



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE  
CENTRO DE ENSINO SUPERIOR DO SERIDÓ  
DEPARTAMENTO DE COMPUTAÇÃO E TECNOLOGIA  
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

**UMA ANÁLISE DA INFLUÊNCIA DO PARÂMETRO DE CONTROLE  
DO LIMIAR NO MÉTODO DE APRENDIZADO  
SEMISSUPERVISIONADO *FLEXCON-C***

**ARTHUR COSTA GORGÔNIO**

Caicó - RN  
2018

**ARTHUR COSTA GORGÔNIO**

**UMA ANÁLISE DA INFLUÊNCIA DO PARÂMETRO DE CONTROLE  
DO LIMIAR NO MÉTODO DE APRENDIZADO  
SEMISSUPERVISIONADO *FLEXCON-C*:**

**Trabalho de Conclusão de Curso II**, apresentado ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal do Rio Grande do Norte, como parte dos requisitos para obtenção do título de Bacharel em Sistemas de Informação.

Orientador(a): MSc. Amarildo Jeele Ferreira de Lucena.

Co-orientador(a): MSc. Karliane Medeiros Ovidio Vale.

Caicó - RN  
2018

## RESUMO

Algoritmos com a capacidade de aprendizado são ferramentas eficazes e eficientes para o processamento de um grande volume de dados. Entretanto, bases de dados das aplicações do mundo real não são completamente rotuladas, isso dificulta o desenvolvimento de um modelo pelos modos tradicionais de aprendizado de máquina. O aprendizado de máquina semissupervisionado surge para realizar o treinamento de algoritmos capazes de aprender com base de dados parcialmente rotuladas. A confiabilidade do processo de classificação depende de vários fatores que incluem o tipo de classificador utilizado e um conjunto de parâmetros que os customiza, além da disposição e/ou formatação da própria base de dados. Um fator importante deste tipo de aprendizado é a seleção de exemplos a serem incluídos no conjunto de dados rotulados. Uma forma de realizar essa seleção é por meio de um limiar que determina quais instâncias são incluídas por iteração, permitindo rotular apenas instâncias classificadas com alto valor de confiança. O método **FlexCon-C** – derivado do algoritmo *Self-Training* – faz uso dessa estratégia e o objeto de estudo deste trabalho foram as três técnicas variantes do método **FlexCon-C** (*FlexCon-C1 (s)*, *FlexCon-C1 (v)*, *FlexCon-C2*). No qual foi analisado diferentes valores para a variação do limiar ( $cr$ ) mensurando o impacto na classificação de dados semissupervisionados. Os resultados demonstraram não haver um valor para o parâmetro  $cr$  que seja superior aos demais em todos os casos, uma vez que o melhor valor depende de diferentes configurações do experimento, tais como: técnica, classificador e percentual de dados inicialmente rotulados. Analisando a acurácia por classificador, percebeu-se que Naïve Bayes e *rpartXse* não apresentaram diferenças significativas no valor da acurácia ao se variar o parâmetro  $cr$ . Entretanto, o **RIPPER** obteve melhores resultados fixando-se o valor do  $cr > 5\%$ , enquanto que o **k-NN** alcançou melhores acurácias com o  $cr < 5\%$ .

**Palavras-chave:** Aprendizado semissupervisionado; *Self-Training*; Ajuste de parâmetros.

## ABSTRACT

Learning algorithms are effective and efficient tools for processing large volumes of data. However, real-world application databases are not fully labeled, this difficult the development of a model through traditional modes of machine learning. The semi-supervised machine learning arises to perform the training of algorithms capable of learning with partially labeled databases. The confidence of the classification process depends on several factors that include the type of the classifier and a set of parameters that customize them, besides the layout and/or the dataset's characteristics. An important factor in this type of learning is the selection of examples to be included in the labeled data set. A way to make this selection is using a threshold that determinate the included instances for each iteration, allowing to label only the instances with high confidence value. The **FlexCon-C** method – derived from the Self-Training algorithm – make use of this strategy and the object of study of this paper were the three variations of the **FlexCon-C** (FlexCon-C1 (s), FlexCon-C1 (v), FlexCon-C2). In this research were analyzed different values for the threshold variation ( $cr$ ), measuring the impact on the classification of semi-supervised learning. The results showed that there is no value for the parameter  $cr$  that is superior to the other in all cases, the best value depends on different configurations of the experiment, such as: technique, classifier and percentage of initially labeled data. Analyzing the accuracy by classifier, it was observed that Naïve Bayes and rpartXse did not present significant differences in the value of accuracy when the parameter  $cr$  was changed. However, the **RIPPER** obtain the best results by setting the value of  $cr > 5\%$ , while the **k-NN** classifier achieved better accuracy with  $cr < 5\%$ .

**Keywords:** Semi-Supervised Learning; Self-Training; Parameter setting.

## LISTA DE FIGURAS

Figura 1 – Divisão do Aprendizado. . . . .	18
Figura 2 – Categorias dos Modelos de Aprendizagem. . . . .	21
Figura 3 – Gráfico do Classificador Naïve Bayes e a Técnica <i>FlexCon-C1(s)</i> . . . . .	34
Figura 4 – Gráfico do Classificador Naïve Bayes e a Técnica <i>FlexCon-C1(v)</i> . . . . .	34
Figura 5 – Gráfico do Classificador Naïve Bayes e a Técnica <i>FlexCon-C2</i> . . . . .	35
Figura 6 – Gráfico do Classificador <i>rpartXse</i> e a Técnica <i>FlexCon-C1(s)</i> . . . . .	35
Figura 7 – Gráfico do Classificador <i>rpartXse</i> e a Técnica <i>FlexCon-C1(v)</i> . . . . .	36
Figura 8 – Gráfico do Classificador <i>rpartXse</i> e a Técnica <i>FlexCon-C2</i> . . . . .	36
Figura 9 – Gráfico do Classificador RIPPER e a Técnica <i>FlexCon-C1(s)</i> . . . . .	37
Figura 10 – Gráfico do Classificador RIPPER e a Técnica <i>FlexCon-C1(v)</i> . . . . .	37
Figura 11 – Gráfico do Classificador RIPPER e a Técnica <i>FlexCon-C2</i> . . . . .	38
Figura 12 – Gráfico do Classificador <i>k</i> -NN e a Técnica <i>FlexCon-C1(s)</i> . . . . .	38
Figura 13 – Gráfico do Classificador <i>k</i> -NN e a Técnica <i>FlexCon-C1(v)</i> . . . . .	39
Figura 14 – Gráfico do Classificador <i>k</i> -NN e a Técnica <i>FlexCon-C2</i> . . . . .	39

## LISTA DE TABELAS

Tabela 1 – Bases de Dados que serão utilizadas nos experimentos. . . . .	27
Tabela 2 – Resultado da acurácia com o classificador Naïve Bayes. . . . .	30
Tabela 3 – Resultado da acurácia com o classificador <i>rpartXse</i> . . . . .	31
Tabela 4 – Resultado da acurácia com o classificador RIPPER. . . . .	32
Tabela 5 – Resultado da acurácia com o classificador <i>k</i> -NN. . . . .	32
Tabela 6 – Desempenho associado a cada valor do parâmetro <i>cr</i> . . . . .	33
Tabela 7 – Resultados do teste de <i>Shapiro-Will</i> do classificador Naïve Bayes. . . . .	41
Tabela 8 – Resultados do teste de <i>Shapiro-Will</i> do classificador <i>rpartXse</i> . . . . .	41
Tabela 9 – Resultados do teste de <i>Shapiro-Will</i> do classificador RIPPER. . . . .	42
Tabela 10 – Resultados do teste de <i>Shapiro-Will</i> do classificador <i>k</i> -NN. . . . .	42
Tabela 11 – Resultados do teste de <i>Friedman</i> do classificador Naïve Bayes. . . . .	43
Tabela 12 – Resultados do teste de <i>Friedman</i> do classificador <i>rpartXse</i> . . . . .	43
Tabela 13 – Resultados do teste de <i>Friedman</i> do classificador RIPPER. . . . .	44
Tabela 14 – Resultados do teste de <i>Friedman</i> do classificador <i>k</i> -NN. . . . .	44
Tabela 15 – Resultado da aplicação da Técnica FlexCon-C1(s) utilizando o classificador Naïve Bayes. . . . .	50
Tabela 16 – Resultado da aplicação da Técnica FlexCon-C1(v) utilizando o classificador Naïve Bayes. . . . .	55
Tabela 17 – Resultado da aplicação da Técnica FlexCon-C2 utilizando o classificador Naïve Bayes. . . . .	60
Tabela 18 – Resultado da aplicação da Técnica FlexCon-C1(s) utilizando o classificador <i>rpartXse</i> . . . . .	65
Tabela 19 – Resultado da aplicação da Técnica FlexCon-C1(v) utilizando o classificador <i>rpartXse</i> . . . . .	70
Tabela 20 – Resultado da aplicação da Técnica FlexCon-C2 utilizando o classificador <i>rpartXse</i> . . . . .	75
Tabela 21 – Resultado da aplicação da Técnica Flexcon-C1(s) utilizando o classificador RIPPER. . . . .	80
Tabela 22 – Resultado da aplicação da Técnica Flexcon-C1(v) utilizando o classificador RIPPER. . . . .	85
Tabela 23 – Resultado da aplicação da Técnica Flexcon-C2 utilizando o classificador RIPPER. . . . .	90
Tabela 24 – Resultado da aplicações da Técnica Flexcon-C1(s) utilizando o classificador <i>k</i> -NN. . . . .	95
Tabela 25 – Resultado da aplicação da Técnica Flexcon-C1(v) utilizando o classificador <i>k</i> -NN. . . . .	100

Tabela 26 – Resultado da aplicação da Técnica FlexCon-C2 utilizando o classificador $k$ -NN.	105
Tabela 27 – Desempenho dos Valores do $cr$ em cada Técnica. . . . .	110

## LISTA DE ABREVIATURAS E SIGLAS

**AM** Aprendizado de Máquina

**FlexCon-C** *Flexible Confidence with Classifier*

**k-NN** *k-Nearest Neighbors*

**IREP** *Incremental Reduced Error Pruning*

**rpart** *Recursive Partitioning and Regression Trees*

**REP** *Reduced Error Pruning*

**RIPPER** *Repeated Incremental Pruning to Produce Error Reduction*

**SSL** *Semi-Supervised Learning*



## LISTA DE SÍMBOLOS

$L$	Conjunto dos dados rotulados
$U$	Conjunto dos dados não rotulados
$X$	Conjunto dos exemplos
$Y$	Conjunto das classes
$f$	Classificador
$x_i$	$i$ -ésimo exemplo
$y_i$	$i$ -ésima classe
$(x_i, y_i)$	Par do $i$ -ésimo exemplo com sua respectiva classe

## LISTA DE PSEUDOCÓDIGOS

1	Pseudocódigo do <i>Self-Training</i> . . . . .	19
2	Pseudocódigo do método FlexCon-C. . . . .	20
3	<i>Design</i> de Experimento utilizado. . . . .	28

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>12</b>
<b>1.1</b>	<b>Contextualização e Problema</b>	<b>12</b>
<b>1.2</b>	<b>Objetivos</b>	<b>14</b>
1.2.1	Objetivo Geral	14
1.2.2	Objetivos Específicos	14
<b>1.3</b>	<b>Delimitação do Estudo</b>	<b>14</b>
<b>1.4</b>	<b>Justificativa</b>	<b>15</b>
<b>1.5</b>	<b>Apresentação do Trabalho</b>	<b>15</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>16</b>
<b>2.1</b>	<b>Aprendizado de Máquina</b>	<b>16</b>
<b>2.2</b>	<b>Aprendizado Semissupervisionado</b>	<b>18</b>
2.2.1	<i>Self-Training</i>	19
2.2.2	FlexCon-C	19
<b>2.3</b>	<b>Modelos de Aprendizagem</b>	<b>21</b>
2.3.1	Baseado no Método Bayesiano	21
2.3.2	Baseado em Árvores de Decisão	22
2.3.3	Regras de Associação	22
2.3.4	Baseado em Instâncias	22
<b>2.4</b>	<b>Trabalhos Relacionados</b>	<b>23</b>
<b>3</b>	<b>DESENVOLVIMENTO DA PESQUISA</b>	<b>25</b>
<b>3.1</b>	<b>Questões de Pesquisa</b>	<b>25</b>
<b>3.2</b>	<b>Proposta de Solução</b>	<b>25</b>
<b>3.3</b>	<b><i>Design</i> de Experimento</b>	<b>26</b>
<b>4</b>	<b>RESULTADOS</b>	<b>30</b>
<b>4.1</b>	<b>Análise de Eficácia</b>	<b>30</b>
<b>4.2</b>	<b>Análise Estatística</b>	<b>39</b>
<b>5</b>	<b>CONCLUSÃO</b>	<b>45</b>
<b>5.1</b>	<b>Discussão</b>	<b>45</b>
<b>5.2</b>	<b>Trabalhos Futuros</b>	<b>45</b>
	<b>REFERÊNCIAS</b>	<b>47</b>

APÊNDICE A – RESULTADO DA TÉCNICA <i>FLEXCON-C1(S)</i> E CLASSIFICADOR NAÏVE BAYES . . . . .	50
APÊNDICE B – RESULTADO DA TÉCNICA <i>FLEXCON-C1(V)</i> E CLASSIFICADOR NAÏVE BAYES . . . . .	55
APÊNDICE C – RESULTADO DA TÉCNICA <i>FLEXCON-C2</i> E CLASSIFICADOR NAÏVE BAYES . . . . .	60
APÊNDICE D – RESULTADO DA TÉCNICA <i>FLEXCON-C1(S)</i> E CLASSIFICADOR <i>RPARTXSE</i> . . . . .	65
APÊNDICE E – RESULTADO DA TÉCNICA <i>FLEXCON-C1(V)</i> E CLASSIFICADOR <i>RPARTXSE</i> . . . . .	70
APÊNDICE F – RESULTADO DA TÉCNICA <i>FLEXCON-C2</i> E CLASSIFICADOR <i>RPARTXSE</i> . . . . .	75
APÊNDICE G – RESULTADO DA TÉCNICA <i>FLEXCON-C1(S)</i> E CLASSIFICADOR RIPPER . . . . .	80
APÊNDICE H – RESULTADO DA TÉCNICA <i>FLEXCON-C1(V)</i> E CLASSIFICADOR RIPPER . . . . .	85
APÊNDICE I – RESULTADO DA TÉCNICA <i>FLEXCON-C2</i> E CLASSIFICADOR RIPPER . . . . .	90
APÊNDICE J – RESULTADO DA TÉCNICA <i>FLEXCON-C1(S)</i> E CLASSIFICADOR <i>K-NN</i> . . . . .	95
APÊNDICE K – RESULTADO DA TÉCNICA <i>FLEXCON-C1(V)</i> E CLASSIFICADOR <i>K-NN</i> . . . . .	100
APÊNDICE L – RESULTADO DA TÉCNICA <i>FLEXCON-C2</i> E CLASSIFICADOR <i>K-NN</i> . . . . .	105
APÊNDICE M – DESEMPENHO DOS VALORES DO <i>CR</i> EM CADA TÉCNICA . . . . .	110

# 1 Introdução

Por um tempo, o computador foi utilizado prioritariamente para efetuar cálculos e retornar resultados. A evolução da tecnologia proporcionou facilidades para a humanidade, incluindo a popularização dos meios de informação e comunicação, melhoria da qualidade de vida e a automatização nos processos. Diante deste cenário, ocorreram avanços na coleta, armazenamento e processamento de dados, além do desenvolvimento de algoritmos capazes de formular modelos e aprender padrões. As máquinas tornaram-se estratégicas no auxílio ao processo de tomada de decisão, pois a sua capacidade de processar dados é significativamente maior, comparada a dos humanos.

## 1.1 Contextualização e Problema

O Aprendizado de Máquina ([AM](#)) pode ser definido como o processo computacional relacionado ao desenvolvimento de algoritmos que automaticamente melhoram o seu desempenho em determinada tarefa a partir de experiências anteriores ([MITCHELL, 1997](#)). Assim, tais algoritmos aprendem a realizar escolhas com base em um histórico de escolhas anteriores, devidamente validadas.

O [AM](#) apresenta-se como uma possível solução para a análise de grandes volumes de dados, uma vez que estas análises possuem algumas dificuldades quando desenvolvidas por seres humanos ([RASCHKA, 2015](#)). Grandes volumes de dados demandam muito tempo para humanos, além de haver a necessidade de especialistas sobre o domínio dos dados para decorram análises mais confiantes. Por sua vez, um algoritmo que possua a característica de aprendizado pode realizar esse processo de análise mais rápida e com um custo reduzido. Assim, surgiu a necessidade de desenvolver algoritmos capazes de aprender padrões e realizar tais tarefas.

A literatura descreve que o [AM](#) possui tarefas de classificação, agrupamento e regressão. [Alpaydin \(2004\)](#) e [Mohri, Rostamizadeh e Talwalkar \(2012\)](#) descrevem como:

- i. Classificação – cada objeto é associado a um rótulo que representa uma das classes existentes, a partir de instâncias cujo rótulo seja conhecido;
- ii. Agrupamento – os objetos são agrupados de acordo com suas características, de forma que um objeto não pertença simultaneamente a mais de um agrupamento;
- iii. Regressão – a partir de funções matemáticas, busca-se encontrar relações algébricas entre atributos;

O AM é tradicionalmente dividido em dois tipos: o aprendizado supervisionado e o aprendizado não supervisionado. O aprendizado supervisionado faz uso de classificação ou regressão, enquanto o aprendizado não supervisionado faz uso de agrupamentos (CHAPELLE; SCHÖLKOPF; ZIEN, 2006).

A técnica de agrupamentos é capaz de realizar uma separação dos objetos de acordo com suas similaridades baseando-se em um conjunto de características. Essa técnica pode ser utilizada quando não existe a necessidade de identificar a qual classe um determinado objeto pertence. Na classificação de dados, os algoritmos auxiliam no processo de encontrar relações entre os preditores e preditos em grandes quantidades de dados. Essa técnica é utilizada quando pretende-se treinar o algoritmo para identificar a categoria a qual pertence um determinado objeto.

O treinamento é uma etapa decisiva na obtenção de um modelo que seja capaz de classificar corretamente os objetos com base em suas características. Nesta etapa, o algoritmo de classificação é aplicado sobre um conjunto de dados com a finalidade de gerar um classificador. A partir deste, é possível desempenhar a tarefa de classificação sobre novos conjuntos de dados.

Um dos desafios do AM está em como realizar esse treinamento, pois nem sempre as bases de dados possuem objetos suficientes para realizá-lo pelas formas tradicionais. Por isso, uma área tem ganhado muita atenção dos pesquisadores, o Aprendizado Semissupervisionado (*Semi-Supervised Learning (SSL)*) que, a partir de uma pequena parcela de objetos classificados, consegue atribuir classes aos demais objetos contidos na base. Ou seja, a necessidade de possuir um grande conjunto de dados com o seu devido rótulo deixa de ser prioridade, pois o conjunto de dados classificado passa a ser maior ao fim de cada iteração do algoritmo.

Os algoritmos do SSL são capazes de realizar o treinamento a partir de um conjunto pequeno de instâncias inicialmente rotuladas e normalmente classificam  $n$  instâncias por iteração. Há variações destes algoritmos que fazem uso de limiares (confiança mínima) com a intenção de inserir no conjunto dos dados rotulados somente as instâncias que possuem a confiança atribuída pelo classificador, maior que o limiar (RODRIGUES; SANTOS; CANUTO, 2013). Isto é, o limiar é um fator de inclusão que torna variável o número de instâncias que são incluídas por iteração, pois só são selecionadas as instâncias com alto valor de confiança.

O uso do limiar fixo pode resultar na não classificação de todo o conjunto de dados. Sendo assim, uma das formas de resolver esse problema é utilizando uma taxa de controle, que auxilie na variação do limiar, ora aumentando, ora diminuindo esse valor, de forma a permitir a inclusão de novas instâncias.

O método *Flexible Confidence with Classifier (FlexCon-C)*, apresentado em Vale *et al.* (2018) faz uso do limiar e de um fator ( $cr$ ) que altera-o de maneira mais lenta. Esse método possui duas técnicas de implementação, são elas: *FlexCon-C1* e *FlexCon-C2*. O *FlexCon-C1* faz uma comparação entre a predição da iteração atual com a predição da primeira iteração. O

*FlexCon-C2* compara a predição atual com a predição de um classificador supervisionado que é treinado com o conjunto de instâncias que estão inicialmente rotuladas. Este trabalho pretende analisar o impacto na classificação dos dados quando a taxa de variação do limiar é manipulada.

## 1.2 Objetivos

Neste tópico são apresentados os objetivos desta pesquisa, divididos em geral e específicos.

### 1.2.1 Objetivo Geral

Este trabalho tem por objetivo analisar a influência da taxa de variação do limiar na acurácia da classificação de dados semissupervisionados quando utilizado o método proposto por [VALE \*et al.\*](#), o *FlexCon-C*, em suas duas variações *FlexCon-C1* e o *FlexCon-C2*.

### 1.2.2 Objetivos Específicos

- a) Definir os valores mínimos e máximos para a variação do limiar;
- b) Identificar métricas que permitam avaliar os experimentos a serem realizados;
- c) Selecionar um conjunto de bases de dados para a aplicação dos experimentos;
- d) Implementar os algoritmos, modificando-os para inclusão das variações descritas na proposta;
- e) Validar os resultados obtidos através da comparação empírica dos algoritmos modificados.

## 1.3 Delimitação do Estudo

Neste trabalho serão analisados os efeitos da taxa de variação do limiar sobre o método *FlexCon-C* do *SSL*, uma vez que ele possui como restrição a variação do limiar pré-fixada em 5%. Para isso, apenas a taxa de variação do limiar será modificada ao longo da execução dos experimentos, mantendo-se inalterados os demais parâmetros do algoritmo. Assim como apresentado em [Vale \*et al.\* \(2018\)](#), também serão utilizados os mesmos percentuais de exemplos inicialmente rotulados (5%, 10%, 15%, 20% e 25%) e classificadores (Naïve Bayes).

Na implementação das técnicas propostas *FlexCon-C1* e *FlexCon-C2*, foram utilizados quatro diferentes classificadores, que possibilitam a exploração dos diversos tipos de dados existentes nas bases de dados, a saber: a partir do método Bayesiano (Naïve Bayes), fundamentado em Árvores de Decisão (*rpartXse*), baseado em Instâncias (*k-Nearest Neighbors (k-NN)*) e com suporte em Regras de Associação (*Repeated Incremental Pruning to Produce Error*

*Reduction* ([RIPPER](#)). Essas técnicas foram submetidas a um conjunto de 31 bases de dados, realizando uma comparação entre a média das acurácias obtidas pelas técnicas quando aplicada a proposta.

## 1.4 Justificativa

Este trabalho se justifica por realizar um estudo na área do [AM](#) fazendo o uso do [SSL](#) com a finalidade de classificar dados a partir de um conjunto pequeno de instâncias conhecidas. Essa área do conhecimento vem ganhando muita atenção dos pesquisadores por fazer o uso de algoritmos de classificação capazes de desempenhar essa tarefa com uma pequena parcela dos dados rotulados.

A classificação de todos os dados de uma base é uma tarefa complexa e custosa, sendo que, algumas vezes, essas informações são impossíveis de serem obtidas; então, como alternativa, o [SSL](#) é empregado. Por sua vez, estes algoritmos normalmente possuem um limiar associado e algumas variações apresentam uma taxa de variação do limiar. Então, a proposta apresentada no trabalho possui relevância, visto que apresenta uma análise nas técnicas tornam variável o limiar de inclusão.

## 1.5 Apresentação do Trabalho

Este trabalho está organizado da seguinte maneira: no [Capítulo 1](#) foi apresentada uma visão geral da pesquisa, abordando o tema, contextualização do problema, objetivos da pesquisa e justificativa para a elaboração deste trabalho. O [Capítulo 2](#) trata sobre a fundamentação teórica, onde são apresentados os conceitos de [AM](#), [SSL](#) e modelos de aprendizagem.

No [Capítulo 3](#) é demonstrado como essa pesquisa foi guiada, incluindo e explanando exatamente o que é proposto a se resolver. Além disso, apresenta a solução proposta e como foi desenvolvida. No [Capítulo 4](#) são descritos os resultados deste trabalho e são conduzidas análises de eficácia e estatística. O [Capítulo 5](#) apresenta as considerações finais deste trabalho. Por fim, as referências bibliográficas que suportam a elaboração desta pesquisa.



## 2 Fundamentação Teórica

Neste capítulo é realizada uma revisão na literatura sobre o tema do trabalho, iniciando-se pelo [AM](#). Após contextualizar a área geral do trabalho, serão introduzidas as particularidades escolhidas que são: o [SSL](#), aplicado para a tarefa de classificação de dados. Por fim, são apresentados alguns trabalhos relacionados a esta linha de pesquisa.

### 2.1 Aprendizado de Máquina

Atualmente, o grande volume das bases de dados contribui para tornar mais complexa a obtenção de informações que deveriam ser consideradas no processo de tomada de decisão. Entretanto, quando os humanos realizam esse processamento, os resultados (informações) são obtidos mais lentamente. Para isso, surge a necessidade de utilizar processos computacionais com a finalidade de agilizar a aquisição das informações.

Diante deste desafio, o [AM](#) surge como um campo de estudo que está relacionado principalmente à capacidade de desenvolver nos computadores um aprendizado semelhante ao dos seres vivos. Desta forma, a máquina torna-se capaz de aprender a partir de experiências passadas, conseguindo melhorar o seu desempenho no futuro. Essa área se destaca pela aplicação nos carros autodirigidos, na forma como o *feed* de notícias é ordenado pelo *Facebook* ou em sugestões de outros produtos quando se está realizando uma compra online.

No contexto do [AM](#) existem diversos termos que são utilizados no decorrer deste trabalho, o Quadro 1 apresenta os termos utilizados no decorrer do trabalho e o seu significado.

[Russell e Norvig \(2009\)](#) defendem que o [AM](#) é uma sub-área da Inteligência Artificial preocupada com programas que aprendam com a experiência. Para tornar ágeis os processos que seriam complexos ou impossíveis de serem realizados por humanos, podem ser empregados métodos computacionais que aprendam a partir de experiências ([MITCHELL, 1997](#); [MOHRI](#);

Quadro 1 – Alguns termos do [AM](#).

<b>Termo</b>	<b>Significado</b>
Atributo	Cada coluna da base de dados
Classe/Rotulo	Categoria ao qual o exemplo pertence
Classificador	Algoritmo utilizado na tarefa do <a href="#">AM</a>
Exemplo/Instância	Uma linha da base de dados
Base de treinamento	Conjunto de exemplos aplicados ao classificador a fim de treiná-lo
Base de teste	Conjunto de exemplos no qual o classificador é testado

**Fonte:** Adaptado de ([MOHRI](#); [ROSTAMIZADEH](#); [TALWALKAR, 2012](#); [GOLLAPUDI, 2016](#))

ROSTAMIZADEH; TALWALKAR, 2012).

Um exemplo de aplicação é a automatização do processo de concessão de empréstimos bancários, no qual o AM é responsável por aprender os perfis de clientes para os quais o banco rejeitaria ou concederia esse tipo de movimentação financeira. Baseando-se em registros anteriores, o algoritmo é capaz de criar um classificador e gerar a inteligência que definiria se um novo cliente está enquadrado no perfil de clientes para receber empréstimo ou não.

O AM possui diversas sub-áreas e sua escolha está relacionada à base de dados, se ela possuir o atributo classe e o problema a ser tratado for de classificação ou regressão deve-se optar pelo aprendizado supervisionado. Porém, há casos nos quais a base possui poucas instâncias rotuladas, então o aprendizado semissupervisionado se torna mais efetivo. Também há situações em que os exemplos não são rotulados, entretanto existe a possibilidade de separá-los a partir de suas características, quando isso ocorre deve-se utilizar o aprendizado não supervisionado. Para Gollapudi (2016) e Wiley (2016) a área do AM está dividida em algumas sub-áreas, sendo elas:

- i. Aprendizado Supervisionado – Todos os dados possuem o atributo classe, isso é, para cada exemplo na base de dados existe uma classe relacionada;
- ii. Aprendizado Não Supervisionado – Os dados não possuem uma classificação, porém existe a possibilidade de agrupa-los de modo que os exemplos do mesmo agrupamento sejam similares;
- iii. Aprendizado Semissupervisionado – Essa sub-área está na intercessão dos aprendizados Supervisionado e Não Supervisionado, ou seja, um sub-conjunto dos dados possui classe e o restante não, porém, a partir deste conjunto rotulado, consegue-se classificar os demais dados;

O aprendizado é dito supervisionado, quando para cada exemplo existente na base de dados existe uma classe associada ao mesmo, ou seja,  $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)$  com  $\mathbf{x}_i \in X$  e  $y_i \in Y$ , onde  $X$  e  $Y$  são os conjuntos dos exemplos e das classes, respectivamente. A Equação 1 apresenta, de forma geral, o aprendizado supervisionado,  $\mathbf{x}_i$  representa o  $i$ -ésimo exemplo,  $y_i$  refere-se a  $i$ -ésima classe,  $i = 1$  o primeiro elemento variando até o  $l$ -ésimo elemento do conjunto dos rotulados ( $L$ ).

Por outro lado, o aprendizado não Supervisionado, ocorre quando os dados não possuem uma classificação, ou seja,  $Y = \emptyset$ . Entretanto existe a possibilidade de dividi-los em agrupamentos cujos elementos de um mesmo agrupamento são semelhantes entre si. A Equação 2 demonstra o  $j$ -ésimo exemplo existente dentro do conjunto dos dados não rotulados ( $U$ ), iniciado em  $j = 1$  e finalizando no  $u$ -ésimo elemento.

$$L = \{(\mathbf{x}_i, y_i)\}_{i=1}^l | x \in X, y \in Y, 1 \leq i \leq l \quad (1)$$

$$U = \{\mathbf{x}_j\}_{j=1}^u | x \in X, 1 \leq j \leq u \quad (2)$$

Por fim, as Equações 3 e 4 descrevem de forma geral o aprendizado Semissupervisionado, onde  $L$  e  $U$  representam os conjuntos de dados rotulados e não rotulados, respectivamente.  $\{(\mathbf{x}_i, y_i)\}$  o par do  $i$ -ésimo exemplo e respectiva classe,  $i = 1$  o primeiro exemplo variando até o  $l$ -ésimo exemplo rotulado e  $\{\mathbf{x}_j\}$  o  $j$ -ésimo elemento iniciado em  $j = l + 1$  finalizando em  $l + u$  que é todo o conjunto de dados retirando a parte rotulada.

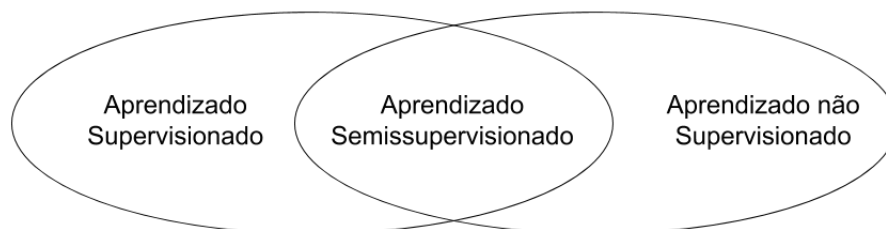
$$L = \{(\mathbf{x}_i, y_i)\}_{i=1}^l | x \in X, y \in Y, 1 \leq i \leq l \quad (3)$$

$$U = \{\mathbf{x}_j\}_{j=l+1}^{l+u} | x \in X, y \in Y, l + 1 \leq j \leq l + u \quad (4)$$

## 2.2 Aprendizado Semissupervisionado

Conforme mencionado anteriormente e apresentado na Figura 1, o **SSL** está na interseção dos aprendizados supervisionado e não supervisionado, onde uma parte dos exemplos possui o rótulo e outra não.

Figura 1 – Divisão do Aprendizado.



**Fonte:** O Próprio Autor (2018)

A abordagem supervisionada exige uma base de dados com todos os exemplos rotulados para gerar um classificador eficaz. Porém, a abordagem semissupervisionada vem demonstrando relevância para muitas aplicações no mundo real, pois nem sempre é possível obter a base com todos os dados rotulados, ou dispor de recursos financeiros e do conhecimento de profissionais com a finalidade de treinar um classificador com uma eficácia maior. Então, com o **SSL** é possível reduzir os custos e esforços necessários para executar a rotulagem em um conjunto de dados (BOUCHACHIA, 2012; HADY; SCHWENKER, 2013).

Além disto, essa abordagem pode ser utilizada tanto com tarefas de classificação quanto as tarefas de agrupamento. Para a tarefa de classificação, os classificadores são utilizados para realizar previsões sobre o conjunto de dados não rotulados  $U$ , adicionado-os ao conjunto dos dados rotulados  $L$ . Quando utilizada a tarefa de agrupamento os exemplos rotulados

auxiliarão a definir os núcleos destes agrupamentos, obtendo assim uma eficácia melhor no agrupamento (CHAPELLE; SCHÖLKOPF; ZIEN, 2006).

### 2.2.1 Self-Training

Dentre os diversos algoritmos do aprendizado semissupervisionado destaca-se *Self-Training* (ZHU, 2008). Um algoritmo do SSL, cuja execução é dada por: i) treinamento do classificador com os exemplos do conjunto dos rotulados  $L$ ; ii) aplicação do classificador nos dados não rotulados; iii) seleção dos exemplos que serão classificados na iteração; iv) adição os exemplos classificados no conjunto dos rotulados  $L$  (GRANDVALET; BENGIO, 2004; CHAPELLE; SCHÖLKOPF; ZIEN, 2006; ZHU; GOLDBERG, 2009). No Pseudocódigo 1 encontra-se o algoritmo do *Self-Training*,

---

#### Pseudocódigo 1: Pseudocódigo do *Self-Training*.

---

**entrada** dados rotulados  $\{(\mathbf{x}_i, y_i)\}_{i=1}^l$ , dados não rotulados  $\{\mathbf{x}_j\}_{j=l+1}^{l+u}$

**garanta**  $L \leftarrow \{(\mathbf{x}_i, y_i)\}_{i=1}^l$  e  $U \leftarrow \{\mathbf{x}_j\}_{j=l+1}^{l+u}$

1: **repita**

2:     *Treinamento  $f$  com os exemplos de  $L$  usando aprendizado supervisionado*

3:     *Aplicação  $f$  as instâncias não rotuladas em  $U$*

4:     *Remoção um subconjunto  $S$  de  $U$*

5:     *Adição  $\{(\mathbf{x}, f(\mathbf{x})) | \mathbf{x} \in S\}$  a  $L$*

6: **até que**  $U = \emptyset$

---

**Fonte:** Adaptado de (ZHU; GOLDBERG, 2009)

onde,  $L$  representa o conjunto de dados inicialmente rotulados,  $U$  representa o restante da base de dados contendo os exemplos não rotulados,  $f$  é o classificador que vai realizar o aprendizado de modo supervisionado a partir do conjunto  $L$ ,  $S$  representa o subconjunto de  $U$  que será incluído em  $L$ ,  $\{(\mathbf{x}_i, y_i)\}_{i=1}^l$  o  $i$ -ésimo par elemento e sua respectiva classe no conjunto dos rotulados e  $\{\mathbf{x}_j\}_{j=l+1}^{l+u}$  representa o  $j$ -ésimo elemento que pertence ao conjunto dos não rotulados.

### 2.2.2 FlexCon-C

Nesta sub-seção é descrito o método *FlexCon-C* proposto por (VALE *et al.*, 2018). O Pseudocódigo 2 apresenta a ideia geral do *FlexCon-C*, cuja execução é dada por: i) separação dos exemplos rotulados dos não-rotulados; ii) treinamento do classificador  $f$  utilizando o conjunto dos rotulados  $L$ ; iii) aplicação de  $f$  no conjunto dos não-rotulados  $U$ , gerando assim uma matriz com o exemplo, sua respectiva classe e confiança indicada pelo classificador; iv) seleção de todos os exemplos dado a regra de inserção; v) adição de todos os exemplos de  $S$  em  $L$ ; vi) utilização do conjunto  $S$  para treinar um classificador  $f'$ ; vii) aplicação do classificador  $f'$  em  $L$ ; viii) atualização do valor do limiar de inserção com base na acurácia obtida em  $f'$ .

O **FlexCon-C** decide se o exemplo deve ir para o conjunto dos rotulados realizando uma varredura nas regras de inserção de exemplos que são: i) as predições convergem entre as classes para o exemplo e as confianças são maiores que o limiar; ii) as predições convergem entre as classes, porém apenas uma de suas confianças é maior que o limiar; iii) as predições divergem entre as classes e ambas as confianças são maiores que o limiar; e iv) as predições divergem entre as classes e uma das confianças é maior que o limiar. A regra de inserção subsequente é executada se, e somente se, a verificação anterior não selecionar exemplos para inclusão. Caso todas as regras falhem, a confiança é alterada para o maior valor da predição da iteração corrente.

---

**Pseudocódigo 2:** Pseudocódigo do método **FlexCon-C**.

---

**entrada** dados rotulados  $\{(\mathbf{x}_i, y_i)\}_{i=1}^l$ , dados não rotulados  $\{\mathbf{x}_j\}_{j=l+1}^{l+u}$

**garanta**  $L \leftarrow \{(\mathbf{x}_i, y_i)\}_{i=1}^l$  e  $U \leftarrow \{\mathbf{x}_j\}_{j=l+1}^{l+u}$

1: **repita**

2:     *Treinamento  $f$  com os exemplos de  $L$  usando aprendizado supervisionado*

3:     *Aplicação  $f$  as instâncias não rotuladas em  $U$*

4:     *Remoção um subconjunto  $S$  de  $U$  tal que o exemplo se enquadre com uma das regras inclusão de novos exemplos*

5:     *Adição  $\{(\mathbf{x}, f(\mathbf{x})) | \mathbf{x} \in S\}$  a  $L$*

6:     *Treinamento  $f'$  com os exemplos de  $S$  utilizando aprendizado supervisionado*

7:     *Aplicação  $f'$  as instâncias em  $L$*

8:     *Atualização o limiar de inclusão a partir da acurácia em  $f'$*

9: **até que**  $U = \emptyset$

---

**Fonte:** Adaptado de (VALE *et al.*, 2018)

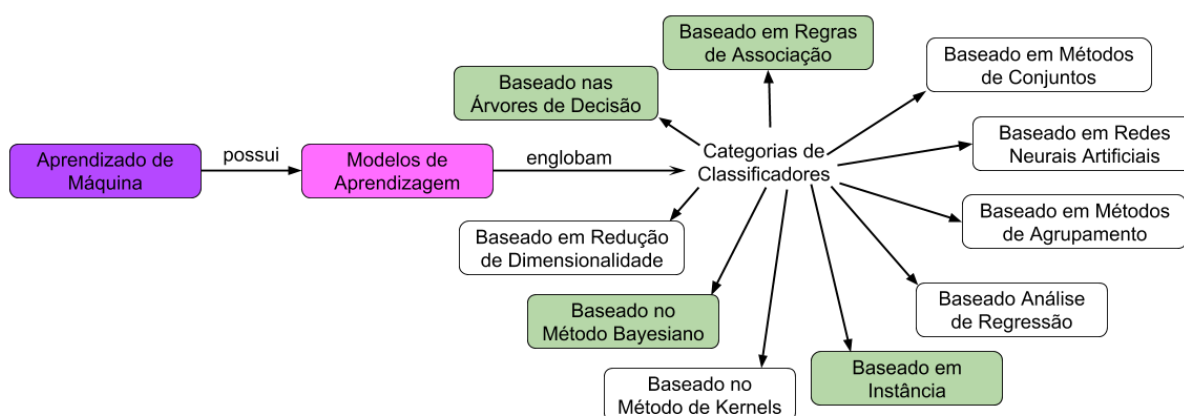
O **FlexCon-C** possui duas técnicas diferentes – *FlexCon-C1* e *FlexCon-C2* – para decidir qual rótulo será atribuído a cada exemplo. A técnica *FlexCon-C1* guarda a predição da primeira iteração para comparar com a predição da iteração corrente, quando as predições convergem para a mesma classe o exemplo é adicionado. Caso o exemplo seja selecionado para ser incluído no conjunto dos exemplos rotulados e as predições não convergirem, a classe é definida por um critério de seleção, isto gerou duas variações desta mesma técnica: o *FlexCon-C1(s)* utiliza a soma das confianças obtidas desde a primeira iteração até a atual para realizar essa decisão, enquanto que o *FlexCon-C1(v)* faz uso de votação, onde são contabilizados somente as ocorrências do exemplo em cada classe.

A técnica *FlexCon-C2* compara as predições do classificador da iteração corrente com a predição de um classificador supervisionado que é treinado com as instâncias inicialmente rotuladas. Quando as predições convergem entre as classes, o exemplo é adicionado; caso o exemplo selecionado para a inclusão não possua a mesma classe em ambas as predições, o rótulo é selecionado a partir do classificador supervisionado.

## 2.3 Modelos de Aprendizagem

Os Modelos de Aprendizagem possuem categorias de classificadores que são responsáveis por aprender os padrões existentes no conjunto de treinamento. Isto é, o classificador é gerado a partir da aplicação do conjunto de treinamento em um algoritmo de AM. Há, na literatura, diversas formas de agrupar os classificadores, a Figura 2 apresenta algumas das subdivisões existentes considerando as diversas sub-áreas de aprendizado, bem como, as categorias do AM. Foram selecionados 4 algoritmos de classificação (Naïve Bayes, *rpartXse*, *RIPPER* e *k-NN*), cada um pertencente a uma das categorias destacadas (Árvore de Decisão, Regras de Associação, Método Bayesiano e Instâncias).

Figura 2 – Categorias dos Modelos de Aprendizagem.



Fonte: Adaptado de (GOLLAPUDI, 2016)

### 2.3.1 Baseado no Método Bayesiano

A partir do teorema de Bayes, oriundo da estatística, a chance de um evento  $A$  acontecer, dado  $B$  ter ocorrido, supondo que  $A$  e  $B$  são eventos distintos, é apresentado na Equação 5. Um classificador Bayesiano calcula a probabilidade do exemplo pertencer a cada uma das classes existentes na base, e associa-o à classe que obtiver o melhor resultado (ZELLNER, 1996; ALPAYDIN, 2004; AU, 2017).

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \quad (5)$$

O Naïve Bayes é um dos classificadores desta categoria. O seu modo de associar uma classe a um exemplo é calculando a probabilidade a partir da contagem da frequência e das combinações dos valores nos atributos. O classificador Naïve Bayes assume que todos os atributos são independentes entre si e a classe como dependente, entretanto, são raros os casos nos quais uma base tem todos atributos independentes são independentes entre si. Ainda assim

esse classificador é bastante utilizado, uma vez que sua execução é eficiente e rápida para muitos dos problemas supervisionados (DIMITOGLOU; ADAMS; JIM, 2012).

### 2.3.2 Baseado em Árvores de Decisão

Árvores de Decisão são baseadas no método da inferência indutiva, onde cada nó da árvore representa um atributo a ser examinado, suas ramificações correspondem aos possíveis valores que os atributos podem assumir e as folhas correspondem as classes. A medida que uma instância vai percorrendo a árvore ela é classificada (QUINLAN, 1986; MITCHELL, 1997). Árvores de decisão possuem desempenho efetivo sobre bases de dados cujo conjunto de atributos pode ser numérico e/ou categórico.

Dessa categoria, será utilizado o algoritmo *rpartXse*, o qual Torgo (2013) descreve como uma modificação do algoritmo *Recursive Partitioning and Regression Trees* (*rpart*), que realiza uma única chamada do classificador juntamente da regra da poda da árvore. Por sua vez o *rpart* é uma implementação de uma árvore baseando-se no livro *Classification and Regression Trees*, cuja regra da poda da árvore (1-SE) é descrita (BREIMAN *et al.*, 1984; THERNEAU; E.J., 2000).

### 2.3.3 Regras de Associação

O aprendizado baseado nesta sub-divisão funciona de forma que os algoritmos realizam uma avaliação dos atributos da base de dados, gerando um conjunto de regras. Em seguida, este conjunto de regras será submetido a testes para verificar o quão válidas estas regras são para a base, a fim de validar este conhecimento (GOLLAPUDI, 2016). Esse processo se repete até que todas as classes da base de dados possua um conjunto de regras.

Nessa categoria, existe o algoritmo RIPPER (do inglês, *Repeated Incremental Pruning to Produce Error Reduction*), que é uma evolução desenvolvida do IREP (do inglês *Incremental Reduced Error Pruning*), que por sua vez, baseia-se na técnica REP (do inglês, *Reduced Error Pruning*) já existente nos algoritmos baseados em árvores (COHEN, 1995). A sua execução é dada por uma análise inicial nas classes, gerando um conjunto de regras, a partir do qual esse algoritmo analisa se todos os exemplos de determinada classe estão cobertos pelo conjunto de regras da mesma. Finalizada essa etapa, o algoritmo troca a classe e reinicia o processo, até que todas as classes obtenham os seus conjuntos de regras (RAJPUT *et al.*, 2011).

### 2.3.4 Baseado em Instâncias

Tal categoria também é conhecida como Aprendizado Preguiçoso, ou Aprendizado Baseado em Memória. Possuem essa definição por utilizarem todos os dados para gerar cada previsão, encontrando um conjunto de vizinhos próximos, baseando-se em funções matemáticas de distâncias entre pontos onde os vizinhos mais próximos tem uma relevância maior (ATKESON;

MOORE; SCHAAU, 1997; RUSSELL; NORVIG, 2009). Os classificadores baseados em instâncias possuem o melhor desempenho sobre bases de dados com atributos numéricos e, preferencialmente normalizados, uma vez que utilizam medidas de distância com o objetivo de identificar as instâncias próximas entre si.

O  $k$ -NN é um dos algoritmos do aprendizado em instâncias, cuja ideia é utilizar uma quantidade de instâncias  $k$  que sejam as mais próximas, mediante a distância, do exemplo que está sendo avaliado. Existem diversas métricas utilizadas na realização desse cálculo, dentre elas: a distância Euclidiana, a distância de *Chebyshev* também conhecida como valor máximo, a distância de *Manhattan* que é calculada através da diferença entre os pares de pontos dados (AHA; KIBLER; ALBERT, 1991; MULAK; TALHAR, 2015).

## 2.4 Trabalhos Relacionados

Rodrigues, Santos e Canuto (2013), Rodrigues, Canuto e Santos (2014), Tao *et al.* (2018) e Vale *et al.* (2018) propõem pesquisas que façam uso do algoritmo *Self-Training* para problemas de classificação de rótulo único ou multirrótulo. Enquanto que Tanha, Someren e Afsarmanesh (2017) desenvolveram diversas modificações em um classificador aliado ao algoritmo *Self-Training*.

Rodrigues, Santos e Canuto (2013) baseou-se no algoritmo *Self-Training* e sugeriu um novo método para a classificação multirrótulo. O trabalho propôs reduzir a aleatoriedade da escolha dos exemplos durante o processo de rotulagem, fazendo uso de um limiar, aliado ao fator de confiança (o pertencimento de um exemplo a sua classe) para decidir se a instância deveria ir para o conjunto dos rotulados, ou não. Por fim, foi realizada uma análise comparativa entre os três métodos, utilizando-se cinco bases com seis métricas de avaliação.

Rodrigues, Canuto e Santos (2014) identificaram que o uso do limiar de inclusão causou uma limitação no número de instâncias adicionadas, uma vez que uma quantidade bem menor de instâncias é adicionada, necessitando assim de uma quantidade maior de iterações. Como solução, foi proposto retirar esse limiar e fazer uso apenas do fator de confiança, ordenando os exemplos de forma decrescente com base nesse valor e selecionando  $\chi$  exemplos por iteração.

Em Tanha, Someren e Afsarmanesh (2017), os autores mostraram que uma árvore de decisão (classificador supervisionado) não produz boas estimativas de probabilidade para as previsões quando aplicadas em conjunto com o *Self-Training*. Foram desenvolvidas várias modificações nesse classificador com a finalidade de produzir uma estimativa melhor. As técnicas utilizadas foram: i) não poda da árvore; ii) correção de *Laplace*; iii) uma medida baseada em distância.

Em Tao *et al.* (2018), foi proposto uma modificação do *Self-Training*, na qual o algoritmo utiliza dados rotulados e não rotulados para realizar a classificação dos dados, além de utilizar



duas medidas para auxiliar o método proposto, sendo elas: i) um método baseado na caminhada aleatória, utilizado junto ao classificador para gerar previsões confiáveis nos dados não rotulados; ii) o sub-conjunto ótimo  $U'$  que gera mais instâncias no conjunto de treinamento.

No trabalho de [Vale et al. \(2018\)](#), foi realizada uma comparação entre o algoritmo *Self-Training* e três métodos propostos pelos autores. Esses métodos diferenciam-se por realizarem uma nova forma de calcular o limiar controlado dinamicamente, incluindo assim novos exemplos no conjunto dos rotulados. Os resultados das variações propostas obtiveram taxas de acerto melhores que o método proposto por [Rodrigues, Canuto e Santos \(2014\)](#).

O foco desta pesquisa é desenvolver uma análise no método proposto em [Vale et al. \(2018\)](#), o *FlexCon-C*, que utiliza um classificador para avaliar se o limiar deve ser alterado ou não. Diferindo-se por realizar uma análise dos efeitos de diferentes valores ao parâmetro que altera o limiar de inclusão, para avaliar as acurácias obtidas pelos classificadores: Naïve Bayes, *rpartXse*, *k-NN* e *RIPPER*.

## 3 Desenvolvimento da Pesquisa

Neste capítulo será descrito em detalhes o que foi desenvolvido e como a ideia foi executada. Ao fim do capítulo, está descrito o design de experimento utilizado nesta pesquisa.

### 3.1 Questões de Pesquisa

Como explanado na Subseção 1.2.1, este trabalho conduziu uma análise no comportamento do método **FlexCon-C**, realizando manipulações exclusivamente na taxa de variação do limiar, mantendo a proposta original discutida em (VALE *et al.*, 2018), com a taxa de variação fixa em 5%. Na presente pesquisa, foi avaliada a influência do parâmetro que varia o limiar na classificação de dados semissupervisionados. A faixa de variação do parâmetro considerada foi entre 2% e 8%, comparando os resultados obtidos de cada um destes valores com o valor base 5%.

A validação da proposta deste trabalho foi realizada por meio de uma comparação das médias das acurácias e variâncias obtidas pelo respectivo classificador. Para realizar uma análise do ponto de vista estatístico foi aplicado o teste de *Friedman*, pois esse ignora a suposição de normalidade dos dados.

### 3.2 Proposta de Solução

Nesta seção são descritas as alterações propostas no método **FlexCon-C** e as bases de dados utilizadas nesta pesquisa. O limiar da iteração subsequente é calculado verificando se o conjunto dos exemplos adicionados na iteração atual possui uma quantidade mínima de exemplos de todas as classes, em caso positivo o classificador treina com os exemplos adicionados da iteração e testa com os exemplos inicialmente rotulados. Em caso negativo o conjunto de treino fica se acumulando até possuir a quantidade mínima de exemplos por classe.

O limiar de inclusão da primeira iteração é de 95%, isto é, são selecionadas as instâncias que na primeira iteração possuem uma confiança igual ou superior a este valor para compor o conjunto dos dados rotulados. A partir da segunda iteração, o algoritmo realiza o cálculo do novo limiar de inclusão considerando a acurácia do classificador e esse limiar vai receber uma variação conforme representado na Equação 6 proposta em (VALE *et al.*, 2018),

$$conf(t_{i+1}) = \begin{cases} conf(t_i) - cr, & \text{se } acc \geq mp + e \\ conf(t_i), & \text{se } mp - e < acc < mp + e \\ conf(t_i) + cr, & \text{se } acc \leq mp - e \end{cases} \quad (6)$$

onde,  $conf(t_{i+1})$  é o limiar da próxima iteração,  $conf(t_i)$  é o limiar da iteração atual,  $cr$  é a taxa de variação na qual o limiar é aplicado,  $acc$  é a acurácia do classificador,  $mp$  precisão mínima e  $e$  é uma variação de precisão aceitável que é permitida para definir uma estabilização em precisão definida em 1%.

O parâmetro  $acc$  é calculado a partir das instâncias treinadas pelo classificador ( $f'$ ) do conjunto selecionado para a inclusão ( $S$ ).  $mp + e$  é o valor mínimo de  $acc$  que o classificador  $f'$  deve atingir para reduzir o limiar de inclusão, para a inclusão de um número maior de exemplos.  $mp - e$

Em [Vale et al. \(2018\)](#) fixou o parâmetro  $cr$  em 5%, isto é, o limiar variava em 5 pontos percentuais por iteração. Entretanto, este trabalho pretende utilizar diversos valores [2%, 8%] do parâmetro  $cr$  para analisar o seu efeito na classificação dos dados.

Com a finalidade de avaliar a eficácia das variações propostas, foi realizada uma análise empírica. A partir da aplicação do método [FlexCon-C](#) aliado a 4 classificadores sobre um conjunto de 31 bases de dados. A Tabela 1 descreve os conjuntos de dados previamente selecionados, em termos dos números de exemplos (Instâncias), de atributos e a quantidade de classes para cada conjunto de dados, além de indicar os tipos de dados como categórico (C), inteiro (I) ou real (R).

A solução proposta foi desenvolvida na linguagem de programação R, em virtude de ser um software livre (*open source*) que vem apresentando crescimento ao longo dos anos. Além disso, a referida linguagem fornece diversas ferramentas e bibliotecas para facilitar as análises estatísticas.

### 3.3 Design de Experimento

Nesta seção está descrito o *design* (modelo lógico) utilizado para a execução dos experimentos, conforme apresentado no Pseudocódigo 3.

Cada um dos classificadores selecionados para esse estudo (Naïve Bayes, *rpartXse*, *k-NN* e *RIPPER*) é avaliado individualmente, utilizando-se a técnica de validação cruzada (do inglês, *cross-validation*). Essa técnica é utilizada para a validação de dados obtidos em experimentos, dividido-se a base de dados em  $k$  *folds* e mantendo-se a estratificação (proporção de cada classe) nos mesmos. Além disto, cada *fold* é utilizado como conjunto de teste para verificar a acurácia do classificador, sendo utilizados  $k - 1$  *folds* para treinamento e 1 *fold* para teste do classificador ([WONG, 2015](#)).

O algoritmo recebe por parâmetro o classificador e a execução segue os seguintes passos: i) para cada base de dados, carregue-a em memória; ii) separe os *folds* na base selecionada; iii) para cada valor do parâmetro  $cr$ , itere os percentuais de exemplos inicialmente rotulados; iv) para cada um dos percentuais de exemplos inicialmente rotulados, percorra os *folds* criados; v)

Tabela 1 – Bases de Dados que serão utilizadas nos experimentos.

Base de dados	Instâncias	Atributos	Classes	Tipo
Balance Scale	625	4	3	I
Blood Transfusion Service	748	5	2	R
Bupa	345	7	2	C, I, R
Car	1728	6	4	C, R
Cnae-9	1080	857	9	I
Connectionist-mines-vs-rocks	208	60	2	R
Flare	1389	10	6	C
Haberman	306	4	2	I
Handwritten Digits	10992	16	10	I
Hill Valley	606	101	2	R
Image Segmentation	2310	19	7	R
Indian Liver Patient	582	10	2	I, R
Iris	150	5	3	R
King-Rook vs King Pawn	3196	36	2	C
Leukemia Haslinger	100	50	2	R
Mammographic Mass	961	6	2	I
Mfeat-karhunen	2000	64	10	I, R
Mushroom	8124	22	2	C
Musk	6598	168	2	I
Ozone Level Detection	2536	73	2	R
Phishing	2456	30	3	I
Pima	768	9	2	I, R
Planning-relax	182	13	2	R
Seeds	210	7	3	R
Semeion	1593	256	10	I
Spectf Heart	267	14	2	I
Tic-tac-toe	958	9	2	C
Twonorm	7400	21	2	R
Vehicle	946	18	4	I
Waveform	5000	40	3	R
Wilt	4839	6	2	R

Fonte: O Próprio Autor (2018)

para cada um dos *folds*, separe os conjuntos de treino e teste; vi) gere os modelos a partir do método **FlexCon-C**; vii) teste a acurácia dos modelos com o conjunto selecionado para teste; viii) grave as informações em arquivo.

Com este *design*, garante-se a mesma distribuição dos *folds* para as técnicas implementadas, quando fixados os valores do parâmetro *cr* e da taxa de exemplos inicialmente rotulados. Além disto, o **FlexCon-C** teve que ser adaptado para receber dois parâmetros, o valor do *cr* e o

classificador a ser utilizado.

---

**Pseudocódigo 3:** *Design* de Experimento utilizado.

---

**entrada** *classificador cl*

```

1: para cada base em base_de_dados faça
2:     base_carregada ← carregar_base(base)
3:     folds ← separar_folds(base_carregada)
4:     para cada cr em change_rate faça
5:         para i ← 1 até 5 faça
6:             ini_rot ← i * 5
7:             para cada fold em folds faça
8:                 teste ← fold
9:                 treinamento ← folds - fold
10:                L ← selecionar_exemplos(treinamento, ini_rot)
11:                U ← treinamento - L
12:                modelo_c1_s ← flexcon_c1_s(L, U, cr, cl)
13:                modelo_c1_v ← flexcon_c1_v(L, U, cr, cl)
14:                modelo_c2 ← flexcon_c2(L, U, cr, cl)
15:                acc_c1_s[i] ← testar_acc(modelo_c1_s, teste)
16:                acc_c1_v[i] ← testar_acc(modelo_c1_v, teste)
17:                acc_c2[i] ← testar_acc(modelo_c2, teste)
18:            fim para
19:        fim para
20:        gravar_arquivo(acc_c1_s, acc_c1_v, acc_c2)
21:    fim para
22: fim para

```

---

**Fonte:** O Próprio Autor (2018)

Após essa etapa de separação dos *folds* para treinamento do classificador, foi selecionado um percentual de exemplos para permanecer rotulado. Uma vez separados os dados em dois conjuntos: rotulados e não rotulados, estes são aplicados ao método para realização da tarefa de classificação, utilizando-se os classificadores selecionados.

Nos experimentos foram utilizados os seguintes parâmetros:

- i.  $k = 10$  *folds* para o *cross-validation*, conforme proposto por (KOHAVI, 1995; BRAGANETO; DOUGHERTY, 2004; FUSHIKI, 2011; WONG, 2015);
- ii. Percentuais de 5%, 10%, 15%, 20% e 25% instâncias inicialmente rotuladas, conforme descrito em (VALE *et al.*, 2018);

- iii. Classificadores Naïve Bayes, *rpartXse*, e **RIPPER**, os mesmos utilizados no trabalho de (VALE *et al.*, 2018), além da inclusão do classificador *k*-NN.

O valor de *k* (quantidade de instâncias mais próximas) definido para o *k*-NN é descrito pela Equação 7, conforme proposto em (BHATTACHARYA; GHOSH; CHOWDHURY, 2012):

$$k = \sqrt{N} \quad (7)$$

onde, *N* representa a quantidade total de instâncias na base de dados.

## 4 Resultados

Neste capítulo são discutidos os resultados obtidos após a aplicação da proposta apresentada no Capítulo 3. A abordagem possui dois momentos: no primeiro, descrito na Seção 4.1, são apresentados os resultados obtidos por cada classificador individualmente; no segundo momento, Seção 4.2, é conduzida uma análise estatística dos resultados obtidos.

### 4.1 Análise de Eficácia

As Tabelas 2 - 5 apresentam os desempenhos obtidos pelo método *FlexCon-C*, considerando-se cada uma das técnicas abordadas, *FlexCon-C1 (s)*, *FlexCon-C1 (v)* e *FlexCon-C2*. Nestas tabelas, a coluna 1 descreve a técnica que está sendo analisada, na coluna 2 representa os diversos percentuais (5, 10, 15, 20 e 25) de exemplos inicialmente rotulados e as colunas 3 a 9 representam os resultados (média e desvio padrão) obtidos com as variações aplicadas ao parâmetro *cr* obtidos pelos classificadores.

Estão destacados em negrito, a maior acurácia obtida pela respectiva técnica para cada um dos percentuais de exemplos inicialmente rotulados. Além disto, também estão sombreadas as células que obtiveram uma acurácia maior que a obtida pelo *cr* de 5%, utilizado no trabalho de Vale *et al.* (2018).

A Tabela 2, gerada a partir das médias dos resultados nos Apêndices A - C, apresenta a média e o desvio padrão utilizando o classificador Naïve Bayes. Conforme pode ser observado

Tabela 2 – Resultado da acurácia com o classificador Naïve Bayes.

Téc.	Exemplos Inicialmente Rotulados	Variação no parâmetro <i>cr</i>						
		2%	3%	4%	5%	6%	7%	8%
FlexCon-C1(s)	05%	72,38 ± 18,19	73,19 ± 17,53	72,59 ± 17,70	73,10 ± 17,43	72,85 ± 17,85	<b>73,26 ± 17,42</b>	73,03 ± 17,86
	10%	73,71 ± 16,84	73,16 ± 17,02	<b>74,35 ± 16,81</b>	73,60 ± 17,54	72,89 ± 16,73	73,85 ± 17,33	73,73 ± 17,02
	15%	73,23 ± 17,66	<b>73,70 ± 17,41</b>	73,35 ± 17,55	73,40 ± 17,68	73,47 ± 17,57	72,96 ± 17,96	73,41 ± 17,34
	20%	73,24 ± 17,41	73,49 ± 17,07	72,98 ± 17,52	<b>74,02 ± 16,96</b>	73,99 ± 16,32	73,80 ± 17,52	73,65 ± 17,29
	25%	73,68 ± 17,33	73,74 ± 17,00	73,89 ± 17,06	73,47 ± 17,34	<b>74,11 ± 16,88</b>	73,51 ± 17,47	73,38 ± 17,26
FlexCon-C1(v)	05%	72,38 ± 18,19	73,19 ± 17,53	72,59 ± 17,70	73,10 ± 17,43	72,85 ± 17,85	<b>73,26 ± 17,42</b>	73,03 ± 17,86
	10%	73,71 ± 16,84	73,16 ± 17,02	<b>74,35 ± 16,81</b>	73,60 ± 17,54	72,89 ± 16,73	73,85 ± 17,33	73,73 ± 17,02
	15%	73,23 ± 17,66	<b>73,70 ± 17,41</b>	73,35 ± 17,55	73,40 ± 17,68	73,47 ± 17,57	72,96 ± 17,96	73,41 ± 17,34
	20%	73,24 ± 17,41	73,49 ± 17,07	72,98 ± 17,52	<b>74,02 ± 16,96</b>	73,99 ± 16,32	73,80 ± 17,52	73,65 ± 17,29
	25%	73,68 ± 17,33	73,74 ± 17,00	73,89 ± 17,06	73,47 ± 17,34	<b>74,11 ± 16,88</b>	73,51 ± 17,47	73,38 ± 17,26
FlexCon-C2	05%	72,38 ± 18,19	73,19 ± 17,53	72,59 ± 17,70	73,10 ± 17,43	72,85 ± 17,85	<b>73,26 ± 17,42</b>	73,03 ± 17,86
	10%	73,71 ± 16,84	73,16 ± 17,02	<b>74,35 ± 16,81</b>	73,60 ± 17,54	72,89 ± 16,73	73,85 ± 17,33	73,73 ± 17,02
	15%	73,23 ± 17,66	<b>73,70 ± 17,41</b>	73,35 ± 17,55	73,40 ± 17,68	73,47 ± 17,57	72,96 ± 17,96	73,41 ± 17,34
	20%	73,24 ± 17,41	73,49 ± 17,07	72,98 ± 17,52	<b>74,02 ± 16,96</b>	73,99 ± 16,32	73,80 ± 17,52	73,65 ± 17,29
	25%	73,68 ± 17,33	73,74 ± 17,00	73,89 ± 17,06	73,47 ± 17,34	<b>74,11 ± 16,88</b>	73,51 ± 17,47	73,38 ± 17,26

Fonte: O Próprio Autor (2018)

(valores em negrito) todas as técnicas apresentam melhor desempenho quando o valor do

$cr$  é definido entre 3% e 7%. Realizando uma análise mais detalhada, observa-se que as acurácias maiores do que aquelas usando o  $cr$  com valor de 5% estão bastante dispersas (valores sombreados). De maneira geral, quando o valor do parâmetro  $cr$  é definido entre 3% e 7% os resultados são distribuídos de forma igualitária. Além disto, o classificador em questão, obteve os mesmos percentuais de acerto quando aplicado à todas as técnicas estudadas.

A Tabela 3, gerada a partir das médias dos resultados nos Apêndices D - F, apresenta os resultados obtidos pelo classificador *rpartXse*. Para as técnicas *FlexCon-C1 (s)* e *FlexCon-C1 (v)* observou-se melhores acurácias ao utilizar valores mais elevados (acima de 5%) para o parâmetro  $cr$ . Além disto, quando o valor é prefixado em 6% obtém resultados mais significativos que o  $cr$  de 5% em 10 dos 15 casos. Na técnica *FlexCon-C2* o valor de 5% utilizado no trabalho de [Vale et al. \(2018\)](#) obtém os melhores resultados em 2 dos 5 casos. Entretanto, para essa combinação de classificador e técnica os resultados obtidos estão bastante dispersos, parte dos melhores valores concentram-se na faixa intermediária 4% e 5%, e o restante se divide entre 2% e 8%. Em geral, observando-se todas as combinações classificador e técnica, os melhores resultados tendem para valores iguais ou superiores a 5%.

Tabela 3 – Resultado da acurácia com o classificador *rpartXse*.

Téc.	Exemplos Inicialmente Rotulados	Variação no parâmetro $cr$						
		2%	3%	4%	5%	6%	7%	8%
FlexCon-C1 (s)	05%	<b>76,17</b> ± 14,75	75,58 ± 15,32	75,49 ± 15,61	75,23 ± 16,12	75,92 ± 15,01	75,45 ± 15,91	75,67 ± 15,64
	10%	75,73 ± 15,72	75,87 ± 15,50	75,25 ± 15,82	76,11 ± 15,25	<b>76,18</b> ± 15,43	75,64 ± 15,39	76,00 ± 15,15
	15%	75,77 ± 15,41	75,71 ± 15,48	75,47 ± 15,71	75,31 ± 15,79	75,60 ± 15,90	75,95 ± 15,28	<b>76,06</b> ± 15,29
	20%	75,73 ± 15,44	75,81 ± 15,39	<b>75,99</b> ± 15,26	75,80 ± 15,27	75,89 ± 15,39	75,66 ± 15,22	75,91 ± 14,93
	25%	75,75 ± 15,36	75,49 ± 15,21	75,98 ± 15,24	75,84 ± 15,20	75,47 ± 15,29	<b>76,00</b> ± 15,15	75,15 ± 15,65
FlexCon-C1 (v)	05%	76,10 ± 14,73	75,47 ± 15,37	75,41 ± 15,68	75,43 ± 15,78	<b>76,13</b> ± 14,75	75,59 ± 15,74	75,69 ± 15,86
	10%	75,79 ± 15,50	75,59 ± 15,74	75,48 ± 15,61	<b>76,13</b> ± 15,23	76,04 ± 15,58	75,63 ± 15,40	75,90 ± 15,21
	15%	75,67 ± 15,45	75,59 ± 15,66	75,38 ± 15,95	75,54 ± 15,53	75,74 ± 15,62	75,68 ± 15,38	<b>76,13</b> ± 15,32
	20%	75,84 ± 15,39	75,75 ± 15,48	75,88 ± 15,48	75,66 ± 15,39	<b>75,90</b> ± 15,45	75,61 ± 15,23	75,78 ± 15,04
	25%	75,84 ± 15,25	75,61 ± 15,18	75,99 ± 15,25	75,70 ± 15,43	75,59 ± 15,14	<b>76,05</b> ± 15,12	75,50 ± 15,21
FlexCon-C2	05%	<b>76,05</b> ± 14,90	75,65 ± 15,22	75,47 ± 15,57	75,63 ± 15,64	75,92 ± 14,91	75,61 ± 15,84	75,88 ± 15,53
	10%	75,89 ± 15,56	75,71 ± 15,68	75,53 ± 15,58	<b>76,13</b> ± 15,36	76,00 ± 15,45	75,74 ± 15,43	75,96 ± 15,23
	15%	75,70 ± 15,34	75,81 ± 15,57	75,49 ± 15,83	75,43 ± 15,70	75,70 ± 15,85	75,85 ± 15,37	<b>76,06</b> ± 15,36
	20%	75,85 ± 15,42	75,83 ± 15,47	<b>76,03</b> ± 15,39	75,74 ± 15,22	75,97 ± 15,27	75,75 ± 15,26	75,69 ± 15,06
	25%	75,94 ± 15,21	75,53 ± 15,37	75,97 ± 15,30	<b>76,08</b> ± 15,08	75,58 ± 15,26	75,95 ± 15,22	75,48 ± 15,21

Fonte: O Próprio Autor (2018)

Na Tabela 4, gerada a partir das médias das tabelas apresentadas nos Apêndices G - I, estão expostos os resultados do classificador *RIPPER*. Para as 3 técnicas analisadas neste trabalho observa-se que todas obtiveram comportamentos semelhantes atingindo os mesmos resultados quando se mantém fixo os percentuais de exemplos inicialmente rotulados e o valor do  $cr$ . No geral, as maiores acurácias estão localizadas quando o valor do  $cr > 5%$  resultando em 12 dos 15 casos, sendo que o valor de 7% alcançou os melhores resultados em 6 dos 15 casos. Considerando os percentuais de exemplos inicialmente rotulados (5, 10, 15 e 20) observa-se que a medida que este percentual aumenta, obtém melhor eficácia reduzindo o valor do parâmetro  $cr$ . Assim, verifica-se uma relação inversamente proporcional entre o parâmetro  $cr$  e o percentual de exemplos inicialmente rotulados.



Tabela 4 – Resultado da acurácia com o classificador RIPPER.

Téc.	Exemplos Inicialmente Rotulados	Variação no parâmetro $cr$						
		2%	3%	4%	5%	6%	7%	8%
FlexCon-C1(s)	05%	73,43 ± 15,31	73,46 ± 15,24	73,82 ± 15,23	73,75 ± 15,12	73,63 ± 15,24	73,54 ± 15,54	<b>73,92</b> ± 14,44
	10%	74,69 ± 14,13	74,51 ± 14,51	74,45 ± 14,73	74,73 ± 14,14	74,58 ± 14,57	<b>74,77</b> ± 14,66	74,66 ± 15,09
	15%	74,41 ± 14,24	73,95 ± 14,99	73,91 ± 14,90	74,07 ± 14,65	<b>74,70</b> ± 14,65	74,25 ± 14,81	74,28 ± 14,66
	20%	74,06 ± 15,36	73,98 ± 15,04	<b>74,41</b> ± 15,38	74,24 ± 15,19	74,07 ± 14,90	73,76 ± 14,71	73,43 ± 15,49
	25%	73,83 ± 15,37	73,54 ± 14,70	73,55 ± 15,36	74,04 ± 15,03	73,96 ± 15,11	<b>74,23</b> ± 15,08	73,71 ± 15,55
FlexCon-C1(v)	05%	73,43 ± 15,31	73,46 ± 15,24	73,82 ± 15,23	73,75 ± 15,12	73,63 ± 15,24	73,54 ± 15,54	<b>73,92</b> ± 14,44
	10%	74,69 ± 14,13	74,51 ± 14,51	74,45 ± 14,73	74,73 ± 14,14	74,58 ± 14,57	<b>74,77</b> ± 14,66	74,66 ± 15,09
	15%	74,41 ± 14,24	73,95 ± 14,99	73,91 ± 14,90	74,07 ± 14,65	<b>74,70</b> ± 14,65	74,25 ± 14,81	74,28 ± 14,66
	20%	74,06 ± 15,36	73,98 ± 15,04	<b>74,41</b> ± 15,38	74,24 ± 15,19	74,07 ± 14,90	73,76 ± 14,71	73,43 ± 15,49
	25%	73,83 ± 15,37	73,54 ± 14,70	73,55 ± 15,36	74,04 ± 15,03	73,96 ± 15,11	<b>74,23</b> ± 15,08	73,71 ± 15,55
FlexCon-C2	05%	73,43 ± 15,31	73,46 ± 15,24	73,82 ± 15,23	73,75 ± 15,12	73,63 ± 15,24	73,54 ± 15,54	<b>73,92</b> ± 14,44
	10%	74,69 ± 14,13	74,51 ± 14,51	74,45 ± 14,73	74,73 ± 14,14	74,58 ± 14,57	<b>74,77</b> ± 14,66	74,66 ± 15,09
	15%	74,41 ± 14,24	73,95 ± 14,99	73,91 ± 14,90	74,07 ± 14,65	<b>74,70</b> ± 14,65	74,25 ± 14,81	74,28 ± 14,66
	20%	74,06 ± 15,36	73,98 ± 15,04	<b>74,41</b> ± 15,38	74,24 ± 15,19	74,07 ± 14,90	73,76 ± 14,71	73,43 ± 15,49
	25%	73,83 ± 15,37	73,54 ± 14,70	73,55 ± 15,36	74,04 ± 15,03	73,96 ± 15,11	<b>74,23</b> ± 15,08	73,71 ± 15,55

Fonte: O Próprio Autor (2018)

A Tabela 5, foi gerada a partir das médias dos Apêndices J - L, apresenta os resultados obtidos pelo classificador  $k$ -NN. Para as técnicas  $FlexCon-C1(s)$  e  $FlexCon-C1(v)$  os melhores

Tabela 5 – Resultado da acurácia com o classificador  $k$ -NN.

Téc.	Exemplos Inicialmente Rotulados	Variação no parâmetro $cr$						
		2%	3%	4%	5%	6%	7%	8%
FlexCon-C1(s)	05%	79,21 ± 13,30	<b>79,30</b> ± 13,80	78,89 ± 14,17	79,12 ± 13,59	78,81 ± 13,94	79,01 ± 13,45	79,27 ± 14,04
	10%	78,95 ± 13,59	78,69 ± 13,91	78,89 ± 13,53	78,26 ± 13,92	<b>78,96</b> ± 13,81	78,50 ± 14,18	78,91 ± 13,80
	15%	79,53 ± 13,55	79,25 ± 13,54	79,29 ± 13,67	<b>79,77</b> ± 13,34	79,05 ± 13,78	79,44 ± 13,71	79,34 ± 13,65
	20%	79,07 ± 13,38	78,99 ± 13,43	<b>79,08</b> ± 13,38	79,07 ± 13,44	79,05 ± 13,68	78,91 ± 13,27	78,91 ± 13,67
	25%	<b>79,96</b> ± 13,41	79,75 ± 13,24	79,46 ± 13,63	79,64 ± 13,72	79,72 ± 13,27	79,25 ± 13,62	79,78 ± 13,35
FlexCon-C1(v)	05%	79,21 ± 13,30	<b>79,30</b> ± 13,80	78,89 ± 14,17	79,12 ± 13,59	78,81 ± 13,94	79,01 ± 13,45	79,27 ± 14,04
	10%	78,95 ± 13,59	78,69 ± 13,91	78,89 ± 13,53	78,26 ± 13,92	<b>78,96</b> ± 13,81	78,50 ± 14,18	78,91 ± 13,80
	15%	79,53 ± 13,55	79,25 ± 13,54	79,29 ± 13,67	<b>79,77</b> ± 13,34	79,05 ± 13,78	79,44 ± 13,71	79,34 ± 13,65
	20%	79,07 ± 13,38	78,99 ± 13,43	<b>79,08</b> ± 13,38	79,07 ± 13,44	79,05 ± 13,68	78,91 ± 13,27	78,91 ± 13,67
	25%	<b>79,96</b> ± 13,41	79,75 ± 13,24	79,46 ± 13,63	79,64 ± 13,72	79,72 ± 13,27	79,25 ± 13,62	79,78 ± 13,35
FlexCon-C2	05%	79,17 ± 13,35	<b>79,27</b> ± 13,82	78,95 ± 14,13	79,26 ± 13,55	78,59 ± 14,21	78,98 ± 13,57	79,26 ± 13,96
	10%	78,96 ± 13,54	78,67 ± 13,84	78,79 ± 13,60	78,15 ± 13,96	78,82 ± 14,01	78,53 ± 14,17	<b>79,02</b> ± 13,73
	15%	<b>79,66</b> ± 13,42	79,33 ± 13,43	79,30 ± 13,59	79,59 ± 13,60	79,02 ± 13,75	79,45 ± 13,74	79,35 ± 13,55
	20%	<b>79,14</b> ± 13,27	78,95 ± 13,39	79,08 ± 13,43	79,10 ± 13,38	79,02 ± 13,68	78,91 ± 13,27	78,88 ± 13,73
	25%	<b>79,82</b> ± 13,42	79,65 ± 13,30	79,53 ± 13,67	79,62 ± 13,79	79,77 ± 13,35	79,17 ± 13,75	<b>79,82</b> ± 13,26

Fonte: O Próprio Autor (2018)

resultados estão distribuídos entre 2% e 6%. Enquanto que na técnica  $FlexCon-C2$  os melhores resultados concentram-se entre 2% e 3% com 4 dos 5 casos, usando 25% dos exemplos rotulados inicialmente ambos os  $cr(2%$  e  $8%)$  atingem a mesma acurácia, entretanto o valor de 8% possui um desvio padrão menor que o valor de 2%. Observando os percentuais de exemplos inicialmente rotulados (10, 15, 20 e 25) os melhores resultados concentram-se em 2%, 4%, 5% e 6%, resultando em uma relação inversamente proporcional entre o valor do  $cr$  e o percentual de exemplos inicialmente rotulados. De modo geral, o melhor valor de  $cr$  para essa técnica específica é o valor de 2%, pois este obtém os melhores resultados em 5 dos 15 casos considerando todas as técnicas analisadas.

A Tabela 6, gerada a partir do Apêndice M, apresenta uma síntese do desempenho dos classificadores nos experimentos, onde cada linha representa um classificador e cada uma das colunas indica um valor para o parâmetro  $cr$ . Em cada célula, da tabela está representado a quantidade de vezes que o valor de  $cr$  obteve o melhor desempenho com o classificador analisado. Nesta tabela, observa-se que em geral quando o  $cr$  é definido em 6% ou 7% obtém-se os melhores resultados em 18,33% dos casos. E ao utilizar o valor do  $cr$  fixado em 5% ele é superior em 13,34% dos casos, obtendo o mesmo valor para o  $cr$  de 8%. Para o classificador Naïve Bayes o parâmetro  $cr$  deve ser pré-fixar entre 3% e 7%, o desempenho é efetivo no classificador *rpartXse* fixando o  $cr$  entre 5% e 8%. Por sua vez, o classificador RIPPER obtém os melhores resultados fixando o valor do  $cr$  em 7% (6 dos 15 casos) e definindo o  $cr$  em 2% resultou nas melhores acurácias em 5 casos para o classificador *k-NN*.

Tabela 6 – Desempenho associado a cada valor do parâmetro  $cr$ .

Classificador	Variação do $cr$						
	2%	3%	4%	5%	6%	7%	8%
Naïve Bayes	0	3	3	3	3	3	0
<i>rpartXse</i>	2	0	2	3	3	2	3
RIPPER	0	0	3	0	3	6	3
<i>k-NN</i>	5	3	2	2	2	0	2
<b>Total</b>	7/60 - 11,67%	6/60 - 10,00%	10/60 - 16,67%	8/60 - 13,34%	11/60 - 18,33%	11/60 - 18,33%	8/60 - 13,33%

Fonte: O Próprio Autor (2018).

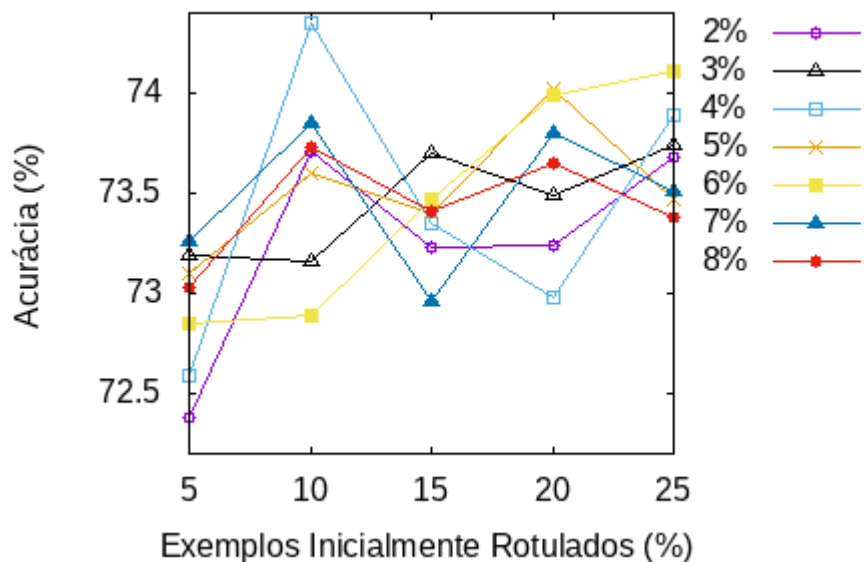
As Figuras 3 - 14 demonstram, por meio de gráficos, as informações contadas nas Tabelas 2 - 5. Para cada uma das referidas tabelas, foram geradas 3 gráficos, um para cada técnica analisada.

Cada uma das linhas, nos gráficos, representa um valor de  $cr$  analisado, o eixo horizontal representa os diferentes percentuais de exemplos inicialmente rotulados e o eixo vertical representa as acurácias. As formas geométricas (■▲●) representam os valores do parâmetro  $cr$  entre 6% e 8%, respectivamente; por sua vez, as formas geométricas (○△□) representam os valores entre 2% e 4%, respectivamente; por fim, o valor de 5% utilizado em Vale *et al.* (2018) está representado por ×.

As Figuras 3 - 5 representam as técnicas utilizadas neste trabalho aliadas ao classificador Naïve Bayes. A partir de uma análise dos gráficos, conclui-se que ao se fixar os valores de  $cr$  em 3% e em 5% são obtidos resultados mais estáveis para todos os percentuais de exemplos inicialmente rotulados avaliados na pesquisa. Ao selecionar valores para o parâmetro  $cr > 5%$ , são obtidos resultados superiores ao  $cr$  de 5% em 6 dos 15 casos e ao considerar os valores de  $cr < 5%$  obtém-se resultados superiores em 7 dos 15 casos. Analisando o gráfico de modo geral observa-se que o valor de  $cr$  fixado previamente em 5% é superior aos outros valores em 17 dos 30 casos, representando 56,67%.

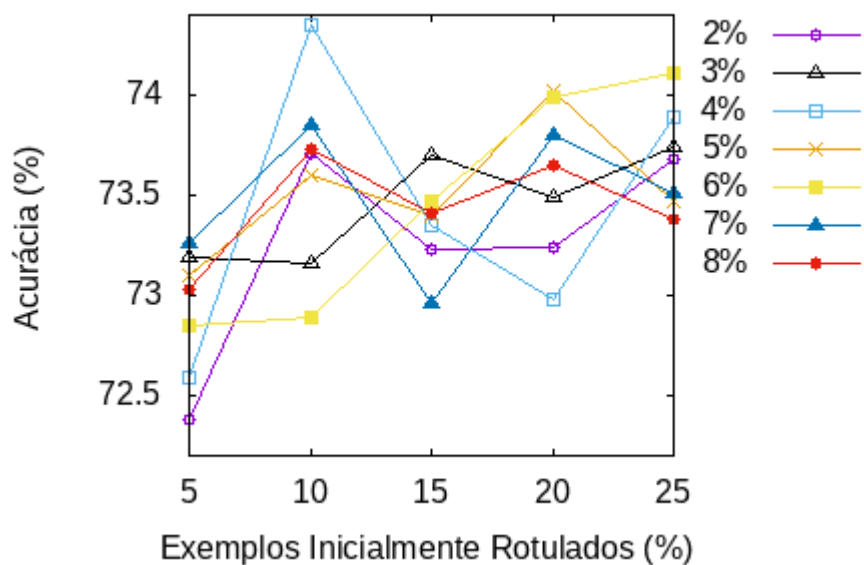
Os gráficos das técnicas utilizando o classificador *rpartXse* estão dispostos nas Figuras 6 - 8. Para a Técnica *FlexCon-C1(s)* (Figura 6) percebe-se uma inconstância quando se

Figura 3 – Gráfico do Classificador Naïve Bayes e a Técnica *FlexCon-C1(s)*.



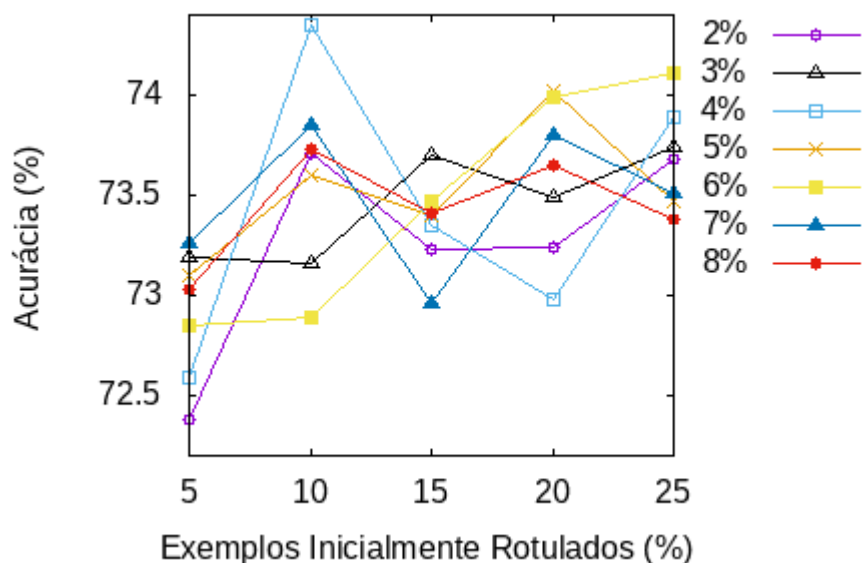
Fonte: O Próprio Autor (2018).

Figura 4 – Gráfico do Classificador Naïve Bayes e a Técnica *FlexCon-C1(v)*.



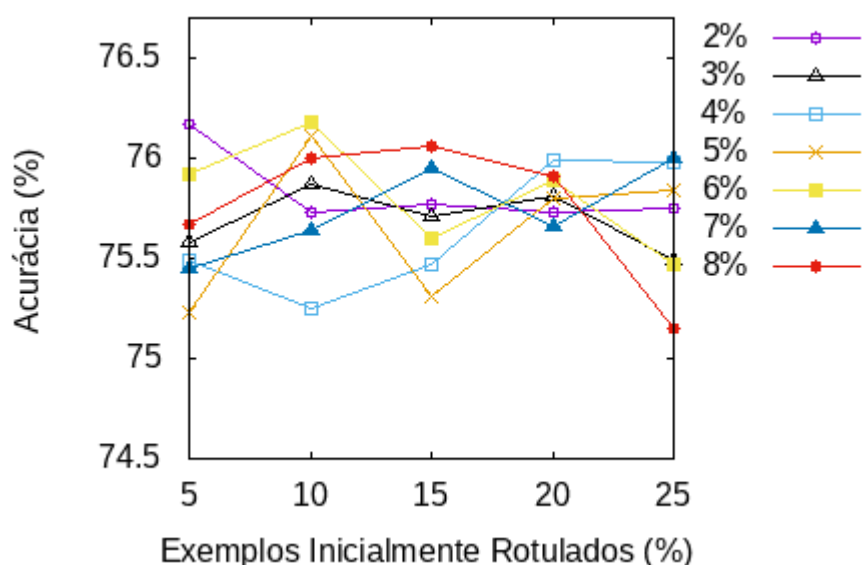
Fonte: O Próprio Autor (2018).

Figura 5 – Gráfico do Classificador Naïve Bayes e a Técnica *FlexCon-C2*.



Fonte: O Próprio Autor (2018).

Figura 6 – Gráfico do Classificador *rpartXse* e a Técnica *FlexCon-C1(s)*.



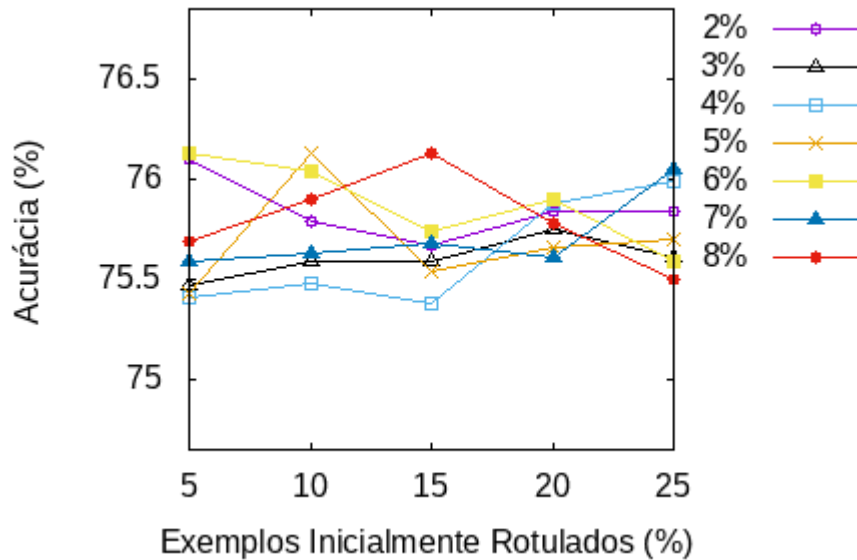
Fonte: O Próprio Autor (2018).

determina o valor do parâmetro  $cr = 5\%$  usando inicialmente 5%, 10% e 15% dos exemplos rotulados. Além disto, tanto para 5% como 15% todos os outros valores para o parâmetro  $cr$  obtiveram resultados mais efetivos.

A Figura 7 apresenta os resultados obtidos com a utilização da Técnica *FlexCon-C1(v)* que ao se fixar o valor do parâmetro  $cr$  em 2% obtém melhores resultados que quando fixa o  $cr$  em 5% em 4 dos 5 percentuais (5, 15, 20 e 25) de exemplos inicialmente rotulados avaliados.

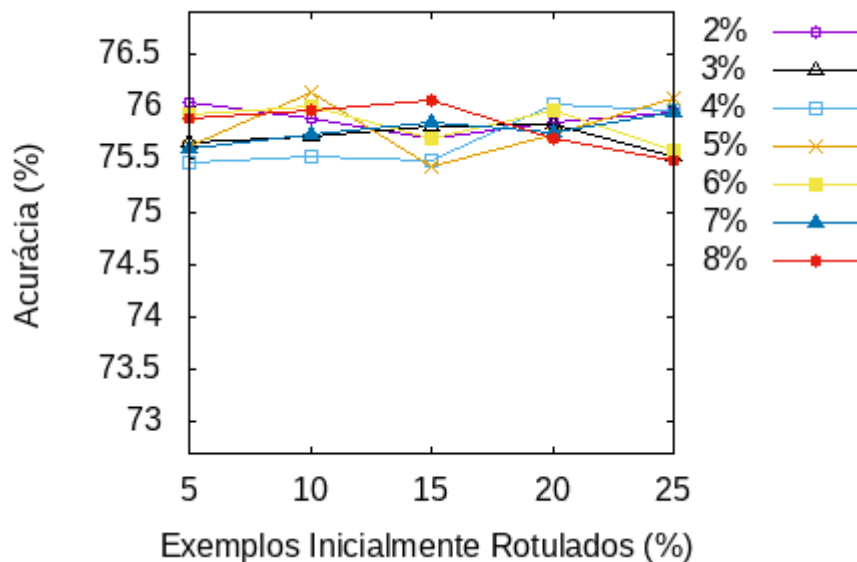
Na Técnica *FlexCon-C2* (Figura 8) analisando o *cr* de 4% observa-se que entre 5% e 15% dos

Figura 7 – Gráfico do Classificador *rpartXse* e a Técnica *FlexCon-C1(v)*.



Fonte: O Próprio Autor (2018).

Figura 8 – Gráfico do Classificador *rpartXse* e a Técnica *FlexCon-C2*.

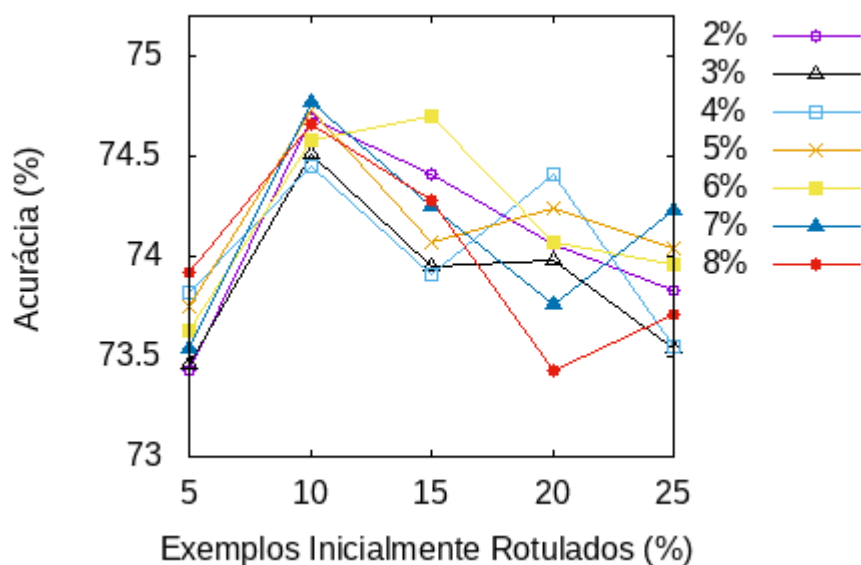


Fonte: O Próprio Autor (2018).

exemplos inicialmente rotulados são obtidos os menores valores de acurácia. Por outro lado, quando utiliza-se 20% e 25% dos exemplos rotulados esse *cr* obtém valores elevados de acurácia. De modo geral, essa técnica não demonstra um comportamento estável para os valores de *cr* considerados.

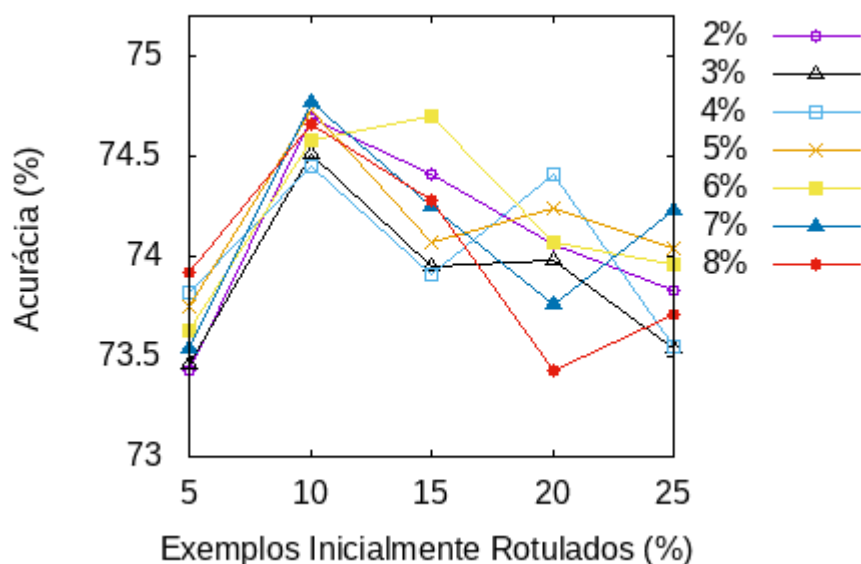
As Figuras 9 - 11 representam as técnicas analisadas aliadas ao classificador **RIPPER**.

Figura 9 – Gráfico do Classificador **RIPPER** e a Técnica *FlexCon-C1(s)*.



Fonte: O Próprio Autor (2018).

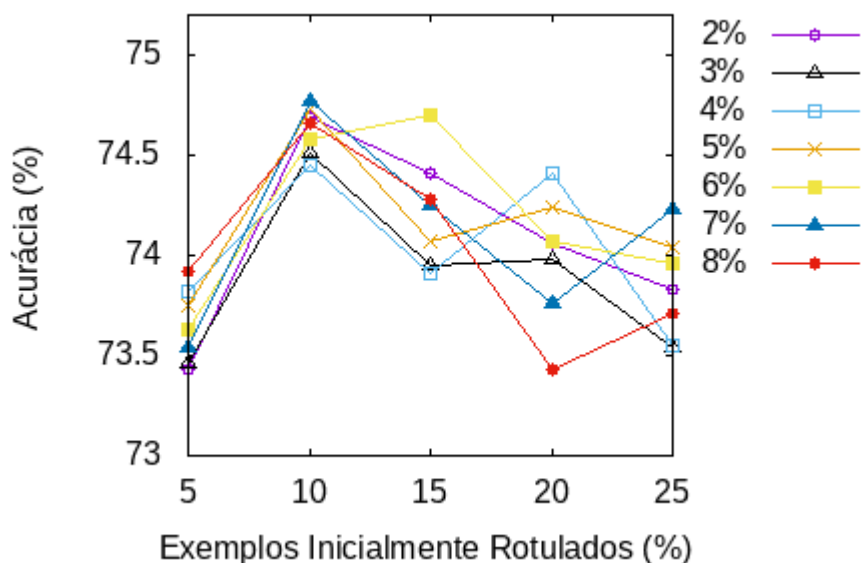
Figura 10 – Gráfico do Classificador **RIPPER** e a Técnica *FlexCon-C1(v)*.



Fonte: O Próprio Autor (2018).

Para as 3 técnicas analisadas, considerando apenas o *cr* de 6% observa-se que entre 5% e 15% de exemplos inicialmente rotulados há crescimento da acurácia, enquanto nos outros *crs* (2%, 3%, 4%, 5%, 7% e 8%) ao utilizar 15% de exemplos inicialmente rotulados há uma queda nas acurácias obtidas. De modo geral, esse classificador não apresenta um comportamento estável para os valores de *cr* considerados.

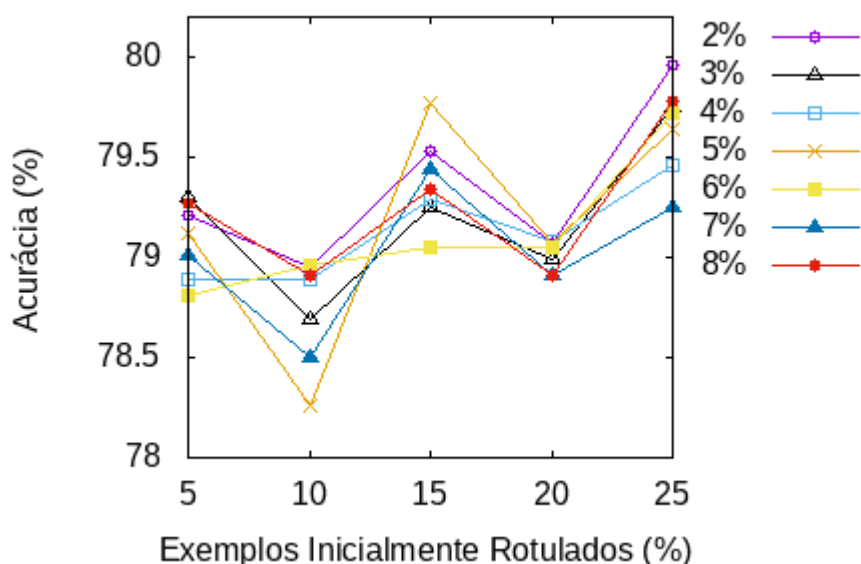
Figura 11 – Gráfico do Classificador RIPPER e a Técnica FlexCon-C2.



Fonte: O Próprio Autor (2018).

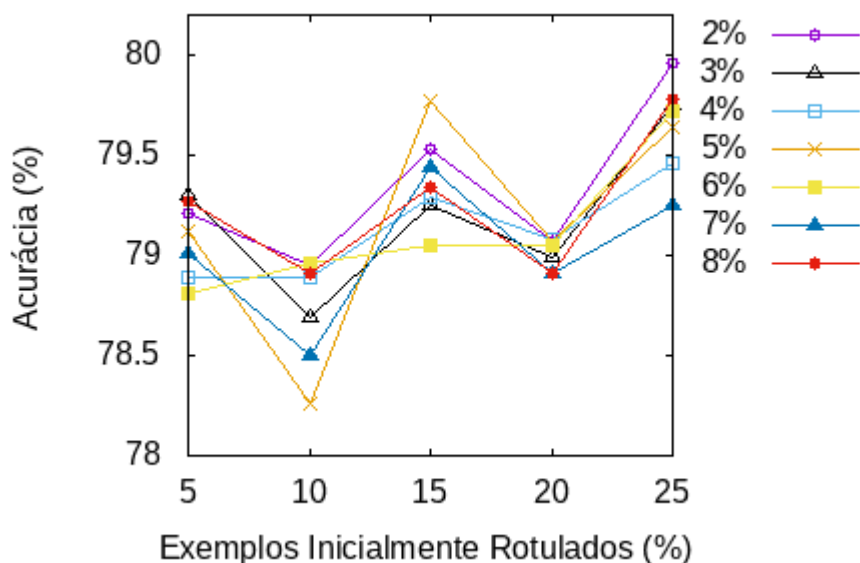
As Figuras 12 - 14 apresentam os gráficos para o classificador *k*-NN. Nas 3 técnicas

Figura 12 – Gráfico do Classificador *k*-NN e a Técnica FlexCon-C1(s).

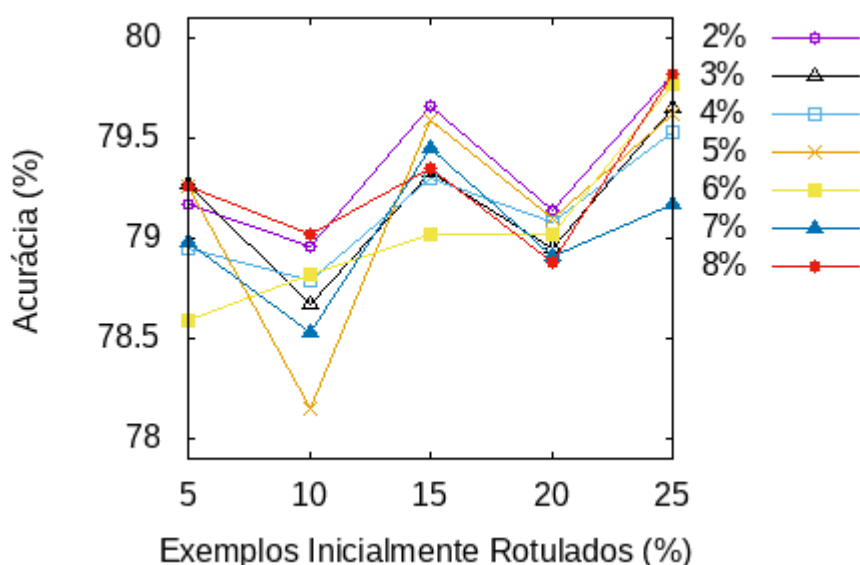


Fonte: O Próprio Autor (2018).

analisadas, ao se utilizar 10% dos exemplos de exemplos inicialmente rotulados o *cr* fixado em 5% sempre obtém os piores resultados. Além disto, os resultados mais estáveis nas técnicas analisadas foram obtidos pelo *cr* de 6%. Para os outros valores de *cr* considerados não há comportamento estável do classificador.

Figura 13 – Gráfico do Classificador  $k$ -NN e a Técnica  $FlexCon-C1(v)$ .

Fonte: O Próprio Autor (2018).

Figura 14 – Gráfico do Classificador  $k$ -NN e a Técnica  $FlexCon-C2$ .

Fonte: O Próprio Autor (2018).

## 4.2 Análise Estatística

Para validar a proposta deste trabalho foi desenvolvida uma análise estatística, para tal, foram aplicados 2 testes, *Shapiro-Will* que avalia a normalidade dos dados e *Friedman* que realiza um ranqueamento dos resultados. Em cada um destes testes é obtido um parâmetro  $p$ -value (valor p) cuja faixa varia entre (0, 1]. O erro do tipo I, simbolizado por  $\alpha$ , é cometido por rejeitar a hipótese nula ela estando correta. Neste trabalho, o nível de significância considerado



para rejeitar a hipótese nula é de  $\alpha = 0.05$ . Em outras palavras, quando o *p-value*  $< 0.05$  é dito que a hipótese nula é rejeitada, caso contrário não rejeita-se a hipótese nula (HOFFMAN, 2015).

Em Razali e Wah (2011) são realizadas comparações entre alguns testes de normalidades. Dentre eles o teste *Shapiro-Will*, que aplicados em bases de dados entre 10 e 2000 instâncias. Considerando o  $\alpha = 0.05$  é demonstrado que o teste de *Shapiro-Will* possui uma melhor eficácia na identificação de não normalidade dos dados.

A escolha do teste de *Friedman* justifica-se por esse teste estatístico realizar um ranqueamento dos dados, para evitar a suposição da normalidade. Esse tipo de análise pode ser empregada no lugar da análise de variância quando há mais de um critério de classificação. A hipótese nula desse teste afirma que o efeito da variável-resposta é nulo entre as populações testadas (FRIEDMAN, 1937).

Antes de prosseguir com a análise estatística, conduziu-se uma análise na distribuição dos dados coletados a fim de validar se as amostras coletadas não são uma distribuição normal para tal, o teste de *Shapiro-Will* foi aplicado. As Tabelas 7 - 10 sumarizam os resultados após efetuar o teste de *Shapiro-Will* para cada um dos classificadores utilizados neste trabalho utilizando a média dos 10 *folds* para cada base, ou seja, cada cédula está computado o valor de 31 resultados.

Na Tabela 7 está sumarizado os resultados do classificador Naïve Bayes e observa-se que para cada técnica estudada os resultados se apresentaram como uma distribuição não normal em 18/35 casos. Por sua vez, as Tabelas 8 - 10 apresentam os resultados das distribuições de resultados para os classificadores *rpartXse*, *RIPPER* e *k-NN*, a partir de uma análise destas tabelas, conclui-se que para as técnicas estudadas o teste não refuta a hipótese de distribuição normal.

As Tabelas 11 - 14 apresentam os resultados da aplicação do teste de *Friedman* tais tabelas foram geradas comparando uma das variações do parâmetro com o valor original do trabalho de Vale *et al.* (2018). Como observado nas tabelas e gráficos da Seção 4.1 os resultados dos diferentes valores para o parâmetro *cr* estão bastante próximos e para todos os casos avaliados pelas Tabelas 11 - 14 não foi possível refutar a hipótese que o efeito do parâmetro *cr* influencia diretamente na acurácia da classificação de dados semissupervisionados com as técnicas estudadas.

Tabela 7 – Resultados do teste de *Shapiro-Will* do classificador Naïve Bayes.

Téc.	Exemplos Inicialmente Rotulados	Variação no parâmetro <i>cr</i>						
		2%	3%	4%	5%	6%	7%	8%
FlexCon-C1(s)	5%	0.0513	0.0766	0.0530	0.0820	<b>0.0358</b>	<b>0.0485</b>	<b>0.0491</b>
	10%	<b>0.0301</b>	<b>0.0279</b>	<b>0.0137</b>	<b>0.0215</b>	<b>0.0421</b>	<b>0.0097</b>	<b>0.0117</b>
	15%	0.0535	0.0552	<b>0.0456</b>	<b>0.0373</b>	0.1344	<b>0.0356</b>	<b>0.0477</b>
	20%	0.0949	0.0845	<b>0.0305</b>	0.0662	0.1364	<b>0.0346</b>	0.0688
	25%	0.0798	<b>0.0407</b>	0.0630	<b>0.0097</b>	0.0673	0.0754	0.0957
FlexCon-C1(v)	5%	0.0513	0.0766	0.0530	0.0820	<b>0.0358</b>	<b>0.0485</b>	<b>0.0491</b>
	10%	<b>0.0301</b>	<b>0.0279</b>	<b>0.0137</b>	<b>0.0215</b>	<b>0.0421</b>	<b>0.0097</b>	<b>0.0117</b>
	15%	0.0535	0.0552	<b>0.0456</b>	<b>0.0373</b>	0.1344	<b>0.0356</b>	<b>0.0477</b>
	20%	0.0949	0.0845	<b>0.0305</b>	0.0662	0.1364	<b>0.0346</b>	0.0688
	25%	0.0798	<b>0.0407</b>	0.0630	<b>0.0097</b>	0.0673	0.0754	0.0957
FlexCon-C2	5%	0.0513	0.0766	0.0530	0.0820	<b>0.0358</b>	<b>0.0485</b>	<b>0.0491</b>
	10%	<b>0.0301</b>	<b>0.0279</b>	<b>0.0137</b>	<b>0.0215</b>	<b>0.0421</b>	<b>0.0097</b>	<b>0.0117</b>
	15%	0.0535	0.0552	<b>0.0456</b>	<b>0.0373</b>	0.1344	<b>0.0356</b>	<b>0.0477</b>
	20%	0.0949	0.0845	<b>0.0305</b>	0.0662	0.1364	<b>0.0346</b>	0.0688
	25%	0.0798	<b>0.0407</b>	0.0630	<b>0.0097</b>	0.0673	0.0754	0.0957

Fonte: O Próprio Autor (2018).

Tabela 8 – Resultados do teste de *Shapiro-Will* do classificador *rpartXse*.

Téc.	Exemplos Inicialmente Rotulados	Variação no parâmetro <i>cr</i>						
		2%	3%	4%	5%	6%	7%	8%
FlexCon-C1(s)	5%	0.3899	0.2046	0.2310	0.2671	0.2127	0.1857	0.1894
	10%	0.3330	0.1499	0.1935	0.2525	0.3100	0.2500	0.3014
	15%	0.2095	0.3001	0.2431	0.2800	0.2639	0.3027	0.1965
	20%	0.3613	0.2549	0.3392	0.2072	0.4164	0.4248	0.1798
	25%	0.1746	0.0833	0.1914	0.1403	0.1562	0.1887	0.1996
FlexCon-C1(v)	5%	0.3943	0.2552	0.2306	0.3036	0.1910	0.2058	0.2045
	10%	0.3818	0.1655	0.2327	0.2452	0.2983	0.2215	0.2835
	15%	0.2205	0.3164	0.2470	0.2721	0.4090	0.3672	0.2489
	20%	0.3400	0.2882	0.3896	0.2418	0.4642	0.3516	0.1579
	25%	0.1662	0.1489	0.1656	0.2905	0.1931	0.2053	0.1096
FlexCon-C2	5%	0.3481	0.2188	0.2174	0.2942	0.1572	0.1459	0.2520
	10%	0.3775	0.2288	0.1495	0.2686	0.3287	0.1915	0.2650
	15%	0.2561	0.2673	0.2432	0.3017	0.3572	0.2906	0.2275
	20%	0.4043	0.2823	0.4352	0.2191	0.3822	0.4256	0.1648
	25%	0.2124	0.1294	0.1603	0.2001	0.1269	0.2269	0.2184

Fonte: O Próprio Autor (2018).

Tabela 9 – Resultados do teste de *Shapiro-Will* do classificador **RIPPER**.

Téc.	Exemplos Inicialmente Rotulados	Variação no parâmetro <i>cr</i>						
		2%	3%	4%	5%	6%	7%	8%
FlexCon-C1(s)	5%	0.7214	0.4082	0.7631	0.8438	0.4290	0.7012	0.7426
	10%	0.4461	0.5548	0.7312	0.5410	0.3928	0.6463	0.7572
	15%	0.6672	0.7941	0.5565	0.4929	0.5545	0.6784	0.5976
	20%	0.7713	0.7645	0.5644	0.7167	0.8765	0.5716	0.5468
	25%	0.3051	0.6066	0.5976	0.7546	0.7917	0.7977	0.6659
FlexCon-C1(v)	5%	0.7214	0.4082	0.7631	0.8438	0.4290	0.7012	0.7426
	10%	0.4461	0.5548	0.7312	0.5410	0.3928	0.6463	0.7572
	15%	0.6672	0.7941	0.5565	0.4929	0.5545	0.6784	0.5976
	20%	0.7713	0.7645	0.5644	0.7167	0.8765	0.5716	0.5468
	25%	0.3051	0.6066	0.5976	0.7546	0.7917	0.7977	0.6659
FlexCon-C2	5%	0.7214	0.4082	0.7631	0.8438	0.4290	0.7012	0.7426
	10%	0.4461	0.5548	0.7312	0.5410	0.3928	0.6463	0.7572
	15%	0.6672	0.7941	0.5565	0.4929	0.5545	0.6784	0.5976
	20%	0.7713	0.7645	0.5644	0.7167	0.8765	0.5716	0.5468
	25%	0.3051	0.6066	0.5976	0.7546	0.7917	0.7977	0.6659

Fonte: O Próprio Autor (2018).

Tabela 10 – Resultados do teste de *Shapiro-Will* do classificador **k-NN**.

Téc.	Exemplos Inicialmente Rotulados	Variação no parâmetro <i>cr</i>						
		2%	3%	4%	5%	6%	7%	8%
FlexCon-C1(s)	5%	0.3912	0.1775	0.1636	0.2378	0.4142	0.2815	0.1538
	10%	0.2332	0.3160	0.2760	0.4056	0.3372	0.2399	0.2891
	15%	0.2004	0.2877	0.4063	0.2799	0.4996	0.3473	0.3859
	20%	0.2771	0.3651	0.3540	0.2641	0.3899	0.2535	0.3281
	25%	0.2284	0.4916	0.4427	0.3321	0.4892	0.3545	0.4158
FlexCon-C1(v)	5%	0.3912	0.1775	0.1636	0.2378	0.4142	0.2815	0.1538
	10%	0.2332	0.3160	0.2760	0.4056	0.3372	0.2399	0.2891
	15%	0.2004	0.2877	0.4063	0.2799	0.4996	0.3473	0.3859
	20%	0.2771	0.3651	0.3540	0.2641	0.3899	0.2535	0.3281
	25%	0.2284	0.4916	0.4427	0.3321	0.4892	0.3545	0.4158
FlexCon-C2	5%	0.4007	0.1970	0.1800	0.2702	0.3514	0.2503	0.1832
	10%	0.2618	0.3372	0.2685	0.3702	0.3040	0.2536	0.2791
	15%	0.2069	0.2841	0.4097	0.2393	0.4385	0.3661	0.4051
	20%	0.3110	0.3892	0.3187	0.2966	0.3761	0.2037	0.3009
	25%	0.2378	0.4853	0.4375	0.2789	0.4310	0.3357	0.3988

Fonte: O Próprio Autor (2018).

Tabela 11 – Resultados do teste de *Friedman* do classificador Naïve Bayes.

Téc.	Exemplos Inicialmente Rotulados	Variação no parâmetro <i>cr</i>					
		2%	3%	4%	6%	7%	8%
FlexCon-C1(s)	5%	0.9136	0.9992	0.9543	0.9952	0.5527	0.9996
	10%	0.9972	0.6715	0.2772	0.1210	0.9136	0.4928
	15%	1.0000	1.0000	0.9709	1.0000	0.9984	0.9857
	20%	0.1820	0.9232	0.9857	0.4343	0.7630	0.7630
	25%	0.8119	0.8415	0.9999	0.9979	0.9034	0.9604
FlexCon-C1(v)	5%	0.9136	0.9992	0.9543	0.9952	0.5527	0.9996
	10%	0.9972	0.6715	0.2772	0.1210	0.9136	0.4928
	15%	1.0000	1.0000	0.9709	1.0000	0.9984	0.9857
	20%	0.1820	0.9232	0.9857	0.4343	0.7630	0.7630
	25%	0.8119	0.8415	0.9999	0.9979	0.9034	0.9604
FlexCon-C2	5%	0.9136	0.9992	0.9543	0.9952	0.5527	0.9996
	10%	0.9972	0.6715	0.2772	0.1210	0.9136	0.4928
	15%	1.0000	1.0000	0.9709	1.0000	0.9984	0.9857
	20%	0.1820	0.9232	0.9857	0.4343	0.7630	0.7630
	25%	0.8119	0.8415	0.9999	0.9979	0.9034	0.9604

Fonte: O Próprio Autor (2018).

Tabela 12 – Resultados do teste de *Friedman* do classificador *rpartXse*.

Téc.	Exemplos Inicialmente Rotulados	Variação no parâmetro <i>cr</i>					
		2%	3%	4%	6%	7%	8%
FlexCon-C1(s)	5%	1.0000	0.9034	0.9401	0.9709	0.8415	0.6128
	10%	0.9401	0.9963	0.7962	0.8415	0.9979	0.9992
	15%	0.8925	0.9923	0.9904	0.9979	0.9979	0.9998
	20%	1.0000	0.9999	0.9034	1.0000	0.9857	0.9543
	25%	0.9923	0.9999	0.8925	0.9979	1.0000	0.3090
FlexCon-C1(v)	5%	0.9995	0.9475	1.0000	0.7276	0.8119	0.2473
	10%	0.8119	1.0000	1.0000	0.9792	0.9996	1.0000
	15%	0.9992	1.0000	1.0000	0.9984	1.0000	0.9753
	20%	0.9999	1.0000	1.0000	0.9963	0.9998	0.7962
	25%	0.8553	1.0000	0.9475	0.9709	1.0000	0.9939
FlexCon-C2	5%	0.9923	0.9963	0.9034	0.9792	0.7276	0.3428
	10%	0.9753	0.9999	1.0000	0.9792	0.9996	1.0000
	15%	0.9995	1.0000	0.9939	0.9857	1.0000	0.9984
	20%	0.9401	0.9999	1.0000	1.0000	1.0000	0.7093
	25%	1.0000	0.9963	0.9232	0.9996	1.0000	0.9753

Fonte: O Próprio Autor (2018).

Tabela 13 – Resultados do teste de *Friedman* do classificador *RIPPER*.

Téc.	Exemplos Inicialmente Rotulados	Variação no parâmetro <i>cr</i>					
		2%	3%	4%	6%	7%	8%
FlexCon-C1(s)	5%	0.9992	0.9984	0.9999	1.0000	1.0000	0.9963
	10%	0.9320	0.9660	0.9992	1.0000	0.9660	0.9904
	15%	1.0000	0.9232	0.9989	0.2065	0.9999	0.8684
	20%	0.9923	1.0000	0.8925	0.9034	0.9989	0.8925
	25%	0.9034	0.5527	0.9923	0.8415	0.3603	0.8270
FlexCon-C1(v)	5%	0.9992	0.9984	0.9999	1.0000	1.0000	0.9963
	10%	0.9320	0.9660	0.9992	1.0000	0.9660	0.9904
	15%	1.0000	0.9232	0.9989	0.2065	0.9999	0.8684
	20%	0.9923	1.0000	0.8925	0.9034	0.9989	0.8925
	25%	0.9034	0.5527	0.9923	0.8415	0.3603	0.8270
FlexCon-C2	5%	0.9992	0.9984	0.9999	1.0000	1.0000	0.9963
	10%	0.9320	0.9660	0.9992	1.0000	0.9660	0.9904
	15%	1.0000	0.9232	0.9989	0.2065	0.9999	0.8684
	20%	0.9923	1.0000	0.8925	0.9034	0.9989	0.8925
	25%	0.9034	0.5527	0.9923	0.8415	0.3603	0.8270

Fonte: O Próprio Autor (2018).

Tabela 14 – Resultados do teste de *Friedman* do classificador *k-NN*.

Téc.	Exemplos Inicialmente Rotulados	Variação no parâmetro <i>cr</i>					
		2%	3%	4%	6%	7%	8%
FlexCon-C1(s)	5%	0.9952	1.0000	0.9827	1.0000	1.0000	1.0000
	10%	1.0000	0.8925	0.9998	0.6906	0.9475	0.6522
	15%	0.9979	0.9543	0.9660	0.6325	1.0000	1.0000
	20%	0.9979	1.0000	1.0000	0.9827	1.0000	1.0000
	25%	1.0000	0.9792	0.9827	0.9996	0.9904	0.9939
FlexCon-C1(v)	5%	0.9952	1.0000	0.9827	1.0000	1.0000	1.0000
	10%	1.0000	0.8925	0.9998	0.6906	0.9475	0.6522
	15%	0.9979	0.9543	0.9660	0.6325	1.0000	1.0000
	20%	0.9979	1.0000	1.0000	0.9827	1.0000	1.0000
	25%	1.0000	0.9792	0.9827	0.9996	0.9904	0.9939
FlexCon-C2	5%	0.9939	1.0000	0.9972	0.9963	0.9996	1.0000
	10%	0.9999	0.8415	0.9992	0.8553	0.9989	0.3603
	15%	0.9857	0.9660	0.9972	0.5928	0.9984	0.9979
	20%	0.9984	0.9923	1.0000	0.9232	0.9979	1.0000
	25%	0.9998	0.9999	0.9999	1.0000	0.9543	0.9923

Fonte: O Próprio Autor (2018).

# 5 Conclusão

## 5.1 Discussão

Observa-se que é bastante difícil definir um valor para o parâmetro  $cr$  que seja único, pois não há garantias que apenas um determinado valor apresente os melhores resultados. Diferentes configurações do experimento demonstraram que o melhor valor obtido depende do conjunto: técnica - classificador - percentual inicialmente rotulado.

Na Seção 4.1 foram apresentados resultados da variação do parâmetro  $cr$ , percentuais de exemplos inicialmente rotulados e classificadores e observou-se que o método [FlexCon-C](#) está sujeito a diversas variações de resultados a partir dos valores selecionados para cada um dos parâmetros. Entretanto, quando se pré-fixa os demais parâmetros observa-se que nem sempre o valor utilizado por [Vale et al. \(2018\)](#),  $cr$  pré-fixado em 5%, obtém os melhores resultados para cada uma das técnicas estudadas. Neste trabalho, foi analisada a influência do parâmetro  $cr$  na acurácia da classificação de dados semissupervisionados.

Ao selecionar a técnica *FlexCon-C1(s)*, é recomendado utilizar os valores de  $cr$  pré-fixos em 4%, 6% e 7%, pois, cada um destes valores obtém os melhores resultados em 4 dos 20 casos. Para a técnica *FlexCon-C1(v)* os melhores valores de  $cr$  são 6% e 7% com 5 e 4 vitórias, respectivamente. Por fim, para a técnica *FlexCon-C2* os melhores valores estão em ambos os extremos 2% e 8%, onde cada um possui os melhores resultados em 4 dos 20 casos.

Analisando a acurácia por classificador, percebe-se que para os classificadores Naïve Bayes e *rpartXse* houve uma distribuição igualitária entre os melhores resultados, considerando todas as técnicas. Porém, para o classificador [RIPPER](#), as maiores acurácias foram obtidas fixando o valor do parâmetro  $cr > 5%$  e, para o classificador *k-NN*, as maiores acurácias foram alcançadas por valores do parâmetro  $cr < 5%$ .

De modo geral, ao utilizar os valores de  $cr$  pré-fixos em 4%, 6% e 7% obtém-se bons resultados em 10 e 11 dos 60 casos, respectivamente, enquanto o  $cr$  pré-fixado em 5% obtém as melhores acurácias em 8 dos 60 casos. Informações detalhadas desta análise estão no Apêndice [M](#).

## 5.2 Trabalhos Futuros

Em trabalhos futuros, sugere-se a adição de outros classificadores, pode-se comparar os novos resultados obtidos com os já anteriores para avaliar se o comportamento apresentado se manteve. Além disto, pode avaliar melhor os domínios dos dados, em virtude das bases de dados

conterem os mais variados tipos de dados (categóricos, inteiros e/ou reais).

Ampliando a escala do parâmetro  $cr$  para uma escala contínua pode-se analisar melhor as faixas exploradas previamente definidas, podendo encontrar um valor para o parâmetro  $cr$  que seja, de fato, o melhor. Assim, realiza-se uma exploração mais detalhada na faixa de valores considerada.

Neste trabalho, a variação aplicada ao limiar, através do uso do parâmetro  $cr$  foi de maneira fixa  $x\%$  por iteração. Porém, essa variação fixa pode não resultar na melhor acurácia. Então, como trabalhos futuros, pode-se desenvolver uma estratégia para alterar o  $cr$  dinamicamente por iteração, tornando o método **FlexCon-C** autoajustável.

# Referências

- AHA, D. W.; KIBLER, D.; ALBERT, M. K. Instance-based learning algorithms. *Machine Learning*, v. 6, n. 1, p. 37–66, Jan 1991. Citado na página [23](#).
- ALPAYDIN, E. *Introduction to Machine Learning*. Cambridge, Massachusetts, London: The MIT Press, 2004. Citado 4 vezes nas páginas [12](#) e [21](#).
- ATKESON, C. G.; MOORE, A. W.; SCHAAU, S. Locally weighted learning. *Artificial Intelligence Review*, Springer Science+Business Media, B.V., Van Godewijckstraat, Netherlands, v. 11, n. 1, p. 11–73, Feb 1997. Citado na página [23](#).
- AU, S.-K. *Operational Modal Analysis: Modeling, Bayesian Inference, Uncertainty Laws*. Gateway East, Singapore: Springer Nature Singapore Pte Ltd., 2017. Citado na página [21](#).
- BHATTACHARYA, G.; GHOSH, K.; CHOWDHURY, A. S. An affinity-based new local distance function and similarity measure for knn algorithm. *Pattern Recognition Letters*, v. 33, n. 3, p. 356 – 363, 2012. ISSN 0167-8655. Citado na página [29](#).
- BOUCHACHIA, A. Learning with partional supervision. In: *Machine Learning: Concepts, methodologies, tools and applications*. Hershey PA 17033: Information Science Reference, 2012. v. 3, p. 1880–1888. Citado na página [18](#).
- BRAGA-NETO, U. M.; DOUGHERTY, E. R. Is cross-validation valid for small-sample microarray classification? *Bioinformatics*, Oxford University Press, v. 20, n. 3, p. 374–380, 2004. Citado na página [28](#).
- BREIMAN, L.; FRIEDMAN, J.; OLSHEN, R. A.; STONE, C. J. *Classification and Regression Trees*. New York: CHAPMAN & HALL/CRC, 1984. Citado na página [22](#).
- CHAPELLE, O.; SCHÖLKOPF, B.; ZIEN, A. *Semi-Supervised Learning*. Cambridge, Massachusetts, London: The MIT Press, 2006. Citado 5 vezes nas páginas [13](#) e [19](#).
- COHEN, W. W. Fast effective rule induction. In: *Machine Learning: Proceedings of the twelfth international conference on machine learning*. Tahoe City, California, USA: Morgan Kaufmann Publishers, 1995. (ML95), p. 115–123. Citado na página [22](#).
- DIMITOGLOU, G.; ADAMS, J. A.; JIM, C. M. Comparison of the c4.5 and a naïve bayes classifier for the prediction of lung cancer survivability. *Journal of Computing*, v. 4, n. 8, 2012. Citado na página [22](#).
- FRIEDMAN, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, v. 32, n. 200, p. 675–701, 1937. Citado na página [40](#).
- FUSHIKI, T. Estimation of prediction error by using k-fold cross-validation. *Statistics and Computing*, v. 21, n. 2, p. 137–146, 2011. Citado na página [28](#).
- GOLLAPUDI, S. *Practical Machine Learning*. Livery Place: Packt Publishing Ltd., 2016. ISBN 9781784399689. Citado 5 vezes nas páginas [16](#), [17](#), [21](#) e [22](#).



- GRANDVALET, Y.; BENGIO, Y. Semi-supervised learning by entropy minimization. In: *Proceedings of the 17th International Conference on Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2004. (NIPS'04), p. 529–536. Citado na página 19.
- HADY, M. F. A.; SCHWENKER, F. Semi-supervised learning. In: *Handbook on Neural Information Processing*. Berlin Heidelberg: Springer, 2013. p. 215–239. ISBN 978-3-642-36657-4. Citado na página 18.
- HOFFMAN, J. I. E. *Biostatistics for medical and biomedical practitioners: an interpretative guide for medicine and biology*. Tiburon, California, USA: Elsevier Ltd, Academic Press, 2015. 153–158 p. ISBN 978-0-12-802387-7. Citado na página 40.
- KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence*, v. 14, n. 2, p. 1137–1145, 1995. Citado na página 28.
- MITCHELL, T. M. *Machine Learning*. Boston, Massachusetts, London: McGraw-Hill, 1997. ISBN 0-07-042807-7. Citado 5 vezes nas páginas 12, 16, 17 e 22.
- MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. *Foundations of Machine Learning*. Cambridge, Massachusetts, London: MIT Press, 2012. ISBN 9780262018258. Citado 5 vezes nas páginas 12, 16 e 17.
- MULAK, P.; TALHAR, N. Analysis of distance measures using k-nearest neighbor algorithm on kdd dataset. *International Journal of Science and Research*, v. 4, n. 7, p. 2101–2104, 2015. Citado na página 23.
- QUINLAN, J. R. Induction of decision trees. *Machine Learning*, Kluwer Academic Publishers, v. 1, n. 1, p. 81 – 106, Mar 1986. Citado na página 22.
- RAJPUT, A.; AHARWAL, R. P.; DUBEY, M.; SAXENA, S.; RAGHUVANSHI, M. J48 and jrip rules for e-governance data. *International Journal of Computer Science and Security (IJCSS)*, v. 5, n. 2, p. 201 – 207, 2011. Citado na página 22.
- RASCHKA, S. *Python Machine Learning*. first. Birmingham, UK: Packt Publishing, 2015. Citado na página 12.
- RAZALI, N. M.; WAH, Y. B. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics*, v. 2, n. 01, p. 21 – 33, 2011. Citado na página 40.
- RODRIGUES, F. M.; CANUTO, A. M. P.; SANTOS, A. M. Confidence factor and feature selection for semi-supervised multi-label classification methods. *International Joint Conference on Neural Networks*, p. 864 – 871, 2014. Citado 5 vezes nas páginas 23 e 24.
- RODRIGUES, F. M.; SANTOS, A. M.; CANUTO, A. M. P. Using confidence values in multi-label classification problems with semi-supervised learning. *International Joint Conference on Neural Networks*, p. 1 – 8, 2013. Citado 5 vezes nas páginas 13 e 23.
- RUSSELL, S. J.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 3. ed. Englewood Cliffs, New Jersey: Prentice Hall, 2009. ISBN 978-0136042594. Citado 3 vezes nas páginas 16 e 23.

- TANHA, J.; SOMEREN, M. van; AFSARMANESH, H. Semi-supervised self-training for decision tree classifiers. *International Journal of Machine Learning and Cybernetics*, Springer, v. 8, n. 1, p. 355–370, 2017. Citado na página 23.
- TAO, Y.; ZHANG, D.; CHENG, S.; TANG, X. Improving semi-supervised self-training with embedded manifold transduction. *Transactions of the Institute of Measurement and Control*, v. 40, n. 2, p. 363–374, 2018. Citado na página 23.
- THERNEAU, T. M.; E.J., A. *An introduction to recursive partitioning using the RPART routines*. Section of Statistics, Mayo Clinic, 2000. Disponível em: <<http://www.cs.uu.nl/docs/vakken/adm/rpartmini.pdf>>. Citado na página 22.
- TORGO, L. Package 'dmwr'. *Comprehensive R Archive Network*, 2013. Citado na página 22.
- VALE, K. M. O.; CANUTO, A. M. de P.; SANTOS, A. de M.; GORGÔNIO, F. da Luz e; TAVARES, A. de M.; GORGÔNIO, A. C.; ALVES, C. T. Automatic adjustment of confidence values in self-training semi-supervised method. *International Joint Conference on Neural Networks*, 2018. Citado 18 vezes nas páginas 13, 14, 19, 20, 23, 24, 25, 26, 28, 29, 30, 31, 33, 40 e 45.
- WILEY, J. F. *R Deep Learning Essentials*. Livery Place: Packt Publishing Ltd., 2016. ISBN 978-1-78528-058-0. Citado na página 17.
- WONG, T.-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, v. 48, n. 9, p. 2839 – 2846, 2015. Citado 5 vezes nas páginas 26 e 28.
- ZELLNER, A. *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley-Interscience, 1996. ISBN 0-471-98165-6. Citado na página 21.
- ZHU, X. *Semi-Supervised Learning Literature Survey*. Computer Sciences, University of Wisconsin-Madison, 2008. Disponível em: <[pages.cs.wisc.edu/~jerryzhu/pub/ssl\\_survey.pdf](http://pages.cs.wisc.edu/~jerryzhu/pub/ssl_survey.pdf)>. Citado na página 19.
- ZHU, X.; GOLDBERG, A. B. *Introduction to semi-supervised learning*. 3. ed. San Rafael, California: Morgan & Claypool Publishers, 2009. ISBN 9781598295481. Citado na página 19.

# APÊNDICE A – Resultado da Técnica *FlexCon-C1(s)* e Classificador Naïve Bayes

Tabela 15 – Resultado da aplicação da Técnica *FlexCon-C1(s)* utilizando o classificador Naïve Bayes.

Base de dados	Exemplos Ini. Rot.	Variação do <i>cr</i>						
		2	3	4	5	6	7	8
Balance Scale	5%	80,33	80,78	83,18	79,69	82,08	79,37	80,31
	10%	81,59	82,54	84,13	82,38	80,48	83,97	84,44
	15%	79,67	80,32	80,02	84,65	80,31	80,14	83,67
	20%	77,74	81,61	78,71	81,94	76,94	80,97	76,29
	25%	83,85	84,65	83,53	83,37	81,77	83,86	83,05
Blood Transfusion Service	5%	75,87	74,80	74,67	72,93	70,80	74,00	72,53
	10%	75,87	75,60	76,93	74,67	74,93	75,07	74,53
	15%	73,96	73,02	74,63	72,35	72,47	72,19	74,36
	20%	71,13	71,66	73,42	73,03	72,34	69,65	71,54
	25%	75,20	75,87	77,73	75,73	76,93	74,13	75,60
Bupa	5%	55,14	54,29	50,29	55,43	50,29	56,86	56,29
	10%	55,02	49,22	55,34	55,60	50,09	55,32	53,87
	15%	55,88	54,41	54,12	54,71	55,29	54,41	53,24
	20%	60,57	57,14	56,57	55,43	58,00	57,14	57,43
	25%	54,41	50,59	57,94	50,29	54,12	54,71	51,47
Car	5%	79,82	79,82	79,82	79,77	80,40	80,69	79,30
	10%	81,28	80,71	80,93	81,00	81,11	80,71	81,00
	15%	81,69	81,16	80,17	82,38	81,10	80,99	80,99
	20%	80,94	81,47	81,69	82,27	81,23	80,88	81,87
	25%	81,09	80,34	80,40	83,05	82,13	81,38	81,49
Cnae-9	5%	25,00	29,26	30,28	25,74	23,24	30,00	26,76
	10%	30,65	29,07	28,24	26,11	29,63	25,83	27,22
	15%	28,89	29,54	28,61	25,65	30,37	24,44	28,33
	20%	29,54	26,76	24,63	29,07	30,00	22,87	27,50
	25%	29,63	30,93	30,00	24,63	26,48	29,07	30,65
Connectionist Mines vs Rocks	5%	56,17	59,50	55,67	57,57	61,52	60,52	63,93
	10%	64,76	61,90	67,62	67,14	65,24	65,71	64,29
	15%	67,14	69,05	60,48	62,86	62,38	65,24	61,90
	20%	53,64	61,50	63,98	66,02	60,57	64,83	62,21

	25%	72,64	72,82	64,73	71,45	74,77	70,68	75,14
Flare	5%	40,43	42,43	42,29	45,57	45,86	41,29	40,29
	10%	46,78	47,50	49,96	41,79	46,14	46,00	51,91
	15%	41,58	43,46	51,69	47,57	54,73	43,23	51,97
	20%	43,38	46,33	39,78	44,03	47,48	45,76	46,19
	25%	48,29	46,92	50,90	44,96	51,23	50,68	50,40
Haberman	5%	70,52	73,41	72,74	69,88	71,80	72,42	72,69
	10%	71,14	69,25	76,45	70,49	69,15	78,10	74,11
	15%	65,92	70,81	67,17	67,82	67,16	67,17	67,81
	20%	72,00	74,67	70,00	72,33	72,00	71,00	69,33
	25%	63,52	69,57	66,76	67,42	68,57	70,22	65,10
Handwritten Digits	5%	83,74	84,02	84,27	83,83	84,15	83,83	84,06
	10%	84,01	84,37	84,18	84,64	84,28	84,95	84,41
	15%	85,66	85,99	85,72	86,00	85,82	86,11	85,42
	20%	85,17	85,30	85,81	85,65	85,42	85,47	85,18
	25%	85,73	85,80	85,51	85,73	85,71	85,58	85,45
Hill Valley	5%	49,14	52,83	46,76	53,72	49,22	50,36	49,22
	10%	51,07	51,63	51,06	49,43	50,83	48,92	48,84
	15%	51,03	49,79	50,45	50,12	50,12	50,12	50,37
	20%	48,43	47,36	49,42	50,25	51,16	48,60	49,59
	25%	50,33	50,66	50,65	50,24	51,74	50,50	50,34
Image Segmentation	5%	76,23	76,02	73,38	76,10	77,32	76,71	76,06
	10%	78,92	79,91	79,57	80,09	80,13	78,35	80,91
	15%	75,97	76,67	78,66	79,39	76,54	76,75	78,40
	20%	76,84	77,58	76,71	77,66	77,58	76,49	76,97
	25%	79,00	79,78	78,74	79,31	77,79	79,78	76,93
Indian Liver Patient	5%	54,51	55,37	55,03	55,38	55,39	53,18	53,14
	10%	52,41	53,10	53,79	53,97	53,28	52,93	51,55
	15%	56,55	57,07	57,41	55,00	57,07	57,59	55,86
	20%	62,93	63,09	63,76	65,47	66,32	65,99	67,86
	25%	62,04	57,46	58,99	63,27	63,09	59,36	60,89
Iris	5%	89,33	92,67	92,67	94,67	93,33	89,33	90,67
	10%	94,00	86,67	89,33	90,67	89,33	90,67	86,00
	15%	88,67	89,33	88,00	89,33	95,33	87,33	93,33
	20%	90,67	94,00	88,00	92,00	92,00	95,33	90,00
	25%	92,67	88,67	91,33	88,00	96,00	89,33	90,00
King Ropck vs King Pawn	5%	83,07	83,39	84,70	85,80	82,45	83,95	84,61
	10%	83,38	82,88	82,88	84,70	83,29	84,63	84,70
	15%	86,97	88,44	86,38	86,94	86,94	87,56	86,75
	20%	85,48	85,01	85,07	85,04	84,32	86,60	84,07

	25%	86,53	85,12	83,84	84,47	84,09	85,88	84,03
Leukemia Haslinger	5%	68,00	72,00	75,00	73,00	71,00	76,00	73,00
	10%	82,00	80,00	88,00	82,00	80,00	84,00	82,00
	15%	79,00	86,00	83,00	82,00	82,00	81,00	82,00
	20%	66,00	68,00	71,00	66,00	70,00	75,00	69,00
	25%	79,00	80,00	81,00	81,00	77,00	75,00	72,00
Mammographic Mass	5%	80,31	81,03	80,00	80,62	80,00	80,21	78,45
	10%	81,46	78,63	79,16	79,99	78,95	79,58	79,36
	15%	82,10	80,75	83,14	82,31	81,99	82,40	81,99
	20%	81,48	80,53	80,85	80,64	80,65	78,01	79,60
	25%	77,62	78,34	78,75	77,30	78,44	77,72	78,54
Mfeat Karhunen	5%	90,45	89,70	90,35	89,85	90,20	90,95	90,20
	10%	88,45	89,15	89,95	89,05	89,10	89,35	88,90
	15%	88,60	89,45	87,75	88,70	87,80	87,30	87,25
	20%	88,15	88,20	87,65	88,30	87,55	89,70	88,70
	25%	88,50	88,60	89,50	88,70	89,35	89,95	90,05
Mushroom	5%	94,97	94,72	94,77	95,29	94,79	95,18	94,97
	10%	94,86	94,53	95,02	94,81	94,34	94,59	94,70
	15%	94,66	94,95	95,03	94,74	95,37	94,93	95,11
	20%	94,32	94,15	94,20	94,00	93,79	94,04	93,92
	25%	93,05	93,39	93,51	93,10	93,04	92,91	93,35
Musk	5%	83,70	82,49	82,56	83,93	82,61	83,55	82,76
	10%	84,44	85,02	84,91	85,71	84,45	85,05	84,47
	15%	85,19	83,90	84,37	86,22	84,91	85,19	83,81
	20%	83,62	85,55	84,92	84,00	84,00	85,00	85,24
	25%	83,74	82,74	83,33	81,32	82,45	83,20	83,24
Ozone Level Detection	5%	78,78	79,49	79,13	80,75	78,31	80,24	78,50
	10%	77,52	78,50	79,61	77,56	77,48	81,38	81,73
	15%	82,24	80,20	80,11	79,13	79,88	79,41	80,16
	20%	79,49	75,67	75,74	76,77	76,69	76,76	78,78
	25%	77,71	79,53	81,90	80,55	78,89	79,29	79,45
Phishing	5%	82,67	82,52	82,07	83,78	82,22	82,22	82,15
	10%	84,78	85,15	85,51	85,07	85,66	85,66	85,15
	15%	83,83	83,83	84,72	83,24	84,28	83,54	83,98
	20%	83,48	81,41	82,00	83,04	84,44	82,30	81,63
	25%	83,85	81,78	83,56	82,15	82,81	83,78	84,74
Pima	5%	69,40	70,85	69,78	70,70	71,75	71,09	69,54
	10%	70,65	72,47	72,34	71,30	71,04	71,17	70,26
	15%	72,21	69,87	72,99	72,99	72,08	71,30	71,82
	20%	82,08	77,27	79,22	80,00	80,26	79,09	80,65

	25%	70,07	70,33	71,37	70,46	71,37	70,46	72,68
Planning Relax	5%	62,78	58,89	59,44	58,89	61,11	63,33	63,33
	10%	63,89	61,67	57,78	63,89	61,11	64,44	62,78
	15%	61,73	63,25	57,37	61,75	55,32	61,05	57,89
	20%	59,91	59,56	56,35	64,36	64,91	62,75	57,51
	25%	60,00	67,22	57,22	63,33	63,89	57,22	62,22
Seeds	5%	86,67	90,00	88,57	89,52	89,05	90,48	91,43
	10%	81,43	82,38	84,29	84,29	76,19	80,95	83,81
	15%	87,62	90,95	89,05	85,24	88,57	87,62	86,67
	20%	91,43	90,48	90,95	93,33	90,95	89,52	92,86
	25%	91,43	91,43	91,43	87,14	92,86	89,05	89,05
Semeion	5%	64,15	64,85	63,41	65,73	65,17	63,34	66,62
	10%	64,02	67,55	67,40	66,59	67,80	64,63	65,50
	15%	60,06	59,62	61,38	63,08	59,25	59,62	62,64
	20%	64,73	64,78	64,30	61,94	66,37	65,09	69,13
	25%	62,72	61,53	65,71	68,53	61,69	65,83	61,16
Spectf Heart	5%	81,69	78,90	79,16	77,75	81,71	79,76	82,29
	10%	77,66	73,64	74,47	75,04	74,17	75,04	74,72
	15%	73,64	72,88	73,64	71,06	70,51	71,99	72,01
	20%	74,00	75,14	78,57	78,00	73,43	75,71	75,43
	25%	73,06	69,56	72,18	70,74	70,72	70,41	67,83
Tic Tac Toe	5%	68,18	69,84	69,43	68,49	69,85	70,26	69,12
	10%	70,58	68,48	70,27	68,28	69,12	69,95	69,43
	15%	66,35	66,25	65,42	65,21	65,00	68,85	66,56
	20%	69,69	68,44	67,81	68,33	67,92	68,54	69,37
	25%	68,32	67,68	67,89	68,21	67,89	66,11	67,47
Twonorm	5%	97,80	97,73	97,76	97,59	97,74	97,77	97,85
	10%	97,31	97,14	97,35	97,32	97,39	97,30	97,35
	15%	97,83	97,83	97,72	97,73	97,81	97,91	97,65
	20%	98,00	97,86	97,83	97,77	97,94	97,90	97,90
	25%	97,74	97,80	97,78	97,77	97,73	97,78	97,82
Vehicle	5%	45,55	46,61	45,30	46,14	45,07	46,50	44,01
	10%	45,48	49,13	49,11	47,25	49,34	45,84	48,67
	15%	43,06	45,06	44,12	44,71	45,29	46,12	46,47
	20%	46,16	48,65	44,97	48,52	50,16	45,80	47,32
	25%	44,07	47,18	46,95	46,72	47,18	44,21	44,07
Waveform	5%	79,28	79,58	80,08	80,18	79,86	80,04	79,28
	10%	79,74	79,68	79,20	79,44	79,36	79,50	79,16
	15%	80,30	80,54	80,32	80,18	80,56	80,66	80,30
	20%	79,44	79,72	79,32	79,90	79,52	79,72	79,44

	25%	80,10	80,30	80,36	80,24	79,88	80,20	80,44
Wilt	5%	90,02	91,18	87,73	87,89	89,94	87,54	90,68
	10%	89,92	90,58	90,02	91,45	86,26	89,88	89,89
	15%	92,12	90,26	90,34	92,20	91,23	89,49	87,13
	20%	89,98	89,26	89,15	89,44	89,75	91,26	90,58
	25%	88,22	89,44	87,13	88,37	87,67	90,58	89,96

**Fonte:** O Próprio Autor (2018)

# APÊNDICE B – Resultado da Técnica *FlexCon-C1(v)* e Classificador Naïve Bayes

Tabela 16 – Resultado da aplicação da Técnica *FlexCon-C1(v)* utilizando o classificador Naïve Bayes.

Base de dados	Exemplos Ini. Rot.	Variação do <i>cr</i>						
		2	3	4	5	6	7	8
Balance Scale	5%	80,33	80,78	83,18	79,69	82,08	79,37	80,31
	10%	81,59	82,54	84,13	82,38	80,48	83,97	84,44
	15%	79,67	80,32	80,02	84,65	80,31	80,14	83,67
	20%	77,74	81,61	78,71	81,94	76,94	80,97	76,29
	25%	83,85	84,65	83,53	83,37	81,77	83,86	83,05
Blood Transfusion Service	5%	75,87	74,80	74,67	72,93	70,80	74,00	72,53
	10%	75,87	75,60	76,93	74,67	74,93	75,07	74,53
	15%	73,96	73,02	74,63	72,35	72,47	72,19	74,36
	20%	71,13	71,66	73,42	73,03	72,34	69,65	71,54
	25%	75,20	75,87	77,73	75,73	76,93	74,13	75,60
Bupa	5%	55,14	54,29	50,29	55,43	50,29	56,86	56,29
	10%	55,02	49,22	55,34	55,60	50,09	55,32	53,87
	15%	55,88	54,41	54,12	54,71	55,29	54,41	53,24
	20%	60,57	57,14	56,57	55,43	58,00	57,14	57,43
	25%	54,41	50,59	57,94	50,29	54,12	54,71	51,47
Car	5%	79,82	79,82	79,82	79,77	80,40	80,69	79,30
	10%	81,28	80,71	80,93	81,00	81,11	80,71	81,00
	15%	81,69	81,16	80,17	82,38	81,10	80,99	80,99
	20%	80,94	81,47	81,69	82,27	81,23	80,88	81,87
	25%	81,09	80,34	80,40	83,05	82,13	81,38	81,49
Cnae-9	5%	25,00	29,26	30,28	25,74	23,24	30,00	26,76
	10%	30,65	29,07	28,24	26,11	29,63	25,83	27,22
	15%	28,89	29,54	28,61	25,65	30,37	24,44	28,33
	20%	29,54	26,76	24,63	29,07	30,00	22,87	27,50
	25%	29,63	30,93	30,00	24,63	26,48	29,07	30,65
Connectionist Mines vs Rocks	5%	56,17	59,50	55,67	57,57	61,52	60,52	63,93
	10%	64,76	61,90	67,62	67,14	65,24	65,71	64,29
	15%	67,14	69,05	60,48	62,86	62,38	65,24	61,90
	20%	53,64	61,50	63,98	66,02	60,57	64,83	62,21



	25%	72,64	72,82	64,73	71,45	74,77	70,68	75,14
Flare	5%	40,43	42,43	42,29	45,57	45,86	41,29	40,29
	10%	46,78	47,50	49,96	41,79	46,14	46,00	51,91
	15%	41,58	43,46	51,69	47,57	54,73	43,23	51,97
	20%	43,38	46,33	39,78	44,03	47,48	45,76	46,19
	25%	48,29	46,92	50,90	44,96	51,23	50,68	50,40
Haberman	5%	70,52	73,41	72,74	69,88	71,80	72,42	72,69
	10%	71,14	69,25	76,45	70,49	69,15	78,10	74,11
	15%	65,92	70,81	67,17	67,82	67,16	67,17	67,81
	20%	72,00	74,67	70,00	72,33	72,00	71,00	69,33
	25%	63,52	69,57	66,76	67,42	68,57	70,22	65,10
Handwritten Digits	5%	83,74	84,02	84,27	83,83	84,15	83,83	84,06
	10%	84,01	84,37	84,18	84,64	84,28	84,95	84,41
	15%	85,66	85,99	85,72	86,00	85,82	86,11	85,42
	20%	85,17	85,30	85,81	85,65	85,42	85,47	85,18
	25%	85,73	85,80	85,51	85,73	85,71	85,58	85,45
Hill Valley	5%	49,14	52,83	46,76	53,72	49,22	50,36	49,22
	10%	51,07	51,63	51,06	49,43	50,83	48,92	48,84
	15%	51,03	49,79	50,45	50,12	50,12	50,12	50,37
	20%	48,43	47,36	49,42	50,25	51,16	48,60	49,59
	25%	50,33	50,66	50,65	50,24	51,74	50,50	50,34
Image Segmentation	5%	76,23	76,02	73,38	76,10	77,32	76,71	76,06
	10%	78,92	79,91	79,57	80,09	80,13	78,35	80,91
	15%	75,97	76,67	78,66	79,39	76,54	76,75	78,40
	20%	76,84	77,58	76,71	77,66	77,58	76,49	76,97
	25%	79,00	79,78	78,74	79,31	77,79	79,78	76,93
Indian Liver Patient	5%	54,51	55,37	55,03	55,38	55,39	53,18	53,14
	10%	52,41	53,10	53,79	53,97	53,28	52,93	51,55
	15%	56,55	57,07	57,41	55,00	57,07	57,59	55,86
	20%	62,93	63,09	63,76	65,47	66,32	65,99	67,86
	25%	62,04	57,46	58,99	63,27	63,09	59,36	60,89
Iris	5%	89,33	92,67	92,67	94,67	93,33	89,33	90,67
	10%	94,00	86,67	89,33	90,67	89,33	90,67	86,00
	15%	88,67	89,33	88,00	89,33	95,33	87,33	93,33
	20%	90,67	94,00	88,00	92,00	92,00	95,33	90,00
	25%	92,67	88,67	91,33	88,00	96,00	89,33	90,00
King Ropck vs King Pawn	5%	83,07	83,39	84,70	85,80	82,45	83,95	84,61
	10%	83,38	82,88	82,88	84,70	83,29	84,63	84,70
	15%	86,97	88,44	86,38	86,94	86,94	87,56	86,75
	20%	85,48	85,01	85,07	85,04	84,32	86,60	84,07

	25%	86,53	85,12	83,84	84,47	84,09	85,88	84,03
Leukemia Haslinger	5%	68,00	72,00	75,00	73,00	71,00	76,00	73,00
	10%	82,00	80,00	88,00	82,00	80,00	84,00	82,00
	15%	79,00	86,00	83,00	82,00	82,00	81,00	82,00
	20%	66,00	68,00	71,00	66,00	70,00	75,00	69,00
	25%	79,00	80,00	81,00	81,00	77,00	75,00	72,00
Mammographic Mass	5%	80,31	81,03	80,00	80,62	80,00	80,21	78,45
	10%	81,46	78,63	79,16	79,99	78,95	79,58	79,36
	15%	82,10	80,75	83,14	82,31	81,99	82,40	81,99
	20%	81,48	80,53	80,85	80,64	80,65	78,01	79,60
	25%	77,62	78,34	78,75	77,30	78,44	77,72	78,54
Mfeat Karhunen	5%	90,45	89,70	90,35	89,85	90,20	90,95	90,20
	10%	88,45	89,15	89,95	89,05	89,10	89,35	88,90
	15%	88,60	89,45	87,75	88,70	87,80	87,30	87,25
	20%	88,15	88,20	87,65	88,30	87,55	89,70	88,70
	25%	88,50	88,60	89,50	88,70	89,35	89,95	90,05
Mushroom	5%	94,97	94,72	94,77	95,29	94,79	95,18	94,97
	10%	94,86	94,53	95,02	94,81	94,34	94,59	94,70
	15%	94,66	94,95	95,03	94,74	95,37	94,93	95,11
	20%	94,32	94,15	94,20	94,00	93,79	94,04	93,92
	25%	93,05	93,39	93,51	93,10	93,04	92,91	93,35
Musk	5%	83,70	82,49	82,56	83,93	82,61	83,55	82,76
	10%	84,44	85,02	84,91	85,71	84,45	85,05	84,47
	15%	85,19	83,90	84,37	86,22	84,91	85,19	83,81
	20%	83,62	85,55	84,92	84,00	84,00	85,00	85,24
	25%	83,74	82,74	83,33	81,32	82,45	83,20	83,24
Ozone Level Detection	5%	78,78	79,49	79,13	80,75	78,31	80,24	78,50
	10%	77,52	78,50	79,61	77,56	77,48	81,38	81,73
	15%	82,24	80,20	80,11	79,13	79,88	79,41	80,16
	20%	79,49	75,67	75,74	76,77	76,69	76,76	78,78
	25%	77,71	79,53	81,90	80,55	78,89	79,29	79,45
Phishing	5%	82,67	82,52	82,07	83,78	82,22	82,22	82,15
	10%	84,78	85,15	85,51	85,07	85,66	85,66	85,15
	15%	83,83	83,83	84,72	83,24	84,28	83,54	83,98
	20%	83,48	81,41	82,00	83,04	84,44	82,30	81,63
	25%	83,85	81,78	83,56	82,15	82,81	83,78	84,74
Pima	5%	69,40	70,85	69,78	70,70	71,75	71,09	69,54
	10%	70,65	72,47	72,34	71,30	71,04	71,17	70,26
	15%	72,21	69,87	72,99	72,99	72,08	71,30	71,82
	20%	82,08	77,27	79,22	80,00	80,26	79,09	80,65

	25%	70,07	70,33	71,37	70,46	71,37	70,46	72,68
Planning Relax	5%	62,78	58,89	59,44	58,89	61,11	63,33	63,33
	10%	63,89	61,67	57,78	63,89	61,11	64,44	62,78
	15%	61,73	63,25	57,37	61,75	55,32	61,05	57,89
	20%	59,91	59,56	56,35	64,36	64,91	62,75	57,51
	25%	60,00	67,22	57,22	63,33	63,89	57,22	62,22
Seeds	5%	86,67	90,00	88,57	89,52	89,05	90,48	91,43
	10%	81,43	82,38	84,29	84,29	76,19	80,95	83,81
	15%	87,62	90,95	89,05	85,24	88,57	87,62	86,67
	20%	91,43	90,48	90,95	93,33	90,95	89,52	92,86
	25%	91,43	91,43	91,43	87,14	92,86	89,05	89,05
Semeion	5%	64,15	64,85	63,41	65,73	65,17	63,34	66,62
	10%	64,02	67,55	67,40	66,59	67,80	64,63	65,50
	15%	60,06	59,62	61,38	63,08	59,25	59,62	62,64
	20%	64,73	64,78	64,30	61,94	66,37	65,09	69,13
	25%	62,72	61,53	65,71	68,53	61,69	65,83	61,16
Spectf Heart	5%	81,69	78,90	79,16	77,75	81,71	79,76	82,29
	10%	77,66	73,64	74,47	75,04	74,17	75,04	74,72
	15%	73,64	72,88	73,64	71,06	70,51	71,99	72,01
	20%	74,00	75,14	78,57	78,00	73,43	75,71	75,43
	25%	73,06	69,56	72,18	70,74	70,72	70,41	67,83
Tic Tac Toe	5%	68,18	69,84	69,43	68,49	69,85	70,26	69,12
	10%	70,58	68,48	70,27	68,28	69,12	69,95	69,43
	15%	66,35	66,25	65,42	65,21	65,00	68,85	66,56
	20%	69,69	68,44	67,81	68,33	67,92	68,54	69,37
	25%	68,32	67,68	67,89	68,21	67,89	66,11	67,47
Twonorm	5%	97,80	97,73	97,76	97,59	97,74	97,77	97,85
	10%	97,31	97,14	97,35	97,32	97,39	97,30	97,35
	15%	97,83	97,83	97,72	97,73	97,81	97,91	97,65
	20%	98,00	97,86	97,83	97,77	97,94	97,90	97,90
	25%	97,74	97,80	97,78	97,77	97,73	97,78	97,82
Vehicle	5%	45,55	46,61	45,30	46,14	45,07	46,50	44,01
	10%	45,48	49,13	49,11	47,25	49,34	45,84	48,67
	15%	43,06	45,06	44,12	44,71	45,29	46,12	46,47
	20%	46,16	48,65	44,97	48,52	50,16	45,80	47,32
	25%	44,07	47,18	46,95	46,72	47,18	44,21	44,07
Waveform	5%	79,28	79,58	80,08	80,18	79,86	80,04	79,28
	10%	79,74	79,68	79,20	79,44	79,36	79,50	79,16
	15%	80,30	80,54	80,32	80,18	80,56	80,66	80,30
	20%	79,44	79,72	79,32	79,90	79,52	79,72	79,44

	25%	80,10	80,30	80,36	80,24	79,88	80,20	80,44
Wilt	5%	90,02	91,18	87,73	87,89	89,94	87,54	90,68
	10%	89,92	90,58	90,02	91,45	86,26	89,88	89,89
	15%	92,12	90,26	90,34	92,20	91,23	89,49	87,13
	20%	89,98	89,26	89,15	89,44	89,75	91,26	90,58
	25%	88,22	89,44	87,13	88,37	87,67	90,58	89,96

**Fonte:** O Próprio Autor (2018)

# APÊNDICE C – Resultado da Técnica *FlexCon-C2* e Classificador Naïve Bayes

Tabela 17 – Resultado da aplicação da Técnica *FlexCon-C2* utilizando o classificador Naïve Bayes.

Base de dados	Exemplos Ini. Rot.	Variação do <i>cr</i>						
		2	3	4	5	6	7	8
Balance Scale	5%	80,33	80,78	83,18	79,69	82,08	79,37	80,31
	10%	81,59	82,54	84,13	82,38	80,48	83,97	84,44
	15%	79,67	80,32	80,02	84,65	80,31	80,14	83,67
	20%	77,74	81,61	78,71	81,94	76,94	80,97	76,29
	25%	83,85	84,65	83,53	83,37	81,77	83,86	83,05
Blood Transfusion Service	5%	75,87	74,80	74,67	72,93	70,80	74,00	72,53
	10%	75,87	75,60	76,93	74,67	74,93	75,07	74,53
	15%	73,96	73,02	74,63	72,35	72,47	72,19	74,36
	20%	71,13	71,66	73,42	73,03	72,34	69,65	71,54
	25%	75,20	75,87	77,73	75,73	76,93	74,13	75,60
Bupa	5%	55,14	54,29	50,29	55,43	50,29	56,86	56,29
	10%	55,02	49,22	55,34	55,60	50,09	55,32	53,87
	15%	55,88	54,41	54,12	54,71	55,29	54,41	53,24
	20%	60,57	57,14	56,57	55,43	58,00	57,14	57,43
	25%	54,41	50,59	57,94	50,29	54,12	54,71	51,47
Car	5%	79,82	79,82	79,82	79,77	80,40	80,69	79,30
	10%	81,28	80,71	80,93	81,00	81,11	80,71	81,00
	15%	81,69	81,16	80,17	82,38	81,10	80,99	80,99
	20%	80,94	81,47	81,69	82,27	81,23	80,88	81,87
	25%	81,09	80,34	80,40	83,05	82,13	81,38	81,49
Cnae-9	5%	25,00	29,26	30,28	25,74	23,24	30,00	26,76
	10%	30,65	29,07	28,24	26,11	29,63	25,83	27,22
	15%	28,89	29,54	28,61	25,65	30,37	24,44	28,33
	20%	29,54	26,76	24,63	29,07	30,00	22,87	27,50
	25%	29,63	30,93	30,00	24,63	26,48	29,07	30,65
Connectionist Mines vs Rocks	5%	56,17	59,50	55,67	57,57	61,52	60,52	63,93
	10%	64,76	61,90	67,62	67,14	65,24	65,71	64,29
	15%	67,14	69,05	60,48	62,86	62,38	65,24	61,90
	20%	53,64	61,50	63,98	66,02	60,57	64,83	62,21
	25%	72,64	72,82	64,73	71,45	74,77	70,68	75,14

Flare	5%	40,43	42,43	42,29	45,57	45,86	41,29	40,29
	10%	46,78	47,50	49,96	41,79	46,14	46,00	51,91
	15%	41,58	43,46	51,69	47,57	54,73	43,23	51,97
	20%	43,38	46,33	39,78	44,03	47,48	45,76	46,19
	25%	48,29	46,92	50,90	44,96	51,23	50,68	50,40
Haberman	5%	70,52	73,41	72,74	69,88	71,80	72,42	72,69
	10%	71,14	69,25	76,45	70,49	69,15	78,10	74,11
	15%	65,92	70,81	67,17	67,82	67,16	67,17	67,81
	20%	72,00	74,67	70,00	72,33	72,00	71,00	69,33
	25%	63,52	69,57	66,76	67,42	68,57	70,22	65,10
Handwritten Digits	5%	83,74	84,02	84,27	83,83	84,15	83,83	84,06
	10%	84,01	84,37	84,18	84,64	84,28	84,95	84,41
	15%	85,66	85,99	85,72	86,00	85,82	86,11	85,42
	20%	85,17	85,30	85,81	85,65	85,42	85,47	85,18
	25%	85,73	85,80	85,51	85,73	85,71	85,58	85,45
Hill Valley	5%	49,14	52,83	46,76	53,72	49,22	50,36	49,22
	10%	51,07	51,63	51,06	49,43	50,83	48,92	48,84
	15%	51,03	49,79	50,45	50,12	50,12	50,12	50,37
	20%	48,43	47,36	49,42	50,25	51,16	48,60	49,59
	25%	50,33	50,66	50,65	50,24	51,74	50,50	50,34
Image Segmentation	5%	76,23	76,02	73,38	76,10	77,32	76,71	76,06
	10%	78,92	79,91	79,57	80,09	80,13	78,35	80,91
	15%	75,97	76,67	78,66	79,39	76,54	76,75	78,40
	20%	76,84	77,58	76,71	77,66	77,58	76,49	76,97
	25%	79,00	79,78	78,74	79,31	77,79	79,78	76,93
Indian Liver Patient	5%	54,51	55,37	55,03	55,38	55,39	53,18	53,14
	10%	52,41	53,10	53,79	53,97	53,28	52,93	51,55
	15%	56,55	57,07	57,41	55,00	57,07	57,59	55,86
	20%	62,93	63,09	63,76	65,47	66,32	65,99	67,86
	25%	62,04	57,46	58,99	63,27	63,09	59,36	60,89
Iris	5%	89,33	92,67	92,67	94,67	93,33	89,33	90,67
	10%	94,00	86,67	89,33	90,67	89,33	90,67	86,00
	15%	88,67	89,33	88,00	89,33	95,33	87,33	93,33
	20%	90,67	94,00	88,00	92,00	92,00	95,33	90,00
	25%	92,67	88,67	91,33	88,00	96,00	89,33	90,00
King Ropck vs King Pawn	5%	83,07	83,39	84,70	85,80	82,45	83,95	84,61
	10%	83,38	82,88	82,88	84,70	83,29	84,63	84,70
	15%	86,97	88,44	86,38	86,94	86,94	87,56	86,75
	20%	85,48	85,01	85,07	85,04	84,32	86,60	84,07
	25%	86,53	85,12	83,84	84,47	84,09	85,88	84,03

Leukemia Haslinger	5%	68,00	72,00	75,00	73,00	71,00	76,00	73,00
	10%	82,00	80,00	88,00	82,00	80,00	84,00	82,00
	15%	79,00	86,00	83,00	82,00	82,00	81,00	82,00
	20%	66,00	68,00	71,00	66,00	70,00	75,00	69,00
	25%	79,00	80,00	81,00	81,00	77,00	75,00	72,00
Mammographic Mass	5%	80,31	81,03	80,00	80,62	80,00	80,21	78,45
	10%	81,46	78,63	79,16	79,99	78,95	79,58	79,36
	15%	82,10	80,75	83,14	82,31	81,99	82,40	81,99
	20%	81,48	80,53	80,85	80,64	80,65	78,01	79,60
	25%	77,62	78,34	78,75	77,30	78,44	77,72	78,54
Mfeat Karhunen	5%	90,45	89,70	90,35	89,85	90,20	90,95	90,20
	10%	88,45	89,15	89,95	89,05	89,10	89,35	88,90
	15%	88,60	89,45	87,75	88,70	87,80	87,30	87,25
	20%	88,15	88,20	87,65	88,30	87,55	89,70	88,70
	25%	88,50	88,60	89,50	88,70	89,35	89,95	90,05
Mushroom	5%	94,97	94,72	94,77	95,29	94,79	95,18	94,97
	10%	94,86	94,53	95,02	94,81	94,34	94,59	94,70
	15%	94,66	94,95	95,03	94,74	95,37	94,93	95,11
	20%	94,32	94,15	94,20	94,00	93,79	94,04	93,92
	25%	93,05	93,39	93,51	93,10	93,04	92,91	93,35
Musk	5%	83,70	82,49	82,56	83,93	82,61	83,55	82,76
	10%	84,44	85,02	84,91	85,71	84,45	85,05	84,47
	15%	85,19	83,90	84,37	86,22	84,91	85,19	83,81
	20%	83,62	85,55	84,92	84,00	84,00	85,00	85,24
	25%	83,74	82,74	83,33	81,32	82,45	83,20	83,24
Ozone Level Detection	5%	78,78	79,49	79,13	80,75	78,31	80,24	78,50
	10%	77,52	78,50	79,61	77,56	77,48	81,38	81,73
	15%	82,24	80,20	80,11	79,13	79,88	79,41	80,16
	20%	79,49	75,67	75,74	76,77	76,69	76,76	78,78
	25%	77,71	79,53	81,90	80,55	78,89	79,29	79,45
Phishing	5%	82,67	82,52	82,07	83,78	82,22	82,22	82,15
	10%	84,78	85,15	85,51	85,07	85,66	85,66	85,15
	15%	83,83	83,83	84,72	83,24	84,28	83,54	83,98
	20%	83,48	81,41	82,00	83,04	84,44	82,30	81,63
	25%	83,85	81,78	83,56	82,15	82,81	83,78	84,74
Pima	5%	69,40	70,85	69,78	70,70	71,75	71,09	69,54
	10%	70,65	72,47	72,34	71,30	71,04	71,17	70,26
	15%	72,21	69,87	72,99	72,99	72,08	71,30	71,82
	20%	82,08	77,27	79,22	80,00	80,26	79,09	80,65
	25%	70,07	70,33	71,37	70,46	71,37	70,46	72,68

Planning Relax	5%	62,78	58,89	59,44	58,89	61,11	63,33	63,33
	10%	63,89	61,67	57,78	63,89	61,11	64,44	62,78
	15%	61,73	63,25	57,37	61,75	55,32	61,05	57,89
	20%	59,91	59,56	56,35	64,36	64,91	62,75	57,51
	25%	60,00	67,22	57,22	63,33	63,89	57,22	62,22
Seeds	5%	86,67	90,00	88,57	89,52	89,05	90,48	91,43
	10%	81,43	82,38	84,29	84,29	76,19	80,95	83,81
	15%	87,62	90,95	89,05	85,24	88,57	87,62	86,67
	20%	91,43	90,48	90,95	93,33	90,95	89,52	92,86
	25%	91,43	91,43	91,43	87,14	92,86	89,05	89,05
Semeion	5%	64,15	64,85	63,41	65,73	65,17	63,34	66,62
	10%	64,02	67,55	67,40	66,59	67,80	64,63	65,50
	15%	60,06	59,62	61,38	63,08	59,25	59,62	62,64
	20%	64,73	64,78	64,30	61,94	66,37	65,09	69,13
	25%	62,72	61,53	65,71	68,53	61,69	65,83	61,16
Spectf Heart	5%	81,69	78,90	79,16	77,75	81,71	79,76	82,29
	10%	77,66	73,64	74,47	75,04	74,17	75,04	74,72
	15%	73,64	72,88	73,64	71,06	70,51	71,99	72,01
	20%	74,00	75,14	78,57	78,00	73,43	75,71	75,43
	25%	73,06	69,56	72,18	70,74	70,72	70,41	67,83
Tic Tac Toe	5%	68,18	69,84	69,43	68,49	69,85	70,26	69,12
	10%	70,58	68,48	70,27	68,28	69,12	69,95	69,43
	15%	66,35	66,25	65,42	65,21	65,00	68,85	66,56
	20%	69,69	68,44	67,81	68,33	67,92	68,54	69,37
	25%	68,32	67,68	67,89	68,21	67,89	66,11	67,47
Twonorm	5%	97,80	97,73	97,76	97,59	97,74	97,77	97,85
	10%	97,31	97,14	97,35	97,32	97,39	97,30	97,35
	15%	97,83	97,83	97,72	97,73	97,81	97,91	97,65
	20%	98,00	97,86	97,83	97,77	97,94	97,90	97,90
	25%	97,74	97,80	97,78	97,77	97,73	97,78	97,82
Vehicle	5%	45,55	46,61	45,30	46,14	45,07	46,50	44,01
	10%	45,48	49,13	49,11	47,25	49,34	45,84	48,67
	15%	43,06	45,06	44,12	44,71	45,29	46,12	46,47
	20%	46,16	48,65	44,97	48,52	50,16	45,80	47,32
	25%	44,07	47,18	46,95	46,72	47,18	44,21	44,07
Waveform	5%	79,28	79,58	80,08	80,18	79,86	80,04	79,28
	10%	79,74	79,68	79,20	79,44	79,36	79,50	79,16
	15%	80,30	80,54	80,32	80,18	80,56	80,66	80,30
	20%	79,44	79,72	79,32	79,90	79,52	79,72	79,44
	25%	80,10	80,30	80,36	80,24	79,88	80,20	80,44



Wilt	5%	90,02	91,18	87,73	87,89	89,94	87,54	90,68
	10%	89,92	90,58	90,02	91,45	86,26	89,88	89,89
	15%	92,12	90,26	90,34	92,20	91,23	89,49	87,13
	20%	89,98	89,26	89,15	89,44	89,75	91,26	90,58
	25%	88,22	89,44	87,13	88,37	87,67	90,58	89,96

**Fonte:** O Próprio Autor (2018)

# APÊNDICE D – Resultado da Técnica *FlexCon-C1(s)* e Classificador *rpartXse*

Tabela 18 – Resultado da aplicação da Técnica *FlexCon-C1(s)* utilizando o classificador *rpartXse*.

Base de dados	Exemplos Ini. Rot.	Variação do <i>cr</i>						
		2	3	4	5	6	7	8
Balance Scale	5%	70,95	67,78	68,89	65,71	69,05	68,89	65,40
	10%	69,96	70,60	71,71	68,83	69,77	69,77	68,33
	15%	72,98	69,77	66,55	70,06	68,16	67,50	69,31
	20%	64,33	69,29	64,32	68,46	70,24	65,93	68,02
	25%	70,81	72,26	73,06	74,19	69,52	72,10	70,81
Blood Transfusion Service	5%	76,27	76,27	75,60	76,27	76,67	75,47	76,27
	10%	76,00	76,00	77,73	75,60	76,67	76,00	76,27
	15%	76,51	76,51	76,38	76,24	76,11	76,51	76,78
	20%	76,25	75,84	76,25	76,11	76,78	76,11	76,38
	25%	75,33	76,13	75,47	74,53	74,67	75,73	74,53
Bupa	5%	59,71	59,14	58,55	59,45	62,28	55,90	58,85
	10%	57,92	62,10	59,13	56,55	57,68	57,39	59,76
	15%	61,44	61,77	53,37	58,30	52,95	57,39	56,53
	20%	57,94	58,53	61,47	59,41	61,76	60,59	57,35
	25%	56,29	60,86	58,86	57,14	58,57	57,43	57,14
Car	5%	87,80	88,15	86,47	85,78	87,40	84,74	86,88
	10%	87,92	87,51	85,61	88,44	88,21	88,27	85,78
	15%	87,69	84,37	86,70	84,43	85,02	86,41	86,88
	20%	87,54	87,77	85,10	85,45	85,57	84,76	84,70
	25%	88,10	88,10	85,40	85,00	88,05	86,90	85,92
Cnae-9	5%	66,30	66,76	65,37	64,26	66,67	65,46	65,37
	10%	65,28	67,31	67,50	67,69	64,17	71,02	65,46
	15%	69,54	67,22	69,54	69,17	66,57	68,70	68,24
	20%	68,15	68,70	69,35	69,17	70,37	68,98	74,54
	25%	70,65	68,33	68,70	68,98	68,33	70,19	70,65
Connectionist Mines vs Rocks	5%	62,38	60,00	56,19	61,90	64,76	56,19	56,67
	10%	63,14	57,02	59,62	57,00	65,86	60,55	57,07
	15%	61,26	60,05	55,60	60,17	68,83	62,45	69,40
	20%	60,95	58,10	63,81	66,67	65,71	65,24	61,90
	25%	63,33	59,05	57,62	63,33	60,00	62,86	54,29

Flare	5%	71,26	71,19	71,18	70,62	71,98	71,83	71,47
	10%	72,80	72,73	72,08	71,45	73,53	72,66	72,88
	15%	71,84	71,62	72,41	71,55	72,41	72,20	70,26
	20%	73,01	73,66	74,75	71,79	73,66	73,95	72,38
	25%	72,06	71,84	72,08	72,13	72,64	73,80	72,58
Haberman	5%	73,76	73,08	72,80	69,17	73,10	74,43	73,76
	10%	73,76	70,54	73,76	73,76	71,18	70,53	72,18
	15%	71,81	74,09	72,75	70,22	73,76	73,76	73,76
	20%	73,10	73,76	73,43	73,09	71,77	73,76	72,76
	25%	72,60	73,25	73,23	73,23	71,35	72,94	72,94
Handwritten Digits	5%	87,81	89,15	88,34	88,25	88,25	88,40	88,84
	10%	88,43	89,08	87,84	88,20	88,03	88,30	87,74
	15%	87,73	87,38	88,14	88,11	86,81	87,69	87,68
	20%	89,02	88,73	88,81	89,18	89,21	89,34	88,79
	25%	89,04	89,33	89,24	89,02	88,59	88,38	88,23
Hill Valley	5%	51,90	50,33	51,24	50,99	51,90	53,06	52,48
	10%	50,45	50,86	51,43	52,17	51,02	49,29	51,10
	15%	51,16	53,88	51,82	50,00	52,31	53,72	53,47
	20%	50,38	49,45	50,46	50,53	49,22	48,29	53,03
	25%	51,31	52,54	50,82	49,67	50,41	51,39	50,16
Image Segmentation	5%	89,26	89,87	89,65	89,91	89,70	90,09	89,91
	10%	89,83	90,39	90,61	89,48	89,26	89,31	90,26
	15%	90,39	89,22	88,79	89,74	89,83	88,83	89,91
	20%	89,96	90,56	88,05	90,30	90,56	88,83	91,04
	25%	90,22	89,09	90,48	90,17	89,13	90,35	89,00
Indian Liver Patient	5%	69,86	70,66	71,56	70,18	69,10	71,56	70,33
	10%	70,42	71,80	71,29	71,46	71,80	69,56	71,80
	15%	70,34	71,19	71,19	70,34	69,66	70,34	70,68
	20%	70,43	71,48	70,78	71,30	72,34	71,31	70,96
	25%	70,94	70,94	71,28	70,94	71,11	70,43	69,75
Iris	5%	80,67	78,67	80,00	83,33	80,67	78,67	79,33
	10%	78,67	77,33	80,67	81,33	80,67	76,67	80,00
	15%	75,33	77,33	80,67	76,67	81,33	78,67	83,33
	20%	82,67	79,33	81,33	80,00	82,00	82,00	78,67
	25%	76,67	72,67	76,00	75,33	75,33	77,33	74,67
King Ropck vs King Pawn	5%	95,37	95,49	95,96	96,09	95,65	95,81	96,43
	10%	95,18	95,37	95,65	96,03	95,87	95,46	94,99
	15%	94,15	95,81	95,24	94,52	94,65	95,52	94,99
	20%	96,47	95,97	96,59	95,91	95,97	96,41	96,59
	25%	95,59	95,84	95,49	95,68	95,93	95,81	95,18

Leukemia Haslinger	5%	57,00	59,00	57,00	54,00	60,00	54,00	59,00
	10%	63,54	68,99	58,79	67,17	68,99	64,04	73,23
	15%	58,00	60,00	58,00	57,00	60,00	58,00	63,00
	20%	66,00	65,00	68,00	56,00	58,00	60,00	64,00
	25%	61,00	67,00	70,00	70,00	60,00	65,00	64,00
Mammographic Mass	5%	82,81	80,50	82,62	82,72	80,30	83,45	83,25
	10%	82,58	81,65	80,94	82,18	82,58	82,06	82,06
	15%	80,73	82,09	78,56	81,66	79,39	79,18	79,60
	20%	79,99	79,06	80,50	79,16	79,78	79,57	80,72
	25%	79,18	77,61	79,80	75,80	79,17	79,80	79,07
Mfeat Karhunen	5%	68,50	69,15	66,30	70,15	68,30	66,95	68,35
	10%	68,50	67,20	65,85	66,95	66,20	67,90	65,00
	15%	65,00	65,75	63,35	64,60	64,85	66,00	63,80
	20%	66,20	67,60	65,20	67,25	64,50	64,00	64,65
	25%	63,45	64,45	62,35	63,35	61,95	64,70	64,05
Mushroom	5%	99,40	99,56	99,29	99,67	99,61	99,42	99,53
	10%	99,85	99,68	99,74	99,84	99,77	99,63	99,67
	15%	99,75	99,77	99,53	99,70	99,57	99,79	99,73
	20%	99,62	99,52	99,50	99,54	99,43	99,52	99,57
	25%	99,45	99,43	99,47	99,50	99,59	99,35	99,70
Musk	5%	98,42	98,59	98,18	98,35	98,41	98,42	98,44
	10%	98,33	98,55	98,41	98,24	98,38	98,44	98,21
	15%	98,67	98,98	99,09	98,68	98,82	98,64	99,01
	20%	98,33	98,80	98,85	98,91	99,08	98,83	98,74
	25%	98,67	98,42	98,50	98,52	98,82	98,42	98,39
Ozone Level Detection	5%	97,05	97,05	97,05	97,05	97,05	97,05	97,05
	10%	97,23	97,23	97,23	97,23	97,23	97,08	97,23
	15%	97,05	97,05	97,05	97,05	97,05	97,05	97,05
	20%	97,04	97,04	97,04	97,04	97,04	97,00	96,69
	25%	97,24	97,24	97,24	97,04	97,04	97,24	97,24
Phishing	5%	82,65	82,97	83,03	82,81	82,44	83,04	82,66
	10%	81,03	81,33	81,11	81,11	81,41	80,80	81,26
	15%	79,33	79,33	80,30	79,04	80,30	81,41	79,26
	20%	82,43	81,98	83,24	83,17	81,98	82,28	82,87
	25%	83,93	82,22	83,85	82,81	83,93	83,93	84,00
Pima	5%	68,05	69,35	68,83	65,84	70,13	69,87	68,83
	10%	69,17	69,69	68,78	69,81	69,42	68,12	72,83
	15%	73,19	74,92	74,26	73,60	74,51	71,92	71,75
	20%	66,62	70,65	70,65	68,31	68,70	67,79	68,96
	25%	68,70	66,75	68,57	69,87	69,35	71,95	67,79

Planning Relax	5%	71,67	70,56	72,22	72,22	71,67	72,22	72,22
	10%	70,56	72,22	70,56	72,22	72,22	72,22	72,22
	15%	70,32	70,32	69,21	69,21	70,32	70,32	69,21
	20%	70,32	70,32	70,32	70,32	65,53	70,32	70,32
	25%	72,22	72,22	71,67	72,22	70,56	72,22	70,00
Seeds	5%	79,05	71,43	70,48	72,38	70,95	72,38	73,33
	10%	71,43	70,95	71,43	74,29	74,76	70,95	70,95
	15%	75,71	75,71	80,48	73,81	75,24	82,86	76,67
	20%	73,81	70,00	72,38	71,90	78,10	73,81	73,33
	25%	77,62	73,81	78,10	76,19	78,57	75,71	71,43
Semeion	5%	55,53	54,40	56,67	54,65	56,48	52,26	56,10
	10%	51,31	53,75	53,81	53,69	53,56	53,62	54,81
	15%	54,32	51,56	54,00	54,07	54,14	53,38	55,65
	20%	53,50	55,56	54,82	57,33	55,86	57,95	56,79
	25%	57,22	53,53	56,66	54,52	55,51	56,26	53,81
Spectf Heart	5%	74,00	76,00	74,86	74,57	71,71	75,43	74,86
	10%	73,73	73,42	67,89	72,61	72,03	73,71	75,15
	15%	72,77	71,34	72,48	74,19	69,24	71,88	70,45
	20%	70,36	72,58	72,61	73,20	71,78	73,19	72,08
	25%	69,14	70,86	71,43	71,71	71,43	70,57	71,14
Tic Tac Toe	5%	73,71	72,03	74,76	73,81	75,18	77,27	73,41
	10%	75,61	73,11	70,68	78,22	76,45	74,56	71,63
	15%	72,89	73,09	72,89	72,68	75,46	76,29	72,99
	20%	75,92	75,71	72,86	73,19	73,51	71,71	71,73
	25%	72,65	72,15	74,65	71,83	72,14	69,95	75,90
Twonorm	5%	79,19	78,95	79,43	78,68	78,89	78,97	79,46
	10%	79,31	79,46	79,36	79,03	79,28	78,81	80,08
	15%	79,97	79,72	80,28	79,59	80,24	79,81	79,70
	20%	80,11	79,86	79,72	79,54	79,88	80,08	79,26
	25%	79,37	78,85	78,65	80,67	80,15	79,80	79,22
Vehicle	5%	62,38	59,52	60,36	55,00	58,21	58,69	59,05
	10%	56,71	57,18	55,29	60,35	57,06	59,06	60,24
	15%	59,73	57,41	59,21	59,26	55,40	56,47	58,83
	20%	58,35	57,18	56,35	58,12	55,88	55,65	58,12
	25%	55,56	57,28	56,98	57,93	58,94	55,46	58,35
Waveform	5%	71,98	70,76	70,74	71,70	70,60	72,62	71,50
	10%	72,06	71,84	71,44	71,68	71,62	72,04	71,54
	15%	72,07	72,51	74,87	73,69	73,59	74,69	72,81
	20%	72,73	72,00	73,05	73,25	72,10	72,20	72,42
	25%	72,61	71,25	72,99	72,71	71,21	72,85	72,78

Wilt	5%	96,66	96,72	96,66	96,62	96,57	96,43	96,61
	10%	96,79	97,02	96,79	96,65	96,92	96,98	96,38
	15%	97,19	97,17	96,86	97,42	96,98	96,96	97,17
	20%	96,11	95,99	95,99	96,07	96,19	96,15	95,86
	25%	97,31	96,84	96,80	96,86	97,46	97,02	97,02

**Fonte:** O Próprio Autor (2018)

# APÊNDICE E – Resultado da Técnica *FlexCon-C1(v)* e Classificador *rpartXse*

Tabela 19 – Resultado da aplicação da Técnica *FlexCon-C1(v)* utilizando o classificador *rpartXse*.

Base de dados	Exemplos Ini. Rot.	Variação do <i>cr</i>						
		2	3	4	5	6	7	8
Balance Scale	5%	71,43	68,73	68,10	66,51	69,68	70,48	64,76
	10%	69,96	68,84	69,79	70,59	70,41	69,62	67,69
	15%	72,18	69,46	66,54	70,22	67,52	66,55	70,89
	20%	64,18	68,97	63,84	67,66	70,57	66,56	69,44
	25%	71,45	74,52	75,16	73,87	70,48	71,45	72,58
Blood Transfusion Service	5%	76,27	75,73	75,07	76,27	76,00	75,73	76,00
	10%	76,00	76,13	77,33	75,73	76,80	76,00	76,13
	15%	76,65	76,51	76,11	76,24	76,11	76,25	76,92
	20%	76,25	75,84	76,38	76,25	76,91	76,11	76,51
	25%	75,33	76,13	75,33	75,07	75,07	75,47	73,07
Bupa	5%	59,72	57,98	58,84	58,27	63,20	56,21	60,01
	10%	58,55	60,92	59,13	55,68	56,50	58,26	57,98
	15%	60,28	62,35	52,76	58,30	57,05	56,81	57,41
	20%	58,82	58,53	59,71	60,00	59,71	60,00	57,35
	25%	56,29	62,00	57,14	59,71	58,00	58,57	58,57
Car	5%	86,65	87,86	87,28	86,30	87,23	84,74	87,57
	10%	87,51	87,51	85,49	87,86	87,75	87,40	85,90
	15%	87,69	86,06	86,70	84,61	85,77	84,90	86,24
	20%	87,19	86,78	85,10	85,74	85,57	84,35	83,60
	25%	87,99	86,90	85,40	85,75	87,99	87,99	86,38
Cnae-9	5%	65,93	66,57	65,46	64,26	65,28	65,00	65,83
	10%	65,28	67,31	67,04	69,63	65,56	69,54	65,56
	15%	69,35	64,72	70,19	69,63	66,39	68,98	68,61
	20%	69,07	69,17	67,50	69,35	70,00	69,91	74,44
	25%	69,63	68,80	69,44	67,50	67,87	70,83	70,37
Connectionist Mines vs Rocks	5%	64,76	61,90	56,67	63,33	66,19	56,67	56,67
	10%	61,64	57,50	62,00	58,00	63,95	63,98	56,07
	15%	57,76	60,60	55,10	62,55	68,83	63,45	68,45
	20%	61,43	57,62	65,24	64,76	63,81	66,67	60,48
	25%	63,33	57,62	57,62	61,43	60,48	62,86	56,19

Flare	5%	71,55	71,05	71,19	71,91	71,98	71,54	71,19
	10%	72,52	72,66	73,02	72,38	73,60	72,16	73,24
	15%	71,98	71,84	72,41	72,27	72,20	72,20	70,40
	20%	73,22	73,01	74,17	71,22	73,95	73,88	72,24
	25%	72,50	71,85	72,66	72,13	72,64	74,09	72,73
Haberman	5%	73,76	72,43	72,80	68,84	72,10	74,13	73,76
	10%	73,76	70,54	73,76	73,76	71,18	70,22	72,18
	15%	71,81	72,75	69,44	70,22	73,76	73,76	73,76
	20%	74,10	73,76	73,10	73,09	71,77	73,76	71,75
	25%	72,60	73,25	72,60	73,23	72,02	72,62	73,27
Handwritten Digits	5%	88,15	89,03	88,34	88,05	88,21	88,63	88,81
	10%	88,29	88,94	87,97	88,27	87,98	88,43	87,47
	15%	87,44	87,40	88,02	88,13	86,78	87,48	88,04
	20%	88,94	88,75	88,83	89,11	89,44	89,64	88,74
	25%	89,03	89,29	89,27	89,02	88,56	88,72	88,07
Hill Valley	5%	51,82	49,34	52,07	50,58	52,40	52,15	52,73
	10%	50,69	48,89	51,02	51,93	50,61	48,80	51,35
	15%	50,66	53,80	51,65	50,99	51,65	51,90	52,23
	20%	50,38	49,29	50,62	50,04	49,64	48,30	53,20
	25%	51,56	50,74	51,15	48,11	50,49	50,66	52,21
Image Segmentation	5%	89,39	89,87	89,57	90,00	89,61	90,43	89,70
	10%	89,70	90,61	90,61	89,96	89,22	89,57	90,35
	15%	90,74	88,53	89,09	89,61	89,61	89,00	89,74
	20%	90,04	90,91	88,35	90,43	90,39	89,13	90,87
	25%	90,26	88,57	90,65	90,04	89,31	89,96	89,22
Indian Liver Patient	5%	68,44	70,15	71,56	70,88	71,56	72,24	71,56
	10%	71,63	72,14	70,42	71,80	71,11	69,91	71,80
	15%	68,64	71,19	71,19	71,53	69,66	70,34	71,19
	20%	70,43	70,96	71,13	71,30	71,65	71,31	70,96
	25%	70,43	70,77	70,94	70,94	70,42	69,73	71,62
Iris	5%	80,67	78,67	80,00	83,33	80,67	78,67	79,33
	10%	78,67	77,33	80,67	81,33	80,67	76,67	80,00
	15%	75,33	77,33	80,67	76,67	81,33	78,67	83,33
	20%	82,67	79,33	81,33	80,00	82,00	82,00	78,67
	25%	76,67	72,67	76,00	75,33	75,33	77,33	74,67
King Ropck vs King Pawn	5%	95,24	95,65	95,87	95,93	95,74	96,00	96,40
	10%	95,24	95,53	95,40	95,84	95,71	95,46	94,68
	15%	94,43	95,43	95,27	94,27	94,68	95,30	94,90
	20%	96,78	96,44	96,09	96,06	95,91	96,75	96,31
	25%	95,49	95,87	95,65	95,77	94,99	95,81	94,71



Leukemia Haslinger	5%	57,00	59,00	57,00	56,00	61,00	54,00	59,00
	10%	63,54	68,99	58,79	67,17	68,99	61,31	73,23
	15%	59,00	60,00	57,00	60,00	61,00	58,00	63,00
	20%	66,00	65,00	68,00	56,00	57,00	60,00	64,00
	25%	61,00	67,00	70,00	70,00	61,00	65,00	63,00
Mammographic Mass	5%	82,81	80,51	83,35	81,88	81,45	82,30	83,25
	10%	82,58	80,73	80,84	82,59	82,27	82,06	82,48
	15%	80,11	82,92	78,56	81,66	80,02	78,14	79,50
	20%	79,98	79,06	80,40	80,29	80,09	78,85	81,03
	25%	79,28	77,82	80,11	78,33	79,07	79,69	79,81
Mfeat Karhunen	5%	68,95	68,15	66,25	69,85	68,00	67,00	66,85
	10%	68,55	67,45	66,15	65,25	67,05	68,20	65,70
	15%	65,10	64,70	63,75	64,95	64,50	65,55	65,40
	20%	66,90	66,50	64,30	67,30	65,00	64,65	65,35
	25%	63,45	64,10	61,85	62,95	61,60	65,65	64,95
Mushroom	5%	99,42	99,56	99,52	99,63	99,62	99,42	99,57
	10%	99,74	99,62	99,78	99,82	99,77	99,63	99,67
	15%	99,75	99,77	99,50	99,58	99,63	99,78	99,73
	20%	99,57	99,66	99,52	99,46	99,43	99,57	99,47
	25%	99,37	99,43	99,52	99,50	99,59	99,34	99,70
Musk	5%	98,42	98,59	98,18	98,35	98,41	98,42	98,44
	10%	98,33	98,55	98,41	98,24	98,38	98,44	98,21
	15%	98,67	98,98	99,09	98,68	98,82	98,64	99,01
	20%	98,33	98,80	98,85	98,91	99,08	98,83	98,74
	25%	98,67	98,42	98,50	98,52	98,82	98,42	98,39
Ozone Level Detection	5%	97,05	97,05	97,05	97,05	97,05	97,05	97,05
	10%	97,04	97,23	97,23	97,23	97,23	97,23	97,23
	15%	97,05	97,05	97,05	96,58	97,05	97,05	97,05
	20%	97,04	97,04	97,04	97,04	97,04	97,04	97,04
	25%	97,24	97,24	97,24	97,04	97,24	97,24	97,24
Phishing	5%	82,58	82,00	82,73	83,10	82,22	83,18	82,81
	10%	81,70	81,26	82,06	80,96	81,70	80,58	81,55
	15%	79,48	79,63	80,59	78,44	79,56	80,07	79,85
	20%	82,43	82,28	83,31	82,80	82,50	81,69	82,13
	25%	84,07	84,00	83,63	82,81	84,67	83,93	83,85
Pima	5%	68,18	69,35	67,66	66,23	70,26	69,35	69,74
	10%	68,51	69,43	69,70	69,69	69,68	69,57	72,57
	15%	75,42	73,74	76,47	74,12	76,60	70,98	72,39
	20%	68,44	69,61	70,13	68,96	69,09	67,27	68,96
	25%	68,70	67,53	67,66	68,44	69,35	72,21	68,57

Planning Relax	5%	71,67	70,56	72,22	72,22	71,67	72,22	72,22
	10%	70,56	72,22	70,56	72,22	71,67	72,22	72,22
	15%	70,32	67,69	69,21	69,21	70,32	70,32	69,21
	20%	70,32	70,32	70,32	68,74	66,64	70,32	70,32
	25%	72,22	72,22	71,67	71,11	70,56	72,22	71,11
Seeds	5%	77,62	71,43	70,48	71,90	70,95	72,38	73,33
	10%	71,43	70,95	70,95	74,29	75,24	71,90	70,95
	15%	75,71	75,71	80,48	73,81	75,24	82,86	76,67
	20%	73,81	70,00	72,38	71,43	78,10	73,81	73,33
	25%	77,62	73,81	78,10	76,19	78,57	75,71	71,43
Semeion	5%	55,28	54,34	54,91	54,59	55,66	54,40	55,72
	10%	51,38	53,38	54,62	53,81	52,56	53,50	55,38
	15%	56,02	51,50	53,56	55,27	53,13	52,44	55,78
	20%	53,25	55,32	54,64	57,64	54,68	55,88	58,48
	25%	56,91	53,84	57,73	54,77	55,95	55,94	55,20
Spectf Heart	5%	74,00	76,00	76,00	74,86	72,00	75,43	75,14
	10%	73,73	73,43	70,25	72,29	70,56	73,99	74,56
	15%	73,91	70,75	72,77	73,02	68,66	72,47	71,30
	20%	70,08	72,88	72,61	73,48	70,93	73,48	71,49
	25%	71,43	71,14	72,00	68,86	72,00	71,14	72,00
Tic Tac Toe	5%	73,19	73,61	74,23	72,97	75,48	76,54	75,81
	10%	76,97	70,92	72,47	76,66	76,13	73,52	71,10
	15%	72,16	73,61	73,20	72,37	74,95	76,49	72,47
	20%	74,02	76,76	75,60	73,50	74,03	70,58	72,36
	25%	73,17	72,15	73,39	71,93	72,66	69,73	73,93
Twonorm	5%	78,81	78,81	79,23	78,47	79,61	79,30	78,86
	10%	79,34	79,38	78,80	79,16	79,64	79,27	79,58
	15%	79,82	79,95	79,84	79,66	79,88	79,42	79,30
	20%	80,09	80,57	80,26	79,46	80,22	79,62	79,38
	25%	79,78	79,53	78,72	80,77	80,01	79,91	79,24
Vehicle	5%	62,38	58,10	59,52	58,21	59,29	60,12	56,31
	10%	57,76	56,71	57,65	59,53	56,47	58,47	60,71
	15%	58,41	59,30	59,43	57,61	56,95	57,05	56,94
	20%	58,24	56,71	55,18	57,65	57,65	55,88	54,82
	25%	56,27	58,00	57,21	57,34	59,66	55,71	58,71
Waveform	5%	71,40	70,66	70,24	72,00	71,16	73,36	71,34
	10%	72,20	71,20	71,18	71,68	72,04	71,60	71,04
	15%	72,85	72,59	74,01	74,21	73,63	74,39	73,23
	20%	73,07	72,30	72,34	72,02	73,93	72,06	71,54
	25%	71,91	71,51	72,73	73,19	71,59	72,80	72,62

Wilt	5%	96,61	96,82	96,37	96,62	96,47	96,24	96,59
	10%	96,71	97,08	96,75	96,75	96,84	97,02	96,44
	15%	97,02	97,36	97,02	97,29	96,78	96,98	97,07
	20%	96,09	96,17	96,05	95,93	96,11	96,13	96,09
	25%	97,42	97,11	96,76	97,00	97,33	96,90	97,00

**Fonte:** O Próprio Autor (2018)

# APÊNDICE F – Resultado da Técnica *FlexCon-C2* e Classificador *rpartXse*

Tabela 20 – Resultado da aplicação da Técnica *FlexCon-C2* utilizando o classificador *rpartXse*.

Base de dados	Exemplos Ini. Rot.	Variação do <i>cr</i>						
		2	3	4	5	6	7	8
Balance Scale	5%	71,90	68,57	67,78	65,87	69,21	71,27	65,71
	10%	69,01	68,99	69,63	69,31	70,41	69,94	68,01
	15%	72,02	70,57	67,36	69,42	69,11	69,28	71,05
	20%	65,13	69,77	63,84	68,13	68,97	67,21	66,57
	25%	71,29	72,26	74,84	74,19	69,68	72,10	72,58
Blood Transfusion Service	5%	76,27	76,00	75,07	76,13	76,00	75,73	76,40
	10%	76,00	76,53	77,33	75,73	76,00	75,87	76,40
	15%	76,51	76,38	76,11	76,24	76,25	76,51	76,78
	20%	76,25	75,84	76,25	75,71	76,78	77,19	76,51
	25%	75,33	76,13	75,47	75,07	75,20	75,47	73,87
Bupa	5%	62,03	59,45	59,43	58,27	59,42	55,33	59,72
	10%	59,98	62,10	59,13	55,07	56,50	58,57	57,13
	15%	62,03	60,62	52,18	58,01	53,83	54,82	56,53
	20%	63,24	60,00	60,00	58,53	62,35	60,59	57,35
	25%	57,43	62,86	58,86	58,86	58,29	57,14	58,57
Car	5%	86,94	87,11	86,01	86,36	86,99	85,61	87,05
	10%	87,63	87,63	86,01	87,98	86,82	87,86	86,53
	15%	87,69	85,60	86,70	84,31	85,42	85,48	86,41
	20%	86,78	87,77	86,67	85,79	85,45	85,51	84,06
	25%	87,24	87,70	85,80	84,71	88,56	87,53	86,03
Cnae-9	5%	64,91	65,83	65,19	64,17	65,93	65,37	65,65
	10%	64,44	66,02	67,59	69,91	64,35	70,65	65,19
	15%	69,35	66,67	70,28	69,72	67,13	68,80	68,15
	20%	68,98	69,07	68,80	69,44	70,37	70,00	73,52
	25%	69,35	68,70	68,52	69,17	69,26	70,09	70,46
Connectionist Mines vs Rocks	5%	61,43	64,29	57,62	63,33	66,19	56,19	58,10
	10%	64,14	57,98	64,38	57,95	65,36	62,45	56,07
	15%	61,12	57,62	56,60	60,17	67,40	63,45	66,45
	20%	60,95	58,10	61,90	66,67	64,29	67,14	59,52
	25%	63,33	58,10	56,67	63,33	60,00	62,86	56,67

Flare	5%	71,12	71,26	70,90	71,55	70,97	71,26	71,19
	10%	73,73	72,66	71,72	71,88	74,03	72,73	72,81
	15%	72,06	71,77	72,41	71,33	72,34	71,98	70,11
	20%	72,86	73,73	74,61	70,64	73,81	73,81	72,67
	25%	72,50	71,99	72,08	72,13	72,64	74,09	72,80
Haberman	5%	73,76	72,43	72,80	71,11	73,10	74,43	73,76
	10%	73,76	70,54	73,76	73,76	71,18	70,53	73,47
	15%	73,43	73,76	70,42	73,76	73,76	73,76	74,73
	20%	73,76	73,76	73,43	73,09	71,77	74,09	73,09
	25%	72,60	73,25	73,23	73,23	71,35	72,94	72,94
Handwritten Digits	5%	88,15	88,98	88,37	88,27	88,31	88,35	88,81
	10%	88,55	88,94	87,80	88,25	88,05	88,27	87,61
	15%	87,71	87,45	87,95	88,08	86,75	87,69	87,44
	20%	88,62	89,33	88,75	89,41	89,32	89,45	88,78
	25%	89,06	89,32	89,32	89,05	88,56	88,74	88,37
Hill Valley	5%	51,65	49,92	52,31	50,58	53,47	53,22	52,56
	10%	50,37	48,63	52,18	51,59	50,69	49,45	51,27
	15%	50,41	54,05	52,23	49,67	51,49	53,06	53,97
	20%	49,06	48,87	49,79	49,54	49,64	47,96	53,20
	25%	50,49	50,33	52,30	49,59	51,48	49,92	50,66
Image Segmentation	5%	89,65	90,00	89,91	89,57	89,18	90,26	89,52
	10%	90,04	90,26	90,95	90,00	88,57	89,70	90,35
	15%	90,43	88,92	89,26	90,04	90,09	88,53	89,87
	20%	89,96	90,91	88,79	89,65	90,61	89,05	90,65
	25%	90,30	89,39	90,78	90,00	89,44	90,43	89,13
Indian Liver Patient	5%	69,33	70,32	71,21	70,88	69,10	72,24	71,56
	10%	71,28	71,27	71,29	71,80	71,80	69,91	71,80
	15%	68,64	71,19	70,17	71,36	69,66	70,34	71,19
	20%	70,26	70,96	70,43	71,30	72,17	71,31	70,96
	25%	70,43	70,25	71,62	71,28	70,59	69,73	69,93
Iris	5%	80,67	78,67	80,00	83,33	80,67	78,67	79,33
	10%	78,67	77,33	80,67	81,33	80,67	76,67	80,00
	15%	75,33	77,33	80,67	76,67	81,33	78,67	83,33
	20%	82,67	79,33	81,33	80,00	82,00	82,00	78,67
	25%	76,67	72,67	76,00	75,33	75,33	77,33	74,67
King Ropck vs King Pawn	5%	95,46	95,71	95,81	95,59	95,65	95,84	95,90
	10%	95,12	96,15	95,65	96,03	95,31	95,71	94,96
	15%	94,43	95,49	95,30	94,37	94,84	95,55	94,90
	20%	96,50	96,06	96,41	95,88	96,41	96,62	96,25
	25%	95,87	95,74	95,68	95,74	95,49	95,49	94,37

Leukemia Haslinger	5%	57,00	61,00	57,00	58,00	61,00	54,00	59,00
	10%	63,54	68,99	58,79	67,17	68,99	65,86	73,23
	15%	58,00	61,00	57,00	60,00	60,00	58,00	63,00
	20%	66,00	65,00	68,00	58,00	58,00	60,00	64,00
	25%	61,00	67,00	71,00	70,00	60,00	65,00	64,00
Mammographic Mass	5%	82,91	80,50	82,31	83,04	81,14	83,35	82,93
	10%	82,48	81,55	80,84	82,18	82,06	82,06	82,06
	15%	80,11	82,71	78,56	80,83	79,60	78,87	79,81
	20%	80,61	78,95	80,40	79,46	80,09	78,54	80,31
	25%	78,76	77,92	79,38	78,54	78,96	79,69	79,28
Mfeat Karhunen	5%	70,00	69,00	66,30	69,80	68,00	68,00	67,50
	10%	68,60	67,05	66,30	65,40	66,90	67,25	66,10
	15%	66,45	66,05	63,50	63,95	64,50	65,55	63,75
	20%	66,70	66,55	65,70	67,15	66,20	63,95	65,45
	25%	63,20	64,30	62,40	63,25	61,30	65,50	65,20
Mushroom	5%	99,40	99,56	99,52	99,70	99,57	99,42	99,53
	10%	99,78	99,68	99,70	99,83	99,82	99,64	99,67
	15%	99,75	99,77	99,50	99,72	99,61	99,79	99,73
	20%	99,62	99,69	99,52	99,54	99,48	99,57	99,43
	25%	99,43	99,43	99,48	99,50	99,62	99,37	99,73
Musk	5%	98,42	98,59	98,18	98,35	98,41	98,42	98,44
	10%	98,33	98,55	98,41	98,24	98,38	98,44	98,21
	15%	98,67	98,98	99,09	98,68	98,82	98,64	99,01
	20%	98,33	98,80	98,85	98,91	99,08	98,83	98,74
	25%	98,67	98,42	98,50	98,52	98,82	98,42	98,39
Ozone Level Detection	5%	97,05	97,05	97,05	97,05	97,05	97,05	97,05
	10%	97,04	97,23	97,23	97,23	97,23	97,08	97,23
	15%	97,05	97,05	97,05	97,05	97,05	97,05	97,05
	20%	97,04	97,04	97,04	97,04	97,04	97,00	97,04
	25%	97,24	97,24	97,24	97,04	97,04	97,24	97,24
Phishing	5%	82,43	82,44	82,73	82,88	82,23	83,18	82,66
	10%	81,62	81,33	81,55	80,89	81,55	80,80	81,40
	15%	79,26	79,78	80,15	79,56	79,26	80,89	79,26
	20%	83,38	82,43	83,09	82,35	81,98	82,43	82,72
	25%	84,74	83,85	83,41	82,67	84,37	83,93	84,00
Pima	5%	68,18	69,48	68,18	67,01	70,52	69,22	70,39
	10%	69,95	69,68	70,60	70,60	69,29	68,77	71,78
	15%	73,73	74,92	76,60	73,86	76,47	73,48	74,11
	20%	67,92	70,26	70,65	69,35	69,87	67,79	69,48
	25%	70,39	67,14	66,62	69,87	69,74	71,82	67,92

Planning Relax	5%	71,67	70,00	72,22	72,22	71,67	72,22	72,22
	10%	70,56	72,22	70,56	72,22	71,11	72,22	72,22
	15%	70,32	70,32	69,21	68,68	70,32	70,32	69,21
	20%	68,22	70,32	70,32	70,32	66,64	70,32	68,74
	25%	72,22	72,22	71,67	72,22	70,56	72,22	71,11
Seeds	5%	77,62	71,43	70,48	72,86	70,95	72,38	73,33
	10%	71,43	70,95	70,95	74,29	74,76	71,43	70,95
	15%	75,71	75,71	80,48	73,81	75,24	82,86	76,67
	20%	73,81	70,00	72,38	71,90	78,10	73,81	73,33
	25%	77,62	73,81	78,10	76,19	78,57	75,71	71,43
Semeion	5%	54,78	55,09	55,85	55,85	55,35	53,02	55,91
	10%	52,12	52,31	54,50	53,62	52,31	52,56	55,00
	15%	54,82	52,70	53,62	55,02	54,33	52,88	55,71
	20%	52,94	55,63	54,52	55,89	55,74	56,70	58,04
	25%	56,47	53,40	55,22	54,20	55,32	56,01	54,57
Spectf Heart	5%	74,00	75,14	74,86	74,86	71,43	73,71	76,57
	10%	73,73	73,71	69,36	72,57	71,73	73,69	74,26
	15%	74,20	70,45	72,48	71,87	69,24	70,71	71,30
	20%	70,38	72,88	73,73	73,48	71,49	73,77	71,49
	25%	70,00	69,14	72,57	71,43	72,00	71,14	71,43
Tic Tac Toe	5%	71,00	71,82	74,98	72,66	75,61	76,96	76,02
	10%	76,34	74,14	71,42	78,22	75,71	73,20	71,10
	15%	70,10	73,51	73,81	74,02	74,95	75,98	71,55
	20%	75,50	75,61	74,54	72,77	71,10	71,30	72,46
	25%	74,96	72,77	73,71	73,41	72,14	69,11	76,22
Twonorm	5%	78,39	79,35	79,31	79,04	79,81	79,89	79,04
	10%	78,92	78,70	79,58	78,95	79,77	79,54	80,20
	15%	79,35	80,41	80,51	79,97	80,26	79,57	79,74
	20%	80,15	80,22	80,22	79,97	79,91	79,65	79,09
	25%	79,74	79,26	78,52	80,59	79,47	79,66	79,35
Vehicle	5%	62,26	58,45	60,12	57,38	59,05	59,40	58,81
	10%	56,59	57,65	56,82	60,00	58,12	58,71	61,76
	15%	57,59	59,42	59,67	56,56	57,28	57,52	57,17
	20%	57,65	55,53	58,12	60,24	57,41	54,71	55,88
	25%	57,83	58,25	57,34	59,26	59,66	55,82	58,96
Waveform	5%	72,28	70,86	71,34	72,06	70,98	73,56	70,88
	10%	72,14	71,36	70,12	71,72	71,80	71,28	71,44
	15%	73,19	72,53	74,39	74,37	73,53	74,17	72,69
	20%	71,84	72,14	72,73	71,86	72,89	71,86	72,38
	25%	72,69	71,47	72,07	73,09	72,05	72,70	73,00

Wilt	5%	96,76	96,90	96,59	96,66	96,59	96,37	96,57
	10%	96,69	96,79	96,69	96,46	96,82	97,21	96,57
	15%	97,25	97,38	96,84	97,31	96,71	97,13	97,13
	20%	96,17	96,22	95,99	95,93	95,99	96,13	96,07
	25%	97,31	97,13	96,80	96,90	97,48	97,13	96,94

**Fonte:** O Próprio Autor (2018)



# APÊNDICE G – Resultado da Técnica *FlexCon-C1(s)* e Classificador RIPPER

Tabela 21 – Resultado da aplicação da Técnica Flexcon-C1(s) utilizando o classificador RIPPER.

Base de dados	Exemplos Ini. Rot.	Variação do <i>cr</i>						
		2	3	4	5	6	7	8
Balance Scale	5%	70,48	71,27	71,43	71,43	73,17	70,16	73,65
	10%	72,54	77,46	72,38	74,13	72,06	75,87	71,59
	15%	74,60	74,60	73,81	74,13	74,60	72,70	73,33
	20%	73,98	73,69	74,79	73,49	72,03	71,39	73,34
	25%	73,73	69,81	72,06	72,75	71,76	73,73	73,12
Blood Transfusion Service	5%	73,84	72,91	74,51	74,64	73,70	74,64	75,85
	10%	76,00	75,73	73,73	75,87	75,60	76,13	72,80
	15%	73,70	74,09	75,04	75,20	74,78	73,97	74,37
	20%	76,67	74,67	75,33	72,53	70,93	74,80	72,13
	25%	77,20	75,60	76,40	76,53	76,67	76,40	78,13
Bupa	5%	55,95	55,64	57,69	57,99	55,63	57,71	59,95
	10%	59,71	57,43	60,00	61,14	57,14	55,14	59,43
	15%	61,14	66,00	61,43	62,00	61,71	60,29	60,29
	20%	57,65	59,41	52,35	60,88	58,82	58,24	53,82
	25%	58,24	63,53	57,06	56,18	60,59	61,18	56,47
Car	5%	71,94	74,03	75,48	74,78	73,27	75,36	72,69
	10%	73,29	72,83	73,01	74,74	73,41	74,22	74,39
	15%	73,29	73,87	73,12	73,30	74,91	74,84	74,68
	20%	73,83	77,18	76,72	75,79	76,09	74,64	77,70
	25%	75,29	73,31	74,13	75,87	75,58	73,60	74,77
Cnae-9	5%	67,04	65,37	66,30	67,78	66,20	63,89	67,69
	10%	66,85	68,24	68,80	67,96	69,35	64,07	67,31
	15%	65,09	63,24	62,59	64,72	65,56	65,93	67,96
	20%	67,69	67,41	66,94	67,41	66,94	67,13	69,54
	25%	65,56	66,76	68,06	67,41	67,78	64,26	62,13
Connectionist Mines vs Rocks	5%	66,67	66,19	63,33	63,81	65,24	63,33	63,33
	10%	66,67	63,33	61,90	60,48	68,10	65,71	67,14
	15%	62,86	67,62	62,86	64,29	70,00	66,19	61,90
	20%	63,93	62,83	66,36	66,02	66,76	66,33	57,88
	25%	55,02	57,05	62,36	57,14	59,57	65,86	58,10

Flare	5%	66,00	70,71	65,07	65,50	68,36	68,64	68,43
	10%	68,54	67,88	69,20	67,74	67,96	71,17	66,13
	15%	63,23	61,51	64,81	64,30	65,80	64,02	65,53
	20%	68,11	66,97	66,26	67,82	64,67	65,47	64,72
	25%	68,47	68,05	69,99	65,72	69,96	68,68	67,46
Haberman	5%	75,81	70,97	73,23	74,19	74,84	73,23	74,84
	10%	74,19	72,26	73,23	70,65	72,90	73,87	70,65
	15%	72,67	72,00	73,00	71,00	71,00	71,67	74,67
	20%	71,98	69,10	71,04	70,33	70,75	69,46	71,96
	25%	72,00	72,00	71,00	70,33	68,00	71,67	66,00
Handwritten Digits	5%	88,16	88,26	88,34	88,18	88,05	88,33	88,36
	10%	88,11	88,44	87,42	88,45	87,70	87,90	88,16
	15%	88,87	88,65	88,14	88,57	88,25	88,12	89,01
	20%	87,40	86,61	87,23	87,41	87,50	87,71	87,48
	25%	88,05	87,95	87,61	87,18	87,37	87,05	87,72
Hill Valley	5%	49,96	51,70	51,52	53,92	52,10	49,63	52,93
	10%	53,98	52,25	54,33	54,07	50,83	53,00	53,00
	15%	55,07	54,83	54,07	50,66	53,00	50,58	52,00
	20%	45,95	47,77	46,61	46,45	45,54	49,75	47,93
	25%	51,27	51,85	50,28	50,78	51,93	51,68	51,68
Image Segmentation	5%	84,16	86,32	85,50	86,80	85,58	85,67	85,63
	10%	84,11	86,67	82,99	87,36	84,94	83,90	85,19
	15%	84,85	83,46	84,85	85,97	85,32	85,58	85,76
	20%	85,41	84,63	84,50	82,94	85,15	85,32	81,34
	25%	86,45	84,76	83,25	85,63	84,42	86,06	85,50
Indian Liver Patient	5%	69,23	70,78	69,23	67,00	68,22	67,50	70,07
	10%	67,93	71,21	70,52	67,59	70,34	67,07	67,07
	15%	70,34	67,07	67,59	66,21	66,21	68,62	68,62
	20%	68,14	67,80	68,47	67,97	68,81	67,97	71,53
	25%	67,55	64,53	65,16	65,74	66,09	66,91	68,27
Iris	5%	67,33	72,67	76,00	73,33	69,33	72,67	71,33
	10%	79,33	76,00	79,33	80,00	84,67	81,33	78,67
	15%	75,33	78,00	70,67	72,67	76,00	76,00	71,33
	20%	82,00	83,33	82,67	85,33	80,00	72,67	72,67
	25%	80,00	70,67	70,00	78,00	78,67	75,33	81,33
King Ropck vs King Pawn	5%	95,91	94,84	94,81	95,59	95,75	95,94	95,00
	10%	95,18	95,37	94,90	95,43	95,40	94,15	94,99
	15%	95,52	95,09	95,46	95,46	95,87	95,56	95,56
	20%	94,96	93,93	94,90	94,02	94,24	94,37	94,77
	25%	94,99	95,77	94,15	94,40	94,62	95,12	95,77

Leukemia Haslinger	5%	56,00	48,00	56,00	53,00	51,00	57,00	57,00
	10%	56,55	62,00	58,18	62,00	59,36	62,45	62,27
	15%	66,00	58,00	58,00	61,00	62,00	64,00	66,00
	20%	56,00	61,00	69,00	60,00	62,00	57,00	57,00
	25%	52,78	61,00	58,89	63,56	61,33	65,00	59,11
Mammographic Mass	5%	79,48	78,75	78,33	78,65	77,19	78,85	76,46
	10%	81,87	82,60	79,58	82,19	82,19	80,52	83,12
	15%	82,72	82,41	83,79	81,16	83,77	82,53	80,43
	20%	82,71	80,94	82,50	81,56	82,71	82,71	83,33
	25%	80,82	80,52	80,72	80,93	81,03	81,03	79,07
Mfeat Karhunen	5%	61,20	61,75	61,55	60,40	61,90	64,70	61,60
	10%	62,10	62,40	61,55	62,90	64,40	63,00	64,65
	15%	58,90	58,90	62,75	64,25	62,90	59,95	60,70
	20%	62,95	63,60	61,25	62,00	61,25	62,30	62,50
	25%	58,60	59,80	58,85	59,05	57,00	60,70	60,55
Mushroom	5%	82,12	83,62	82,94	82,99	82,86	82,92	82,63
	10%	81,98	82,10	82,96	83,16	81,96	81,46	82,03
	15%	83,92	83,17	83,17	83,12	85,41	83,02	84,45
	20%	83,13	83,26	83,83	83,14	82,23	84,05	83,09
	25%	83,61	82,58	83,11	82,52	83,02	83,45	82,30
Musk	5%	93,71	94,15	93,57	93,49	95,09	93,56	93,89
	10%	94,11	93,29	93,60	94,23	93,41	94,04	92,95
	15%	94,33	93,38	93,94	94,83	94,47	94,07	94,34
	20%	93,42	93,80	92,62	93,80	92,56	94,25	93,95
	25%	94,61	94,45	94,09	94,21	94,98	94,48	94,71
Ozone Level Detection	5%	96,65	96,18	96,17	95,98	96,29	96,25	96,57
	10%	96,50	95,94	95,94	95,87	96,42	96,02	96,22
	15%	96,09	95,78	95,82	95,70	96,41	96,37	96,06
	20%	96,41	96,84	96,57	96,33	96,45	95,94	95,66
	25%	96,33	96,53	96,49	96,57	96,14	96,25	96,49
Phishing	5%	81,11	80,15	82,06	81,99	82,51	80,81	80,45
	10%	85,09	81,47	83,68	82,51	82,37	80,66	84,93
	15%	81,47	82,80	81,10	83,61	81,03	82,36	80,14
	20%	81,92	80,23	80,60	80,96	80,30	80,67	79,04
	25%	81,63	81,85	81,10	80,14	80,43	82,66	81,18
Pima	5%	70,91	71,04	71,82	72,34	70,65	73,64	72,08
	10%	70,34	71,25	71,39	70,47	69,68	72,96	72,04
	15%	68,83	69,09	70,26	69,35	70,00	67,92	69,09
	20%	70,13	67,40	66,62	69,74	72,21	71,95	69,61
	25%	71,12	71,51	75,16	73,20	73,33	71,89	73,59

Planning Relax	5%	71,11	64,44	65,00	62,22	65,56	63,33	70,56
	10%	67,78	62,78	68,33	70,56	72,22	71,11	71,11
	15%	68,22	67,57	67,69	65,47	70,32	68,13	65,88
	20%	67,05	67,05	67,54	65,56	65,53	68,22	65,94
	25%	69,44	72,22	68,89	71,67	69,44	70,56	69,44
Seeds	5%	78,10	75,71	78,57	80,48	75,24	74,76	77,62
	10%	74,29	71,90	80,00	72,86	70,48	76,19	73,81
	15%	80,95	80,48	84,76	80,00	82,38	80,00	81,90
	20%	80,00	76,67	77,14	81,90	83,81	76,67	81,43
	25%	81,90	78,57	76,67	81,90	83,33	78,10	84,29
Semeion	5%	46,00	48,75	48,62	46,50	49,06	47,75	47,31
	10%	52,94	52,94	48,25	52,39	49,52	49,65	49,65
	15%	49,50	47,00	48,75	48,62	49,75	48,19	48,25
	20%	50,38	50,06	49,00	47,88	50,56	51,12	48,69
	25%	50,13	46,92	48,55	45,72	44,21	46,23	45,79
Spectf Heart	5%	66,06	70,73	67,58	71,34	72,13	67,86	67,02
	10%	75,17	72,40	73,21	71,52	72,33	73,22	70,75
	15%	69,25	70,95	69,18	71,49	68,10	67,66	64,02
	20%	73,33	69,51	74,79	72,50	73,07	71,85	71,44
	25%	68,97	74,15	72,97	73,57	71,81	67,82	68,70
Tic Tac Toe	5%	78,54	79,79	80,00	78,23	80,21	81,56	76,15
	10%	77,05	78,43	77,30	78,10	77,28	78,57	78,82
	15%	79,60	73,82	75,91	76,54	77,17	78,33	80,32
	20%	72,19	76,46	80,83	78,02	73,96	76,77	76,46
	25%	74,96	72,57	76,95	78,00	76,62	76,84	75,39
Twonorm	5%	86,16	84,91	85,55	85,51	85,26	85,49	85,11
	10%	85,08	84,73	84,95	84,35	84,86	85,05	84,42
	15%	85,14	84,87	84,14	84,18	84,63	85,49	85,60
	20%	84,32	83,82	85,05	84,98	84,31	84,35	84,65
	25%	83,86	84,01	84,33	84,28	84,47	84,51	84,90
Vehicle	5%	54,58	56,83	56,04	56,09	58,00	54,47	55,53
	10%	55,51	58,08	55,62	57,04	53,56	57,26	58,00
	15%	54,82	53,70	53,84	58,46	54,00	58,71	60,33
	20%	54,79	57,21	54,77	54,81	57,74	54,88	57,50
	25%	54,91	51,88	52,47	56,24	52,71	54,59	52,92
Waveform	5%	74,77	73,95	75,25	74,17	73,25	72,83	74,31
	10%	75,83	75,59	74,49	74,07	74,89	75,46	76,08
	15%	73,75	73,59	73,91	73,29	73,79	74,03	73,01
	20%	72,02	73,05	73,21	72,61	71,92	71,62	72,46
	25%	74,47	72,35	72,47	73,01	72,81	72,55	73,41

Wilt	5%	97,46	96,96	96,80	97,91	96,78	97,29	97,52
	10%	96,65	96,67	97,15	96,76	96,63	96,65	96,98
	15%	96,76	96,92	96,84	96,76	96,55	96,98	97,05
	20%	97,50	97,21	97,19	97,23	97,23	97,11	96,88
	25%	96,73	97,21	96,84	97,04	96,96	97,02	96,63

**Fonte:** O Próprio Autor (2018)

# APÊNDICE H – Resultado da Técnica *FlexCon-C1(v)* e Classificador RIPPER

Tabela 22 – Resultado da aplicação da Técnica Flexcon-C1(v) utilizando o classificador RIPPER.

Base de dados	Exemplos Ini. Rot.	Variação do <i>cr</i>						
		2	3	4	5	6	7	8
Balance Scale	5%	70,48	71,27	71,43	71,43	73,17	70,16	73,65
	10%	72,54	77,46	72,38	74,13	72,06	75,87	71,59
	15%	74,60	74,60	73,81	74,13	74,60	72,70	73,33
	20%	73,98	73,69	74,79	73,49	72,03	71,39	73,34
	25%	73,73	69,81	72,06	72,75	71,76	73,73	73,12
Blood Transfusion Service	5%	73,84	72,91	74,51	74,64	73,70	74,64	75,85
	10%	76,00	75,73	73,73	75,87	75,60	76,13	72,80
	15%	73,70	74,09	75,04	75,20	74,78	73,97	74,37
	20%	76,67	74,67	75,33	72,53	70,93	74,80	72,13
	25%	77,20	75,60	76,40	76,53	76,67	76,40	78,13
Bupa	5%	55,95	55,64	57,69	57,99	55,63	57,71	59,95
	10%	59,71	57,43	60,00	61,14	57,14	55,14	59,43
	15%	61,14	66,00	61,43	62,00	61,71	60,29	60,29
	20%	57,65	59,41	52,35	60,88	58,82	58,24	53,82
	25%	58,24	63,53	57,06	56,18	60,59	61,18	56,47
Car	5%	71,94	74,03	75,48	74,78	73,27	75,36	72,69
	10%	73,29	72,83	73,01	74,74	73,41	74,22	74,39
	15%	73,29	73,87	73,12	73,30	74,91	74,84	74,68
	20%	73,83	77,18	76,72	75,79	76,09	74,64	77,70
	25%	75,29	73,31	74,13	75,87	75,58	73,60	74,77
Cnae-9	5%	67,04	65,37	66,30	67,78	66,20	63,89	67,69
	10%	66,85	68,24	68,80	67,96	69,35	64,07	67,31
	15%	65,09	63,24	62,59	64,72	65,56	65,93	67,96
	20%	67,69	67,41	66,94	67,41	66,94	67,13	69,54
	25%	65,56	66,76	68,06	67,41	67,78	64,26	62,13
Connectionist Mines vs Rocks	5%	66,67	66,19	63,33	63,81	65,24	63,33	63,33
	10%	66,67	63,33	61,90	60,48	68,10	65,71	67,14
	15%	62,86	67,62	62,86	64,29	70,00	66,19	61,90
	20%	63,93	62,83	66,36	66,02	66,76	66,33	57,88
	25%	55,02	57,05	62,36	57,14	59,57	65,86	58,10

Flare	5%	66,00	70,71	65,07	65,50	68,36	68,64	68,43
	10%	68,54	67,88	69,20	67,74	67,96	71,17	66,13
	15%	63,23	61,51	64,81	64,30	65,80	64,02	65,53
	20%	68,11	66,97	66,26	67,82	64,67	65,47	64,72
	25%	68,47	68,05	69,99	65,72	69,96	68,68	67,46
Haberman	5%	75,81	70,97	73,23	74,19	74,84	73,23	74,84
	10%	74,19	72,26	73,23	70,65	72,90	73,87	70,65
	15%	72,67	72,00	73,00	71,00	71,00	71,67	74,67
	20%	71,98	69,10	71,04	70,33	70,75	69,46	71,96
	25%	72,00	72,00	71,00	70,33	68,00	71,67	66,00
Handwritten Digits	5%	88,16	88,26	88,34	88,18	88,05	88,33	88,36
	10%	88,11	88,44	87,42	88,45	87,70	87,90	88,16
	15%	88,87	88,65	88,14	88,57	88,25	88,12	89,01
	20%	87,40	86,61	87,23	87,41	87,50	87,71	87,48
	25%	88,05	87,95	87,61	87,18	87,37	87,05	87,72
Hill Valley	5%	49,96	51,70	51,52	53,92	52,10	49,63	52,93
	10%	53,98	52,25	54,33	54,07	50,83	53,00	53,00
	15%	55,07	54,83	54,07	50,66	53,00	50,58	52,00
	20%	45,95	47,77	46,61	46,45	45,54	49,75	47,93
	25%	51,27	51,85	50,28	50,78	51,93	51,68	51,68
Image Segmentation	5%	84,16	86,32	85,50	86,80	85,58	85,67	85,63
	10%	84,11	86,67	82,99	87,36	84,94	83,90	85,19
	15%	84,85	83,46	84,85	85,97	85,32	85,58	85,76
	20%	85,41	84,63	84,50	82,94	85,15	85,32	81,34
	25%	86,45	84,76	83,25	85,63	84,42	86,06	85,50
Indian Liver Patient	5%	69,23	70,78	69,23	67,00	68,22	67,50	70,07
	10%	67,93	71,21	70,52	67,59	70,34	67,07	67,07
	15%	70,34	67,07	67,59	66,21	66,21	68,62	68,62
	20%	68,14	67,80	68,47	67,97	68,81	67,97	71,53
	25%	67,55	64,53	65,16	65,74	66,09	66,91	68,27
Iris	5%	67,33	72,67	76,00	73,33	69,33	72,67	71,33
	10%	79,33	76,00	79,33	80,00	84,67	81,33	78,67
	15%	75,33	78,00	70,67	72,67	76,00	76,00	71,33
	20%	82,00	83,33	82,67	85,33	80,00	72,67	72,67
	25%	80,00	70,67	70,00	78,00	78,67	75,33	81,33
King Ropck vs King Pawn	5%	95,91	94,84	94,81	95,59	95,75	95,94	95,00
	10%	95,18	95,37	94,90	95,43	95,40	94,15	94,99
	15%	95,52	95,09	95,46	95,46	95,87	95,56	95,56
	20%	94,96	93,93	94,90	94,02	94,24	94,37	94,77
	25%	94,99	95,77	94,15	94,40	94,62	95,12	95,77

Leukemia Haslinger	5%	56,00	48,00	56,00	53,00	51,00	57,00	57,00
	10%	56,55	62,00	58,18	62,00	59,36	62,45	62,27
	15%	66,00	58,00	58,00	61,00	62,00	64,00	66,00
	20%	56,00	61,00	69,00	60,00	62,00	57,00	57,00
	25%	52,78	61,00	58,89	63,56	61,33	65,00	59,11
Mammographic Mass	5%	79,48	78,75	78,33	78,65	77,19	78,85	76,46
	10%	81,87	82,60	79,58	82,19	82,19	80,52	83,12
	15%	82,72	82,41	83,79	81,16	83,77	82,53	80,43
	20%	82,71	80,94	82,50	81,56	82,71	82,71	83,33
	25%	80,82	80,52	80,72	80,93	81,03	81,03	79,07
Mfeat Karhunen	5%	61,20	61,75	61,55	60,40	61,90	64,70	61,60
	10%	62,10	62,40	61,55	62,90	64,40	63,00	64,65
	15%	58,90	58,90	62,75	64,25	62,90	59,95	60,70
	20%	62,95	63,60	61,25	62,00	61,25	62,30	62,50
	25%	58,60	59,80	58,85	59,05	57,00	60,70	60,55
Mushroom	5%	82,12	83,62	82,94	82,99	82,86	82,92	82,63
	10%	81,98	82,10	82,96	83,16	81,96	81,46	82,03
	15%	83,92	83,17	83,17	83,12	85,41	83,02	84,45
	20%	83,13	83,26	83,83	83,14	82,23	84,05	83,09
	25%	83,61	82,58	83,11	82,52	83,02	83,45	82,30
Musk	5%	93,71	94,15	93,57	93,49	95,09	93,56	93,89
	10%	94,11	93,29	93,60	94,23	93,41	94,04	92,95
	15%	94,33	93,38	93,94	94,83	94,47	94,07	94,34
	20%	93,42	93,80	92,62	93,80	92,56	94,25	93,95
	25%	94,61	94,45	94,09	94,21	94,98	94,48	94,71
Ozone Level Detection	5%	96,65	96,18	96,17	95,98	96,29	96,25	96,57
	10%	96,50	95,94	95,94	95,87	96,42	96,02	96,22
	15%	96,09	95,78	95,82	95,70	96,41	96,37	96,06
	20%	96,41	96,84	96,57	96,33	96,45	95,94	95,66
	25%	96,33	96,53	96,49	96,57	96,14	96,25	96,49
Phishing	5%	81,11	80,15	82,06	81,99	82,51	80,81	80,45
	10%	85,09	81,47	83,68	82,51	82,37	80,66	84,93
	15%	81,47	82,80	81,10	83,61	81,03	82,36	80,14
	20%	81,92	80,23	80,60	80,96	80,30	80,67	79,04
	25%	81,63	81,85	81,10	80,14	80,43	82,66	81,18
Pima	5%	70,91	71,04	71,82	72,34	70,65	73,64	72,08
	10%	70,34	71,25	71,39	70,47	69,68	72,96	72,04
	15%	68,83	69,09	70,26	69,35	70,00	67,92	69,09
	20%	70,13	67,40	66,62	69,74	72,21	71,95	69,61
	25%	71,12	71,51	75,16	73,20	73,33	71,89	73,59



Planning Relax	5%	71,11	64,44	65,00	62,22	65,56	63,33	70,56
	10%	67,78	62,78	68,33	70,56	72,22	71,11	71,11
	15%	68,22	67,57	67,69	65,47	70,32	68,13	65,88
	20%	67,05	67,05	67,54	65,56	65,53	68,22	65,94
	25%	69,44	72,22	68,89	71,67	69,44	70,56	69,44
Seeds	5%	78,10	75,71	78,57	80,48	75,24	74,76	77,62
	10%	74,29	71,90	80,00	72,86	70,48	76,19	73,81
	15%	80,95	80,48	84,76	80,00	82,38	80,00	81,90
	20%	80,00	76,67	77,14	81,90	83,81	76,67	81,43
	25%	81,90	78,57	76,67	81,90	83,33	78,10	84,29
Semeion	5%	46,00	48,75	48,62	46,50	49,06	47,75	47,31
	10%	52,94	52,94	48,25	52,39	49,52	49,65	49,65
	15%	49,50	47,00	48,75	48,62	49,75	48,19	48,25
	20%	50,38	50,06	49,00	47,88	50,56	51,12	48,69
	25%	50,13	46,92	48,55	45,72	44,21	46,23	45,79
Spectf Heart	5%	66,06	70,73	67,58	71,34	72,13	67,86	67,02
	10%	75,17	72,40	73,21	71,52	72,33	73,22	70,75
	15%	69,25	70,95	69,18	71,49	68,10	67,66	64,02
	20%	73,33	69,51	74,79	72,50	73,07	71,85	71,44
	25%	68,97	74,15	72,97	73,57	71,81	67,82	68,70
Tic Tac Toe	5%	78,54	79,79	80,00	78,23	80,21	81,56	76,15
	10%	77,05	78,43	77,30	78,10	77,28	78,57	78,82
	15%	79,60	73,82	75,91	76,54	77,17	78,33	80,32
	20%	72,19	76,46	80,83	78,02	73,96	76,77	76,46
	25%	74,96	72,57	76,95	78,00	76,62	76,84	75,39
Twonorm	5%	86,16	84,91	85,55	85,51	85,26	85,49	85,11
	10%	85,08	84,73	84,95	84,35	84,86	85,05	84,42
	15%	85,14	84,87	84,14	84,18	84,63	85,49	85,60
	20%	84,32	83,82	85,05	84,98	84,31	84,35	84,65
	25%	83,86	84,01	84,33	84,28	84,47	84,51	84,90
Vehicle	5%	54,58	56,83	56,04	56,09	58,00	54,47	55,53
	10%	55,51	58,08	55,62	57,04	53,56	57,26	58,00
	15%	54,82	53,70	53,84	58,46	54,00	58,71	60,33
	20%	54,79	57,21	54,77	54,81	57,74	54,88	57,50
	25%	54,91	51,88	52,47	56,24	52,71	54,59	52,92
Waveform	5%	74,77	73,95	75,25	74,17	73,25	72,83	74,31
	10%	75,83	75,59	74,49	74,07	74,89	75,46	76,08
	15%	73,75	73,59	73,91	73,29	73,79	74,03	73,01
	20%	72,02	73,05	73,21	72,61	71,92	71,62	72,46
	25%	74,47	72,35	72,47	73,01	72,81	72,55	73,41

Wilt	5%	97,46	96,96	96,80	97,91	96,78	97,29	97,52
	10%	96,65	96,67	97,15	96,76	96,63	96,65	96,98
	15%	96,76	96,92	96,84	96,76	96,55	96,98	97,05
	20%	97,50	97,21	97,19	97,23	97,23	97,11	96,88
	25%	96,73	97,21	96,84	97,04	96,96	97,02	96,63

**Fonte:** O Próprio Autor (2018)

# APÊNDICE I – Resultado da Técnica *FlexCon-C2* e Classificador RIPPER

Tabela 23 – Resultado da aplicação da Técnica Flexcon-C2 utilizando o classificador RIPPER.

Base de dados	Exemplos Ini. Rot.	Variação do <i>cr</i>						
		2	3	4	5	6	7	8
Balance Scale	5%	70,48	71,27	71,43	71,43	73,17	70,16	73,65
	10%	72,54	77,46	72,38	74,13	72,06	75,87	71,59
	15%	74,60	74,60	73,81	74,13	74,60	72,70	73,33
	20%	73,98	73,69	74,79	73,49	72,03	71,39	73,34
	25%	73,73	69,81	72,06	72,75	71,76	73,73	73,12
Blood Transfusion Service	5%	73,84	72,91	74,51	74,64	73,70	74,64	75,85
	10%	76,00	75,73	73,73	75,87	75,60	76,13	72,80
	15%	73,70	74,09	75,04	75,20	74,78	73,97	74,37
	20%	76,67	74,67	75,33	72,53	70,93	74,80	72,13
	25%	77,20	75,60	76,40	76,53	76,67	76,40	78,13
Bupa	5%	55,95	55,64	57,69	57,99	55,63	57,71	59,95
	10%	59,71	57,43	60,00	61,14	57,14	55,14	59,43
	15%	61,14	66,00	61,43	62,00	61,71	60,29	60,29
	20%	57,65	59,41	52,35	60,88	58,82	58,24	53,82
	25%	58,24	63,53	57,06	56,18	60,59	61,18	56,47
Car	5%	71,94	74,03	75,48	74,78	73,27	75,36	72,69
	10%	73,29	72,83	73,01	74,74	73,41	74,22	74,39
	15%	73,29	73,87	73,12	73,30	74,91	74,84	74,68
	20%	73,83	77,18	76,72	75,79	76,09	74,64	77,70
	25%	75,29	73,31	74,13	75,87	75,58	73,60	74,77
Cnae-9	5%	67,04	65,37	66,30	67,78	66,20	63,89	67,69
	10%	66,85	68,24	68,80	67,96	69,35	64,07	67,31
	15%	65,09	63,24	62,59	64,72	65,56	65,93	67,96
	20%	67,69	67,41	66,94	67,41	66,94	67,13	69,54
	25%	65,56	66,76	68,06	67,41	67,78	64,26	62,13
Connectionist Mines vs Rocks	5%	66,67	66,19	63,33	63,81	65,24	63,33	63,33
	10%	66,67	63,33	61,90	60,48	68,10	65,71	67,14
	15%	62,86	67,62	62,86	64,29	70,00	66,19	61,90
	20%	63,93	62,83	66,36	66,02	66,76	66,33	57,88
	25%	55,02	57,05	62,36	57,14	59,57	65,86	58,10

Flare	5%	66,00	70,71	65,07	65,50	68,36	68,64	68,43
	10%	68,54	67,88	69,20	67,74	67,96	71,17	66,13
	15%	63,23	61,51	64,81	64,30	65,80	64,02	65,53
	20%	68,11	66,97	66,26	67,82	64,67	65,47	64,72
	25%	68,47	68,05	69,99	65,72	69,96	68,68	67,46
Haberman	5%	75,81	70,97	73,23	74,19	74,84	73,23	74,84
	10%	74,19	72,26	73,23	70,65	72,90	73,87	70,65
	15%	72,67	72,00	73,00	71,00	71,00	71,67	74,67
	20%	71,98	69,10	71,04	70,33	70,75	69,46	71,96
	25%	72,00	72,00	71,00	70,33	68,00	71,67	66,00
Handwritten Digits	5%	88,16	88,26	88,34	88,18	88,05	88,33	88,36
	10%	88,11	88,44	87,42	88,45	87,70	87,90	88,16
	15%	88,87	88,65	88,14	88,57	88,25	88,12	89,01
	20%	87,40	86,61	87,23	87,41	87,50	87,71	87,48
	25%	88,05	87,95	87,61	87,18	87,37	87,05	87,72
Hill Valley	5%	49,96	51,70	51,52	53,92	52,10	49,63	52,93
	10%	53,98	52,25	54,33	54,07	50,83	53,00	53,00
	15%	55,07	54,83	54,07	50,66	53,00	50,58	52,00
	20%	45,95	47,77	46,61	46,45	45,54	49,75	47,93
	25%	51,27	51,85	50,28	50,78	51,93	51,68	51,68
Image Segmentation	5%	84,16	86,32	85,50	86,80	85,58	85,67	85,63
	10%	84,11	86,67	82,99	87,36	84,94	83,90	85,19
	15%	84,85	83,46	84,85	85,97	85,32	85,58	85,76
	20%	85,41	84,63	84,50	82,94	85,15	85,32	81,34
	25%	86,45	84,76	83,25	85,63	84,42	86,06	85,50
Indian Liver Patient	5%	69,23	70,78	69,23	67,00	68,22	67,50	70,07
	10%	67,93	71,21	70,52	67,59	70,34	67,07	67,07
	15%	70,34	67,07	67,59	66,21	66,21	68,62	68,62
	20%	68,14	67,80	68,47	67,97	68,81	67,97	71,53
	25%	67,55	64,53	65,16	65,74	66,09	66,91	68,27
Iris	5%	67,33	72,67	76,00	73,33	69,33	72,67	71,33
	10%	79,33	76,00	79,33	80,00	84,67	81,33	78,67
	15%	75,33	78,00	70,67	72,67	76,00	76,00	71,33
	20%	82,00	83,33	82,67	85,33	80,00	72,67	72,67
	25%	80,00	70,67	70,00	78,00	78,67	75,33	81,33
King Ropck vs King Pawn	5%	95,91	94,84	94,81	95,59	95,75	95,94	95,00
	10%	95,18	95,37	94,90	95,43	95,40	94,15	94,99
	15%	95,52	95,09	95,46	95,46	95,87	95,56	95,56
	20%	94,96	93,93	94,90	94,02	94,24	94,37	94,77
	25%	94,99	95,77	94,15	94,40	94,62	95,12	95,77

Leukemia Haslinger	5%	56,00	48,00	56,00	53,00	51,00	57,00	57,00
	10%	56,55	62,00	58,18	62,00	59,36	62,45	62,27
	15%	66,00	58,00	58,00	61,00	62,00	64,00	66,00
	20%	56,00	61,00	69,00	60,00	62,00	57,00	57,00
	25%	52,78	61,00	58,89	63,56	61,33	65,00	59,11
Mammographic Mass	5%	79,48	78,75	78,33	78,65	77,19	78,85	76,46
	10%	81,87	82,60	79,58	82,19	82,19	80,52	83,12
	15%	82,72	82,41	83,79	81,16	83,77	82,53	80,43
	20%	82,71	80,94	82,50	81,56	82,71	82,71	83,33
	25%	80,82	80,52	80,72	80,93	81,03	81,03	79,07
Mfeat Karhunen	5%	61,20	61,75	61,55	60,40	61,90	64,70	61,60
	10%	62,10	62,40	61,55	62,90	64,40	63,00	64,65
	15%	58,90	58,90	62,75	64,25	62,90	59,95	60,70
	20%	62,95	63,60	61,25	62,00	61,25	62,30	62,50
	25%	58,60	59,80	58,85	59,05	57,00	60,70	60,55
Mushroom	5%	82,12	83,62	82,94	82,99	82,86	82,92	82,63
	10%	81,98	82,10	82,96	83,16	81,96	81,46	82,03
	15%	83,92	83,17	83,17	83,12	85,41	83,02	84,45
	20%	83,13	83,26	83,83	83,14	82,23	84,05	83,09
	25%	83,61	82,58	83,11	82,52	83,02	83,45	82,30
Musk	5%	93,71	94,15	93,57	93,49	95,09	93,56	93,89
	10%	94,11	93,29	93,60	94,23	93,41	94,04	92,95
	15%	94,33	93,38	93,94	94,83	94,47	94,07	94,34
	20%	93,42	93,80	92,62	93,80	92,56	94,25	93,95
	25%	94,61	94,45	94,09	94,21	94,98	94,48	94,71
Ozone Level Detection	5%	96,65	96,18	96,17	95,98	96,29	96,25	96,57
	10%	96,50	95,94	95,94	95,87	96,42	96,02	96,22
	15%	96,09	95,78	95,82	95,70	96,41	96,37	96,06
	20%	96,41	96,84	96,57	96,33	96,45	95,94	95,66
	25%	96,33	96,53	96,49	96,57	96,14	96,25	96,49
Phishing	5%	81,11	80,15	82,06	81,99	82,51	80,81	80,45
	10%	85,09	81,47	83,68	82,51	82,37	80,66	84,93
	15%	81,47	82,80	81,10	83,61	81,03	82,36	80,14
	20%	81,92	80,23	80,60	80,96	80,30	80,67	79,04
	25%	81,63	81,85	81,10	80,14	80,43	82,66	81,18
Pima	5%	70,91	71,04	71,82	72,34	70,65	73,64	72,08
	10%	70,34	71,25	71,39	70,47	69,68	72,96	72,04
	15%	68,83	69,09	70,26	69,35	70,00	67,92	69,09
	20%	70,13	67,40	66,62	69,74	72,21	71,95	69,61
	25%	71,12	71,51	75,16	73,20	73,33	71,89	73,59

Planning Relax	5%	71,11	64,44	65,00	62,22	65,56	63,33	70,56
	10%	67,78	62,78	68,33	70,56	72,22	71,11	71,11
	15%	68,22	67,57	67,69	65,47	70,32	68,13	65,88
	20%	67,05	67,05	67,54	65,56	65,53	68,22	65,94
	25%	69,44	72,22	68,89	71,67	69,44	70,56	69,44
Seeds	5%	78,10	75,71	78,57	80,48	75,24	74,76	77,62
	10%	74,29	71,90	80,00	72,86	70,48	76,19	73,81
	15%	80,95	80,48	84,76	80,00	82,38	80,00	81,90
	20%	80,00	76,67	77,14	81,90	83,81	76,67	81,43
	25%	81,90	78,57	76,67	81,90	83,33	78,10	84,29
Semeion	5%	46,00	48,75	48,62	46,50	49,06	47,75	47,31
	10%	52,94	52,94	48,25	52,39	49,52	49,65	49,65
	15%	49,50	47,00	48,75	48,62	49,75	48,19	48,25
	20%	50,38	50,06	49,00	47,88	50,56	51,12	48,69
	25%	50,13	46,92	48,55	45,72	44,21	46,23	45,79
Spectf Heart	5%	66,06	70,73	67,58	71,34	72,13	67,86	67,02
	10%	75,17	72,40	73,21	71,52	72,33	73,22	70,75
	15%	69,25	70,95	69,18	71,49	68,10	67,66	64,02
	20%	73,33	69,51	74,79	72,50	73,07	71,85	71,44
	25%	68,97	74,15	72,97	73,57	71,81	67,82	68,70
Tic Tac Toe	5%	78,54	79,79	80,00	78,23	80,21	81,56	76,15
	10%	77,05	78,43	77,30	78,10	77,28	78,57	78,82
	15%	79,60	73,82	75,91	76,54	77,17	78,33	80,32
	20%	72,19	76,46	80,83	78,02	73,96	76,77	76,46
	25%	74,96	72,57	76,95	78,00	76,62	76,84	75,39
Twonorm	5%	86,16	84,91	85,55	85,51	85,26	85,49	85,11
	10%	85,08	84,73	84,95	84,35	84,86	85,05	84,42
	15%	85,14	84,87	84,14	84,18	84,63	85,49	85,60
	20%	84,32	83,82	85,05	84,98	84,31	84,35	84,65
	25%	83,86	84,01	84,33	84,28	84,47	84,51	84,90
Vehicle	5%	54,58	56,83	56,04	56,09	58,00	54,47	55,53
	10%	55,51	58,08	55,62	57,04	53,56	57,26	58,00
	15%	54,82	53,70	53,84	58,46	54,00	58,71	60,33
	20%	54,79	57,21	54,77	54,81	57,74	54,88	57,50
	25%	54,91	51,88	52,47	56,24	52,71	54,59	52,92
Waveform	5%	74,77	73,95	75,25	74,17	73,25	72,83	74,31
	10%	75,83	75,59	74,49	74,07	74,89	75,46	76,08
	15%	73,75	73,59	73,91	73,29	73,79	74,03	73,01
	20%	72,02	73,05	73,21	72,61	71,92	71,62	72,46
	25%	74,47	72,35	72,47	73,01	72,81	72,55	73,41

Wilt	5%	97,46	96,96	96,80	97,91	96,78	97,29	97,52
	10%	96,65	96,67	97,15	96,76	96,63	96,65	96,98
	15%	96,76	96,92	96,84	96,76	96,55	96,98	97,05
	20%	97,50	97,21	97,19	97,23	97,23	97,11	96,88
	25%	96,73	97,21	96,84	97,04	96,96	97,02	96,63

**Fonte:** O Próprio Autor (2018)

# APÊNDICE J – Resultado da Técnica *FlexCon-C1(s)* e Classificador *k*-NN

Tabela 24 – Resultado da aplicações da Técnica *Flexcon-C1(s)* utilizando o classificador *k*-NN.

Base de dados	Exemplos Ini. Rot.	Variação do <i>cr</i>						
		2	3	4	5	6	7	8
Balance Scale	5%	85,81	85,32	86,29	82,74	84,35	84,84	82,90
	10%	81,75	84,29	85,71	83,65	82,22	86,67	83,02
	15%	84,06	81,64	80,66	85,32	82,72	84,03	82,41
	20%	83,17	81,75	84,76	82,06	80,32	82,54	80,32
	25%	83,68	81,62	84,49	82,74	84,32	83,36	81,93
Blood Transfusion Service	5%	72,80	71,07	72,53	74,27	74,27	72,13	73,07
	10%	76,52	76,37	75,05	75,44	75,58	76,65	75,58
	15%	75,70	77,58	76,77	76,65	75,97	77,05	77,05
	20%	77,20	78,27	77,20	77,07	76,53	75,33	76,00
	25%	78,80	78,27	77,60	76,67	75,87	78,13	75,73
Bupa	5%	56,82	56,23	52,70	57,13	56,23	57,06	54,47
	10%	57,08	59,08	55,68	56,00	61,66	55,61	58,02
	15%	55,07	60,91	57,70	59,40	55,63	54,47	56,16
	20%	55,41	58,01	59,45	58,31	58,02	56,27	57,71
	25%	56,46	62,02	62,06	63,78	59,71	57,70	63,46
Car	5%	79,54	79,71	79,65	80,29	79,31	79,42	79,25
	10%	79,71	80,75	80,58	81,16	80,29	80,06	81,51
	15%	80,58	81,45	79,19	81,10	79,88	80,46	80,00
	20%	77,63	80,17	79,25	79,65	80,63	79,88	78,33
	25%	79,82	80,12	78,32	79,19	79,94	78,26	78,95
Cnae-9	5%	68,70	64,81	66,48	64,72	64,44	69,17	67,96
	10%	60,00	64,17	63,61	64,35	61,11	66,02	65,28
	15%	68,06	63,80	64,07	65,00	65,09	67,59	63,52
	20%	65,28	65,74	63,98	63,52	63,52	67,13	62,31
	25%	63,80	62,04	60,93	58,06	64,07	62,59	61,57
Connectionist Mines vs Rocks	5%	65,00	58,55	63,98	66,79	63,55	59,10	59,93
	10%	67,62	62,86	66,19	63,33	70,00	65,71	61,90
	15%	66,27	64,73	65,45	70,95	61,86	64,91	68,09
	20%	72,88	72,19	69,33	68,88	70,64	65,71	67,38
	25%	71,43	68,57	65,71	70,00	70,95	69,05	71,90



Flare	5%	68,96	69,53	69,89	69,53	68,67	68,53	68,03
	10%	67,67	67,96	69,39	67,17	67,67	67,16	69,53
	15%	70,12	70,10	70,19	69,75	69,04	70,33	70,69
	20%	69,78	69,20	69,93	67,88	69,85	69,12	70,00
	25%	70,58	70,50	70,58	70,00	71,22	67,84	72,30
Haberman	5%	73,77	72,76	72,73	71,82	73,11	73,75	68,83
	10%	70,32	72,58	70,97	72,26	73,23	70,65	71,94
	15%	74,10	71,80	73,41	71,77	74,39	69,75	74,40
	20%	69,67	72,00	69,33	71,67	71,67	69,33	68,33
	25%	71,58	73,21	72,62	71,38	74,52	71,65	71,96
Handwritten Digits	5%	97,48	97,91	97,56	97,85	97,77	97,75	97,74
	10%	97,70	97,60	97,61	97,78	97,82	97,65	98,02
	15%	97,83	97,84	98,03	98,06	98,00	98,24	97,70
	20%	97,77	97,88	97,72	97,54	97,38	97,66	97,76
	25%	98,06	98,20	97,98	98,31	98,15	97,77	97,98
Hill Valley	5%	51,53	52,36	47,88	52,35	51,36	52,69	51,53
	10%	52,01	49,87	51,77	52,01	52,83	49,22	50,29
	15%	52,30	51,89	53,28	49,43	50,82	53,28	50,16
	20%	52,15	51,92	51,65	51,42	51,58	55,05	51,08
	25%	49,17	50,33	51,07	49,59	50,33	49,59	49,26
Image Segmentation	5%	89,61	90,39	90,30	89,96	91,17	89,26	90,43
	10%	88,70	89,09	90,78	89,31	91,04	89,96	90,43
	15%	91,30	91,56	91,73	91,56	91,60	90,43	91,43
	20%	90,26	89,83	88,79	90,61	91,00	90,78	91,43
	25%	92,29	90,91	90,87	90,74	90,82	90,56	90,39
Indian Liver Patient	5%	68,90	71,46	70,60	69,08	69,58	70,77	67,19
	10%	69,42	67,36	71,12	65,64	68,91	68,04	68,02
	15%	70,34	70,34	70,85	68,31	68,98	67,97	65,93
	20%	71,66	68,38	68,18	67,83	66,11	67,29	70,80
	25%	70,52	70,52	67,76	69,66	67,59	68,28	70,69
Iris	5%	90,67	94,00	90,00	88,67	88,67	89,33	95,33
	10%	92,00	90,67	92,67	88,00	90,00	90,67	89,33
	15%	94,00	94,00	88,67	94,67	87,33	90,67	90,00
	20%	93,33	90,67	95,33	90,67	93,33	91,33	93,33
	25%	92,00	86,67	89,33	90,67	92,67	92,67	91,33
King Ropck vs King Pawn	5%	85,10	84,63	83,41	85,92	86,39	84,29	84,70
	10%	84,69	84,26	85,26	83,88	84,69	84,10	84,44
	15%	85,48	84,94	86,13	84,91	84,92	85,63	86,17
	20%	83,38	83,13	82,94	83,91	84,54	83,41	83,88
	25%	84,72	84,97	85,03	84,91	84,47	85,47	84,91

Leukemia Haslinger	5%	81,00	89,00	85,00	84,00	80,00	80,00	92,00
	10%	82,00	76,00	76,00	77,00	76,00	82,00	80,00
	15%	70,00	68,00	66,00	76,00	74,00	73,00	72,00
	20%	83,00	90,00	89,00	91,00	82,00	89,00	87,00
	25%	94,00	89,00	84,00	87,00	87,00	82,00	84,00
Mammographic Mass	5%	76,15	77,81	78,54	78,85	77,08	77,92	77,50
	10%	81,16	81,05	80,11	78,86	81,77	78,55	81,14
	15%	81,88	81,67	81,57	79,91	79,81	80,53	81,99
	20%	78,54	77,81	76,75	79,59	80,01	76,76	77,07
	25%	77,81	79,48	78,65	78,75	78,75	78,96	80,21
Mfeat Karhunen	5%	89,60	89,45	89,95	88,85	91,20	89,55	89,85
	10%	87,55	88,75	87,60	88,00	88,00	87,85	89,15
	15%	87,95	88,75	88,40	88,90	87,90	88,40	88,10
	20%	88,70	88,10	89,25	88,35	88,75	88,05	89,40
	25%	88,55	88,30	88,15	88,75	88,65	88,15	89,15
Mushroom	5%	84,98	84,70	84,03	84,06	84,05	83,10	84,12
	10%	83,37	83,27	82,95	82,61	83,09	82,50	83,10
	15%	84,23	84,37	83,51	83,89	84,60	84,75	83,66
	20%	83,74	83,68	83,61	83,45	84,10	83,73	83,55
	25%	83,31	82,50	82,55	83,74	83,78	83,13	83,37
Musk	5%	95,72	95,45	95,51	95,38	95,60	95,65	95,98
	10%	95,85	95,91	95,76	96,09	96,00	95,71	95,39
	15%	95,67	96,05	95,42	95,76	95,89	95,56	95,92
	20%	95,03	94,83	95,24	95,22	95,36	94,66	95,22
	25%	94,33	95,08	94,67	94,45	94,67	94,42	95,02
Ozone Level Detection	5%	97,15	97,00	97,15	96,92	97,27	96,96	97,15
	10%	97,00	96,96	96,81	97,04	97,08	96,69	96,92
	15%	97,11	96,96	97,23	97,08	96,92	97,00	97,08
	20%	96,66	96,90	96,54	96,62	96,90	96,70	96,66
	25%	97,24	97,13	97,20	97,05	96,97	97,13	97,17
Phishing	5%	84,44	85,11	84,30	84,74	82,59	83,63	81,70
	10%	81,62	82,81	80,89	81,18	84,74	85,33	85,04
	15%	83,41	83,50	84,01	83,93	85,19	83,36	83,50
	20%	83,76	83,25	82,81	85,09	85,41	83,77	84,74
	25%	82,35	81,62	81,76	82,21	81,41	80,30	80,67
Pima	5%	69,54	73,59	71,89	71,62	69,80	70,58	71,11
	10%	68,57	68,83	69,74	71,30	70,52	69,35	69,61
	15%	71,95	70,26	70,13	72,86	70,13	72,99	71,69
	20%	68,10	69,40	69,94	67,84	68,63	69,68	68,76
	25%	68,96	72,21	73,77	70,91	69,87	69,22	69,35

Planning Relax	5%	72,78	68,89	70,00	65,00	65,00	66,67	68,33
	10%	65,50	66,52	70,82	62,11	61,73	62,37	66,52
	15%	68,89	67,22	67,78	67,78	68,33	63,89	68,33
	20%	67,08	62,75	66,05	68,71	64,36	68,16	69,21
	25%	71,11	70,00	71,11	65,56	70,56	68,33	71,11
Seeds	5%	86,67	87,14	87,62	89,05	87,62	89,52	90,95
	10%	93,33	92,86	91,90	90,95	92,38	89,52	92,38
	15%	94,76	90,48	92,38	95,24	92,38	93,81	92,86
	20%	83,33	83,81	82,86	82,38	82,86	80,48	85,71
	25%	89,05	91,90	92,86	90,95	90,48	92,86	91,43
Semeion	5%	80,37	77,92	78,31	79,93	79,74	79,93	78,49
	10%	79,29	78,61	79,11	78,86	78,79	77,27	78,15
	15%	78,69	78,81	80,81	80,31	79,44	80,06	79,81
	20%	79,56	76,37	78,18	78,12	78,37	78,62	77,44
	25%	75,80	77,81	74,48	76,98	76,42	75,17	76,68
Spectf Heart	5%	66,39	66,94	66,39	67,22	66,94	70,00	68,06
	10%	69,13	67,61	66,57	69,40	68,56	67,06	67,65
	15%	73,03	72,20	76,53	70,76	73,29	74,51	74,78
	20%	71,87	71,53	74,15	70,40	73,08	69,80	67,53
	25%	77,07	73,50	76,46	81,12	72,39	77,37	75,29
Tic Tac Toe	5%	83,44	84,06	83,54	84,58	83,13	82,81	84,27
	10%	82,49	80,30	81,96	80,41	81,04	81,76	82,58
	15%	79,79	81,58	81,68	80,63	80,74	82,00	81,05
	20%	81,37	81,03	78,56	80,74	80,51	82,29	82,91
	25%	82,30	81,79	82,50	82,93	80,84	82,09	84,93
Twonorm	5%	98,19	98,19	98,03	98,09	98,22	98,07	98,19
	10%	97,18	97,18	97,08	97,20	97,09	97,16	97,49
	15%	97,26	97,47	97,61	97,41	97,76	97,57	97,70
	20%	96,70	96,80	96,86	96,66	96,64	96,81	96,72
	25%	97,38	97,28	97,20	97,11	96,96	97,28	97,43
Vehicle	5%	58,00	57,06	53,83	56,06	59,04	58,61	61,57
	10%	62,82	59,76	56,24	60,12	58,00	56,12	57,06
	15%	57,79	58,14	59,88	58,26	61,16	62,79	58,37
	20%	57,46	55,53	56,90	57,65	61,08	60,05	58,73
	25%	58,30	59,75	55,60	58,47	59,98	60,11	57,34
Waveform	5%	81,54	82,54	82,32	82,38	82,44	83,30	82,04
	10%	80,64	81,76	81,12	81,04	81,22	81,52	81,96
	15%	83,04	83,70	83,94	82,76	82,84	83,04	84,20
	20%	82,54	83,26	82,54	83,62	83,26	82,10	82,18
	25%	83,54	82,70	83,36	82,68	83,62	82,46	83,26

Wilt	5%	94,73	94,73	95,27	95,02	94,67	94,94	94,69
	10%	94,62	94,23	94,42	94,00	94,54	94,48	94,77
	15%	94,77	94,88	94,96	94,40	93,99	94,71	94,90
	20%	94,17	94,59	94,38	94,61	94,26	94,69	94,67
	25%	94,77	94,42	94,63	94,63	94,40	94,88	94,34

**Fonte:** O Próprio Autor (2018)

# APÊNDICE K – Resultado da Técnica *FlexCon-C1(v)* e Classificador *k*-NN

Tabela 25 – Resultado da aplicação da Técnica *Flexcon-C1(v)* utilizando o classificador *k*-NN.

Base de dados	Exemplos Ini. Rot.	Variação do <i>cr</i>						
		2	3	4	5	6	7	8
Balance Scale	5%	85,81	85,32	86,29	82,74	84,35	84,84	82,90
	10%	81,75	84,29	85,71	83,65	82,22	86,67	83,02
	15%	84,06	81,64	80,66	85,32	82,72	84,03	82,41
	20%	83,17	81,75	84,76	82,06	80,32	82,54	80,32
	25%	83,68	81,62	84,49	82,74	84,32	83,36	81,93
Blood Transfusion Service	5%	72,80	71,07	72,53	74,27	74,27	72,13	73,07
	10%	76,52	76,37	75,05	75,44	75,58	76,65	75,58
	15%	75,70	77,58	76,77	76,65	75,97	77,05	77,05
	20%	77,20	78,27	77,20	77,07	76,53	75,33	76,00
	25%	78,80	78,27	77,60	76,67	75,87	78,13	75,73
Bupa	5%	56,82	56,23	52,70	57,13	56,23	57,06	54,47
	10%	57,08	59,08	55,68	56,00	61,66	55,61	58,02
	15%	55,07	60,91	57,70	59,40	55,63	54,47	56,16
	20%	55,41	58,01	59,45	58,31	58,02	56,27	57,71
	25%	56,46	62,02	62,06	63,78	59,71	57,70	63,46
Car	5%	79,54	79,71	79,65	80,29	79,31	79,42	79,25
	10%	79,71	80,75	80,58	81,16	80,29	80,06	81,51
	15%	80,58	81,45	79,19	81,10	79,88	80,46	80,00
	20%	77,63	80,17	79,25	79,65	80,63	79,88	78,33
	25%	79,82	80,12	78,32	79,19	79,94	78,26	78,95
Cnae-9	5%	68,70	64,81	66,48	64,72	64,44	69,17	67,96
	10%	60,00	64,17	63,61	64,35	61,11	66,02	65,28
	15%	68,06	63,80	64,07	65,00	65,09	67,59	63,52
	20%	65,28	65,74	63,98	63,52	63,52	67,13	62,31
	25%	63,80	62,04	60,93	58,06	64,07	62,59	61,57
Connectionist Mines vs Rocks	5%	65,00	58,55	63,98	66,79	63,55	59,10	59,93
	10%	67,62	62,86	66,19	63,33	70,00	65,71	61,90
	15%	66,27	64,73	65,45	70,95	61,86	64,91	68,09
	20%	72,88	72,19	69,33	68,88	70,64	65,71	67,38
	25%	71,43	68,57	65,71	70,00	70,95	69,05	71,90

Flare	5%	68,96	69,53	69,89	69,53	68,67	68,53	68,03
	10%	67,67	67,96	69,39	67,17	67,67	67,16	69,53
	15%	70,12	70,10	70,19	69,75	69,04	70,33	70,69
	20%	69,78	69,20	69,93	67,88	69,85	69,12	70,00
	25%	70,58	70,50	70,58	70,00	71,22	67,84	72,30
Haberman	5%	73,77	72,76	72,73	71,82	73,11	73,75	68,83
	10%	70,32	72,58	70,97	72,26	73,23	70,65	71,94
	15%	74,10	71,80	73,41	71,77	74,39	69,75	74,40
	20%	69,67	72,00	69,33	71,67	71,67	69,33	68,33
	25%	71,58	73,21	72,62	71,38	74,52	71,65	71,96
Handwritten Digits	5%	97,48	97,91	97,56	97,85	97,77	97,75	97,74
	10%	97,70	97,60	97,61	97,78	97,82	97,65	98,02
	15%	97,83	97,84	98,03	98,06	98,00	98,24	97,70
	20%	97,77	97,88	97,72	97,54	97,38	97,66	97,76
	25%	98,06	98,20	97,98	98,31	98,15	97,77	97,98
Hill Valley	5%	51,53	52,36	47,88	52,35	51,36	52,69	51,53
	10%	52,01	49,87	51,77	52,01	52,83	49,22	50,29
	15%	52,30	51,89	53,28	49,43	50,82	53,28	50,16
	20%	52,15	51,92	51,65	51,42	51,58	55,05	51,08
	25%	49,17	50,33	51,07	49,59	50,33	49,59	49,26
Image Segmentation	5%	89,61	90,39	90,30	89,96	91,17	89,26	90,43
	10%	88,70	89,09	90,78	89,31	91,04	89,96	90,43
	15%	91,30	91,56	91,73	91,56	91,60	90,43	91,43
	20%	90,26	89,83	88,79	90,61	91,00	90,78	91,43
	25%	92,29	90,91	90,87	90,74	90,82	90,56	90,39
Indian Liver Patient	5%	68,90	71,46	70,60	69,08	69,58	70,77	67,19
	10%	69,42	67,36	71,12	65,64	68,91	68,04	68,02
	15%	70,34	70,34	70,85	68,31	68,98	67,97	65,93
	20%	71,66	68,38	68,18	67,83	66,11	67,29	70,80
	25%	70,52	70,52	67,76	69,66	67,59	68,28	70,69
Iris	5%	90,67	94,00	90,00	88,67	88,67	89,33	95,33
	10%	92,00	90,67	92,67	88,00	90,00	90,67	89,33
	15%	94,00	94,00	88,67	94,67	87,33	90,67	90,00
	20%	93,33	90,67	95,33	90,67	93,33	91,33	93,33
	25%	92,00	86,67	89,33	90,67	92,67	92,67	91,33
King Ropck vs King Pawn	5%	85,10	84,63	83,41	85,92	86,39	84,29	84,70
	10%	84,69	84,26	85,26	83,88	84,69	84,10	84,44
	15%	85,48	84,94	86,13	84,91	84,92	85,63	86,17
	20%	83,38	83,13	82,94	83,91	84,54	83,41	83,88
	25%	84,72	84,97	85,03	84,91	84,47	85,47	84,91

Leukemia Haslinger	5%	81,00	89,00	85,00	84,00	80,00	80,00	92,00
	10%	82,00	76,00	76,00	77,00	76,00	82,00	80,00
	15%	70,00	68,00	66,00	76,00	74,00	73,00	72,00
	20%	83,00	90,00	89,00	91,00	82,00	89,00	87,00
	25%	94,00	89,00	84,00	87,00	87,00	82,00	84,00
Mammographic Mass	5%	76,15	77,81	78,54	78,85	77,08	77,92	77,50
	10%	81,16	81,05	80,11	78,86	81,77	78,55	81,14
	15%	81,88	81,67	81,57	79,91	79,81	80,53	81,99
	20%	78,54	77,81	76,75	79,59	80,01	76,76	77,07
	25%	77,81	79,48	78,65	78,75	78,75	78,96	80,21
Mfeat Karhunen	5%	89,60	89,45	89,95	88,85	91,20	89,55	89,85
	10%	87,55	88,75	87,60	88,00	88,00	87,85	89,15
	15%	87,95	88,75	88,40	88,90	87,90	88,40	88,10
	20%	88,70	88,10	89,25	88,35	88,75	88,05	89,40
	25%	88,55	88,30	88,15	88,75	88,65	88,15	89,15
Mushroom	5%	84,98	84,70	84,03	84,06	84,05	83,10	84,12
	10%	83,37	83,27	82,95	82,61	83,09	82,50	83,10
	15%	84,23	84,37	83,51	83,89	84,60	84,75	83,66
	20%	83,74	83,68	83,61	83,45	84,10	83,73	83,55
	25%	83,31	82,50	82,55	83,74	83,78	83,13	83,37
Musk	5%	95,72	95,45	95,51	95,38	95,60	95,65	95,98
	10%	95,85	95,91	95,76	96,09	96,00	95,71	95,39
	15%	95,67	96,05	95,42	95,76	95,89	95,56	95,92
	20%	95,03	94,83	95,24	95,22	95,36	94,66	95,22
	25%	94,33	95,08	94,67	94,45	94,67	94,42	95,02
Ozone Level Detection	5%	97,15	97,00	97,15	96,92	97,27	96,96	97,15
	10%	97,00	96,96	96,81	97,04	97,08	96,69	96,92
	15%	97,11	96,96	97,23	97,08	96,92	97,00	97,08
	20%	96,66	96,90	96,54	96,62	96,90	96,70	96,66
	25%	97,24	97,13	97,20	97,05	96,97	97,13	97,17
Phishing	5%	84,44	85,11	84,30	84,74	82,59	83,63	81,70
	10%	81,62	82,81	80,89	81,18	84,74	85,33	85,04
	15%	83,41	83,50	84,01	83,93	85,19	83,36	83,50
	20%	83,76	83,25	82,81	85,09	85,41	83,77	84,74
	25%	82,35	81,62	81,76	82,21	81,41	80,30	80,67
Pima	5%	69,54	73,59	71,89	71,62	69,80	70,58	71,11
	10%	68,57	68,83	69,74	71,30	70,52	69,35	69,61
	15%	71,95	70,26	70,13	72,86	70,13	72,99	71,69
	20%	68,10	69,40	69,94	67,84	68,63	69,68	68,76
	25%	68,96	72,21	73,77	70,91	69,87	69,22	69,35

Planning Relax	5%	72,78	68,89	70,00	65,00	65,00	66,67	68,33
	10%	65,50	66,52	70,82	62,11	61,73	62,37	66,52
	15%	68,89	67,22	67,78	67,78	68,33	63,89	68,33
	20%	67,08	62,75	66,05	68,71	64,36	68,16	69,21
	25%	71,11	70,00	71,11	65,56	70,56	68,33	71,11
Seeds	5%	86,67	87,14	87,62	89,05	87,62	89,52	90,95
	10%	93,33	92,86	91,90	90,95	92,38	89,52	92,38
	15%	94,76	90,48	92,38	95,24	92,38	93,81	92,86
	20%	83,33	83,81	82,86	82,38	82,86	80,48	85,71
	25%	89,05	91,90	92,86	90,95	90,48	92,86	91,43
Semeion	5%	80,37	77,92	78,31	79,93	79,74	79,93	78,49
	10%	79,29	78,61	79,11	78,86	78,79	77,27	78,15
	15%	78,69	78,81	80,81	80,31	79,44	80,06	79,81
	20%	79,56	76,37	78,18	78,12	78,37	78,62	77,44
	25%	75,80	77,81	74,48	76,98	76,42	75,17	76,68
Spectf Heart	5%	66,39	66,94	66,39	67,22	66,94	70,00	68,06
	10%	69,13	67,61	66,57	69,40	68,56	67,06	67,65
	15%	73,03	72,20	76,53	70,76	73,29	74,51	74,78
	20%	71,87	71,53	74,15	70,40	73,08	69,80	67,53
	25%	77,07	73,50	76,46	81,12	72,39	77,37	75,29
Tic Tac Toe	5%	83,44	84,06	83,54	84,58	83,13	82,81	84,27
	10%	82,49	80,30	81,96	80,41	81,04	81,76	82,58
	15%	79,79	81,58	81,68	80,63	80,74	82,00	81,05
	20%	81,37	81,03	78,56	80,74	80,51	82,29	82,91
	25%	82,30	81,79	82,50	82,93	80,84	82,09	84,93
Twonorm	5%	98,19	98,19	98,03	98,09	98,22	98,07	98,19
	10%	97,18	97,18	97,08	97,20	97,09	97,16	97,49
	15%	97,26	97,47	97,61	97,41	97,76	97,57	97,70
	20%	96,70	96,80	96,86	96,66	96,64	96,81	96,72
	25%	97,38	97,28	97,20	97,11	96,96	97,28	97,43
Vehicle	5%	58,00	57,06	53,83	56,06	59,04	58,61	61,57
	10%	62,82	59,76	56,24	60,12	58,00	56,12	57,06
	15%	57,79	58,14	59,88	58,26	61,16	62,79	58,37
	20%	57,46	55,53	56,90	57,65	61,08	60,05	58,73
	25%	58,30	59,75	55,60	58,47	59,98	60,11	57,34
Waveform	5%	81,54	82,54	82,32	82,38	82,44	83,30	82,04
	10%	80,64	81,76	81,12	81,04	81,22	81,52	81,96
	15%	83,04	83,70	83,94	82,76	82,84	83,04	84,20
	20%	82,54	83,26	82,54	83,62	83,26	82,10	82,18
	25%	83,54	82,70	83,36	82,68	83,62	82,46	83,26



Wilt	5%	94,73	94,73	95,27	95,02	94,67	94,94	94,69
	10%	94,62	94,23	94,42	94,00	94,54	94,48	94,77
	15%	94,77	94,88	94,96	94,40	93,99	94,71	94,90
	20%	94,17	94,59	94,38	94,61	94,26	94,69	94,67
	25%	94,77	94,42	94,63	94,63	94,40	94,88	94,34

**Fonte:** O Próprio Autor (2018)

# APÊNDICE L – Resultado da Técnica *FlexCon-C2* e Classificador *k*-NN

Tabela 26 – Resultado da aplicação da Técnica *FlexCon-C2* utilizando o classificador *k*-NN.

Base de dados	Exemplos Ini. Rot.	Variação do <i>cr</i>						
		2	3	4	5	6	7	8
Balance Scale	5%	85,81	85,32	86,29	82,74	84,35	84,84	82,90
	10%	81,75	84,29	85,71	83,65	82,22	86,67	83,02
	15%	84,06	81,64	80,66	85,32	82,72	84,03	82,41
	20%	83,17	81,75	84,76	82,06	80,32	82,54	80,32
	25%	83,68	81,62	84,49	82,74	84,32	83,36	81,93
Blood Transfusion Service	5%	72,80	71,07	72,53	73,73	74,27	72,27	73,07
	10%	76,38	76,24	75,05	75,58	75,31	76,38	75,58
	15%	75,70	77,58	76,64	76,65	75,97	77,05	77,05
	20%	77,07	77,73	77,20	77,07	76,27	75,47	75,73
	25%	78,00	78,27	77,73	76,67	76,00	78,53	76,00
Bupa	5%	55,97	56,56	53,61	59,16	53,59	55,87	55,05
	10%	56,50	58,52	55,37	56,31	60,52	56,19	59,13
	15%	55,66	62,92	59,13	55,93	55,34	55,03	57,05
	20%	56,87	59,74	60,63	58,88	56,23	55,10	56,85
	25%	56,18	61,74	61,76	61,20	58,55	58,01	63,19
Car	5%	79,54	79,71	79,65	80,29	79,31	79,42	79,25
	10%	79,71	80,75	80,58	81,16	80,29	80,06	81,51
	15%	80,58	81,45	79,19	81,10	79,88	80,46	80,00
	20%	77,63	80,17	79,25	79,65	80,63	79,88	78,33
	25%	79,82	80,12	78,32	79,19	79,94	78,26	78,95
Cnae-9	5%	68,61	65,93	65,74	65,28	64,35	69,07	68,06
	10%	61,11	65,56	63,98	64,26	61,67	65,65	64,81
	15%	68,15	65,56	64,07	65,09	65,46	67,31	63,80
	20%	64,91	65,46	63,98	63,15	65,56	68,15	62,22
	25%	63,70	61,67	61,67	58,80	63,70	62,31	62,50
Connectionist Mines vs Rocks	5%	64,52	59,05	63,50	66,79	61,12	58,12	60,43
	10%	67,62	62,86	66,19	62,38	69,52	68,10	61,43
	15%	66,77	63,32	65,95	70,00	60,41	65,36	67,59
	20%	73,36	71,69	69,36	69,83	69,67	66,71	66,86
	25%	70,48	68,57	64,76	70,95	70,48	66,19	72,86

Flare	5%	68,96	69,53	69,96	69,60	68,74	68,53	68,03
	10%	67,67	68,03	69,39	67,17	67,67	67,24	69,53
	15%	70,12	70,10	70,19	69,82	68,89	70,33	70,55
	20%	69,78	69,20	69,93	67,88	69,71	69,12	69,85
	25%	70,58	70,50	70,58	70,07	71,22	67,84	72,30
Haberman	5%	74,11	72,76	72,73	71,82	73,11	74,09	68,49
	10%	70,32	72,26	71,29	72,58	73,23	70,65	71,94
	15%	74,10	72,12	73,41	71,77	74,39	69,75	74,40
	20%	69,67	72,00	69,33	71,67	71,67	69,33	68,33
	25%	71,25	71,94	72,96	71,06	74,52	71,65	71,96
Handwritten Digits	5%	97,48	97,91	97,56	97,85	97,77	97,75	97,74
	10%	97,70	97,60	97,61	97,78	97,82	97,65	98,02
	15%	97,83	97,84	98,03	98,06	98,00	98,24	97,70
	20%	97,77	97,88	97,72	97,54	97,38	97,66	97,76
	25%	98,06	98,20	97,98	98,31	98,15	97,77	97,98
Hill Valley	5%	51,53	52,28	47,88	52,35	51,27	52,77	51,44
	10%	52,01	49,87	51,94	52,01	52,83	49,13	50,37
	15%	52,30	51,89	53,28	49,51	50,82	53,28	50,16
	20%	52,15	51,92	51,73	51,42	51,74	55,05	51,08
	25%	49,17	50,33	51,07	49,59	50,41	49,67	49,26
Image Segmentation	5%	89,61	90,39	90,00	90,09	91,17	89,22	90,52
	10%	88,83	89,13	90,78	89,35	91,26	90,00	90,56
	15%	91,43	91,56	91,86	91,56	91,60	90,52	91,43
	20%	90,26	90,09	88,87	90,91	90,91	90,91	91,43
	25%	92,34	90,91	90,87	90,69	90,82	90,61	90,61
Indian Liver Patient	5%	67,88	71,46	70,60	68,74	69,41	71,29	66,85
	10%	69,77	67,71	70,61	65,64	68,74	68,38	68,88
	15%	70,17	70,00	70,85	68,81	68,98	68,47	66,61
	20%	71,14	67,86	68,35	67,66	66,29	67,11	69,93
	25%	70,34	70,17	67,93	69,83	67,76	68,10	70,52
Iris	5%	90,67	94,00	91,33	90,00	88,67	90,00	95,33
	10%	92,00	90,67	92,67	88,00	90,00	90,67	89,33
	15%	94,00	94,00	88,67	94,67	86,67	91,33	89,33
	20%	93,33	90,67	95,33	90,67	93,33	91,33	93,33
	25%	92,00	86,67	89,33	90,67	92,67	92,67	91,33
King Ropck vs King Pawn	5%	85,10	84,63	83,41	85,92	86,39	84,29	84,70
	10%	84,69	84,26	85,26	83,88	84,69	84,10	84,44
	15%	85,48	84,94	86,13	84,91	84,92	85,63	86,17
	20%	83,38	83,13	82,94	83,91	84,54	83,41	83,88
	25%	84,72	84,97	85,03	84,91	84,47	85,47	84,91

Leukemia Haslinger	5%	81,00	87,00	85,00	84,00	80,00	80,00	90,00
	10%	80,00	76,00	75,00	77,00	77,00	81,00	81,00
	15%	70,00	68,00	65,00	74,00	74,00	72,00	73,00
	20%	83,00	90,00	89,00	91,00	82,00	89,00	88,00
	25%	92,00	89,00	86,00	87,00	87,00	83,00	83,00
Mammographic Mass	5%	76,15	77,50	78,65	78,85	76,46	78,02	77,40
	10%	81,16	81,05	80,11	78,86	81,77	78,86	81,14
	15%	81,88	82,19	81,57	80,01	79,81	80,01	82,09
	20%	78,65	77,81	76,75	79,38	79,80	76,55	77,28
	25%	77,92	79,48	78,54	78,75	78,44	78,96	80,10
Mfeat Karhunen	5%	89,65	89,50	90,00	88,90	91,30	89,55	89,70
	10%	87,55	88,75	87,65	87,95	87,95	87,95	89,20
	15%	88,00	88,80	88,40	88,85	87,65	88,35	88,05
	20%	88,60	88,15	89,30	88,40	88,45	88,10	89,50
	25%	88,15	87,95	88,30	88,90	88,85	88,10	89,15
Mushroom	5%	84,98	84,70	84,03	84,06	84,05	83,10	84,12
	10%	83,37	83,27	82,95	82,61	83,09	82,50	83,10
	15%	84,23	84,37	83,51	83,89	84,60	84,75	83,66
	20%	83,74	83,68	83,61	83,45	84,10	83,73	83,55
	25%	83,31	82,50	82,55	83,74	83,78	83,13	83,37
Musk	5%	95,74	95,45	95,51	95,36	95,63	95,65	95,97
	10%	95,88	95,94	95,70	96,08	96,00	95,70	95,38
	15%	95,55	96,00	95,42	95,76	95,89	95,52	95,94
	20%	95,03	94,84	95,24	95,22	95,44	94,74	95,22
	25%	94,35	95,06	94,68	94,48	94,70	94,44	95,00
Ozone Level Detection	5%	97,15	97,00	97,15	96,96	97,27	96,96	97,11
	10%	97,00	96,96	96,84	97,04	97,08	96,69	96,92
	15%	97,11	96,96	97,23	97,04	96,96	97,00	97,08
	20%	96,70	96,90	96,54	96,62	96,90	96,70	96,66
	25%	97,24	97,09	97,24	97,05	96,97	97,13	97,13
Phishing	5%	84,44	85,11	84,30	84,74	82,59	83,63	81,70
	10%	81,62	82,81	80,89	81,18	84,74	85,33	85,04
	15%	83,41	83,50	84,01	83,93	85,19	83,36	83,50
	20%	83,76	83,25	82,81	85,09	85,41	83,77	84,74
	25%	82,35	81,62	81,76	82,21	81,41	80,30	80,67
Pima	5%	71,10	73,07	73,58	72,79	69,67	70,97	71,25
	10%	68,31	69,35	70,00	71,30	70,78	69,09	69,74
	15%	73,77	70,52	69,74	73,25	70,39	72,99	71,56
	20%	68,50	68,75	69,03	68,10	69,94	68,63	69,16
	25%	69,35	72,47	73,38	71,04	72,47	68,83	69,87

Planning Relax	5%	72,22	70,00	69,44	65,00	64,44	67,22	68,89
	10%	66,05	65,44	70,26	62,11	60,18	62,37	67,08
	15%	69,44	67,78	68,33	67,78	68,89	63,89	67,78
	20%	67,63	63,27	66,61	68,71	63,80	68,68	69,21
	25%	71,11	69,44	71,11	65,56	71,11	67,78	71,11
Seeds	5%	86,19	86,19	87,14	89,05	87,14	89,05	90,95
	10%	93,81	92,86	90,48	90,00	91,90	89,52	92,86
	15%	94,29	90,00	92,86	95,24	92,38	93,81	92,38
	20%	83,33	83,33	82,86	82,38	82,86	80,48	85,71
	25%	89,05	91,90	92,86	90,48	90,95	92,86	91,43
Semeion	5%	80,37	77,92	78,31	79,93	79,74	79,93	78,49
	10%	79,29	78,61	79,11	78,86	78,79	77,27	78,15
	15%	78,69	78,81	80,81	80,31	79,44	80,06	79,81
	20%	79,56	76,37	78,18	78,12	78,37	78,62	77,44
	25%	75,80	77,81	74,48	76,98	76,42	75,17	76,68
Spectf Heart	5%	66,67	66,94	66,94	67,50	66,94	69,44	68,33
	10%	70,02	67,32	66,26	69,12	67,73	67,89	67,65
	15%	74,20	71,33	75,38	70,47	74,47	74,51	74,50
	20%	72,16	71,53	74,73	70,70	72,50	69,81	67,51
	25%	77,35	73,50	76,46	82,00	73,26	77,36	75,60
Tic Tac Toe	5%	83,44	84,06	83,54	84,58	83,13	82,81	84,27
	10%	82,49	80,30	81,96	80,41	81,04	81,76	82,58
	15%	79,79	81,58	81,68	80,63	80,74	82,00	81,05
	20%	81,37	81,03	78,56	80,74	80,51	82,29	82,91
	25%	82,30	81,79	82,50	82,93	80,84	82,09	84,93
Twonorm	5%	98,19	98,20	98,05	98,07	98,22	98,12	98,18
	10%	97,15	97,26	97,08	97,15	97,12	97,13	97,49
	15%	97,35	97,51	97,57	97,36	97,77	97,55	97,64
	20%	96,77	96,78	96,88	96,55	96,69	96,78	96,73
	25%	97,31	97,34	97,16	97,14	96,99	97,30	97,38
Vehicle	5%	58,00	56,94	53,35	55,46	59,04	58,37	61,81
	10%	62,82	59,76	55,76	58,24	57,06	54,59	57,06
	15%	57,56	58,26	60,12	58,60	60,47	62,33	58,26
	20%	57,46	54,82	55,14	57,41	61,08	59,82	58,73
	25%	58,43	59,29	55,96	57,88	58,91	60,11	56,86
Waveform	5%	81,64	82,42	82,48	82,52	82,56	83,20	82,36
	10%	80,44	81,24	81,44	81,08	81,00	81,48	82,00
	15%	83,14	83,92	83,60	82,56	82,98	83,28	84,44
	20%	82,36	83,02	82,62	83,24	83,26	82,04	82,30
	25%	83,40	82,84	83,38	82,76	83,50	82,26	83,46

Wilt	5%	94,73	94,73	95,27	95,02	94,67	94,96	94,69
	10%	94,62	94,23	94,42	93,98	94,52	94,48	94,77
	15%	94,75	94,88	94,98	94,40	93,99	94,71	94,90
	20%	94,17	94,59	94,38	94,61	94,26	94,69	94,69
	25%	94,75	94,42	94,63	94,65	94,40	94,90	94,34

**Fonte:** O Próprio Autor (2018)

## APÊNDICE M – Desempenho dos Valores do *cr* em cada Técnica

A Tabela 27 apresenta o desempenho para cada técnica em cada um dos classificadores nos experimentos, onde cada linha representa um valor para o parâmetro *cr*, cada coluna representa cada uma das técnicas. Em cada célula, da tabela está representado a quantidade de vezes que o valor de *cr* obteve o melhor desempenho com a combinação de classificador e técnica analisados.

Tabela 27 – Desempenho dos Valores do *cr* em cada Técnica.

<i>cr/técnica</i>	Classificadores												Parcial		Total		
	Naïve Bayes			<i>rpartXse</i>			RIPPER			<i>k</i> -NN							
	C1(s)	C1(v)	C2	C1(s)	C1(v)	C2	C1(s)	C1(v)	C2	C1(s)	C1(v)	C2	C1(s)	C1(v)	C2	vitórias	percentual
2	0	0	0	1	0	1	0	0	0	1	1	3	2	1	4	7/60	11,67%
3	1	1	1	0	0	0	0	0	0	1	1	1	2	2	2	6/60	10,00%
4	1	1	1	1	0	1	1	1	1	1	1	0	4	3	3	10/60	16,67%
5	1	1	1	0	1	2	0	0	0	1	1	0	2	3	3	8/60	13,34%
6	1	1	1	1	2	0	1	1	1	1	1	0	4	5	2	11/60	18,33%
7	1	1	1	1	1	0	2	2	2	0	0	0	4	4	3	11/60	18,33%
8	0	0	0	1	1	1	1	1	1	0	0	2	2	2	4	8/60	13,33%

**Fonte:** O Próprio Autor (2018)