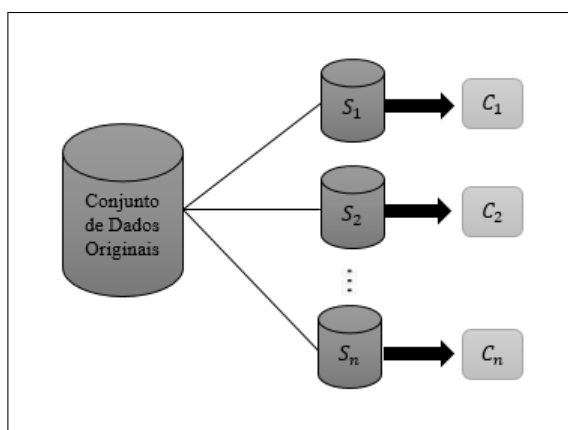


Figura 5. Representação do algoritmo *Bagging* [40] - Adaptado.

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.

#### 4.2 *Boosting*

O método *Boosting* possui estratégia semelhante ao *Bagging*. 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 [41].

Consequentemente, 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.

#### 4.3 *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 [39]. De fato, o *Stacking* tenta aprender quais dos classificadores base são os mais confiáveis e descobrir a melhor maneira de combinar

as suas saídas [42].

*Stacking* [43] 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 [44]. A Figura 6 representa o processo realizado pelo *Stacking*, na primeira fase, ou nível zero, um conjunto de  $n$  classificadores base ( $C_1, C_2, \dots, C_n$ ) é gerado por diferentes algoritmos de aprendizado, a partir de um único conjunto de dados original; na segunda fase, ou no nível um, o meta classificador combina as saídas geradas dos classificadores base em uma predição final [37].

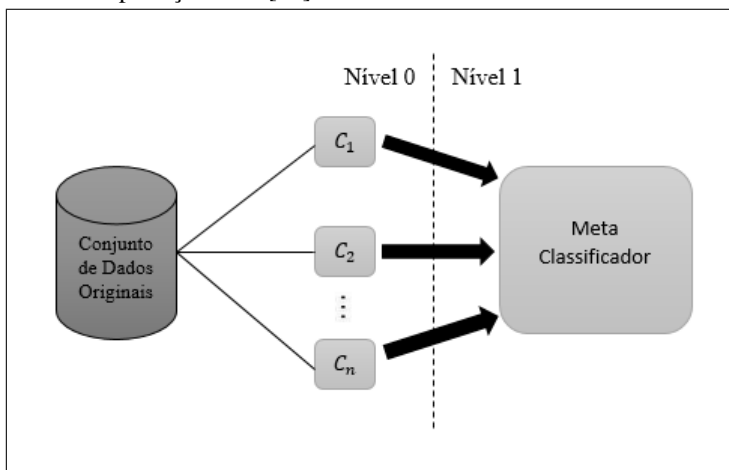


Figura 6. Representação do algoritmo *Stacking* [40] - Adaptado.

## 5 Metodologia dos experimentos

Esta seção apresenta os procedimentos adotados para a realização dos experimentos e está organizado da seguinte forma: a primeira subseção trata das bases de dados utilizadas; a segunda subseção trata das métricas de avaliação; a terceira trata dos testes estatísticos, e por fim, a quarta subseção trata das configurações dos experimentos.

### 5.1 Bases de Dados

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, neste trabalho. A Tabela 1 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 3.** Detalhamentos das Bases de Dados Multirrótulo

Base	Num. Exemplos	Num. Atrib. Numérico	Num. Atrib. Discreto	Rótulos
<i>Birds</i>	323	279	0	19
<i>Emotions</i>	202	78	0	6
<i>Enron</i>	578	0	1054	53
<i>Yeast</i>	917	117	0	14
<i>CAL500</i>	502	68	0	174
<i>Flags</i>	194	09	30	7

Fonte: Autória Própria

## 5.2 Métricas de Avaliação

É possível encontrar várias medidas de classificação multirrótulo na literatura. Segundo (TSOU-MAKAS 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, são apresentadas as medidas de avaliação baseadas em Bipartição *Hamming Loss (HLoss)*, *Precision (Prec)*, *Recall (Rec)*, *FMeasure (FM)* e as medidas baseadas em *Ranking: One-Error (1-Err)*, *Average Precision (AvPrec)* que foram utilizadas durante a realização dos experimentos.

### *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 apresentada a formulação matemática.

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

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  $\Delta$  é a diferença simétrica entre os dois conjuntos.

### *Recall*

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

$$Rec = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i \cap z_i|}{|Y_i|} \quad (2)$$

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.

***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[5]. A seguir é apresentada a formulação matemática.

$$FM = \frac{1}{N} \sum_{i=1}^N \frac{2 |Y_i \cap Z_i|}{|Z_i| + |Y_i|} \quad (3)$$

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.

***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 apresentada a formulação matemática.

$$Prec = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i \cap Z_i|}{|Z_i|} \quad (4)$$

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.

***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 apresentada a formulação matemática.

$$1 - Err = \frac{1}{N} \sum_{i=1}^{|N|} \delta(\arg_{\gamma \in L} \min r_i(\lambda)) \quad (5)$$

Formulação matemática da métrica *One-Error*.

onde:

$$\delta(\lambda) = \{ 1 \text{ se } \lambda \in Y_i \text{ 0 caso contrario} \}. \quad (6)$$

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

### *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 apresentada a formulação matemática.

$$AvPrec(H, |D|) = \frac{1}{|D|} * \sum_{i=1}^{|D|} \frac{1}{Z_i} * \sum_{y \in Z_i} \frac{|y' \in Z_i : r(y')r(y)|}{r(y)} \quad (7)$$

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.

### 5.3 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 foram formuladas duas hipóteses:

- Hipótese nula ( $H_0$ ): Não existe diferença estatística entre o desempenho dos sistemas, ou seja, não se pode 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 correteza. Neste trabalho é adotado o nível de confiança é de 95 % ( $\alpha = 0,05$ ). Existe uma série de testes estatísticos que podem auxiliar as pesquisas. Neste trabalho, utilizamos o teste estatístico de *Wilcoxon* por ser bastante conhecido e utilizado pela comunidade científica. 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 [45].

## 5.4 Configuração dos Experimentos

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). 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 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 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 [46]. 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 [47]. 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. Para a análise dos resultados obtidos pelos diferentes métodos de aprendizado foi utilizado o teste estatísticos de *Wilcoxon*, com o nível de confiança de 95% ( $\alpha = 0,05$ ).

## 6 Resultados dos Experimentos

Esta Seção apresenta uma investigação comparativa dos métodos de classificação multirrótulo utilizando a abordagem supervisionada e semissupervisionada.

Neste contexto, todos os resultados apresentam a média e o desvio padrão das melhores métricas de avaliação, independente do tamanho do comitê e dos classificadores supervisionado, tendo sido sombreados os melhores resultados em cada uma das medidas. Assim, para análise da relevância estatística foi utilizado o teste de *Wilcoxon*. Os resultados nos quais os testes estatísticos mostraram relevância do ponto de vista estatístico tiveram os valores **WIL** da coluna marcados em negrito. 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  $\downarrow$  (o menor valor significa o melhor) e  $\uparrow$  (o maior valor significa o melhor) para representar o comportamento das métricas de avaliação.

### 6.1 Melhores resultados do método BR utilizando a abordagem do aprendizado supervisionado versus método BR utilizando a abordagem do aprendizado semissupervisionado

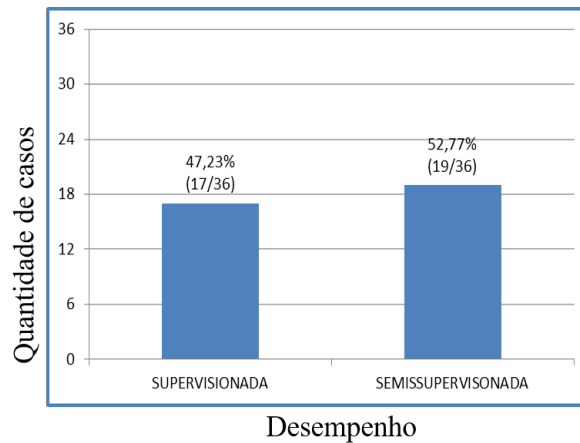
A Tabela 4 e a Figura 7 apresentam a média e o desvio padrão dos melhores resultados do método BR utilizando a abordagem do aprendizado supervisionado versus método BR utilizando a abordagem do aprendizado semissupervisionado.

Tabela 4: 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		
BIRDS					
Medida	Média	Desvio Padrão	Média	Desvio Padrão	WIL
HLoss $\downarrow$	<b>0.0428</b>	0.0111	0.0436	0.0083	0.220

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





**Figura 7.** Média dos melhores resultados do método BR utilizando a abordagem do aprendizado supervisionado *versus* método BR utilizando a abordagem do aprendizado semissupervisionado.

Analisando a Tabela 4, é possível observar que os experimentos utilizando o método BR com a abordagem do aprendizado supervisionado *versus* BR utilizando a abordagem semissupervisionado, a segunda abordagem 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.

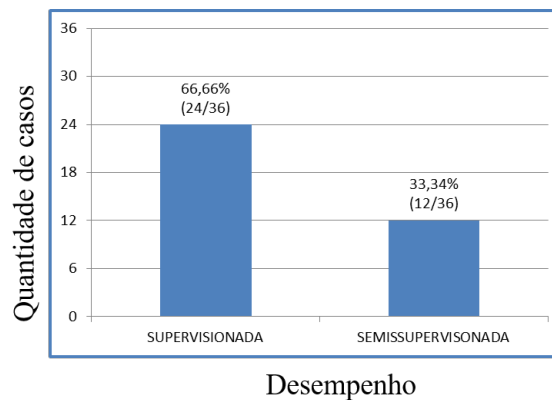
## 6.2 Melhores resultados do método LP utilizando a abordagem do aprendizado supervisionado *versus* método LP utilizando a abordagem do aprendizado semissupervisionado

A Tabela 5 e a Figura 8 apresentam 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.

Tabela 5: Abordagem do aprendizado supervisionado *versus* abordagem do aprendizado semissupervisionado.

Abordagem	SUPERVISIONADA		SEMISSUPERVISIONADA		
Comitê	BAGGING		STACKING		
BIRDS					
Medida	Média	Desvio Padrão	Média	Desvio Padrão	WIL
HLoss↓	<b>0.0483</b>	0.0112	0.0560	0.0134	0.508
PPRec↑	<b>0.5902</b>	0.1681	0.2855	0.1168	0.506
PRec↑	<b>0.4340</b>	0.1216	0.2488	0.1149	0.170
F-M↑	<b>0.4585</b>	0.1061	0.2644	0.1144	0.200
l-Err↓	<b>0.7958</b>	0.0795	0.8670	0.0665	<b>0.003</b>
AvgPPRec↑	<b>0.4456</b>	0.1201	0.3376	0.1154	<b>0.021</b>
EMOTIONS					
HLoss↓	<b>0.1922</b>	0.0220	0.2541	0.0227	<b>0.022</b>
PPRec↑	<b>0.6858</b>	0.0292	0.5925	0.0473	<b>0.021</b>

PRec↑	<b>0.7061</b>	0.0482	0.5878	0.0499	<b>0.020</b>
F-M↑	<b>0.6955</b>	0.0363	0.5899	0.0284	<b>0.024</b>
l-Err↓	<b>0.4081</b>	0.0663	0.4133	0.0411	<b>0.030</b>
AvgPPRec↑	0.6923	0.0355	<b>0.6969</b>	0.0288	<b>0.025</b>
ENRON					
HLoss↓	0.0595	0.0046	<b>0.0207</b>	0.0030	<b>0.002</b>
PPRec↑	0.5449	0.0537	<b>0.6287</b>	0.0459	<b>0.002</b>
PRec↑	0.4112	0.0374	<b>0.5941</b>	0.0643	0.109
F-M↑	0.4442	0.0459	<b>0.6099</b>	0.0516	<b>0.025</b>
l-Err↓	0.5595	0.0407	<b>0.3754</b>	0.0517	<b>0.002</b>
AvgPPRec↑	0.3752	0.0369	<b>0.6298</b>	0.0450	<b>0.002</b>
YEAST					
HLoss↓	<b>0.2130</b>	0.0122	0.2436	0.0000	0.110
PPRec↑	0.6638	0.0236	<b>0.6646</b>	0.0000	<b>0.002</b>
PRec↑	0.6069	0.0327	<b>0.7063</b>	0.0000	0.751
F-M↑	0.6333	0.0237	<b>0.6837</b>	0.0000	0.338
l-Err↓	0.3849	0.0563	<b>0.3767</b>	0.0000	1.000
AvgPPRec↑	<b>0.6340</b>	0.0243	0.7434	0.0000	<b>0.021</b>
FLAGS					
HLoss↓	<b>0.2571</b>	0.0863	0.3257	0.0229	0.184
PPRec↑	<b>0.7309</b>	0.0707	0.6538	0.1129	<b>0.021</b>
PRec↑	<b>0.7248</b>	0.0582	0.6282	0.1929	0.337
F-M↑	<b>0.7215</b>	0.0569	0.6372	0.0416	0.104
l-Err↓	0.4553	0.1101	<b>0.3500</b>	0.0000	0.495
AvgPPRec↑	<b>0.6984</b>	0.1095	0.6906	0.0246	0.349
CAL500					
HLoss↓	<b>0.1931</b>	0.0053	<b>0.1979</b>	0.0060	1.000
PPRec↑	<b>0.3507</b>	0.0196	0.3430	0.0058	1.000
PRec↑	<b>0.3561</b>	0.3488	0.3369	0.0357	0.351
F-M↑	<b>0.3477</b>	0.0228	0.3261	0.0208	1.000
l-Err↓	<b>0.9284</b>	0.0361	0.9421	0.0251	0.108
AvgPPRec↑	<b>0.1883</b>	0.0149	0.1745	0.0256	<b>0.002</b>



**Figura 8.** Média dos melhores resultados do método LP utilizando a abordagem do aprendizado supervisionado *versus* método LP utilizando a abordagem do aprendizado semissupervisionado.

Analisando a Tabela 5, é possível observar que os experimentos utilizando o método LP com a abordagem do aprendizado supervisionado *versus* LP utilizando a abordagem semissupervisionado, a primeira abordagem 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.

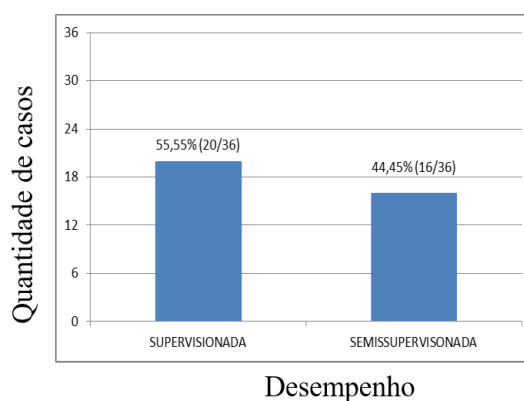
### 6.3 Melhores resultados do método Rakel utilizando a abordagem do aprendizado supervisionado *versus* método Rakel utilizando a abordagem do aprendizado semissupervisionado

A Tabela 6 e a Figura 9 apresentam 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.

Tabela 6: Média dos melhores resultados do método Rakel utilizando a abordagem do aprendizado supervisionado *versus* método Rakel utilizando a abordagem do aprendizado semissupervisionado.

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

F-M↑	<b>0.7572</b>	0.0311	0.7070	0.0309	0.755
1-Err↓	<b>0.2061</b>	0.0476	0.2850	0.0669	0.730
AvgPPRec↑	<b>0.8184</b>	0.0345	0.7944	0.0161	1.000
CAL500					
HLoss↓	0.1815	0.0165	<b>0.1564</b>	0.0048	<b>0.003</b>
PPRec↑	<b>0.7293</b>	0.0457	0.4646	0.0187	<b>0.002</b>
PRec↑	<b>0.7209</b>	0.0513	0.3551	0.0148	<b>0.002</b>
F-M↑	<b>0.7021</b>	0.0239	0.3851	0.0185	<b>0.002</b>
1-Err↓	0.2477	0.0381	<b>0.2230</b>	0.0599	<b>0.002</b>
AvgPPRec↑	<b>0.8021</b>	0.0234	0.4150	0.0136	<b>0.002</b>



**Figura 9.** Média dos melhores resultados do método Rakel utilizando a abordagem do aprendizado supervisionado *versus* método Rakel utilizando a abordagem do aprendizado semissupervisionado.

Analisando a Tabela 6, é possível observar que os experimentos utilizando o método Rakel com a abordagem do aprendizado supervisionado *versus* Rakel utilizando a abordagem semissupervisionado, a primeira abordagem 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.

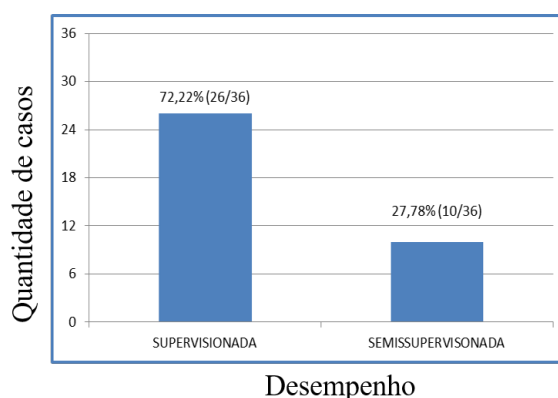
#### 6.4 Melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Bagging* utilizando a abordagem do aprendizado supervisionado *versus* métodos BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado

A Tabela 7 e a Figura 10 apresentam 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.

## Estratégias de Construções de Comitês de Classificadores Multirrotulos no Aprendizado Semissupervisionado Multidescrição

Tabela 7: 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		
BIRDS					
Medida	Média	Desvio Padrão	Média	Desvio Padrão	WIL
HLoss↓	0.0428	0.0111	<b>0.0421</b>	0.0092	1.000
PPRec↑	<b>0.7426</b>	0.1363	0.6454	0.1147	0.111
PRec↑	<b>0.4994</b>	0.0875	0.3742	0.1032	<b>0.039</b>
F-M↑	<b>0.4641</b>	0.0861	0.4620	0.0975	1.000
1-Err↓	0.6836	0.1082	<b>0.1922</b>	0.0274	<b>0.002</b>
AvgPPRec↑	<b>0.6137</b>	0.0772	0.6044	0.0797	0.757
EMOTIONS					
HLoss↓	<b>0.2058</b>	0.0009	0.0444	0.0269	0.758
PPRec↑	<b>0.7120</b>	0.0024	0.6707	0.0513	0.108
PRec↑	<b>0.7544</b>	0.000	0.6757	0.0383	<b>0.002</b>
F-M↑	<b>0.6815</b>	0.0034	0.6695	0.0397	<b>0.334</b>
1-Err↓	<b>0.2083</b>	0.0264	0.258	0.0468	<b>0.040</b>
AvgPPRec↑	<b>0.8091</b>	0.0018	0.7971	0.0274	0.754
ENRON					
HLoss↓	0.0491	0.0016	<b>0.0080</b>	0.0028	<b>0.002</b>
PPRec↑	0.6607	0.0475	<b>0.9679</b>	0.1417	<b>0.002</b>
PRec↑	0.6012	0.0408	<b>0.8335</b>	0.0100	<b>0.002</b>
F-M↑	0.5438	0.0442	<b>0.8468</b>	0.0532	<b>0.003</b>
1-Err↓	0.2642	0.0648	<b>0.0981</b>	0.0640	<b>0.002</b>
AvgPPRec↑	0.6617	0.0414	<b>0.9002</b>	0.0488	<b>0.002</b>
YEAST					
HLoss↓	<b>0.1994</b>	0.0065	0.2045	0.0077	0.338
PPRec↑	<b>0.7139</b>	0.0176	0.6816	0.0179	<b>0.003</b>
PRec↑	<b>0.7265</b>	0.0162	0.6094	0.0171	0.348
F-M↑	<b>0.6962</b>	0.0143	0.6432	0.0143	<b>0.002</b>
1-Err↓	0.2420	0.0202	<b>0.2396</b>	0.0217	0.752
AvgPPRec↑	<b>0.7534</b>	0.0140	0.7422	0.0165	<b>0.002</b>
FLAGS					
HLoss↓	<b>0.2415</b>	0.0328	0.2686	0.0274	<b>0.003</b>
PPRec↑	<b>0.7388</b>	0.0600	0.6637	0.0274	<b>0.023</b>
PRec↑	<b>0.7793</b>	0.0437	0.7733	0.0439	1.000
F-M↑	<b>0.7573</b>	0.0361	0.7070	0.0309	0.109
1-Err↓	<b>0.1955</b>	0.0750	0.2850	0.0669	<b>0.020</b>
AvgPPRec↑	<b>0.8163</b>	0.0268	0.7944	0.0161	<b>0.022</b>
CAL500					
Base					
HLoss↓	<b>0.1375</b>	0.0052	0.1564	0.0048	<b>0.002</b>
PPRec↑	<b>0.6059</b>	0.0258	0.4646	0.0187	<b>0.002</b>
PRec↑	<b>0.5098</b>	0.0389	0.3551	0.0128	<b>0.002</b>
F-M↑	0.3605	0.0174	<b>0.3851</b>	0.0185	<b>0.002</b>
1-Err↓	<b>0.1656</b>	0.0713	0.2230	0.0599	<b>0.019</b>
AvgPPRec↑	<b>0.4692</b>	0.0169	0.4150	0.0136	<b>0.002</b>



**Figura 10.** 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.

Analisando a Tabela 7, é possível observar que os experimentos utilizando os métodos BR, LP e Rakel com o comitê de classificadores *Bagging* com a abordagem do aprendizado supervisionado *versus* método BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado, a primeira abordagem 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.

### 6.5 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

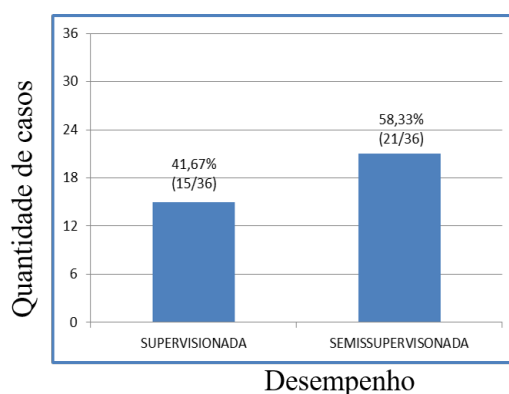
A Tabela 8 e a Figura 11 apresentam a média e o desvio padrão dos melhores resultados dos métodos BR, LP e Rakel com o comitê de classificadores *Boosting* utilizando a abordagem do aprendizado supervisionado *versus* método BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado.

Tabela 8: 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		WIL
	Média	Desvio Padrão	Média	Desvio Padrão	
BIRDS					
HLoss↓	<b>0.0413</b>	0.0079	0.0422	0.0114	0.753
PPRec↑	<b>0.7880</b>	0.1523	0.6391	0.1435	<b>0.022</b>
PRec↑	<b>0.5033</b>	0.1317	0.3791	0.1317	0.111
F-M↑	<b>0.4909</b>	0.1035	0.4655	0.1226	0.757
1-Err↓	0.6869	0.0712	<b>0.6840</b>	0.0712	1.000
AvgPPRec↑	0.5835	0.0830	<b>0.6102</b>	0.1099	0.756
EMOTIONS					

Estratégias de Construções de Comitês de Classificadores Multirrótulos no Aprendizado Semissupervisionado Multidescrição

HLoss↓	<b>0.1815</b>	0.0165	0.2136	0.0114	<b>0.023</b>
PPRec↑	0.7293	0.0457	<b>0.7397</b>	0.0000	0.103
PRec↑	0.7209	0.0513	<b>0.7561</b>	0.0770	0.342
F-M↑	<b>0.7021</b>	0.0240	0.6463	0.0301	<b>0.021</b>
1-Err↓	0.2477	0.0381	<b>0.2167</b>	0.0352	0.179
AvgPPRec↑	0.8021	0.0234	<b>0.8111</b>	0.0000	0.343
ENRON					
Base					
HLoss↓	0.0516	0.0034	<b>0.0105</b>	0.0017	<b>0.002</b>
PPRec↑	0.6266	0.0425	<b>0.8351</b>	0.0519	<b>0.002</b>
PRec↑	0.6007	0.0414	<b>0.7921</b>	0.0246	<b>0.002</b>
F-M↑	0.5328	0.0342	<b>0.8049</b>	0.0266	<b>0.002</b>
1-Err↓	0.2917	0.0737	<b>0.1647</b>	0.0359	<b>0.021</b>
AvgPPRec↑	0.6032	0.0470	<b>0.8687</b>	0.0221	<b>0.002</b>
YEAST					
Base					
HLoss↓	0.2073	0.0154	<b>0.0105</b>	0.0017	<b>0.002</b>
PPRec↑	0.7063	0.0388	<b>0.8351</b>	0.0519	<b>0.001</b>
PRec↑	0.6191	0.0274	<b>0.7921</b>	0.0246	<b>0.002</b>
F-M↑	0.6324	0.0276	<b>0.8049</b>	0.0266	<b>0.002</b>
1-Err↓	0.2497	0.0505	<b>0.1647</b>	0.0359	0.116
AvgPPRec↑	0.7318	0.0286	<b>0.8687</b>	0.0221	<b>0.002</b>
FLAGS					
HLoss↓	0.2416	0.0304	<b>0.1564</b>	0.0250	<b>0.012</b>
PPRec↑	<b>0.7385</b>	0.0442	0.4646	0.0494	<b>0.013</b>
PRec↑	<b>0.7786</b>	0.0367	0.3551	0.0539	0.315
F-M↑	<b>0.7572</b>	0.0311	0.3851	0.0322	0.501
1-Err↓	<b>0.2061</b>	0.0476	0.2230	0.0985	0.267
AvgPPRec↑	<b>0.8184</b>	0.0345	0.4150	0.0370	0.312
CAL500					
Base					
HLoss↓	0.1815	0.0165	<b>0.1515</b>	0.0026	<b>0.001</b>
PPRec↑	<b>0.7293</b>	0.0457	0.4595	0.0082	<b>0.002</b>
PRec↑	<b>0.7209</b>	0.0513	0.4501	0.0372	<b>0.001</b>
F-M↑	<b>0.7021</b>	0.0239	0.4566	0.0264	<b>0.001</b>
1-Err↓	0.2477	0.0381	<b>0.0567</b>	0.0492	<b>0.001</b>
AvgPPRec↑	<b>0.8021</b>	0.0234	0.4281	0.0089	<b>0.001</b>



**Figura 11.** 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.

Analisando a Tabela 8, é possível observar que os experimentos utilizando os métodos BR, LP e Rakel com o comitê de classificadores *Boosting* com a abordagem do aprendizado supervisionado *versus* método BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado, a segunda abordagem 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.

### 6.6 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 9 e a Figura 12 apresentam 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.

Tabela 9: 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.

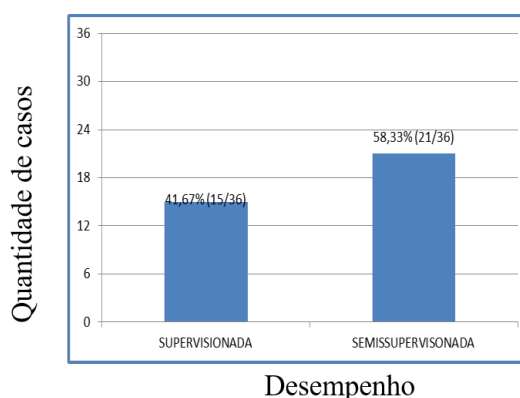
Método	RAKEL		RAKEL		WIL
	Média	Desvio Padrão	Média	Desvio Padrão	
BIRDS					
HLoss↓	0.0433	0.0114	<b>0.0391</b>	0.0014	0.751
PPRec↑	<b>0.6914</b>	0.0696	0.5934	0.1409	<b>0.013</b>
PRec↑	<b>0.6556</b>	0.0796	0.3867	0.0973	<b>0.001</b>
F-M↑	0.4380	0.0760	<b>0.4625</b>	0.0979	0.092
1-Err↓	<b>0.7025</b>	0.0953	0.7243	0.0715	0.241
AvgPPRec↑	<b>0.5999</b>	0.0961	0.5837	0.0906	0.245
EMOTIONS					



Estratégias de Construções de Comitês de Classificadores Multirrotulos no Aprendizado Semissupervisionado Multidescrição

---

HLoss↓	<b>0.1821</b>	0.0100	0.2071	0.0227	<b>0.003</b>
PPRec↑	<b>0.7108</b>	0.0192	0.6834	0.0473	0.051
PRec↑	<b>0.7007</b>	0.0301	0.6938	0.0499	0.408
F-M↑	<b>0.7053</b>	0.0186	0.6545	0.0284	<b>0.003</b>
1-Err↓	<b>0.2460</b>	0.0324	0.2749	0.0411	<b>0.048</b>
AvgPPRec↑	<b>0.8068</b>	0.0177	0.7880	0.0288	<b>0.022</b>
ENRON					
HLoss↓	0.0504	0.0019	<b>0.0102</b>	0.0014	<b>0.001</b>
PPRec↑	0.6647	0.0435	<b>0.8544</b>	0.0432	<b>0.001</b>
PRec↑	0.5094	0.0309	<b>0.7963</b>	0.0481	<b>0.001</b>
F-M↑	0.5227	0.0284	<b>0.8121</b>	0.0290	<b>0.001</b>
1-Err↓	0.3248	0.0603	<b>0.2045</b>	0.0362	<b>0.001</b>
AvgPPRec↑	0.6116	0.0345	<b>0.8191</b>	0.0324	<b>0.001</b>
YEAST					
HLoss↓	0.2298	0.0168	<b>0.0102</b>	0.0014	<b>0.001</b>
PPRec↑	0.6225	0.0304	<b>0.8544</b>	0.0432	<b>0.001</b>
PRec↑	0.6277	0.0185	<b>0.7963</b>	0.0481	<b>0.001</b>
F-M↑	0.6201	0.0270	<b>0.8121</b>	0.0290	<b>0.001</b>
1-Err↓	0.2584	0.0497	<b>0.2045</b>	0.0362	<b>0.005</b>
AvgPPRec↑	0.7243	0.0177	<b>0.8191</b>	0.0324	<b>0.001</b>
FLAGS					
HLoss↓	<b>0.2470</b>	0.0419	0.2571	0.1466	0.496
PPRec↑	0.7322	0.0599	<b>1.1575</b>	0.4596	<b>0.014</b>
PRec↑	0.7895	0.0297	<b>0.8509</b>	0.1929	0.217
F-M↑	0.7515	0.0467	<b>0.9996</b>	0.3228	<b>0.017</b>
1-Err↓	<b>0.2155</b>	0.0682	0.2353	0.1035	0.220
AvgPPRec↑	0.8156	0.0321	<b>1.1896</b>	0.3961	<b>0.012</b>
CAL500					
HLoss↓	<b>0.1411</b>	0.1042	0.1697	0.0050	0.250
PPRec↑	<b>0.4511</b>	0.2117	0.4004	0.0235	0.242
PRec↑	0.3418	0.0272	<b>0.4990</b>	0.0244	<b>0.001</b>
F-M↑	<b>0.3931</b>	0.1031	0.3665	0.0071	<b>0.003</b>
1-Err↓	0.8000	0.2422	<b>0.2227</b>	0.0979	<b>0.001</b>
AvgPPRec↑	0.2440	0.0666	<b>0.3893</b>	0.0053	<b>0.001</b>



**Figura 12.** 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.

Analisando a Tabela 9, é possível observar que os experimentos utilizando os métodos BR, LP e Rakel com o comitê de classificadores *Stacking* com a abordagem do aprendizado supervisionado *versus* método BR, LP e Rakel utilizando a abordagem do aprendizado semissupervisionado, a segunda abordagem 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.

## 7 Considerações Finais

Neste trabalho foram propostas várias estratégias de construções de comitês de classificadores multirrótulo utilizando a abordagem do aprendizado supervisionado e semissupervisionado. Após a realização dos experimentos podemos compreender algumas questões que foram indagadas durante o trabalho como: O uso de comitês classificadores junto ao *Co-training* trouxe melhorias no desempenho? O uso do aprendizado semissupervisionado multidescrição degradou o desempenho? Qual a abordagem de aprendizado de máquina obteve os melhores resultados?.

Ao analisar comparativamente os resultados obtidos por comitês de classificadores multirrótulos construídos através da abordagem supervisionada e da abordagem semissupervisionada multidescrição, verificou-se que, a abordagem semissupervisionada apresentou resultados semelhantes obtidos pela abordagem supervisionada (50% dos casos (03/06)).

Em relação à melhoria de desempenho 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 geral, o aprendizado semissupervisionado multidescrição apresentou resultados competitivos

em relação ao aprendizado supervisionado, e o uso de comitês de classificadores no aprendizado semissupervisionado não degradou o desempenho. O trabalho traz contribuições às pesquisas nas áreas de classificação de dados multirrótulo e aprendizado semissupervisionado multidescrição com a utilização de comitês classificadores.

Como a área de aprendizado de máquina semissupervisionado possui muito campo de pesquisa em aberto, sendo indicado principalmente em problemas onde a quantidade de exemplos não rotulados gerados é alta e a quantidade de exemplos rotulados é muito baixa, muitos estudos ainda precisam ser realizados a fim de compreender melhor a sua utilização.

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

- Utilizar bases de dados de maiores, com mais atributos, rótulos e instâncias, uma vez que no estudo realizado neste trabalho foram utilizadas bases de dados pequenas;
- Combinar a classificação multirrótulo e o aprendizado semissupervisionado com comitês de classificadores utilizando outros métodos de classificação existentes na literatura.

## 8 Contribuições dos Autores

- Wilamis Kleiton Nunes da Silva foi o autor principal do trabalho, desenvolveu a implementação e a escrita do trabalho.
- Araken de Medeiros Santos foi orientador do trabalho contribuindo com o processo de revisão do trabalho.

## Referências

- [1] MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. *Sistemas Inteligentes-Fundamentos e Aplicações*, v. 1, n. 1, p. 1, 2003.
- [2] REZENDE, S. O. *Sistemas inteligentes: fundamentos e aplicações*. Barueri, SP: Editora Manole Ltda, 2003.
- [3] 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*. Amsterdam, The Netherlands: Ios Press, 2005. p. 97–104.
- [4] BLUM, A.; MITCHELL, T. Combining labeled and unlabeled data with co-training. In: *ACM. Proceedings of the eleventh annual conference on Computational learning theory*. New York, USA, 1998. p. 92–100.
- [5] SANTOS, A. d. M. *Investigando a combinação de técnicas de aprendizado semissupervisionado e classificação hierárquica multirrótulo*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Norte, 2012.
- [6] GAMA, J. et al. *Extração de conhecimento de dados: data mining*. Lisboa, Portugal: Edições Silabo, 2015.

- [7] SANTOS, F. D.; BERCHT, M.; WIVES, L. Classificação de alunos desanimados em um aavea: uma proposta a partir da mineração de dados educacionais. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. Porto Alegre, Brasil: SBC, 2015. v. 26, n. 1, p. 1052.
- [8] REAL, E. M.; NICOLETTI, M. do C. *Investigação de algoritmos sequenciais de agrupamento com pré-processamento de dados em aprendizado de máquina*. Dissertação (Mestrado) — Faculdade Campo Limpo Paulista, 2014.
- [9] MONTEIRO, F. P. et al. *ClasSIS: uma metodologia para classificação supervisionada de imagens de satélite em áreas de assentamento localizados na Amazônia*. Dissertação (Mestrado) — Universidade Federal do Pará, 2015.
- [10] ALENCAR, F. E. S. *Desenvolvimento de um sistema para classificação automática de imagens dermatoscópicas para dispositivos móveis*. Dissertação (Mestrado) — Universidade do Estado do Rio Grande do Norte, 2015.
- [11] RAMOS, A. L.; SANTOS, C. N. dos. Combinando algoritmos de classificação para detecção de intrusão em redes de computadores. *Fortaleza CE*, 2015.
- [12] NOGUEIRA, P. S. d. S. *Utilizando comitês de classificadores para predição de rendimento escolar*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Norte, 2015.
- [13] GORGÔNIO, A. et al. Uma proposta de utilização de jogos sérios para auxiliar na identificação da dislalia em crianças. *Universidade Federal do Rio Grande do Norte*, 2016.
- [14] LOCHTER, J. V.; ZANETTI, R. F.; ALMEIDA, T. A. Detecção de opiniao em mensagens curtas usando comitê de classificadores e indexação semântica. *Universidade Federal de São Carlos*, 2016.
- [15] LOPES, U. K. *Redes neurais convolucionais aplicadas ao diagnóstico de tuberculose por meio de imagens radiológicas*. Dissertação (Mestrado) — Universidade do Vale do Rio dos Sinos, 2017.
- [16] 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.
- [17] MANCHEGO, F. E. A. *Anotação automática semissupervisionada de papéis semânticos para o português do Brasil*. Tese (Doutorado) — Universidade de São Paulo, 2013.
- [18] ASSIS, M. S. d. *Classificação multirrótulo com aprendizado semissupervisionado: uma análise multivisão de dados*. Dissertação (Mestrado) — Brasil, 2016.
- [19] KROGEL, M.-A.; SCHEFFER, T. Multi-relational learning, text mining, and semi-supervised learning for functional genomics. *Machine Learning*, Springer, Heidelberg, Alemanha, v. 57, n. 1-2, p. 61–81, 2004.
- [20] SILVA, P. N. D. *Classificação Multirrótulo em Cadeia: Novas Abordagens*. Dissertação (Mestrado) — Universidade Federal Fluminense, 2014.
- [21] VIEIRA, L. N.; ROMÃO, L. M. *Análise do uso de técnicas de pré-processamento de dados em algoritmos para classificação de proteínas*. Dissertação (Mestrado) — Universidade Federal de Minas Gerais, 2016.
- [22] LIMA, A. C. E. S. et al. Mineração de mídias sociais como ferramenta para a análise de tríade da persona virtual. *Universidade Presbiteriana Mackenzie*, 2016.

- [23] PECLAT, R. N. *Avaliação semântica da integração da gestão de riscos de segurança em documentos de software da administração pública*. Dissertação (Mestrado) — Universidade de Brasília, Instituto de Ciências Exatas, Departamento de Ciência da Computação, 2015.
- [24] BLOCKEEL, H. et al. *Decision trees for hierarchical multilabel classification: A case study in functional genomics*. Verlag Berlin Heidelberg: Springer, 2006.
- [25] 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.
- [26] LI, T.; OGIHARA, M. Toward intelligent music information retrieval. *Multimedia, IEEE Transactions on*, IEEE, v. 8, n. 3, p. 564–574, 2006.
- [27] LI, T.; OGIHARA, M. Detecting emotion in music. In: *ISMIR*. [S.l.]: ISMIR, 2003. v. 3, p. 239–240.
- [28] WIECZORKOWSKA, A.; SYNAK, P.; RAŚ, Z. W. Multi-label classification of emotions in music. In: *Intelligent Information Processing and Web Mining*. Heidelberg, Alemanha: Springer, 2006. p. 307–315.
- [29] TROHIDIS, K. et al. Multi-label classification of music into emotions. In: *ISMIR*. [S.l.]: ISMIR, 2008. v. 8, p. 325–330.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] TSOUMAKAS, G.; KATAKIS, I. Multi-label classification: An overview. *Dept. of Informatics, Aristotle University of Thessaloniki, Greece*, 2007.
- [34] RODRIGUES, F. M. *Uso de confiabilidade na rotulação de exemplos em problemas de classificação multirrotulo com aprendizado semissupervisionado*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Norte, 2014.
- [35] METZ, J. *Abordagens para aprendizado semissupervisionado multirrotulo e hierárquico*. Tese (Doutorado) — Universidade de São Paulo, 2011.
- [36] NASIERDING, G.; TSOUMAKAS, G.; KOUZANI, A. Z. Clustering based multi-label classification for image annotation and retrieval. In: *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*. Washington, USA: IEEE, 2009. p. 4514–4519.
- [37] KUNCHEVA, L. I. *Combining pattern classifiers: methods and algorithms*. New Jersey, USA: John Wiley & Sons, 2004.
- [38] BREIMAN, L. Bagging predictors. *Machine learning*, Springer, Heidelberg, Alemanha, v. 24, n. 2, p. 123–140, 1996.
- [39] CANUTO, A. M. d. P. *Combining neural networks and fuzzy logic for applications in character recognition*. Tese (Doutorado) — University of Kent at Canterbury, 2001.
- [40] CANUTO, A. M.; SANTOS, A. M.; VARGAS, R. R. Ensembles of artmap-based neural networks: an experimental study. *Applied Intelligence*, Springer, Heidelberg, Alemanha, v. 35, n. 1, p. 1–17, 2011.

- [41]LIMA, C. A. de M. *Comitê de Máquinas: uma abordagem unificada empregando máquinas de vetores-suporte*. Tese (Doutorado) — Universidade Estadual de Campinas, 2005.
- [42]WITTEN, I. H.; FRANK, E. *Data Mining: Practical machine learning tools and techniques*. Burlington, USA: Morgan Kaufmann, 2005.
- [43]WOLPERT, D. H. Stacked generalization. *Neural networks*, Elsevier, v. 5, n. 2, p. 241–259, 1992.
- [44]MERZ, C. J. Using correspondence analysis to combine classifiers. *Machine Learning*, Springer, Heidelberg, Alemanha, v. 36, n. 1-2, p. 33–58, 1999.
- [45]WILCOXON, F. Individual comparisons by ranking methods. *Biometrics bulletin*, JSTOR, v. 1, n. 6, p. 80–83, 1945.
- [46]MULAN. *Mulan: A Java Library for Multi-Label Learning*. Disponível em: <<http://mulan.sourceforge.net/>>. Acessado em: <04/06/2017>.
- [47]WEKA. *Data Mining Software in Java*. Disponível em: <<http://www.cs.waikato.ac.nz/ml/index.html>>. Acessado em: <11/08/2016>.