UNIVERSIDADE FEDERAL DE OURO PRETO
ESCOLA DE MINAS
CECAU - COLEGIADO DE ENGENHARIA DE CONTROLE E AUTOMAÇÃO

MAYKO ARAUJO DE CARVALHO

PROPOSTA DE UM CLASSIFICADOR HIERÁRQUICO MONORRÓTULO BASEADO EM INSTÂNCIAS

MONOGRAFIA DE GRADUAÇÃO EM ENGENHARIA DE CONTROLE E AUTOMAÇÃO

Ouro Preto, 2015

MAYKO ARAUJO DE CARVALHO

PROPOSTA DE UM CLASSIFICADOR HIERÁRQUICO MONORRÓTULO BASEADO EM INSTÂNCIAS

Monografia apresentada ao Curso de Engenharia de Controle e Automação da Universidade Federal de Ouro Preto como parte dos requisitos para a obtenção do Grau de Engenheira de Controle e Automação.

Orientador: Luiz Henrique de Campos Merschmann

Ouro Preto
Escola de Minas - UFOP
Julho/2015

Monografia defendida e aprovada, em 01 de julho de 2015, pela comissão avaliadora constituída pelos professores:

Prof. Dr. Luiz Henrique de Campos Merschmann - Orientador

Profa. Dra. Amanda Sávio Nascimento e Silva – Professora Convidada

Prof. Dr. Paulo Marcos de Barros Monteiro – Professor Convidado

Aos meus pais e a minha avo Maria Joana.

AGRADECIMENTOS

Primeiramente agradeço a Deus pelo dom da vida. À minha avó, Maria Joana, pelas lições de vida e pelo amor constante. Aos meus pais, Joni e Marina, que são e sempre foram meus grande incentivadores. Se não fosse por vocês eu não estaria aqui hoje e tampouco teria as oportunidades que tive na vida. Agradeço aos meus irmãos, Johnathas e Marielli, pela amizade e inspiração e também a toda família. À minha companheira, Amanda, obrigado pelo amor, compreensão e apoio. Ao meu orientador, Prof. Dr. Luiz Henrique de Campos Merschmann, pela excelente orientação durante os três anos em que trabalhamos juntos. Seus conselhos e recomendações foram essenciais para a realização deste trabalho. Quero agradecer também aos Prof. Paulo Monteiro, Luiz Fernando Rispoli Alves e Sávio Augusto Lopes da Silva pelas excelentes aulas, conhecimentos transmitidos e dedicação ao dever de ensinar. Inspiro-me em vocês para tentar sempre ser uma pessoa melhor e também contribuir na disseminação do conhecimento. Aos demais professores do DECAT e DECOM também. Aos meus colegas do DECAT, pelos frequentes momentos de muito estudo e diversão.

"Quando penso que cheguei ao meu limite, descubro que tenho forças para ir além."

- Ayrton Senna

RESUMO

Na área de Mineração de Dados existem problemas de classificação complexos, conhecidos como problemas de classificação hierárquica, nos quais as classes a serem preditas estão organizadas de acordo com uma hierarquia pré-definida. Na solução de um problema de classificação hierárquica duas abordagens podem ser utilizadas: global e local. Na abordagem global, um único classificador é induzido de forma a tratar todas as classes do problema simultaneamente, ao passo que na abordagem local, tem-se um conjunto de classificadores induzidos, em que cada um deles é responsável por um subconjunto de classes da hierarquia. Além disso, entre os problemas de classificação hierárquica existem aqueles nos quais uma instância pode estar associada a somente uma classe (problemas de classificação hierárquica monorrótulo). Métodos de classificação hierárquica vêm sendo propostos por pesquisadores das áreas de aprendizado de máquina, estatística e mineração de dados, no entanto, técnicas de classificação hierárquica que utilizam a aborgadem global mostram-se pouco exploradas na literatura. Baseado nesse fato, o presente trabalho propôs, implementou e avaliou um novo classificador hierárquico monorrótulo baseado na abordagem de classificação global, o HKNN (*Hierarchical K Nearest Neighbor*). O classificador proposto é uma modificação do KNN (método proposto para resolver problemas de classificação plana) que possibilita que a classificação seja realizada considerando-se a hierarquia de classes do problema. Foram realizados experimentos com 18 bases de dados para avaliar o desempenho preditivo do método proposto e os resultados obtidos por meio deles mostraram que o HKNN supera os desempenhos preditivos dos demais métodos avaliados nesse trabalho.

Palavras-chave: classificação hierárquica, mineração de dados, monorrótulo

ABSTRACT

In Data Mining, there are more complex problems, known as hierarchical classification problems, in which the classes to be predicted are structured according to a predefined hierarchy. For solving a hierarchical classification problem two approaches can be used: global and local. In the global approach, a single classifier is induced in order to treat to all classes simultaneously, while in the local approach there is a set of induced classifiers in which each of them is responsible for a subset of the class hierarchy. Furthermore, among the hierarchical classification problems there are those in which an instance may be associated with only one class (single-label hierarchical classification problems). Hierarchical classification methods have been proposed by researchers from the areas of machine learning, statistical and data mining, however, hierarchical classification techniques using the global approach have been underexplored in the literature. Based on that, this work proposes, developed and evaluated a new single-label hierarchical classifier based on the global approach, the HKNN (Hierarchical K Nearest Neighbor). The proposed new classifier is a modification of KNN (method proposed to solve flat classification problems) which allows the execution of the classification process considering the problem's class hierarchy. Computacional experiments were carried out with 18 databases to assess the predictive performance of the presented method and its results showed that the HKNN overcomes the predictive performance when compared to the other methods evaluated in this work.

Keywords: hierarchical classification, data mining, single-label

Glossário

KNN: K-Nearest Neighbor.

HKNN: Hierarchical K-Nearest Neighbor.

HKNN 1xDIST: Hierarchical K-Nearest Neighbor Distance 1 Time.

HKNN 2xDIST: Hierarchical K-Nearest Neighbor Distance 2 Times.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

1.1 Considerações Iniciais

Segundo Merschmann (2007), a quantidade de dados disponível no mundo, em ambientes computacionais, tem aumentado consideravelmente a cada dia, portanto, a necessidade por ferramentas computacionais capazes de analisar esses dados motivou o surgimento da área de pesquisa e aplicação em ciência da computação conhecida como Mineração de Dados. Processos de mineração de dados permitem a transformação de dados, uma matéria bruta, em informação e conhecimento úteis em diversas áreas de aplicação.

Dentre as várias tarefas de mineração de dados, as preditivas visam realizar inferências sobre dados existentes para predizer o comportamento de novos dados. Como exemplo tem-se a classificação e a regressão. A classificação se sobressai uma vez que tem sido utilizada em diversas áreas devido ao seu potencial de aplicação em vários domínios, tais como: esportivo, gestão, finanças, saúde, educação, marketing, entre outros. A tarefa de classificação corresponde a uma maneira de análise de dados que objetiva a construção de modelos preditivos. Sendo assim, a partir de um conjunto de instâncias com características e classes conhecidas, modelos são construídos e utilizados para classificar instâncias em que os atributos preditores são conhecidos, mas as classes desconhecidas.

No contexto de classificação existem problemas de classificação plana (*flat classification problems*) e problemas de classificação hierárquica (*hierarchical classification problems*). A diferença entre eles é que, enquanto na classificação plana não há qualquer tipo de relacionamento entre as classes, na classificação hierárquica as classes estão relacionadas de acordo com uma hierarquia, o que aumenta a complexidade do problema. Entre os problemas de classificação hierárquica existem aqueles nos quais uma instância pode estar associada a somente uma classe. Tais problemas são chamados de problemas de classificação hierárquica monorrótulo.

Para um problema de classificação hierárquica duas abordagens podem ser utilizadas na sua solução: global e local. Na abordagem global, um único classificador é induzido de forma a tratar todas as classes do problema simultaneamente. Sendo assim, a classificação de novas instâncias é realizada em apenas um passo. Já na abordagem local, tem-se

um conjunto de classificadores induzidos, em que cada um deles é responsável por um subconjunto de classes da hierarquia. Portanto, a classificação de novas instâncias é realizada em vários passos através das predições realizadas pelos vários classificadores construídos.

O método proposto neste trabalho, denominado HKNN, é uma modificação do KNN (método proposto para resolver problemas de classificação plana) que possibilita que a classificação seja realizada considerando-se a hierarquia de classes do problema. Trabalhos nesse sentido têm sido propostos na literatura, como é o caso da técnica de classificação hierárquica *Global Naive Bayes* (SILLA JR, 2011, p. 102), uma extensão do classificador plano denominado *Naive Bayes*.

1.2 Objetivos

Os objetivos deste trabalho são propor, implementar e avaliar uma técnica de classificação hierárquica global monorrótulo. Além disso, a técnica proposta é capaz de lidar eficientemente com bases de dados onde as classes estão hierarquicamente estruturadas de acordo com uma árvore.

1.3 Motivação

Devido à necessidade de uma análise de dados cada vez mais complexos, métodos de classificação hierárquica vêm sendo propostos por pesquisadores das áreas de aprendizado de máquina, estatística e mineração de dados. No entanto, técnicas de classificação hierárquica que utilizam a aborgadem global mostram-se pouco exploradas na literatura. Baseado nesse fato, o presente trabalho propõe o classificador hierárquico global HKNN.

1.4 Metodologia

Inicialmente foi realizado um estudo dirigido a respeito dos seguintes temas: Mineração de Dados e Classificação.

Numa segunda etapa foi feita uma revisão bibliográfica focada nos assuntos que serviram como embasamento para o desenvolvimento de novas abordagens de classificação hierárquica, a saber, abordagens por classificadores locais e globais para resolução de problemas de classificação hierárquica. Nesse momento, foram investigados algumas abordagens

propostas na literatura. Essa investigação facilitou a proposição do método classificação hierárquica deste trabalho.

Por fim, como esse projeto é uma pesquisa experimental, a técnica proposta (HKNN) foi implementada e testada a partir de dois conjuntos de bases de dados, especificadas no Capítulo 4, cujas classes encontram-se organizadas numa hierarquia. Dessa forma, os resultados obtidos a partir do HKNN foram comparados com aqueles obtidos através dos mesmos conjuntos de dados para outros classificadores.

1.5 Organização do Trabalho

O restante do trabalho encontra-se organizado da seguinte forma: o Capítulo 2 traz uma revisão teórica dos temas abordados nesse trabalho. Em seguida, o Capítulo 3 apresenta o classificador proposto. A avaliação empírica com os seus respectivos resultados são apresentados no Capítulo 4. Por fim, o Capítulo 5 traz as conclusões e trabalhos futuros.

2 REVISÃO TEÓRICA

Para se obter um panorama geral e embasar a proposta deste trabalho, primeiramente foi realizado um estudo aprofundado referente aos temas abordados: Tarefas da Mineração de Dados (Seção 2.1), Processo de Classificação (Seção 2.2), Classificação Hierárquica (Seção 2.3), Estimativa de Desempenho Preditivo (Seção 2.4) e KNN (Seção 2.5).

2.1 Tarefas da Mineração de Dados

A vasta quantidade de dados disponível em ambientes computacionais e a falta de ferramentas para análise dos mesmos motivou o surgimento da área de pesquisa e aplicação em ciência da computação conhecida como Mineração de Dados. De forma simples, segundo Merschmann (2007), tarefas em mineração de dados podem ser definidas como processos automatizados de descoberta de novas informações a partir de grandes massas de dados armazenadas em bancos de dados, arquivos de texto, data warehouses, ou em algum outro repositório de dados. Tais tarefas visam transformar grandes quantidades de dados em informações e conhecimentos úteis para diversas áreas de aplicação.

Apesar de alguns autores utilizarem o termo mineração de dados como sinônimo de KDD (Knowledge Discovery in Database) - processo de descoberta de conhecimento em bases de dados - outros consideram que a mineração de dados representa a etapa central desse processo maior denominado KDD. As outras etapas tratam, basicamente, do pré-processamento dos dados (seleção, limpeza e transformação) e pós-processamento da informação minerada (visualização e análise) (MERSCHMANN, 2007, p. 15). Os problemas em mineração de dados foram agrupados em classes de acordo com suas características, dando origem as tarefas de mineração de dados (HAN; KAMBER, 2006). Essas tarefas, classes de problemas definidas através de estudos na área, podem ser divididas em duas categorias:

- **Tarefas Descritivas**: objetivam encontrar padrões que descrevam os dados, possibilitando sua análise. As principais tarefas descritivas são: Extração de Regras de Associação (AGRAWAL; SRIKANT, 1994), Extração de Padrões Sequenciais (AGRAWAL; SRIKANT, 1995) e Agrupamento (*Clustering*) (TAN; STEINBACH; KUMAR, 2005).

- **Tarefas Preditivas**: realizam inferências sobre os dados existentes para prever o comportamento de novos dados. As principais tarefas preditivas são: Classificação (MITRA; ACHARYA, 2003) e Regressão (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

O método proposto neste trabalho está relacionado com a tarefa de classificação.

2.2 O Processo de Classificação

Pode-se dividir o processo de classificação em duas etapas: treinamento e teste. Na primeira etapa, o intuito é construir modelos que sejam capazes de estabelecer relações concretas entre os valores dos atributos preditores e as classes. A construção desses modelos é realizada por meio da análise das instâncias contidas numa base de dados, conhecida como base de dados de treinamento (Figura 2.1).

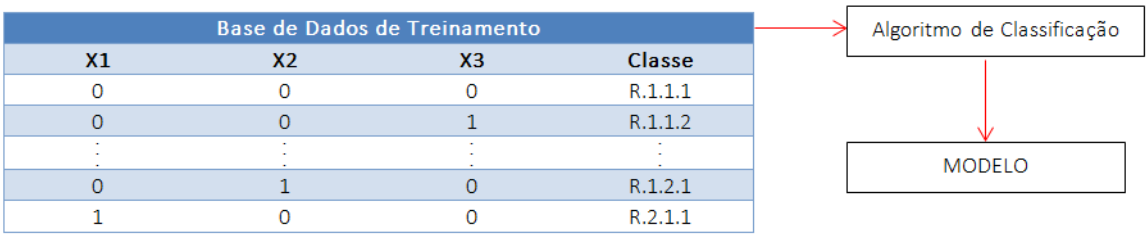


Figura 2.1 – Processo de Treinamento
Fonte: (FIDENCIO, 2014, p. 47)

Na segunda etapa, o modelo construído é avaliado quanto ao seu desempenho preditivo. Caso a estimativa de desempenho forneça resultados aceitáveis para o modelo, ele poderá ser utilizado na classificação de novas instâncias, ou seja, instâncias em que as classes são desconhecidas. A estimativa de desempenho preditivo do modelo é calculada a partir de um conjunto de instâncias com classes conhecidas que formam a base de dados de teste. Assim, através de uma análise comparativa, calcula-se o percentual de instâncias corretamente classificadas, ou seja, a acurácia do modelo para as instâncias em questão.

2.3 Classificação Hierárquica

Em mineração de dados, comumente as tarefas de classificação utilizam as técnicas de classificação plana (*flat classification*), onde cada instância é associada a uma das classes pertencentes a um conjunto pré-definido (e de forma geral pequeno) de classes, sendo que não há qualquer tipo de relacionamento entre essas classes. No entanto, existem problemas

de classificação mais complexos em que as classes a serem preditas estão estruturadas de acordo com uma hierarquia, tal como uma árvore ou GDA (Grafo Direcionado Acíclico). Para resolver esses problemas, conceitos e técnicas de classificação hierárquica vêm sendo apresentados em trabalhos nessa área de pesquisa.

No contexto de classificação, uma instância é a representação de qualquer objeto por meio de suas características individuais: os atributos preditores. Além disso, cada instância pertence a uma classe, a qual é definida por um de seus atributos, conhecido como atributo classe. A Figura 2.2 exemplifica uma base de dados contendo um grupo de instâncias com seus respectivos atributos preditores (X1, X2 e X3) e atributo classe (Classe).

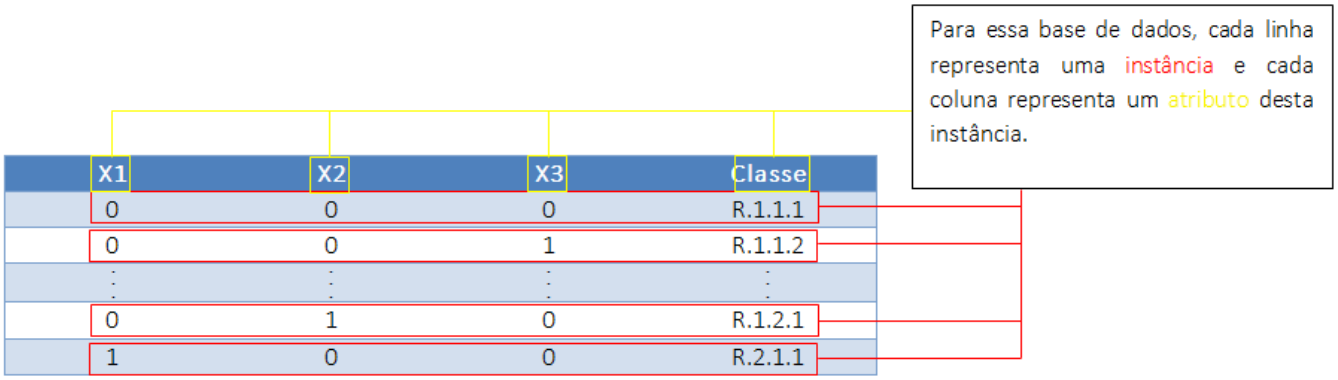


Figura 2.2 – Exemplo de Instância e Atributo
Fonte: (FIDENCIO, 2014, p. 15)

As classes representadas na Figura 2.2 estão organizadas de acordo com a hierarquia apresentada na Figura 2.3. Nesta figura é importante entender a relação entre as classes pais e filhas dentro de uma hierarquia de classes, por exemplo: considerando-se que uma instância possui classe R.1.1, tem-se que essa classe também é R.1, ou seja, a classe R.1 é classe pai da R.1.1 (classe filha).

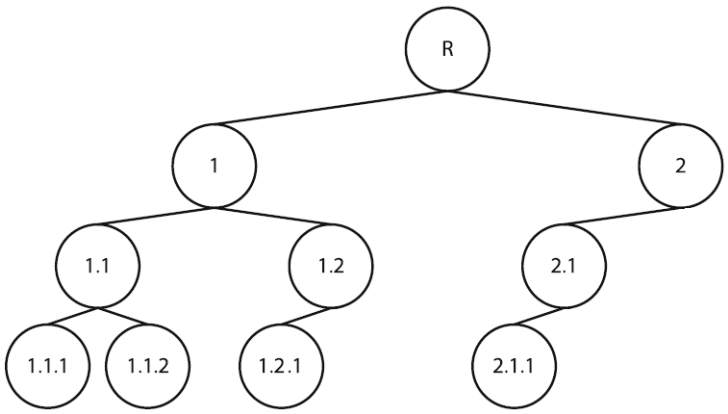
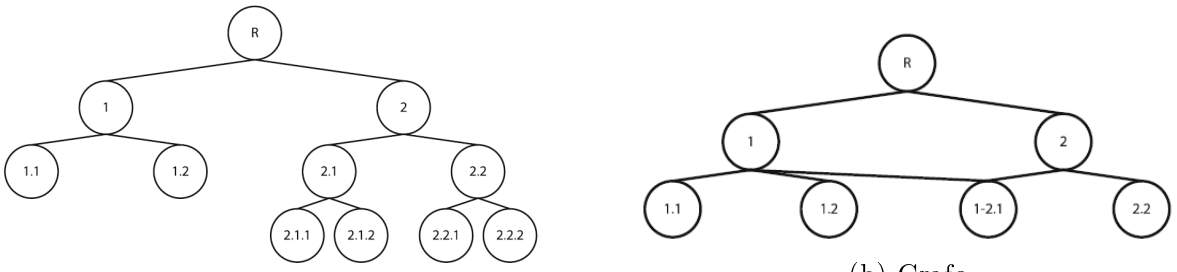


Figura 2.3 – Estrutura Hierárquica das Classes Referente à Figura 2.2

Os métodos de classificação hierárquica podem ser caracterizados de acordo com diferentes aspectos (SILLA JR; FREITAS, 2012). O tipo de estrutura hierárquica (árvore ou GDA) que o método é capaz de processar é o primeiro deles, uma vez que a representação dos relacionamentos entre as classes do problema a ser resolvido é feita por meio dessa estrutura. Em uma árvore cada nó (classe) encontra-se associado a no máximo um nó (classe) pai (Figura 2.4a), enquanto em um GDA um nó (classe) filho pode estar associado a vários nós (classes) pais (Figura 2.4b). Nessas figuras os nós indicados pela letra R representam as raízes das estruturas, sendo assim os demais nós conectados a esses são interpretados como nós filhos e portanto, encontra-se definida uma hierarquia.



(a) Árvore

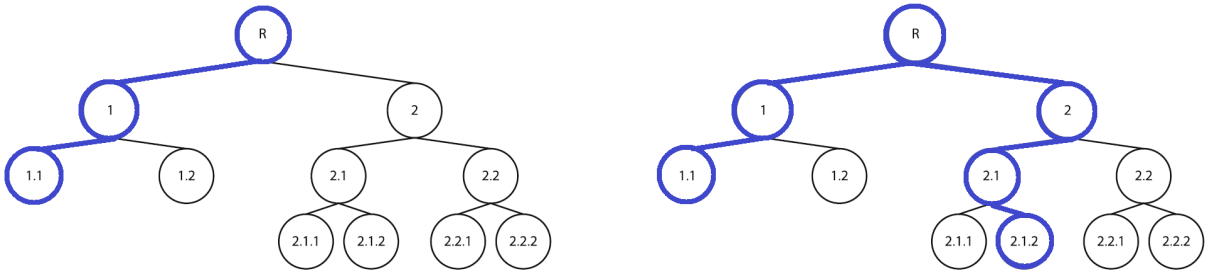
(b) Grafo

Figura 2.4 – Tipos de Estruturas
Fonte: (SILLA JR; FREITAS, 2012, p. 35)

O segundo aspecto diz respeito à profundidade na hierarquia na qual a classificação é realizada. Um método pode realizar predições utilizando somente as classes dos nós folha da hierarquia (*mandatory leaf node prediction*) ou utilizando classes referentes a qualquer nó (interno ou folha) da estrutura hierárquica (*non-mandatory leaf node prediction*). Portanto, a vantagem dos métodos que podem realizar predições referentes a qualquer nó da estrutura hierárquica é a precisão.

O terceiro aspecto diz respeito ao número de diferentes ramos de classes da hierarquia que um método pode dar como resposta para a classificação de uma instância. Um método pode ser capaz de predizer múltiplos ramos da hierarquia de classes para uma determinada instância (*multiple paths of labels* - Figura 2.5b) ou somente um (monorrótulo ou *single path of labels* - Figura 2.5a).

Por fim, o quarto aspecto está relacionado com a maneira adotada pelos métodos para manipular a estrutura hierárquica. Três abordagens diferentes são apresentadas na literatura: classificação plana, na qual a hierarquia de classes é ignorada e as predições são realizadas considerando-se somente as classes dos nós folha da hierarquia; abordagens por classificadores locais, que utilizam um conjunto de classificadores planos tradicionais; e



(a) *Single Path of Labels* - monorrótulo (b) *Multiple Path of Labels* - multirrótulo

Figura 2.5 – Número de Ramos
Fonte: (FIDENCIO, 2014, p. 17)

a abordagem por classificador global, onde um único modelo é construído considerando toda a hierarquia de classes.

2.3.1 Abordagem por Classificador Plano

Essa é a abordagem mais simples para lidar com problemas de classificação hierárquica. Um classificador plano é construído ignorando-se completamente a hierarquia de classes, realizando predições considerando somente as classes dos nós folha da hierarquia (Figura 2.6). Desse modo, ele fornece uma solução indireta para o problema de classificação hierárquica, dado que, se uma classe de um nó folha é atribuída a uma instância, todas as suas classes ancestrais também estão implicitamente atribuídas a essa instância.

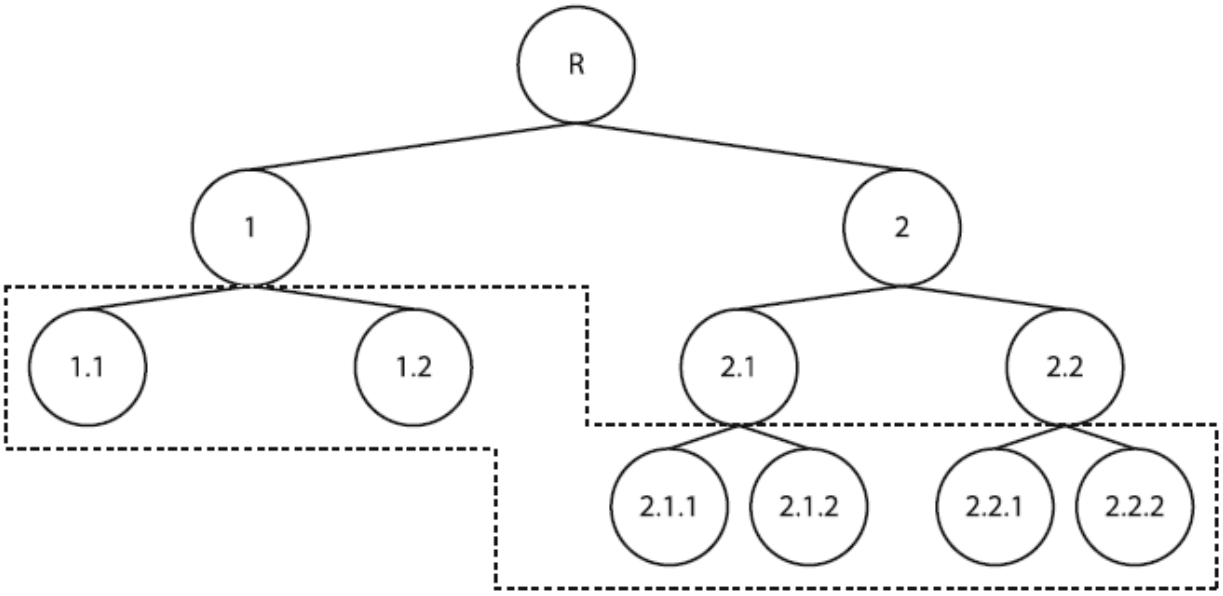


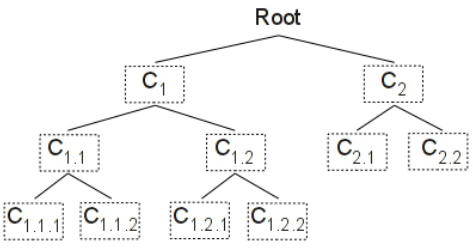
Figura 2.6 – Abordagem por Classificador Plano
Fonte: (SILLA JR; FREITAS, 2012, p. 37)

Apesar da simplicidade dessa abordagem, ela apresenta algumas desvantagens. Uma delas está relacionada com o fato de o classificador perder a oportunidade de explorar

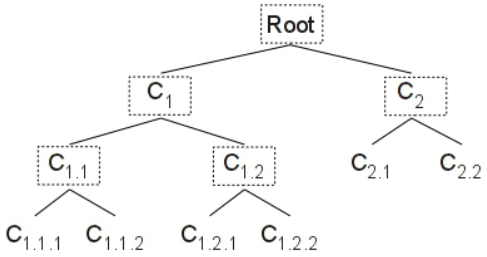
os relacionamentos entre as classes presentes na hierarquia. Além disso, como o número de classes associadas aos nós folha da hierarquia tende a ser muito grande, a tarefa de predição torna-se mais difícil. Essa abordagem também é incapaz de lidar com problemas onde se deseja realizar predições em qualquer nível da hierarquia, uma vez que somente classes dos nós folha podem ser atribuídas às instâncias.

2.3.2 Abordagem por Classificador Local

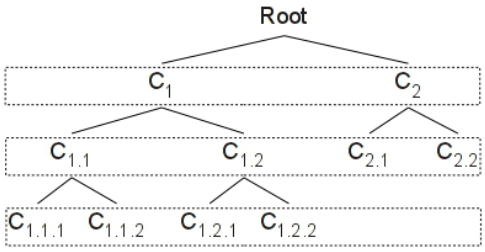
Nessa abordagem, vários classificadores são construídos, cada um com uma visão local do problema, ou seja, a hierarquia de classes é explorada segundo uma perspectiva local. De acordo com as diferentes maneiras de se utilizar essa informação local, os classificadores podem ser agrupados nas seguintes categorias: abordagem local por nó, abordagem local por nó pai e abordagem local por nível. Na Figura 2.7 as categorias mencionadas são ilustradas.



(a) Classificador Local por Nó



(b) Classificador Local por Nó Pai



(c) Classificador Local por Nível

Figura 2.7 – Abordagem por Classificação Local
Fonte: (MERSCHMANN; FREITAS, 2013, p. 162)

Na abordagem local por nó um classificador binário é criado para cada nó da hierarquia de classes (exceto para o nó raiz). Cada classificador prediz se uma instância pertence ou não à classe a qual ele está associado. Os retângulos pontilhados da Figura 2.7 representam os classificadores. Observe que essa abordagem permite que uma instância seja associada a classes que pertencem a diferentes ramos da hierarquia, o que pode resultar numa inconsistência. Por exemplo, considerando a Figura 2.7a, na classificação de uma instância, os classificadores binários podem dar resultado positivo para as classes 1, 2.1 e 1.1.2. Nesse caso, temos uma inconsistência entre a classe predita no nível 2 (classe 2.1) e aquelas preditas nos níveis 1 e 3 (classes 1 e 1.1.2, respectivamente). Para resolver esse

problema, vários métodos de tratamento ou correção de inconsistências foram propostos na literatura.

Na abordagem local por nó pai um classificador plano tradicional é treinado para cada nó pai da hierarquia de classes. Desse modo, para cada classificador construído, somente as classes associadas aos seus nós filhos são consideradas durante o processo de classificação. Essa estratégia, é conhecida como *top-down* e esta sujeita à propagação de erros uma vez que o classificador pode errar a classificação da nova instância nos primeiros níveis hierárquicos de forma que após esse erro ocorrerá sua propagação. Por outro lado, essa abordagem não está sujeita à ocorrência de inconsistências.

Na abordagem local por nível um classificador plano binário é treinado para cada nível da hierarquia de classes. Desse modo, cada um desses classificadores associa uma ou mais classes por nível a fim de serem associadas à instância de classe desconhecida. Esta abordagem está sujeita ao problema de inconsistência da mesma forma em que ocorre na abordagem local por nó.

2.3.3 Abordagem por Classificador Global

Ao invés de construir um conjunto de classificadores, a abordagem global envolve o treinamento de um único classificador que considera toda a hierarquia de classes numa única execução do algoritmo de classificação (Figura 2.8). Portanto, dada uma nova instância, esse tipo de classificador é capaz de apresentar como resultado qualquer classe de qualquer nível da hierarquia.

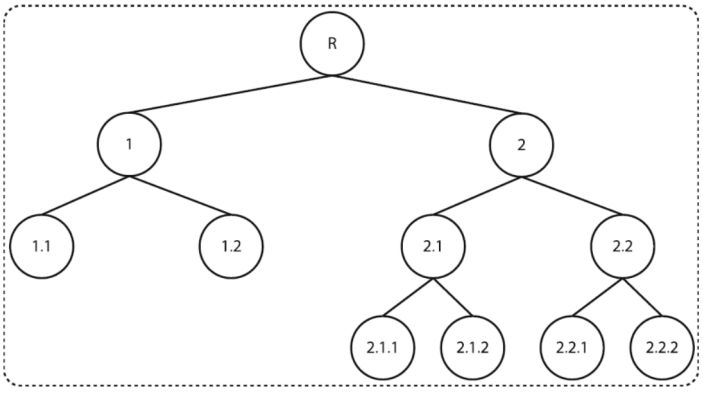


Figura 2.8 – Abordagem por Classificação Global
Fonte: (SILLA JR; FREITAS, 2012, p. 47)

Enquanto as abordagens locais com a estratégia de predição *top-down* têm a desvantagem de propagar o erro de classificação cometido num determinado nível da hierarquia para

os demais níveis mais profundos, a abordagem global evita esse problema realizando a classificação em uma única etapa utilizando um único classificador.

Desse modo, observa-se que a abordagem global perde a natureza modular das abordagens locais, ou seja, a característica de dividir a fase de treinamento em diversos processos, cada um considerando parte da hierarquia de classes. Portanto, o único classificador construído na abordagem global tende a ser mais complexo do que cada classificador individual construído nas abordagens locais. No entanto, essa natureza modular das abordagens locais não necessariamente implica numa superioridade de desempenho preditivo em relação à abordagem global.

2.4 Estimativa de Desempenho Preditivo

As métricas de precisão hierárquica (hP, *hierarchical precision*), revocação hierárquica (hR, *hierarchical recall*) e *f-measure* hierárquica (hF, *hierarchical f-measure*) podem ser utilizadas para estimativa do desempenho preditivo de um classificador hierárquico (SILLA JR, 2011). As equações a seguir definem o cálculo dessas métricas:

$$hP = \frac{\sum_i \mid Pi \cup Ti \mid}{\sum_i \mid Pi \mid} \tag{2.1}$$

$$hR = \frac{\sum_i \mid Pi \cup Ti \mid}{\sum_i \mid Ti \mid} \tag{2.2}$$

$$hF = \frac{2 \times hP \times hR}{hP + hR} \tag{2.3}$$

onde Pi é conjunto que contém a classe mais específica predita para o exemplo de teste i e todas suas classes antecessoras e Ti é o conjunto verdadeiro conhecido da classe mais específica para o exemplo de teste i e todas suas classes antecessoras. O símbolo " $\mid\mid$ " indica a cardinalidade do conjunto envolvido pelo mesmo.

A fim de ilustrar algumas questões associadas com os estimadores hP, hR e hF no contexto de generalização e especialização de erros, considere alguns exemplos hipotéticos a seguir os quais contemplam três casos de generalização de erros:

- A) Classe Predita: R.1, Classe Verdadeira Conhecida: R.1.2
- B) Classe Predita: R.1, Classe Verdadeira Conhecida: R.1.2.1
- C) Classe Predita: R.1, Classe Verdadeira Conhecida: R.1.2.1.1

Nesses casos os valores de hP e hR serão, respectivamente:

- A) hP = 1/1, hR = 1/2
- B) hP = 1/1, hR = 1/3
- C) hP = 1/1, hR = 1/4

Assim, é possível concluir que para uma dada classe verdadeira conhecida, quanto maior o erro de generalização (correspondente a classes verdadeiras conhecidas mais profundas), menor é o valor de hR, enquanto o valor de hP permanece constante.

Considere agora o seguinte caso de erro de especialização:

- Classe Predita: R.2.2, Classe Verdadeira Conhecida: R.2
- Classe Predita: R.2.2.1, Classe Verdadeira Conhecida: R.2
- Classe Predita: R.2.2.1.3, Classe Verdadeira Conhecida: R.2

Nesses casos os valores de hP e hR serão, respectivamente:

- D) hP = 1/2, hR = 1/1
- E) hP = 1/3, hR = 1/1
- F) hP = 1/4, hR = 1/1

Dessa forma, é possível concluir neste último caso que para uma dada classe verdadeira conhecida, quanto maior o erro de especialização (correspondente a um classe predita mais profunda), menor o valor de hP, enquanto o valor de hR permanece constante.

Em síntese, o estimador hF, o qual agrega hP e hR em uma única fórmula, mostra-se apto para penalizar efetivamente ambos os erros (generalização e especialização).

2.5 KNN:K-Nearest Neighbor

A fim de facilitar o entendimento das estratégias adotadas no método proposto, o HKNN, realizou-se o seguinte resumo referente ao funcionamento do classificador KNN uma vez que o HKNN é baseado nele.

O método KNN foi originalmente proposto no início da década de 50. Este método consome muito recurso computacional quando processa bases de treinamento com grandes quantidades de dados, portanto não houve grande popularidade de seu uso até o século 60, quando se aumentou a disponibilidade de máquinas com potencial computacional superior a aquelas até então produzidas (HAN; KAMBER, 2012). Desde então, esse método tem sido largamente utilizado na área de Mineração de Dados. Esse tipo de classificador baseia-se no aprendizado por analogia, ou seja, através da comparação de instâncias. As instâncias da base de dados são descritas por meio de n atributos. Portanto, cada instância representa um ponto em um espaço n-dimensional e pode ser representada por um vetor de tamanho n, ou seja, para uma instância V tem-se o vetor $V = \{v_1, v_2, ..., v_n\}$, onde $v_1, v_2, ..., v_n$ correspondem aos valores dos seus n atributos $A_1, A_2, ..., A_n$.

Desse modo, ao ser fornecida uma instância para ser classificada (uma instância cuja classe é desconhecida), o KNN calcula a proximidade dessa instância em relação àquelas existentes na base de dados (instâncias cujas classes são conhecidas) e contabiliza as k instâncias mais similares (próximas), as quais são conhecidas como os k-vizinhos mais próximos. Posteriormente, a classe mais frequente entre esses k-vizinhos mais próximos é atribuída à instância que se deseja classificar. Esse processo de classificação repete-se sucessivamente para cada nova instância a ser classificada.

A noção de proximidade que se leva em consideração para avaliação dos k-vizinhos mais próximos é definida por meio de uma métrica de distância, como por exemplo, distância Euclidiana, distância de Manhattan, dentre outras. Dessa forma, para calcular a distân-

cia euclidiana entre duas instâncias, $V = \{v_1, ..., v_n\}$ e $Z = \{z_1, ..., z_n\}$, faz-se uso da Equação 2.4.

$$Dist.Euclidiana(V, Z) = \sqrt{\sum_{i}^{n} (v_i - z_i)^2} \tag{2.4}$$

Em outras palavras, o cálculo da distância Euclidiana entre duas instâncias V e Z envolve, para cada atributo, o cálculo da diferença entre os valores correspondentes do atributo para V e Z. Normalmente, os valores dos atributos contínuos são normalizados antes de se aplicar a equação da distância Euclidiana, o que ajuda a prevenir que atributos com escala muito grande sobreponham-se àqueles com escala pequena. A normalização mínimo-máximo, por exemplo, pode ser utilizada para colocar os valores de um atributo contínuo no intervalo [0,1].

No caso atributos categóricos (ou nominais), outra estratégia se faz necessária para calcular a diferença entre os valores dos atributos. Nesse caso, um método simples é comparar os valores correspondentes do atributo para as instâncias V e Z. Se ambos são idênticos, então a diferença entre eles é igual a menor diferença possível. Caso os dois valores sejam diferentes, então a diferença é igual a máxima diferença.

No caso em que se têm valores de atributos desconhecidos, aplica-se a seguinte estratégia: se existem dois valores de atributos para os quais está sendo calculada a diferença e um deles não possui valor definido, considera-se a diferença máxima possível. No caso em que ambos não apresentam valores definidos, também considera-se a diferença máxima possível.

3 O MÉTODO PROPOSTO

O método proposto, o HKNN, é basicamente uma modificação do algoritmo KNN que possibilita que este passe a considerar a hierarquia de classes ao lidar com problemas de classificação hierárquica. A seguir, a Seção 3.1 apresenta o classificador proposto.

3.1 HKNN: Hierarchical K-Nearest Neighbor

O método proposto, denominado HKNN, é capaz de lidar com problemas de classificação hierárquica onde a hierarquia de classes é organizada em árvore. Além disso, considera-se que as predições podem ser realizadas em qualquer nó (classe) da hierarquia (*non-mandatory lead node prediction*) e que a uma dada instância somente pode ser associada uma classe (classificação monorrótulo) (Seção 2.3).

Dessa forma, dada uma nova instância a qual deseja-se classificar, o HKNN executa o mesmo procedimento empregado no KNN, ou seja, utiliza uma métrica de distância para obtenção dos k-vizinhos mais próximos da instância a ser classificada.

A Figura 3.1 ilustra o cálculo do vetor de distância considerando uma base de dados fictícia. Nessa figura, as colunas representam as instâncias da base de dados (variando de TRA_1 à TRA_n) e a linha refere-se à instância que se deseja classificar ($NOVA\_i$).

Figura 3.1 – Valores do Vetor de Distâncias

Posteriormente ao cálculo do vetor de distâncias, realiza-se seu ordenamento a fim de que os k-vizinhos mais próximos possam ser reconhecidos e utilizados na tarefa de classificação da nova instância. Na Figura 3.2 está ilustrado o ordenamento dos valores do vetor de distâncias da Figura 3.1.

Concluído o ordenamento do vetor de distâncias, o HKNN realiza a classificação proces-

	TRA_6	TRA_2	TRA_1	TRA_3	TRA_4	TRA_5
NOVA_i	0.165	1.189	1.500	1.569	2.968	9.699

Figura 3.2 – Vetor de Distâncias Ordenado

sando cada nível hierárquico das classes desses k-vizinhos mais próximos. O processamento inicia-se no primeiro nível hierarquia e segue sua execução fundamentando-se em dois critérios: critério de seleção por classe mais frequente no nível (1º Critério) e critério de seleção por média das distâncias (2º Critério). O segundo critério é utilizado sempre que houver um empate para o primeiro critério.

Para cada nível hierárquico das classes dos k-vizinhos mais próximos, o HKNN tenta realizar a seleção da classe do nível em questão através do 1º Critério, ou seja, se existir somente uma classe mais frequente no nível em análise, esta classe e todas as suas descendentes no conjunto dos k-vizinhos mais próximos são selecionadas para análise do nível seguinte. Caso o nível em análise possua mais de uma classe como sendo a mais frequente (empate), aplica-se o 2º Critério, ou seja, calcula-se a média das distâncias (existentes no vetor da Figura 3.1) para as instâncias associadas à cada uma das classes para as quais ocorreu um empate no 1º Critério. Em seguida, seleciona-se a classe associada à menor distância média e todas as suas descendentes no conjunto dos k-vizinhos mais próximos para análise do nível seguinte. Em caso de empate no 2º Critério, a classe selecionada para continuidade da análise é escolhida de forma aleatória.

Para melhor entendimento do funcionamento dos dois critérios adotados considere o seguinte caso baseado no exemplo da Figura 3.2:

- Quantidade de vizinhos mais próximos igual a quatro (k = 4);
- V' = { TRA_6; TRA_2; TRA_1; TRA_3 }, conjunto dos quatro vizinhos mais próximos ordenado pela distância;
- C' = { R.01.07.01; R.01.01.09.06; R.01.06; R.32.01 }, conjunto de classes dos vizinhos contemplados em V', onde a notação R.01.07.01 indica que a instância em questão está associada à classe 01 no primeiro nível da hierarquia, a classe 07 no segundo nível da hierarquia e a classe 01 no terceiro nível da hierarquia;
- D' = { 0,165; 1,189; 1,500; 1,569 }, conjunto das distâncias das instâncias de V' em relação à instância $NOVA\_i$;

A Figura 3.3 ilustra a aplicação dos critérios de seleção de classes para o exemplo em questão. Inicialmente, através do 1º Critério verifica-se que a classe 01 é mais frequente no primeiro nível, logo as classes R.01.07.01, R.01.01.09.06, R.01.06 são selecionadas para o processamento do segundo nível. No segundo nível hierárquico, quando o 1º Critério é aplicado, verifica-se que existe um empate com relação à frequência das classes 07, 01 e 06, portanto, o 2º Critério é aplicado. De acordo com o 2º Critério, a menor distância ocorre para a instância pertencente à classe 07 (no segundo nível). Neste caso, não houve a necessidade de se calcular médias de distâncias pelo fato de existir apenas uma instância para cada uma das três classes (07, 01 e 06) analisadas no segundo nível. Portanto, para análise do terceiro nível restou apenas a classe R.01.07.01, fazendo com que esta seja a classe atribuída à instância $NOVA\_i$.

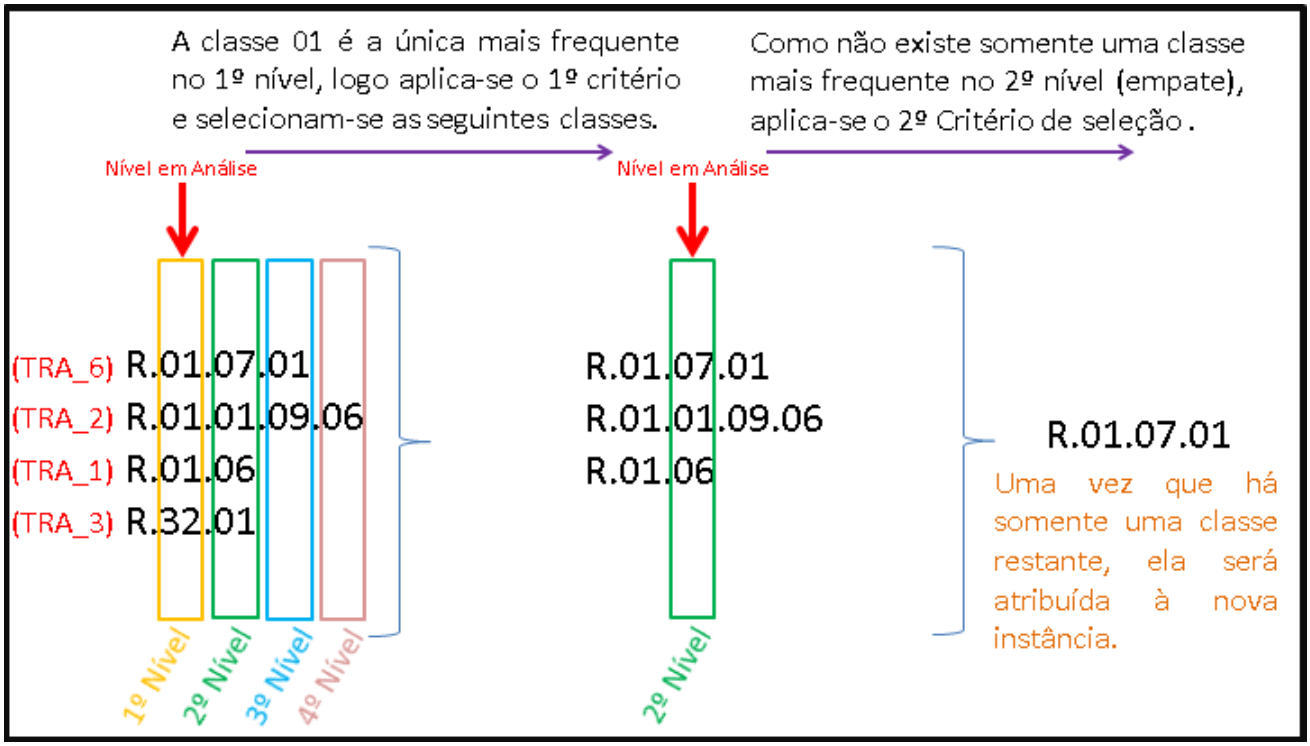


Figura 3.3 – Aplicação dos Critérios de Seleção

3.1.1 HKNN com Critério de Parada

Visando aperfeiçoar o desempenho do HKNN, dois critérios de parada, intitulados "1xDIST" e "2xDIST", foram adicionados ao método. Com esses critérios não necessariamente todos os níveis hierárquicos associados às classes dos k-vizinhos mais próximos serão analisados.

A ideia desses critérios baseia-se na quantidade de vezes que o 2º Critério de seleção é usado. No caso do 1xDIST, admite-se que o 2º Critério seja utilizado no máximo uma vez durante o processamento dos níveis hierárquicos. Sendo assim, o HKNN interrompe

o seu processamento toda vez que existe um empate no 1º Critério pela segunda vez. Já no caso do 2xDIST, o processamento do HKNN é interrompido toda vez que existe um empate no 1º Critério pela terceira vez, ou seja, permite-se que o 2º Critério seja utilizado no máximo duas vezes ao longo do processamento dos níveis hierárquicos.

A Figura 3.4 ilustra a aplicação do HKNN 1xDIST. Considerando que as classes associadas aos 6 vizinhos mais próximos da instância que se deseja classificar são R.01.07.01, R.01.01.09.06, R.32.09, R.32.01, R.03.08.01 e R.03.02, o HKNN 1xDIST funciona da maneira descrita a seguir: inicialmente, através do 1º Critério verifica-se que para o primeiro nível hierárquico existe um empate com relação à frequência das classes 01, 32 e 03, portanto, o 2º Critério é aplicado. De acordo com o 2º Critério, a menor distância ocorre para a instância pertencente à classe 01 (no primeiro nível), portanto as classes R.01.07.01 e R.01.01.09.06 são selecionadas para o processamento do segundo nível. No segundo nível hierárquico, quando o 1º Critério é aplicado, verifica-se que existe novamente um empate em relação à frequência das classes 07 e 01, portanto, como esse é o segundo empate, o HKNN interrompe o seu processamento e atribui a classe R.01, classe mais frequente no nível anterior, à instância $NOVA\_i$.

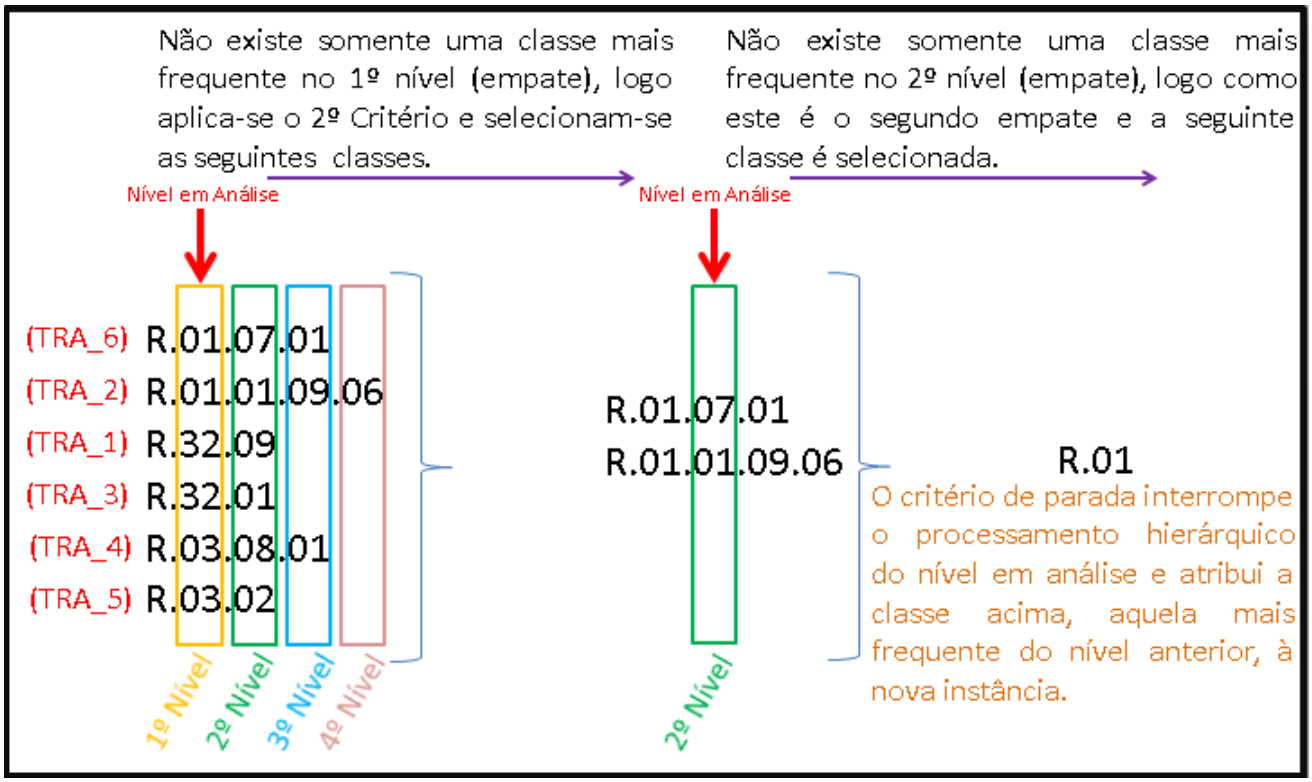


Figura 3.4 – Aplicação do Critério de Parada 1xDIST

4 AVALIAÇÃO EMPÍRICA

Experimentos computacionais foram realizados para se avaliar empiricamente o classificador proposto. A seguir, a Seção 4.1 apresenta as características das bases de dados. Em seguida, a Seção 4.2 descreve os métodos utilizados no estudo comparativo e a Seção 4.3 apresenta a avaliação do HKNN. Por fim, a Seção 4.4 apresenta os resultados da análise comparativa.

4.1 Bases de Dados

Os experimentos foram executados utilizando-se dois conjuntos de bases de dados. O primeiro conjunto é constituído por oito bases de dados hierárquicas de funções de proteínas[1]. Essas oito bases também foram utilizadas em outros trabalhos encontrados na literatura (SILLA JR; FREITAS, 2012; SILLA JR; FREITAS, 2011). As especificações desse primeiro conjunto de bases de dados podem ser visualizadas na Tabela 4.1. Nessa tabela, a primeira coluna contém os nomes das bases de dados, a segunda apresenta a quantidade de atributos, a terceira traz o número de instâncias e, por fim, a quarta apresenta a quantidade de classes por nível.

Tabela 4.1 – Caracterização das Bases de Dados de Funções de Proteínas

Base de Dados	**#Atributos**	**#Instâncias**	**#Classes por Nível**
EC-Interpro	1.217	14.027	18/21/27/14/8/3
EC-Pfam	709	13.987	18/21/27/14/9/3
EC-Prints	383	14.025	18/22/25/14/9/2
EC-Prosite	586	14.041	18/20/24/10/8/3
GPCR-Interpro	451	7.444	18/21/26/14/8/3
GPCR-Pfam	76	7.053	18/21/27/14/8/3
GPCR-Prints	284	5.404	18/21/25/14/8/3
GPCR-Prosite	130	6.246	18/20/24/11/9/2

O segundo conjunto é formado por dez bases de dados que descrevem o genoma de leveduras[2]. Para todas essas bases, os atributos contínuos foram discretizados e normalizados numa etapa de pré-processamento. Trabalhos nos quais essas bases também foram utilizadas podem ser encontrados na literatura, a saber Clare e King (2003) e Vens et al.

[1] Disponível em: https://sites.google.com/site/carlossillajr/resources/
[2] Disponível em: http://dtai.cs.kuleuven.be/clus/hmcdatasets/

(2008). As especificações desse segundo conjunto de bases de dados podem ser verificadas na Tabela 4.2, na qual as informações encontram-se organizadas da mesma forma que a Tabela 4.1.

Tabela 4.2 – Caracterização das Bases de Dados do Genoma de Leveduras

Base de Dados	**#Atributos**	**#Instâncias**	**#Classes por Nível**
CellCycle_single	78	3.757	18/21/27/14/8/3
Church_single	28	3.755	18/21/27/14/9/3
Derisi_single	64	3.725	18/22/25/14/9/2
Eisen_single	80	2.424	18/20/24/10/8/3
Expr_single	552	3.779	18/21/26/14/8/3
Gasch1_single	174	3.764	18/21/27/14/8/3
Gasch2_single	53	3.779	18/21/25/14/8/3
Phenotype_single	70	1.591	18/20/24/11/9/2
Sequence_single	479	3.919	18/22/27/14/9/2
SPO_single	81	3.703	18/22/27/14/9/2

4.2 Métodos Utilizados no Estudo Comparativo

A fim de avaliar o método proposto nesse trabalho (HKNN) e suas variações (HKNN 1xDIST e HKNN 2xDIST), foram implementados outros dois métodos para servirem como base de comparação, a saber "ALEATÓRIO 1" e "ALEATÓRIO 2". Além desses dois métodos, utilizou-se também o KNN, uma técnica proposta para resolver problemas de classificação plana. No caso do KNN, utilizou-se o algoritmo IBK implementado na ferramento Weka (*Waikato Environment for Knowledge Analysis*) (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

O método ALEATÓRIO 1 foi implementado para realizar o processo de classificação aleatóriamente. Dada uma instância a ser classificada, o método escolhe aleatoriamente uma classe da base de dados para ser atribuída à nova instância.

Por sua vez, o método ALEATÓRIO 2 realiza o processo de classificação utilizando a mesma noção de vizinhança adotada pelo KNN. Mais especificamente, dada uma instância a ser classificada, o método seleciona os seus k-vizinhos mais próximos e, a partir deles, escolhe aleatóriamente a classe a ser atribuída à nova instância.

4.3 Avaliação do HKNN

Para avaliação dos classificadores utilizou-se o método 10-validação cruzada estratificada. Segundo Han e Kamber (2006), esse método que é comumente utilizado na avaliação de classificadores, consiste em dividir cada base de dados original em 10 pares de bases de treinamento e teste. Nesse caso, 90% das instâncias contidas na base original compõem a base de treinamento e os 10% restantes a base de teste. Dessa forma, a base de dados de treinamento é utilizada na construção do classificador, enquanto a base de dados de teste é utilizada na estimativa do desempenho do mesmo.

Uma vez que o parâmetro k (número de vizinhos mais próximos) é o único parâmetro de entrada do HKNN, foram realizados experimentos para os seguintes valores de k: 1, 3, 5, 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 150, 300, 450, 600, 750 e 900, totalizando 21 experimentos para cada base de dados.

Por fim, a medida hF (ver Seção 2.4) foi adotada para reportar o desempenho dos classificadores considerados nos experimentos aqui realizados.

4.4 Análise Comparativa dos Resultados

Para avaliação do classificador proposto foram utilizados as bases mencionadas na Seção 4.1. Os gráficos apresentados a seguir (Figura 4.1 até 4.18) mostram os resultados obtidos por cada um dos métodos avaliados para diferentes valores do parâmetro k.

Para maioria das bases avaliadas, o HKNN ou pelo menos uma de suas variações apresentou desempenho preditivo melhor do que os demais métodos (KNN WEKA, ALEATÓRIO 1 e ALEATÓRIO 2) para todos os valores de k avaliados.

Para o primeiro conjunto de bases de dados (Tabela 4.1) os gráficos mostram que, geralmente, os melhores desempenhos preditivos são alcançados pelo HKNN (e suas variações), seguidos pelos métodos ALEATÓRIO 2, KNN WEKA e ALEATÓRIO 1, respectivamente.

Por sua vez, os resultados referentes ao segundo conjunto de bases de dados (ver Tabela 4.2) mostram que, de forma geral, os melhores desempenhos preditivos são obtidos pelo HKNN (e suas variações), seguido pelos métodos KNN WEKA, ALEATÓRIO 2 e ALEATÓRIO 1, respectivamente. Nesse conjunto de bases, uma exceção ocorreu para a

base Church_single (Figura 4.10) onde o desempenho preditivo do KNN WEKA foi superior a de todos os demais métodos. Por fim, as variações do HKNN (1xDIST e 2xDIST) têm desempenho sempre melhor ou igual ao do HKNN. Além disso, geralmente o HKNN 1xDIST tem desempenho melhor que o HKNN 2xDIST.

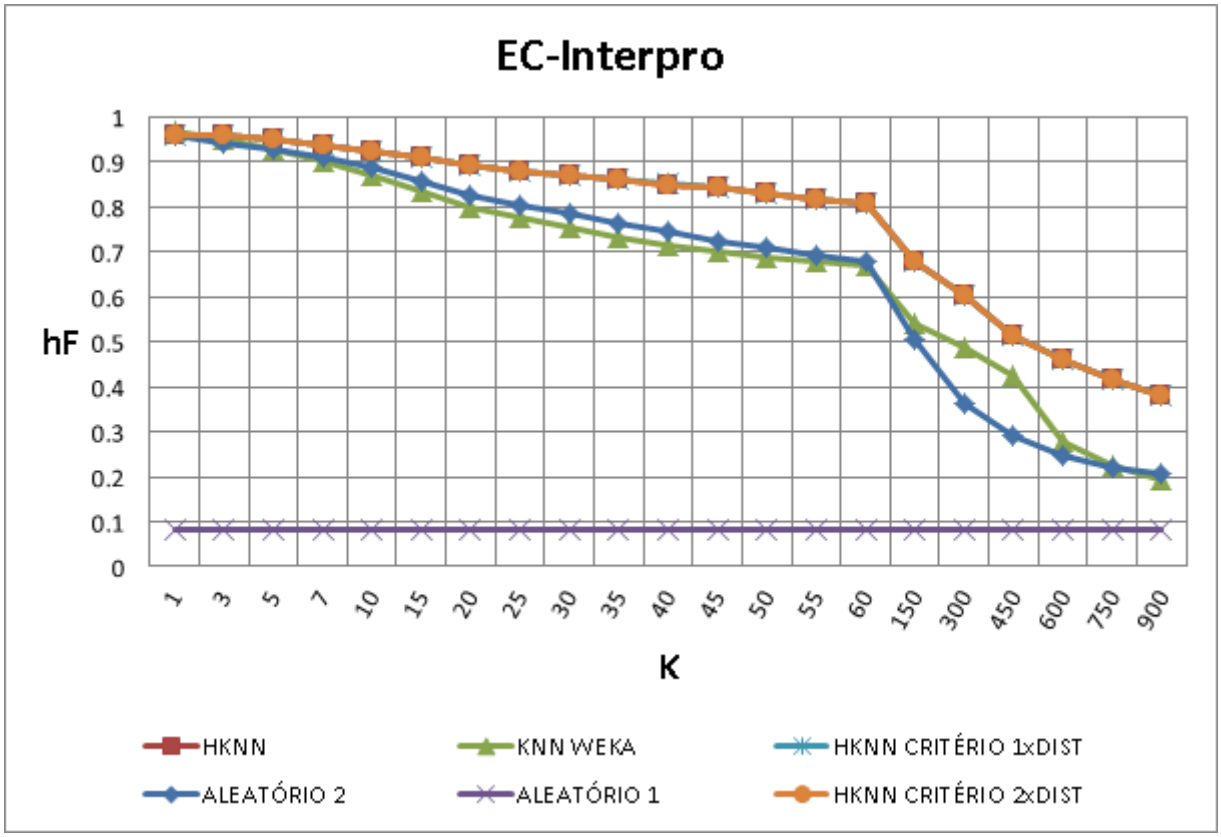


Figura 4.1 – Resultados para base de dados EC-Interpro

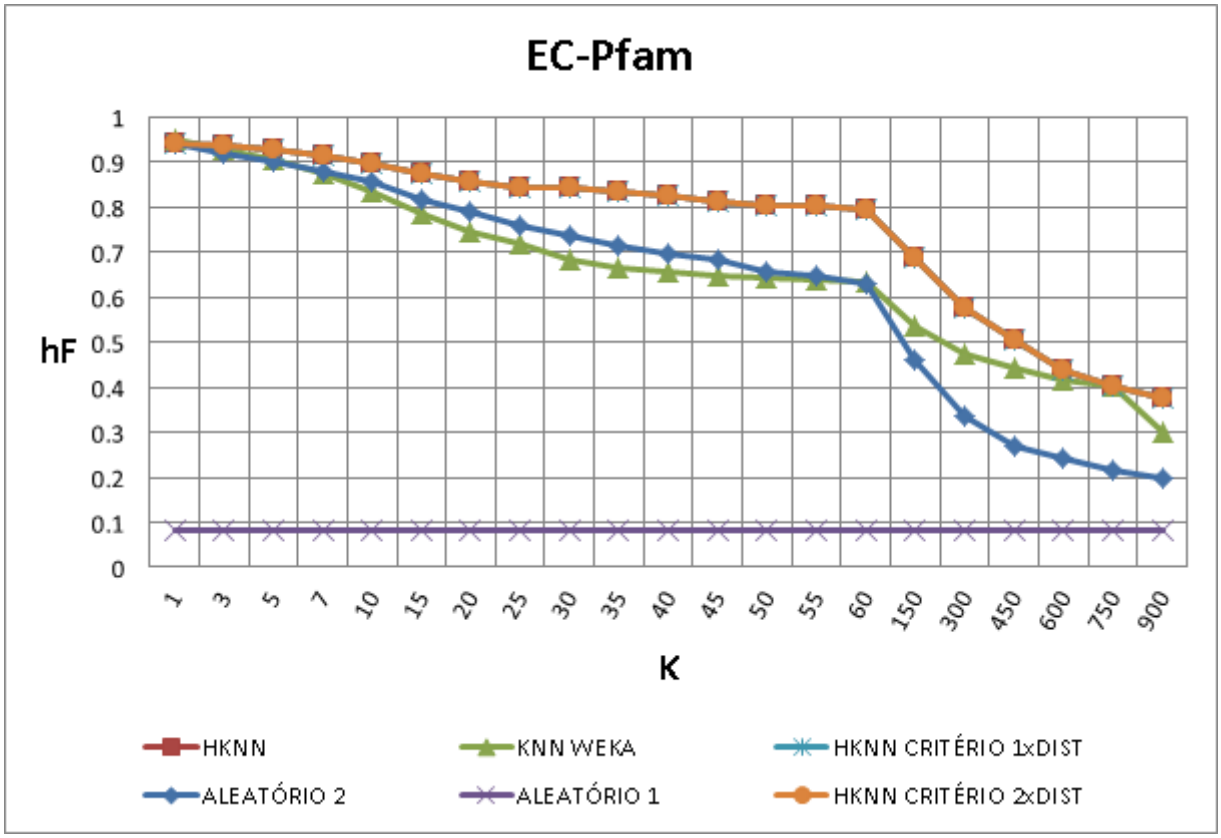


Figura 4.2 – Resultados para base de dados EC-Pfam

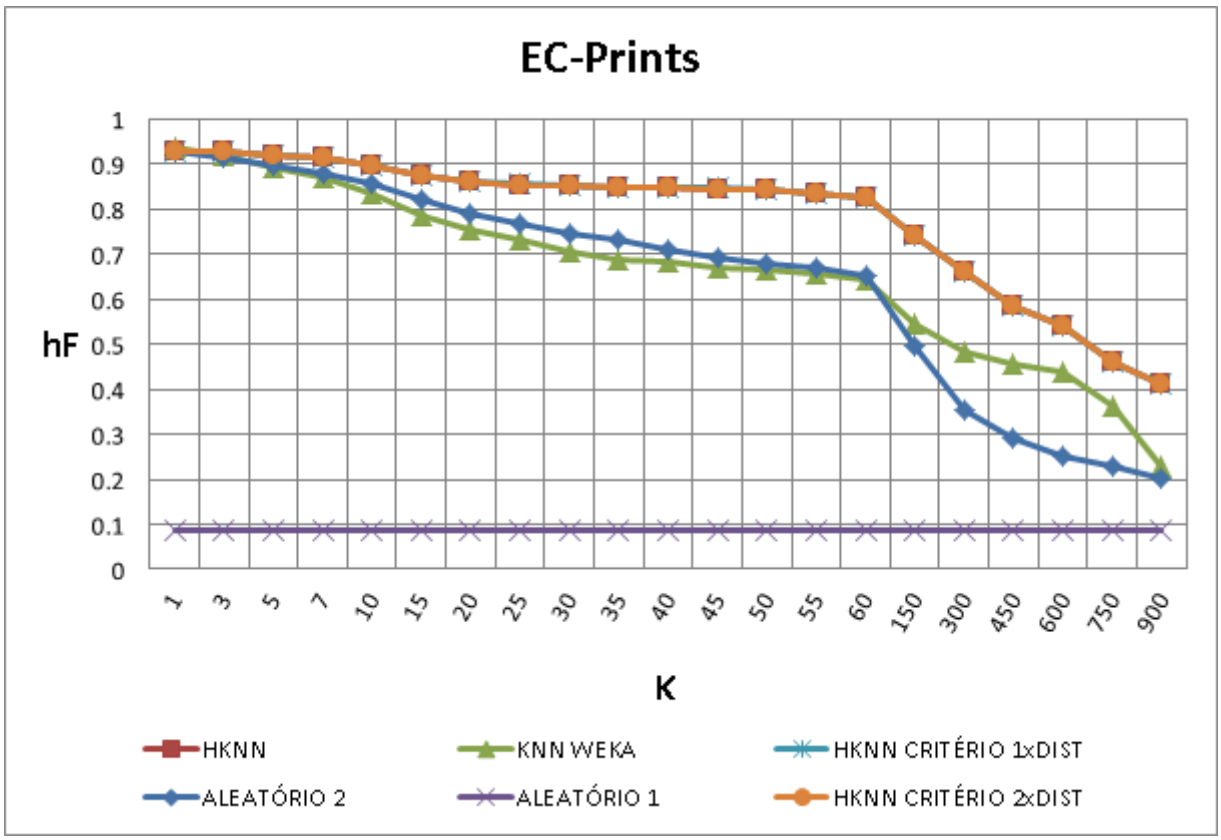


Figura 4.3 – Resultados para base de dados EC-Prints

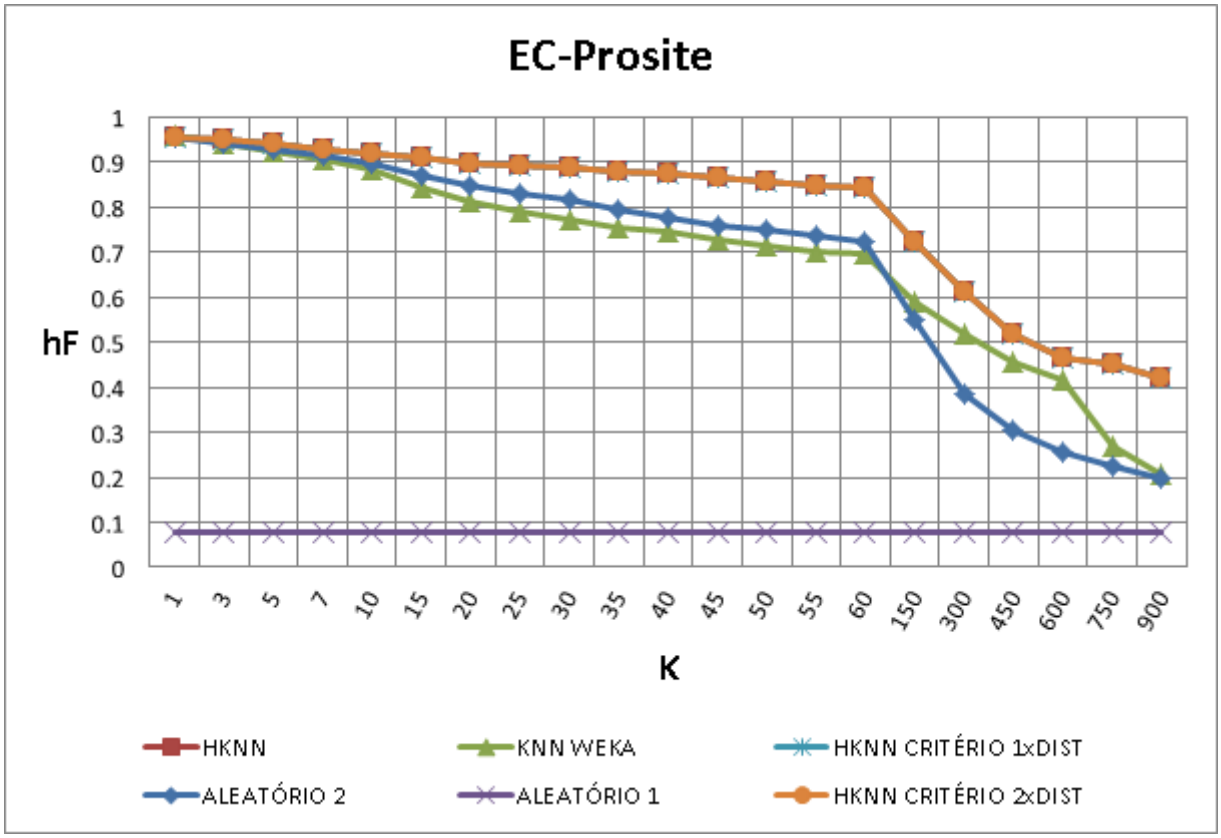


Figura 4.4 – Resultados para base de dados EC-Prosite

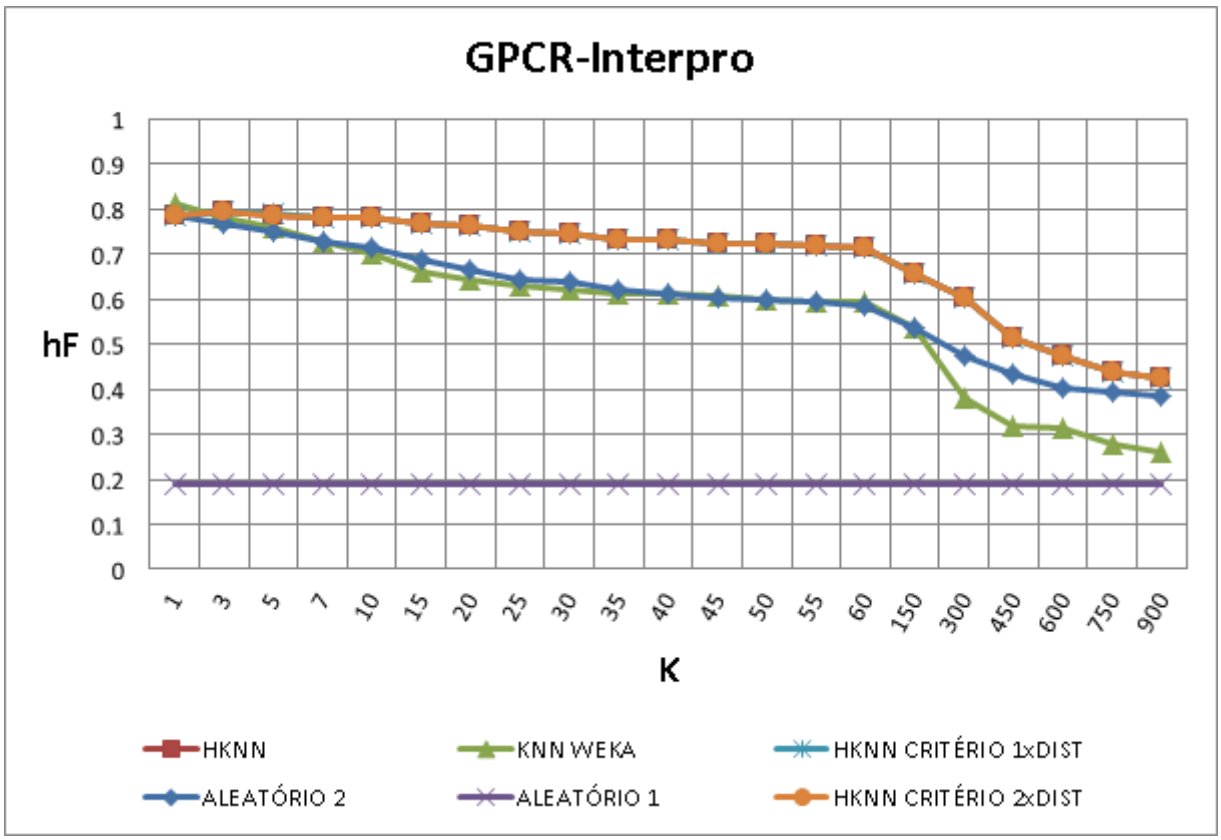


Figura 4.5 – Resultados para base de dados GPCR-Interpro

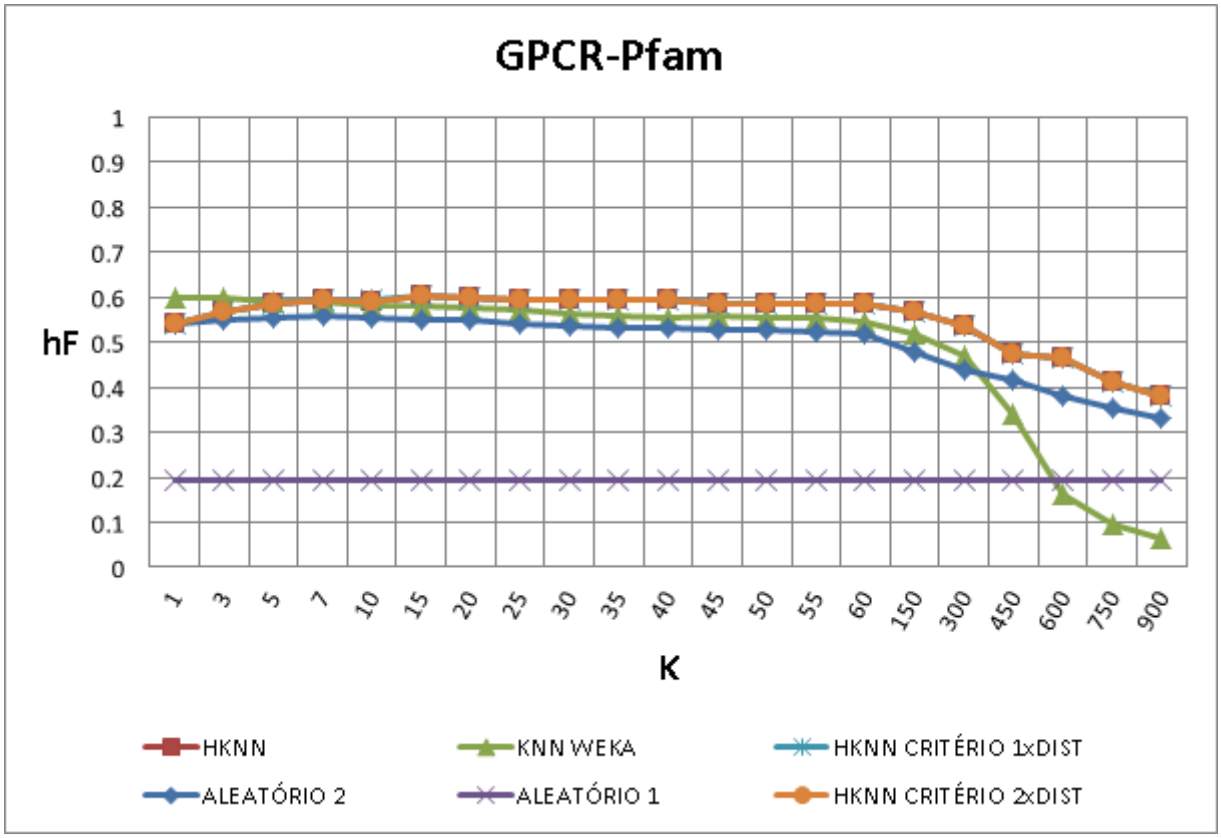


Figura 4.6 – Resultados para base de dados GPCR-Pfam

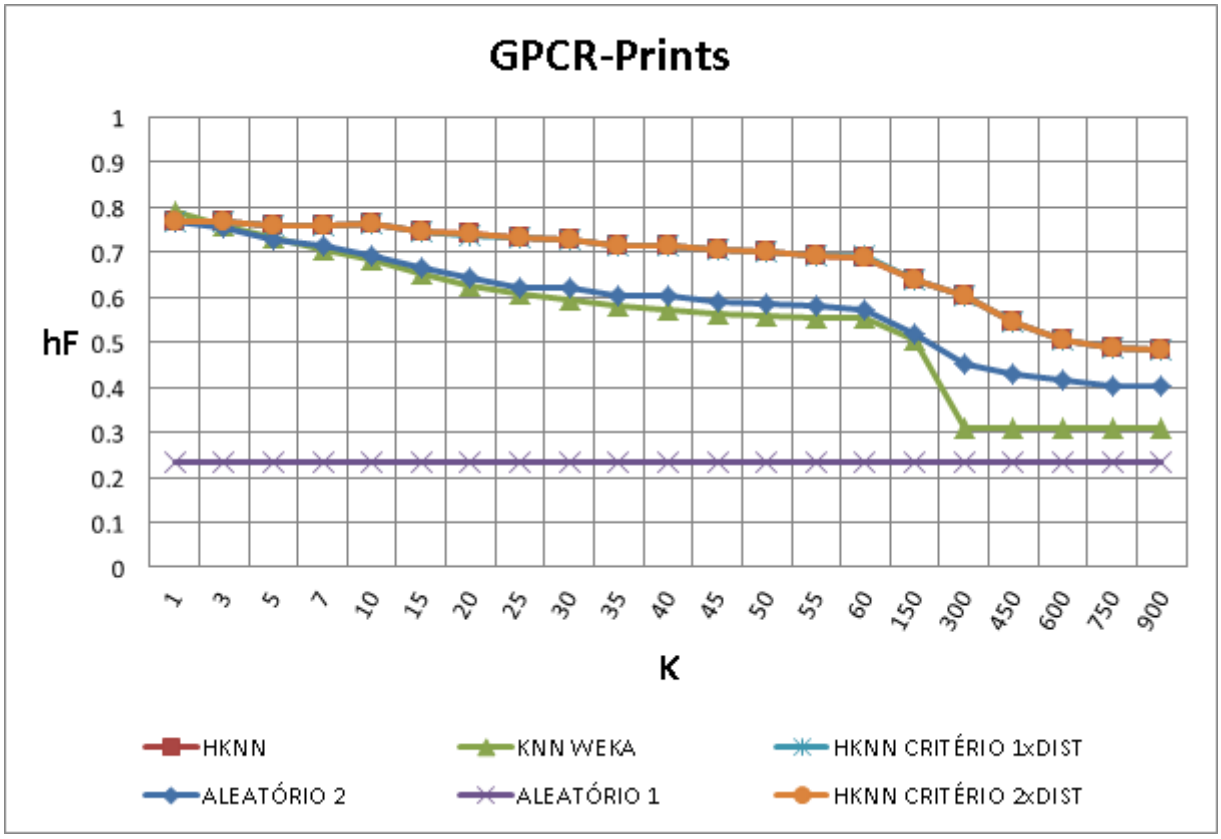


Figura 4.7 – Resultados para base de dados GPCR-Prints

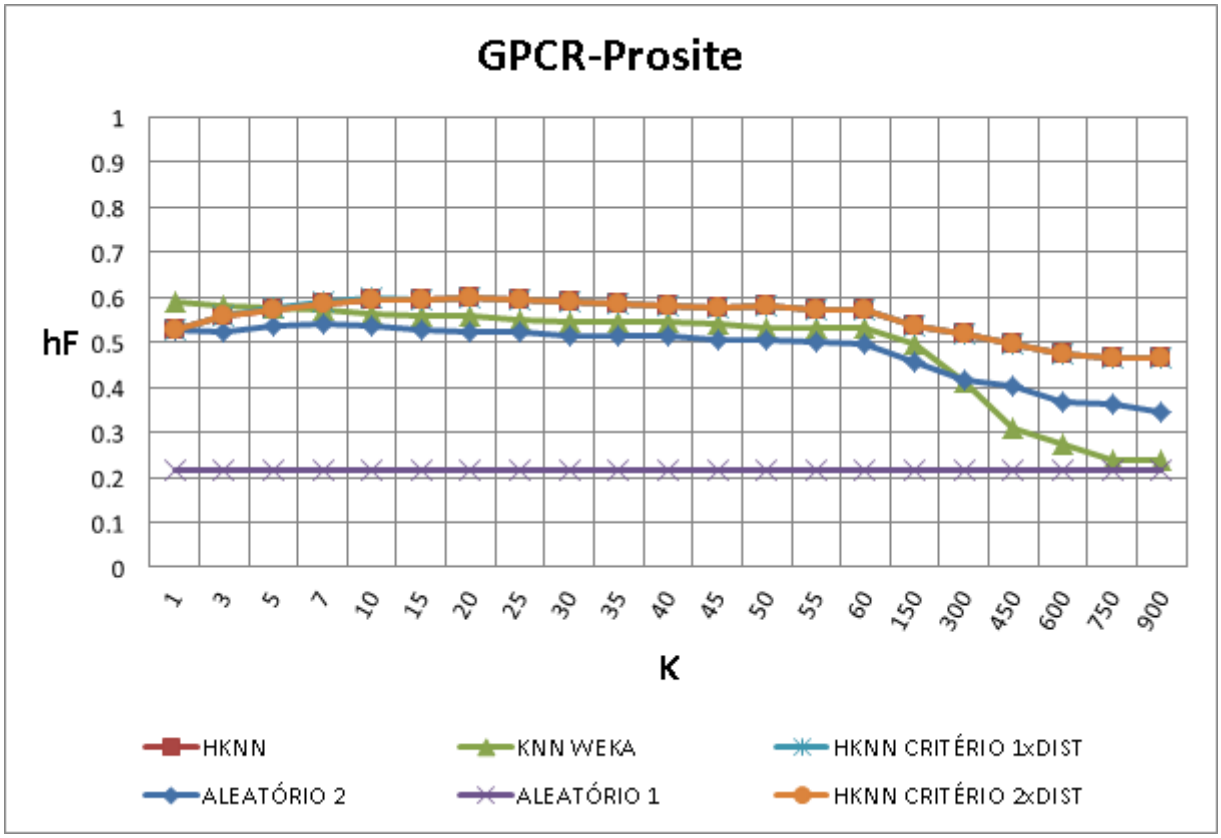


Figura 4.8 – Resultados para base de dados GPCR-Prosite

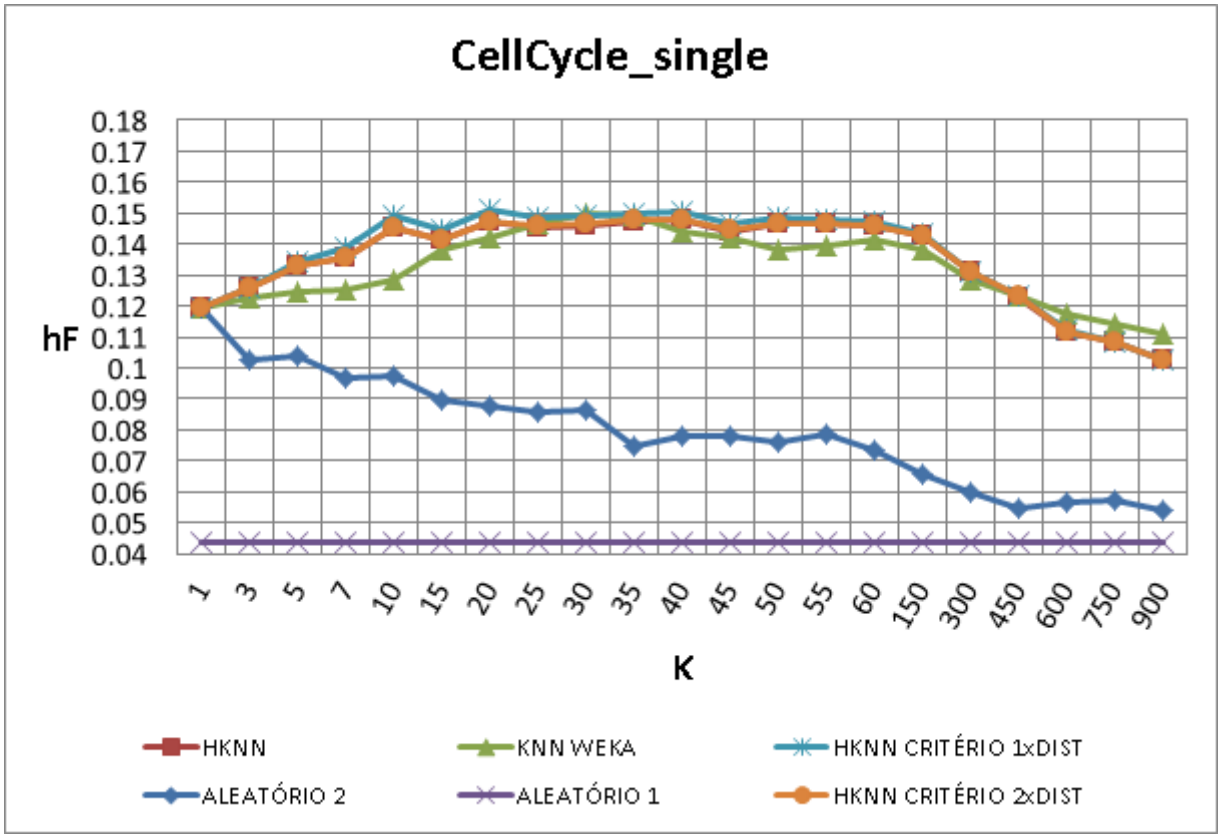


Figura 4.9 – Resultados para base de dados CellCycle_single

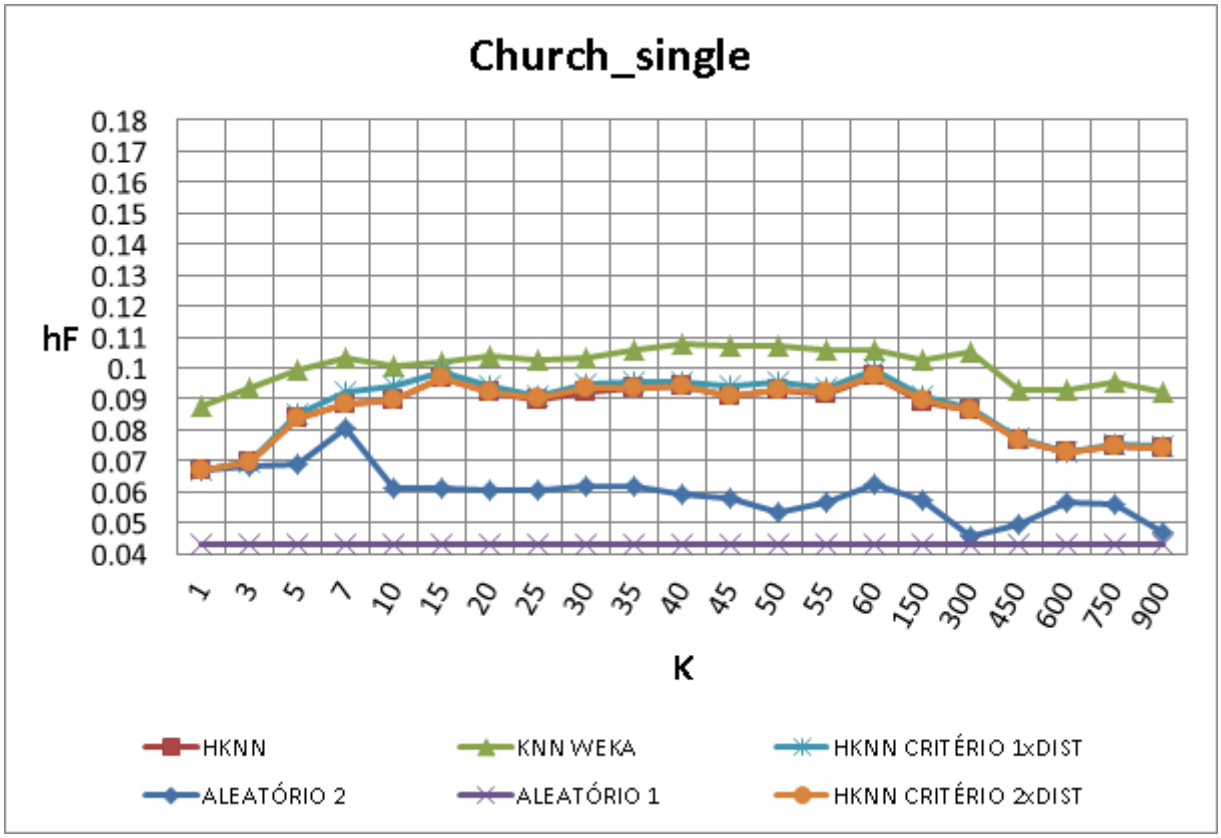


Figura 4.10 – Resultados para base de dados Church_single

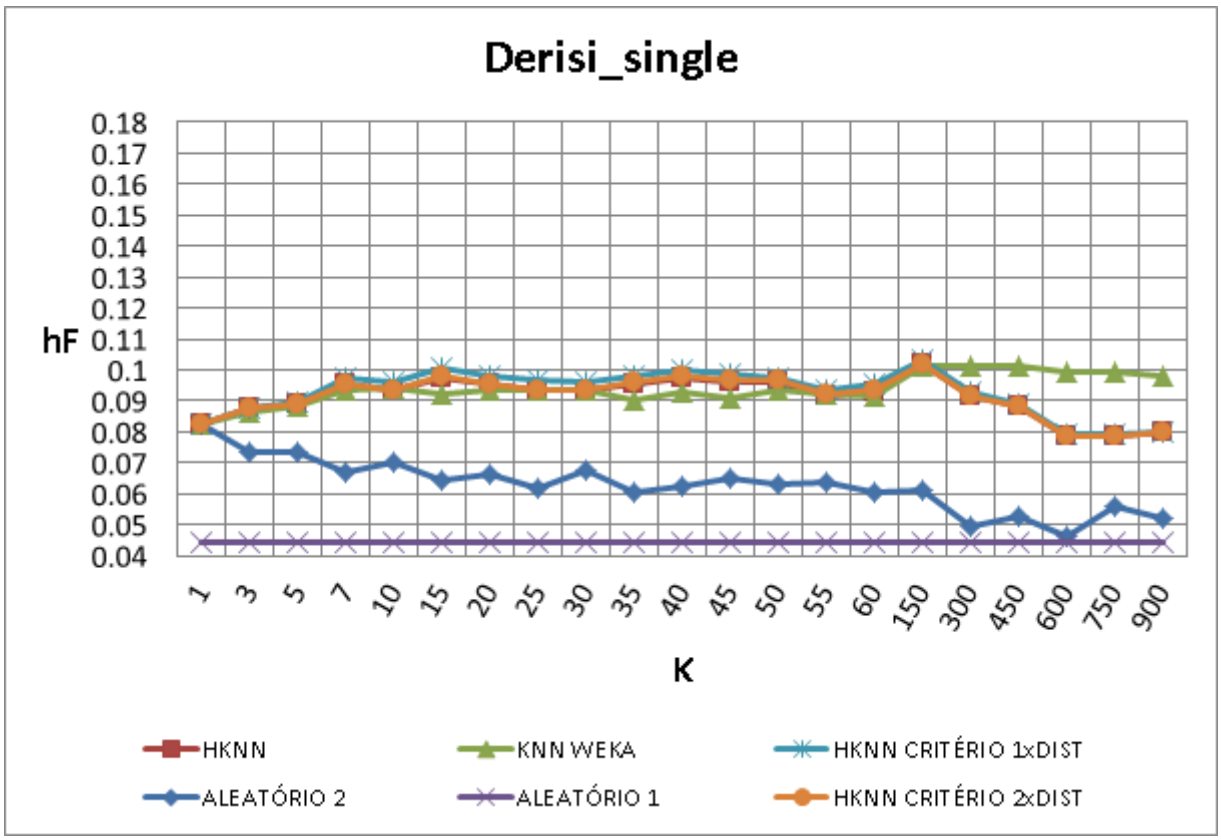


Figura 4.11 – Resultados para base de dados Derisi_single

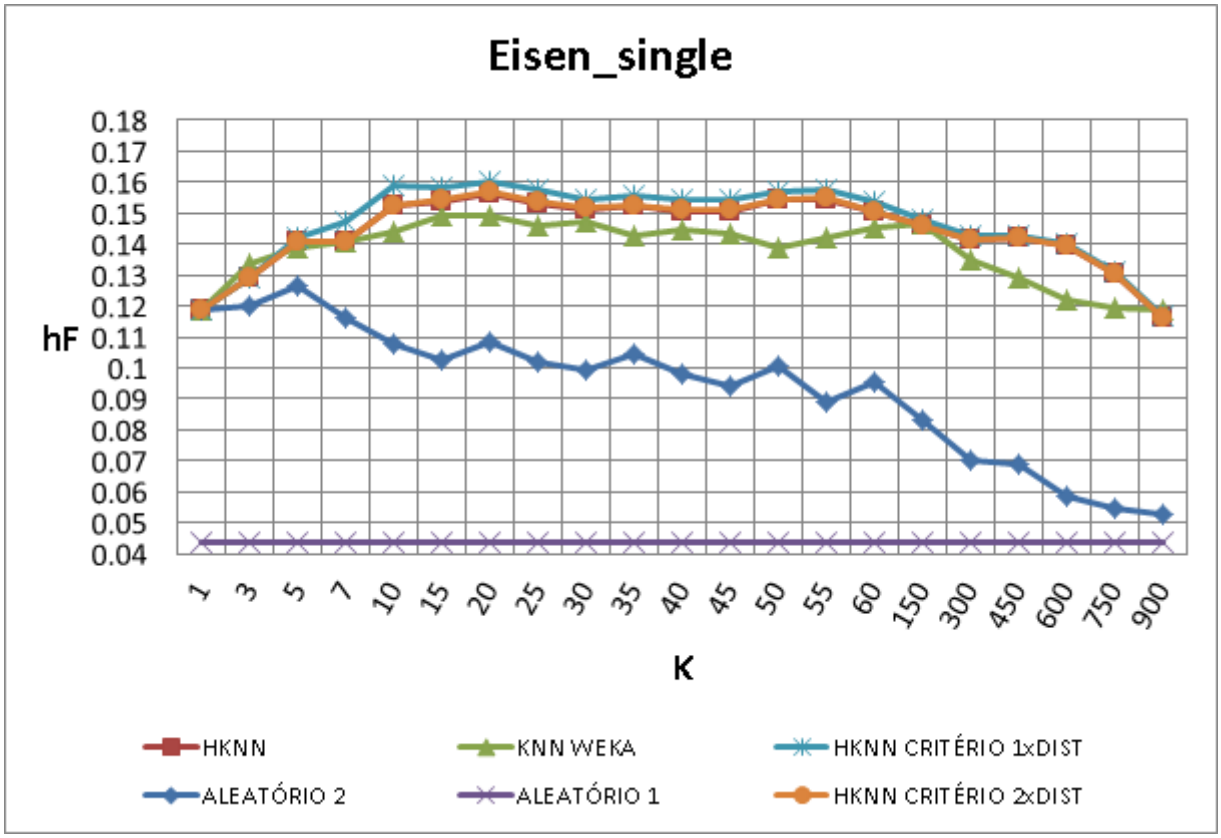


Figura 4.12 – Resultados para base de dados Eisen_single

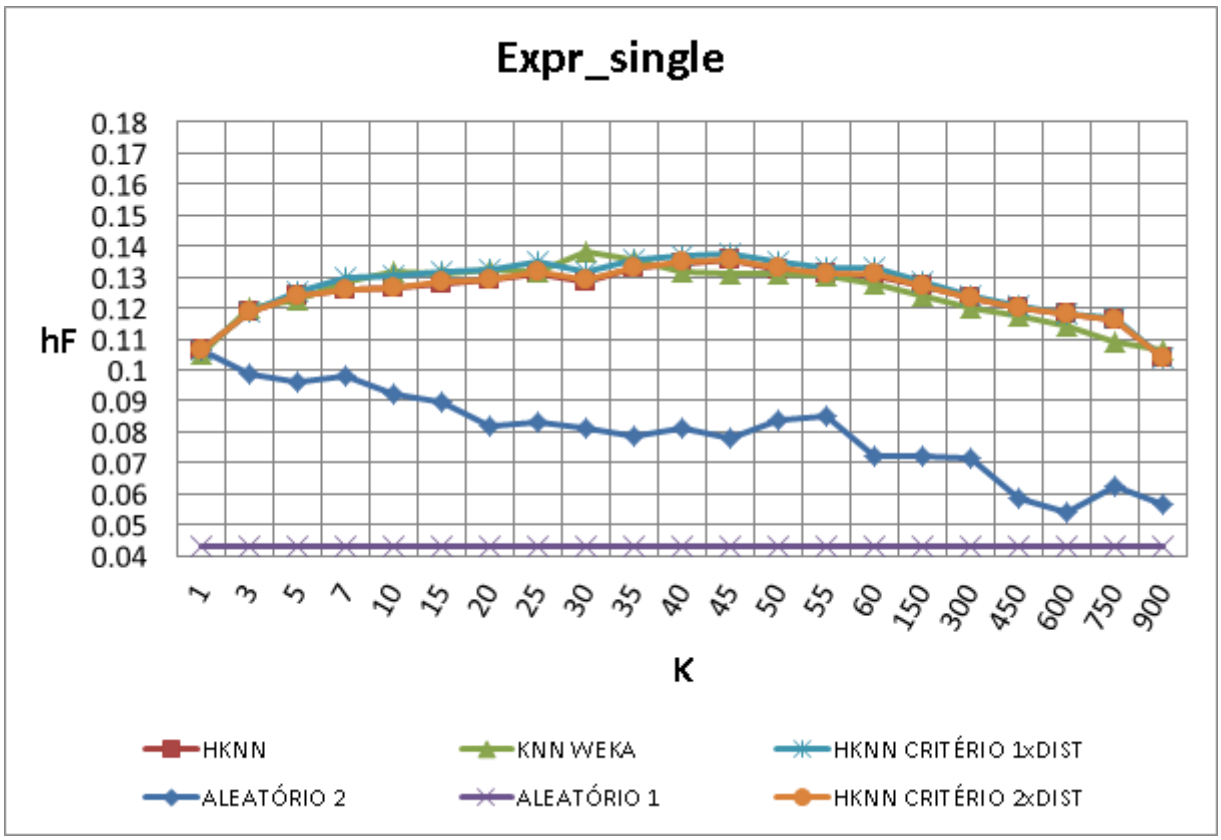


Figura 4.13 – Resultados para base de dados Expr_single

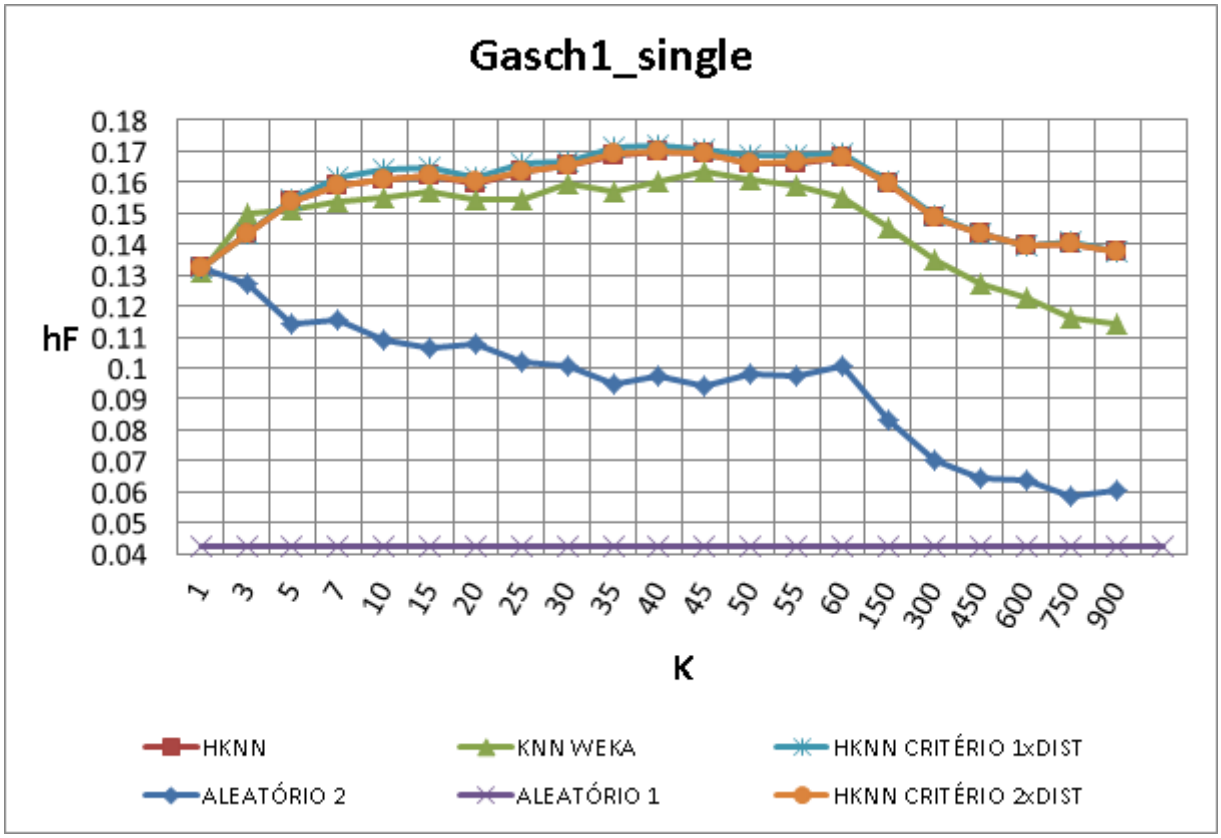


Figura 4.14 – Resultados para base de dados Gasch1_single

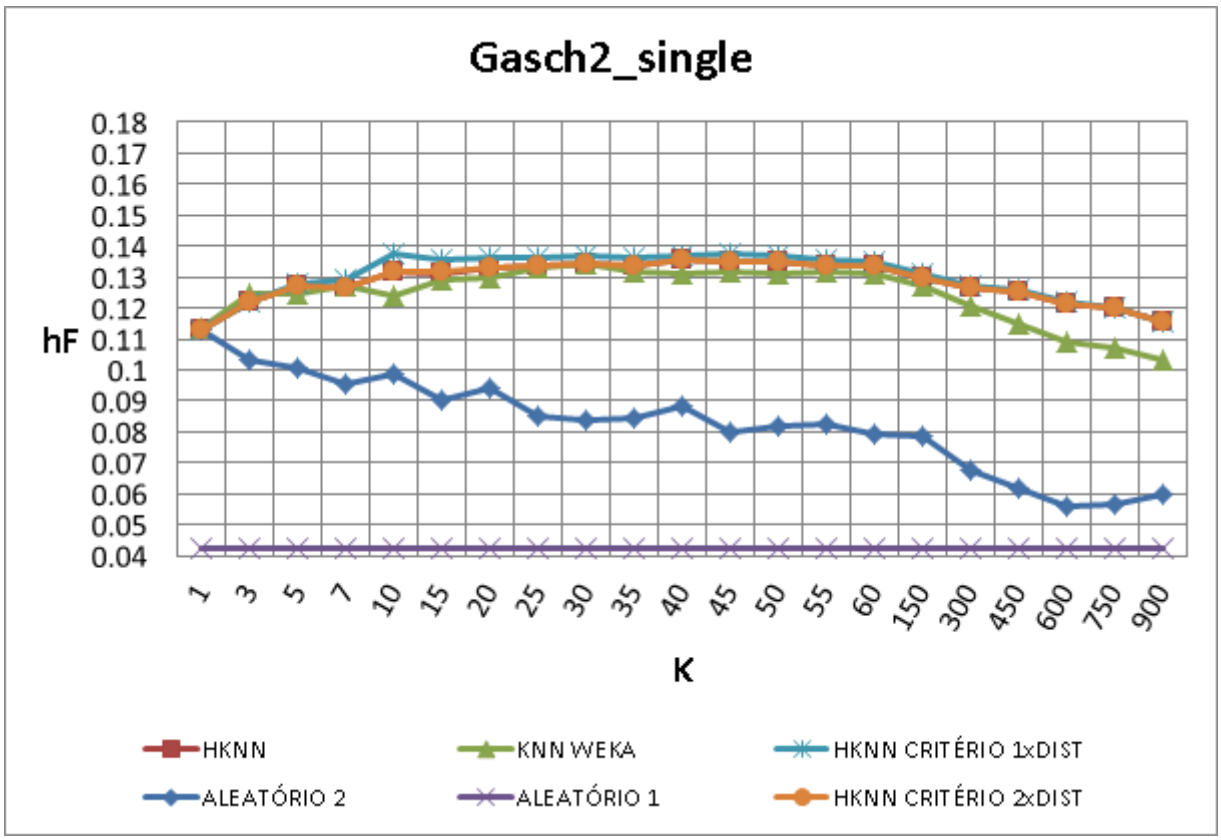


Figura 4.15 – Resultados para base de dados Gasch2_single

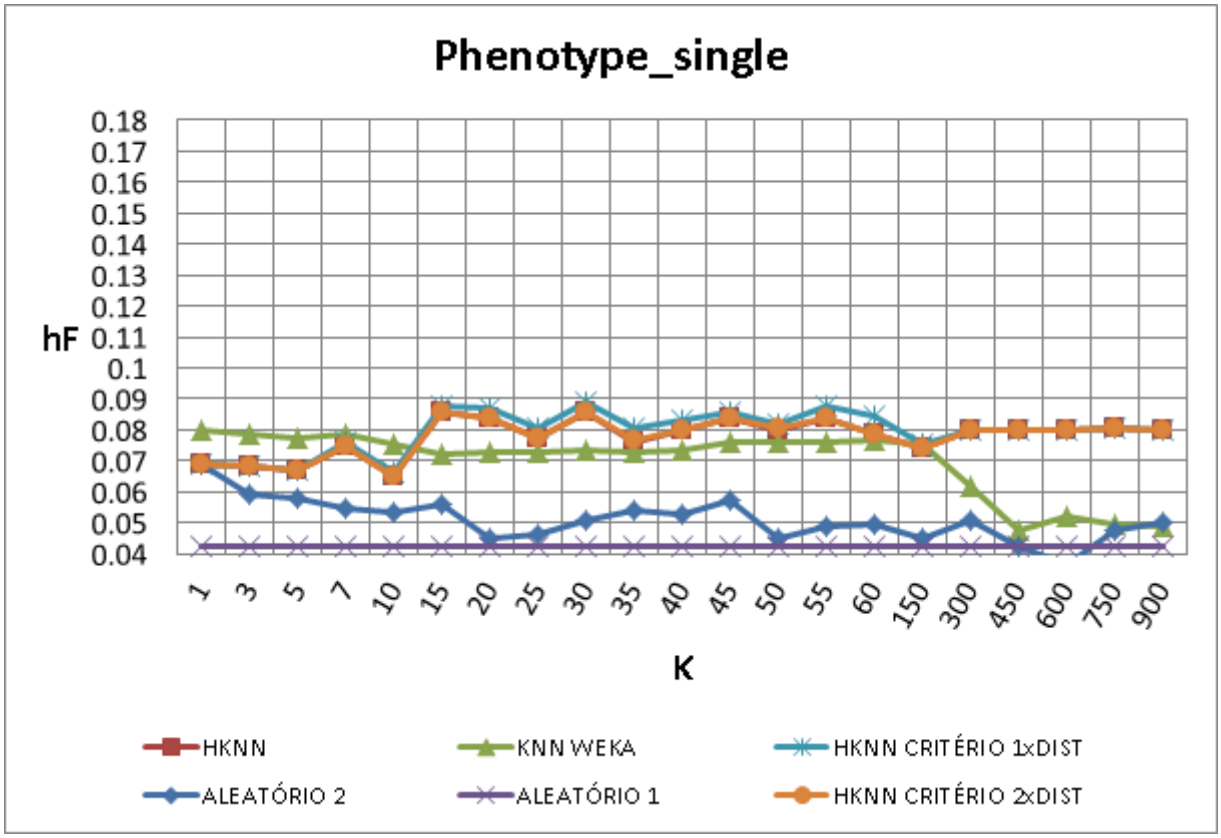


Figura 4.16 – Resultados para base de dados Phenotype_single

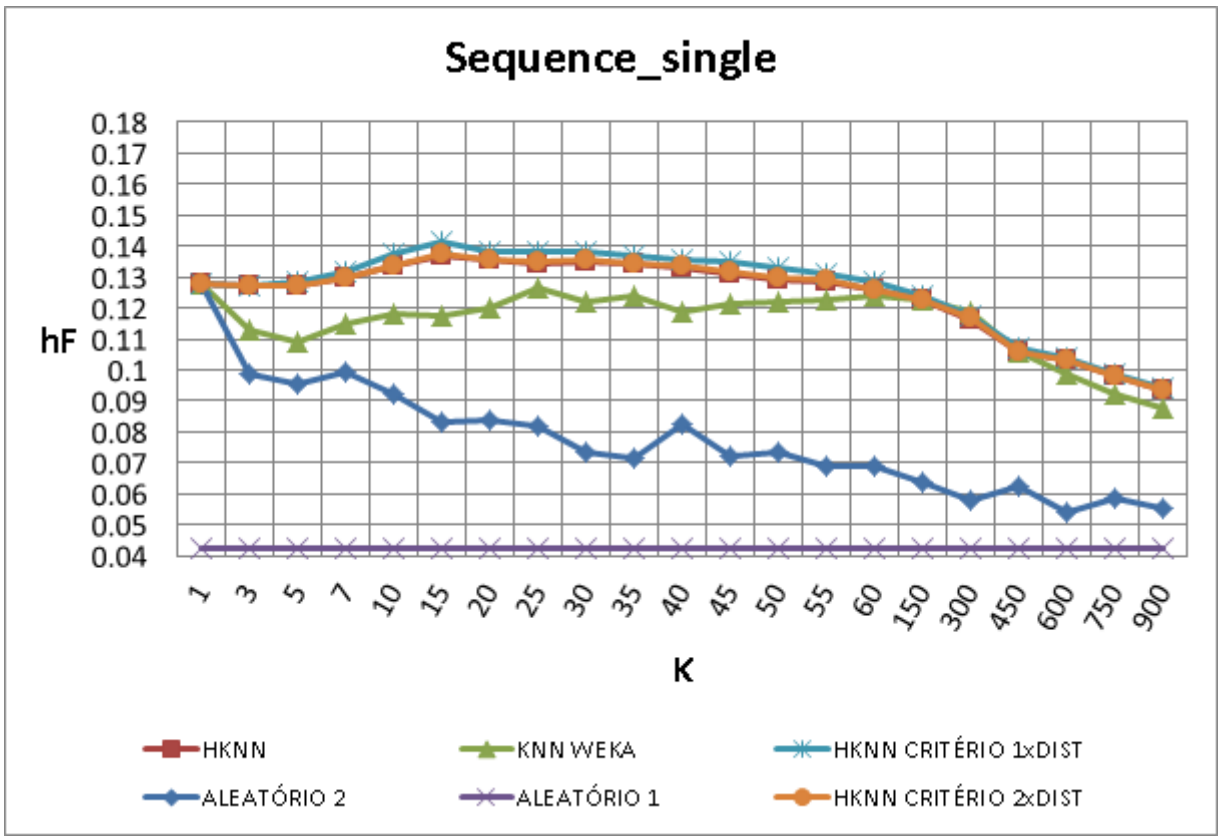


Figura 4.17 – Resultados para base de dados Sequence_single

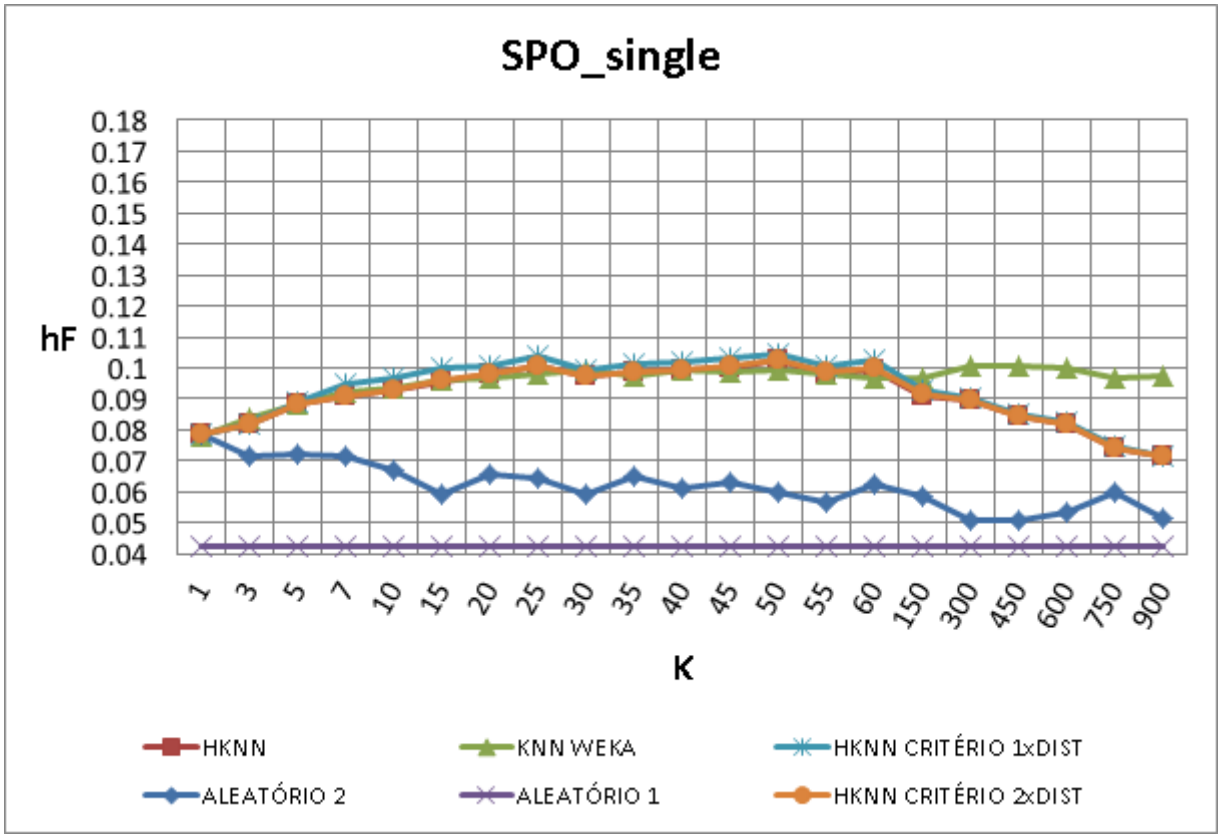


Figura 4.18 – Resultados para base de dados SPO_single

5 CONCLUSÕES

O presente trabalho apresentou um novo classificador hierárquico monorrótulo baseado na abordagem de classificação global. O classificador proposto utiliza a mesma noção de vizinhança adotada no KNN, no entanto, ele faz uso de estratégias que consideram as relações existentes entre as classes do problema.

Para o avaliação do classificador proposto (e suas variações) foram utilizados dois conjuntos de bases de dados: o primeiro refere-se às funções de proteínas e o segundo ao genoma de leveduras.

Como era de se esperar, o HKNN e suas variações obtiveram desempenhos preditivos sempre superiores aos dos métodos ALEATÓRIO 1 e ALEATÓRIO 2. Esse resultado serve para indicar que o método proposto consegue aprender com as instâncias das bases de dados de treinamento. Além disso, de modo geral, o HKNN e suas variações também apresentaram desempenhos superiores ao do KNN (IBK). Esse resultado comprova que a estratégia utilizada pelo HKNN é adequada para problemas de classificação hierárquica.

Como trabalho futuro, propõe-se a comparação do HKNN (classificador global) com classificadores locais que também utilizam o KNN com as abordagens por nó e por nó pai.

REFERÊNCIAS

AGRAWAL, R.; SRIKANT, R. Fast algorithms for mining association rules. In: **Proceedings of International Conference on Very Large Data Bases, VLDB**. Santiago, Chile: [s.n.], 1994.

AGRAWAL, R.; SRIKANT, R. Mining sequential patterns. In: **Proceedings of the International Conference on Data Engineering**. [S.l.: s.n.], 1995. p. 3–14.

CLARE, A.; KING, R. D. Predicting gene function in *saccharomyces cerevisiae*. In: **Proceedings of the European Conference on Computational Biology**. [S.l.: s.n.], 2003.

FIDENCIO, A. X. **Desenvolvimento de uma técnica de classificação hierárquica multirrótulo e sua aplicação em um problema de bioinformática**. 34 p. Monografia (Trabalho Final de Curso em Engenharia de Controle e Automação) — Universidade Federal de Ouro Preto, Escola de Minas, Colegiado do Curso de Engenharia de Controle e Automação, Ouro Preto, 2014.

HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**. Second. USA: Morgan Kaufmann Publishers, 2006.

HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**. Second. USA: Morgan Kaufmann Publishers, 2012.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. Second. [S.l.]: Springer, 2009.

MERSCHMANN, L. H. de C. **CLASSIFICAÇÃO PROBABILÍSTICA BASEADA EM ANÁLISE DE PADRÕES**. 117 p. Tese (Doutorado) — Universidade Federal Fluminense, Programa de Pós-Graduação em Otimização Combinatória, Niterói, 2007.

MERSCHMANN, L. H. de C.; FREITAS, A. A. An extended local hierarchical classifier for prediction of protein and gene functions: Proceedings of the 15th international conference, dawak 2013, prague, czech republic, august 26-29, 2013. In: BELLATRECHE, L.; MOHANIA, M. K. (Ed.). **Data Warehousing and Knowledge Discovery**. Berlin: Springer Berlin Heidelberg, 2013. p. 159–171.

MITRA, S.; ACHARYA, T. **Data Mining: Multimedia, Soft Computing and Bioinformatics**. [S.l.]: John Wiley & Sons, 2003.

SILLA JR, C. N. **Novel Approaches for Hierarchical Classification with Case Studies in Protein Function Prediction**. Tese (phd) — University of Kent, Canterbury, 2011.

SILLA JR, C. N.; FREITAS, A. A. A survey of hierarchical classification across different application domains. data mining and knowledge discovery. In: ZHOU, S.; ZHANG, S.; KARYPIS, G. (Ed.). **Advanced Data Mining and Applications**. Germany: Springer, 2012. v. 22, p. 31–72.

SILLA JR, C. N.; FREITAS, A. A. A. Selecting different protein representations and classification algorithms in hierarchical protein function prediction. **Intelligent Data Analysis Journal,**, v. 15, n. 6, p. 979–999, 2011.

TAN, P.; STEINBACH, M.; KUMAR, V. **Introduction to Data Mining**. [S.l.]: Addison Wesley, 2005.

VENS, C. et al. Decision trees for hierarchical multi-label classification. **Machine Learning**, v. 2, p. 185–214, 2008.