

Faculdade de Tecnologia de São Paulo
Departamento de Tecnologia da Informação
Curso de Análise e Desenvolvimento de Sistemas

José Soares da Silva Neto

Aprendizagem de Máquina Supervisionada:
O Algoritmo K-Vizinhos Mais Próximos

São Paulo
2017

Faculdade de Tecnologia de São Paulo
Departamento de Tecnologia da Informação

José Soares da Silva Neto

Aprendizagem de Máquina Supervisionada:
O Algoritmo K-Vizinhos Mais Próximos

Monografia submetida como exigência parcial
para obtenção do Grau de Tecnólogo em Análise e
Desenvolvimento de Sistemas da Faculdade de
Tecnologia de São Paulo.

Orientador: Prof. Dr. Silvio do Lago Pereira

São Paulo

2017

Faculdade de Tecnologia de São Paulo
Departamento de Tecnologia da Informação

José Soares da Silva Neto

Aprendizagem de Máquina Supervisionada:
O Algoritmo K-Vizinhos Mais Próximos

Monografia submetida como exigência parcial para a obtenção do Grau de
Superior Tecnológico em Análise e Desenvolvimento de Sistemas

Parecer do Professor Orientador: _____

Conceito/Nota Final: _____

Orientador: Prof. Dr. Silvio do Lago Pereira
São Paulo, ____ de _____ de 2017

Dedico esse trabalho a toda a minha família, em especial à Júlio, Cleusa, Rosana, Camilla, Junior e Tatiane, pois sem essas pessoas a realização desse curso não seria possível.

AGRADECIMENTOS

Agradecimentos especiais ao orientador desse trabalho, o professor Dr. Silvio do Lago Pereira, que me apresentou a área de aprendizagem de máquina, sugeriu o tema e me acompanhou ao longo de todo o trabalho. Agradecimentos a minha família, em especial a meu irmão Júlio, a minha irmã Tatiane e aos meus pais, Júlio e Cleusa, que ao longo da minha vida me deram apoio e me aconselharam muito, me incentivando a chegar até aqui. Agradecimentos a todos os professores da Fatec de Mogi das Cruzes e da Fatec São Paulo, que através de cada aula ou conversa informal contribuíram para o meu conhecimento, assim como a todos que ao longo do curso me orientaram em meus estágios. Agradecimentos também a Camilla e a Rosana, que me apoiaram ao decorrer de toda a graduação, e em especial por terem paciência durante todo o tempo que dediquei a esse trabalho.

*“The 20th century was the century of the equation; the 21st century is the century of
the algorithm.”*
Bernard Chazelle

RESUMO

Apesar de estar no foco da mídia há apenas alguns anos, inteligência artificial é uma área que vem sendo estudada há muito tempo. Dentre suas subáreas, existe a aprendizagem de máquina, que é utilizada para diversos fins, e dentre eles a classificação e predição. Um algoritmo de aprendizagem de máquina muito conhecido e amplamente utilizado, é o algoritmo K-Vizinhos mais próximos, que durante esse trabalho será estudado, e aplicado em diversos problemas envolvendo a classificação. Ao final do trabalho, serão comparados os resultados de desempenho desse algoritmo, utilizando diversas métricas diferentes, e serão apontadas quais são as melhores métricas para utilizar esse algoritmo na tarefa de classificação de dados.

Palavras-chave: Aprendizagem de Máquina; Classificação de Dados; KDD; K-Vizinhos mais próximos; Mineração de Dados;

ABSTRACT

Despite being in the focus of the media just a few years ago, artificial intelligence is a field that has been studied for a long time. Among its subareas there is machine learning, which is used for various purposes, including classification and prediction. A widely known and used machine learning algorithm is the K-Nearest Neighbors algorithm, which will be studied during this researching work and applied to several problems involving classification. At the end of the study, the performance results of this algorithm will be compared using several different metrics, and the best ones will be pointed out to use this algorithm in the data classification task.

Keywords: Machine Learning; Data Classification; KDD; K-Nearest Neighbors; Data Mining;

LISTA DE ILUSTRAÇÕES

Figura 1 . Knowledge Discovery in Databases.....	31
Figura 2 . Pré-processamento e transformação de dados.....	33
Figura 3 . Agrupamento em uma base de dados de empréstimo	35
Figura 4 . Modelo de Classificação	36
Figura 5 . Gráfico representando a tarefa de Regressão Linear.....	37
Figura 6 . Representação de um classificador	38
Figura 7 . Árvore de decisão de prováveis compradores de computadores	39
Figura 8 . Classificador Linear	40
Figura 9 . Modelo matemático simples de um neurônio	40
Figura 10 . Matriz de confusão	44
Figura 11 . Representação da aplicação do algoritmo KNN	52
Figura 12 . Resultados de classificação – Bases padrão – Heart Disease	64
Figura 13 . Resultados de classificação – Bases padrão – Mammographic Mass.....	65
Figura 14 . Resultados de classificação – Bases padrão – Winequality Red.....	66
Figura 15 . Resultados de classificação – Bases padrão – Titanic	66
Figura 16 . Resultados de classificação – Bases padrão – Vehicle Silhouettes.....	67
Figura 17 . Resultados de classificação – Bases padrão c/ atr. categóricos – Automobile	68
Figura 18 . Resultados de classificação – Bases padrão c/ atr. categóricos – German Credit..	69
Figura 19 . Resultados de classificação – Bases padrão c/ atr. categóricos – Lymphography.	69
Figura 20 . Resultados de classificação – Bases padrão c/ atr. categóricos – South African Heart	70
Figura 21 . Resultados de classificação – Bases desbalanceadas – Ecoli.....	71
Figura 22 . Resultados de classificação – Bases desbalanceadas – Iris.....	71
Figura 23 . Resultados de classificação – Bases desbalanceadas – Pima.....	72
Figura 24 . Resultados de classificação – Bases desbalanceadas – Vehicle.....	73
Figura 25 . Resultados de classificação – Bases desbalanceadas – Wisconsin	73
Figura 26 . Resultados de classificação – Bases c/ atributos faltantes – Bands	74
Figura 27 . Resultados de classificação – Bases c/ atributos faltantes – Cleveland.....	75
Figura 28 . Resultados de classificação – Bases c/ atributos faltantes – Hepatitis.....	75
Figura 29 . Resultados de classificação – Bases c/ atributos faltantes – Mammographic.....	76
Figura 30 . Resultados de classificação – Bases c/ atributos categóricos – Car.....	77
Figura 31 . Resultados de classificação – Bases c/ atributos categóricos – Tic-Tac-Toe	77
Figura 32 . Resultados de classificação – Bases c/ ruído de atributos – Contraceptive	78
Figura 33 . Resultados de classificação – Bases c/ ruído de atributos – Heart.....	79
Figura 34 . Resultados de classificação – Bases c/ ruído de atributos – Iris	79
Figura 35 . Resultados de classificação – Bases c/ ruído de atributos – Pima	80

Figura 36 . Resultados de classificação – Bases c/ ruído de atributos – Sonar.....	81
Figura 37 . Resultados de classificação – Bases c/ ruído de classes – Contraceptive	82
Figura 38 . Resultados de classificação – Bases c/ ruído de classes – Heart.....	83
Figura 39 . Resultados de classificação – Bases c/ ruído de classes – Iris.....	83
Figura 40 . Resultados de classificação – Bases c/ ruído de classes – Pima.....	84
Figura 41 . Resultados de classificação – Bases c/ ruído de classes – Sonar.....	85
Figura 42 . Resultados de tempo de execução – Bases padrão – Titanic.....	85

LISTA DE ABREVIATURAS E SIGLAS

cos	Cosseno
FN	Falso negativo
FP	Falso positivo
KDD	<i>Knowledge Discovery in Databases</i>
KNN	<i>K-Nearest Neighbor</i>
log	Logaritmo
max	Ponto máximo da função
min	Ponto mínimo da função
SVM	<i>Support Vector Machines</i>
VN	Verdadeiro negativo
VP	Verdadeiro positivo

LISTA DE SÍMBOLOS

δ	Delta minúsculo
\neq	Diferente
\cap	Intersecção
\pm	Mais ou menos
\leq	Menor ou igual
$\sqrt{}$	Raiz quadrada
σ	Sigma minúscula
Σ	Sigma maiúscula (somatório)
θ	Teta
\cup	União

SUMÁRIO

Capítulo 1.....	27
Introdução	27
1.1. Contextualização.....	27
1.2. Motivação	28
1.3. Objetivo	28
1.4. Organização	28
Capítulo 2.....	31
Mineração de Dados	31
2.1. Extração do conhecimento.....	31
2.2. Tarefas e técnicas de mineração de dados	33
2.2.1. Detecção de anomalias	33
2.2.2. Regras de associação	34
2.2.3. Agrupamento (<i>Clustering</i>)	34
2.2.4. Classificação	35
2.2.5. Regressão	35
2.2.6. Sumarização.....	36
2.3. Aprendizagem de máquina	37
2.4. Classificação	38
2.4.1. Árvores de decisão.....	39
2.4.2. Modelos lineares	39
2.4.3. Redes neurais	40
2.4.4. Modelos não paramétricos	41
2.4.5. Máquinas de vetores de suporte.....	41
2.5. Técnicas de treinamento	42
2.5.1. <i>Holdout</i>	42
2.5.2. <i>K-Fold Cross-validation</i>	42
2.5.3. <i>Leave-one-out</i>	43
2.5.4. <i>Bootstrap</i>	43
2.6. Métricas de avaliação de desempenho.....	43

2.6.1. Acurácia	43
2.6.2. Matriz de confusão	44
2.6.3. Sensibilidade ou revocação	45
2.6.4. Precisão	45
2.6.5. <i>F-Score</i>	45
2.6.6. <i>G-Score</i>	45
2.7. Similaridade e Distância	45
2.7.1. Bray Curtis	46
2.7.2. Canberra	46
2.7.3. Chebyshev	46
2.7.4. Euclidiana.....	46
2.7.5. Manhattan.....	46
2.7.6. Minkowski.....	47
2.7.7. Hamming.....	47
2.7.8. Correlacional	47
2.7.9. Cosseno	47
2.7.10. Jaccard.....	48
2.8. Estatística	48
2.8.1. Média.....	48
2.8.2. Variância	48
2.8.3. Desvio padrão.....	48
2.9. Matemática	49
2.9.1. Logaritmo	49
Capítulo 3.....	51
O algoritmo K-vizinhos mais próximos	51
3.1. Definição	51
3.1.1. Algoritmo	51
3.2. Implementação	53
3.2.1. Pré-processamento e transformação.....	53
3.2.2. Treinamento	53

3.2.3. Classificação	54
3.2.4. Escolha do k	54
Capítulo 4.....	55
Bases de dados utilizadas.....	55
4.1. Bases padrão	55
4.1.1. <i>Heart</i>	55
4.1.2. <i>Mammographic Mass</i>	56
4.1.3. <i>Red Wine Quality</i>	56
4.1.4. <i>Titanic</i>	56
4.1.5. <i>Vehicle Silhouettes</i>	56
4.2. Bases padrão com atributos categóricos	56
4.2.1. <i>Automobile</i>	56
4.2.2. <i>German Credit</i>	57
4.2.3. <i>Lymphography</i>	57
4.2.4. <i>South African Heart</i>	57
4.3. Bases desbalanceadas	57
4.3.1. <i>Ecoli</i>	57
4.3.2. <i>Iris</i>	58
4.3.3. <i>Pima</i>	58
4.3.4. <i>Vehicle Silhouettes</i>	58
4.3.5. <i>Wisconsin</i>	58
4.4. Bases com atributos faltantes.....	59
4.4.1. <i>Cylinder Bands</i>	59
4.4.2. <i>Heart Disease (Cleveland)</i>	59
4.4.3. <i>Hepatitis</i>	59
4.4.4. <i>Mammographic Mass</i>	59
4.5. Bases com atributos categóricos	60
4.5.1. <i>Car Evaluation</i>	60
4.5.2. <i>Tic-Tac-Toe Endgame</i>	60
4.6. Bases com ruído de atributos	60

4.6.1. <i>Contraceptive Method Choice</i>	60
4.6.2. <i>Heart</i>	60
4.6.3. <i>Iris</i>	61
4.6.4. <i>Pima</i>	61
4.6.5. <i>Sonar, Mines vs. Rocks</i>	61
4.7. Bases com ruído de classes	61
Capítulo 5.....	63
Experimentos.....	63
5.1. Bases padrão.....	64
5.1.1. <i>Heart</i>	64
5.1.2. <i>Mammographic Mass</i>	64
5.1.3. <i>Red Wine Quality</i>	65
5.1.4. <i>Titanic</i>	66
5.1.5. <i>Vehicle Silhouettes</i>	67
5.2. Bases padrão com atributos categóricos.....	67
5.2.1. <i>Automobile</i>	67
5.2.2. <i>German Credit</i>	68
5.2.3. <i>Lymphography</i>	69
5.2.4. <i>South African Heart</i>	70
5.3. Bases desbalanceadas.....	70
5.3.1. <i>Ecoli</i>	70
5.3.2. <i>Iris</i>	71
5.3.3. <i>Pima</i>	72
5.3.4. <i>Vehicle Silhouettes</i>	72
5.3.5. <i>Wisconsin</i>	73
5.4. Bases com atributos faltantes	74
5.4.1. <i>Cylinder Bands</i>	74
5.4.2. <i>Heart Disease (Cleveland)</i>	74
5.4.3. <i>Hepatitis</i>	75
5.4.4. <i>Mammographic Mass</i>	76

5.5. Bases com atributos categóricos	76
5.5.1. <i>Car Evaluation</i>	76
5.5.2. <i>Tic-Tac-Toe Endgame</i>	77
5.6. Bases com ruído de atributos	78
5.6.1. <i>Contraceptive Method Choice</i>	78
5.6.2. <i>Heart</i>	78
5.6.3. <i>Iris</i>	79
5.6.4. <i>Pima</i>	80
5.6.5. <i>Sonar, Mines vs. Rocks</i>	80
5.7. Bases com ruído de classes	81
5.7.1. <i>Contraceptive Method Choice</i>	81
5.7.2. <i>Heart</i>	82
5.7.3. <i>Iris</i>	83
5.7.4. <i>Pima</i>	84
5.7.5. <i>Sonar, Mines vs. Rocks</i>	84
5.8. Tempo de execução	85
Capítulo 6.....	87
Conclusões	87
6.1. Principais contribuições	87
6.2. Trabalhos futuros	88
Referências Bibliográficas	89
Apêndice A.	91
Código fonte	91
Apêndice B.	93
Conjuntos de dados	93
Apêndice C.	103
Resultados	103

Capítulo 1.

INTRODUÇÃO

1.1. Contextualização

A pesar de estar no foco da mídia há apenas alguns anos, inteligência artificial é uma área proposta e estudada há muito tempo. O primeiro trabalho reconhecido nesta área, publicado por Warren McCulloch e Walter Pitts em 1943, propôs um modelo de neurônios artificiais conhecido como **redes neurais**. McCulloch e Pitts sugeriram nesse trabalho que, se definidas adequadamente, as **redes neurais** seriam capazes de aprender (RUSSEL e NORVIG, 2013).

Desde 1943 até hoje, surgiram diversos trabalhos na área, porém o trabalho que pode ser considerado o mais influente até hoje é o “*Computing Machinery and Intelligence*” de Alan Turing, publicado em 1950. Nesse artigo foi proposto o famoso teste de Turing, a aprendizagem de máquina, os algoritmos genéticos e a aprendizagem por reforço (RUSSEL e NORVIG, 2013).

Dentre essas, a aprendizagem de máquina é uma subárea de inteligência artificial, que se concentra em algoritmos, ou modelos, que têm a capacidade de aprender, melhorando o seu desempenho nas tarefas futuras de aprendizagem após fazer observações sobre o mundo (RUSSEL e NORVIG, 2013).

Paralelamente, o avanço da tecnologia tem viabilizado o aumento de poder computacional e de armazenamento de dados. Com o aumento de armazenamento e diminuição de custos, a quantidade de dados armazenados cresce exponencialmente. Um estudo ¹ feito em 2014, estima que em 2013, 4.4 *zettabytes* era a quantidade de dados armazenados no mundo todo, porém em 2020, é estimado que essa quantidade chegue a cerca de 44 *zettabytes* (GOLDSCHMIDT, PASSOS e BEZERRA, 2015).

Diante desse cenário, é inviável que esses dados sejam analisados pelo homem. Para atender a essa necessidade, em 1989, foi proposta a área conhecida como KDD ², ou

¹ <https://www.emc.com/infographics/digital-universe-2014.htm>

² *Knowledge Discovery in Databases*

Descoberta de Conhecimento em Bases de Dados (GOLDSCHMIDT, PASSOS e BEZERRA, 2015).

O artigo mais popular dessa área foi escrito por FAYYAD, PIATESKY-SHAPIRO e SMYTH (1996), em que o KDD é definido como: “o processo não trivial de identificação de padrões: válidos, novos, potencialmente úteis e em última análise, compreensíveis nos dados”.

O KDD é composto por basicamente três fases, que são: pré-processamento, mineração de dados e pós-processamento. A fase de mineração de dados é a fase em que os dados são extraídos, e portanto, é a fase em que são aplicados os algoritmos de aprendizagem de máquina.

Ainda na fase de mineração de dados, as técnicas variam de acordo com o objetivo da aplicação do KDD, e podem ser: a detecção de anomalias, regras de associação, o agrupamento, a classificação, a regressão e a sumarização (FAYYAD; PIATESKY-SHAPIRO e SMYTH, 1996).

Nesse estudo, será analisado um dos muitos algoritmos de aprendizagem de máquina, que é o algoritmo KNN ³, ou K-Vizinhos Mais Próximos, e será aplicado na tarefa de classificação e descoberta, assuntos que serão discutidos nos capítulos seguintes.

1.2. Motivação

A área de aprendizagem de máquina possui diversos algoritmos ⁴ que podem ser implementados juntamente ao KDD, sendo o KNN um algoritmo de implementação simples e bastante popular.

A motivação foi propor novas ideias na implementação do mesmo, e avaliar a diferença de desempenho utilizando diferentes métricas.

1.3. Objetivo

O principal objetivo deste trabalho é avaliar diferentes métricas e formas de implementação, e como elas impactam no desempenho do algoritmo KNN.

1.4. Organização

Esta monografia é organizada da seguinte maneira:

Capítulo 2. São introduzidos os fundamentos necessários para acompanhar o trabalho. São abordados os conceitos da técnica KDD, suas tarefas e objetivos. Os conceitos de *Data Mining* e aprendizagem de máquina, além dos diferentes tipos de algoritmos que podem ser aplicados no

³ *K-Nearest Neighbor*

⁴ http://www.saedsayad.com/data_mining_map.htm

Data Mining. Além disso, são abordadas as diferentes métricas de similaridades e distância, e de avaliação do desempenho.

Capítulo 3. Nesse capítulo, é abordado o algoritmo KNN, com sua definição, implementação original e como foi implementado nesse trabalho, inclusive com suas propostas e diferenças em relação ao algoritmo original.

Capítulo 4. Nesse capítulo são descritas todas as bases de dados utilizadas nos experimentos, explicando a diferença entre cada tipo de base, e com informações sobre cada uma.

Capítulo 5. Nesse capítulo são descritos os experimentos realizados, e as comparações de cada experimento. Comparando métricas como: acurácia e tempo de execução.

Capítulo 6. No capítulo final, são discutidas as contribuições e as pesquisas futuras que podem ser desenvolvidas à partir desse trabalho.

Capítulo 2.

MINERAÇÃO DE DADOS

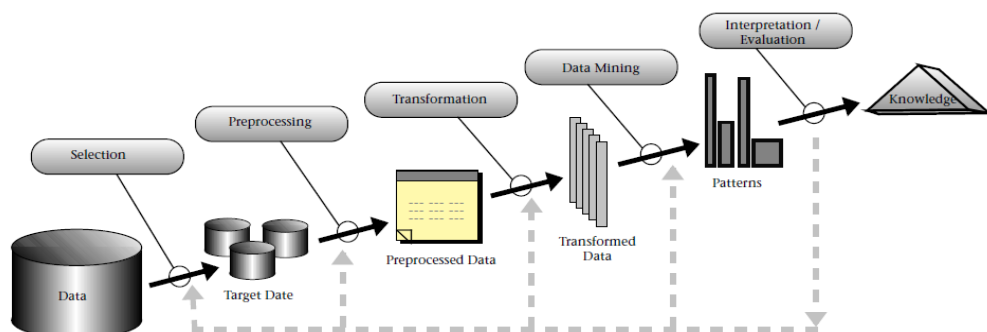
De acordo com [MEIRA JUNIOR e ZAKI \(2014\)](#), a mineração de dados é o processo de descobrir padrões perspicazes, interessantes e inovadores, bem como modelos descritivos, compreensíveis e preditivos à partir de dados de grande escala.

2.1. Extração do conhecimento

De acordo com [FAYYAD, PIATESKY-SHAPIRO e SMYTH \(1996\)](#), a metodologia de extração do conhecimento de bases de dados, ou KDD⁵ é o processo não trivial de identificação de padrões: válidos, novos, potencialmente úteis e em última análise, compreensíveis nos dados. Conforme a [Figura 1](#), a metodologia é composta pelas seguintes etapas:

- Seleção;
- Pré-processamento;
- Transformação;
- Mineração de dados;
- Interpretação;

Figura 1. *Knowledge Discovery in Databases*



([FAYYAD; PIATESKY-SHAPIRO e SMYTH, 1996, 41](#))

⁵ *Knowledge Discovery in Databases*

Na fase de **seleção** é escolhido o conjunto de dados, pertencente a um domínio, que será utilizado no processo de extração do conhecimento. Nessa fase devem ser levados em consideração o domínio da aplicação, o tamanho do conjunto de dados, a quantidade de classes e os tipos de atributos (numéricos, categóricos ou booleanos).

Na fase de **pré-processamento** é realizada a limpeza de dados. Visando o aumento da eficiência da extração, as operações dessa fase incluem: a remoção de ruídos (coleta de informações necessárias para modelar ou contabilizar o ruído), estratégias de manipulação de campos de dados ausentes, técnicas de redução de dimensionalidade, organização em um repositório único (no caso de mais um conjunto de dados) e a eliminação de instâncias repetidas (SILVA; PERES e BOSCAROLI, 2016).

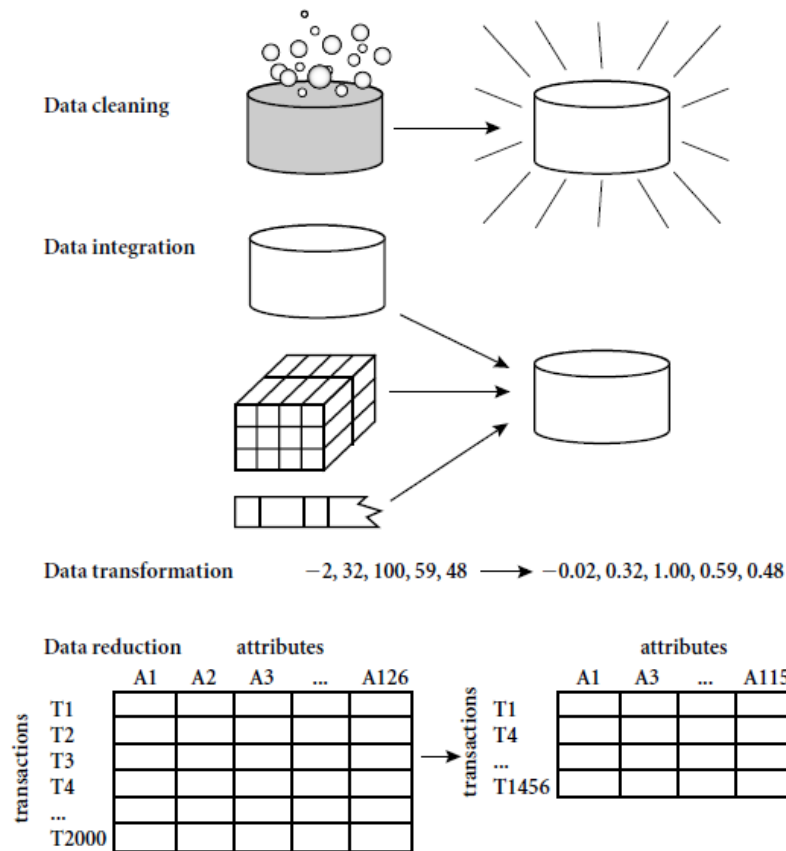
Na fase de **transformação**, os dados são transformados ou consolidados de forma adequada para a aplicação da mineração de dados. Nessa etapa, os dados podem ser consolidados, como exemplo, salários mensais serem agregados para salários anuais. E os atributos podem ser normalizados para que fiquem na mesma escala (0 – 1.0). A quantidade de atributos analisados pode ser diminuída para a aumentar a eficiência da extração (HAN e KAMBER, 2006). Essa fase, e as anteriores podem ser observadas na Figura 2.

A fase de **mineração de dados** consiste na extração de dados em si, a busca de padrões a partir de determinado método de mineração, e é composta de seis tarefas: a detecção de anomalias, regras de associação, o agrupamento, a classificação, a regressão e a sumarização (FAYYAD; PIATESKY-SHAPIRO e SMYTH, 1996). Essas tarefas serão descritas posteriormente.

Por fim, a **avaliação**, ou etapa de **pós-processamento**, consiste no tratamento do conhecimento obtido e padrões descobertos. Se preciso, pode-se retornar aos passos anteriores para ajustar o modelo (FAYYAD; PIATESKY-SHAPIRO e SMYTH, 1996). Segundo GOLDSCHMIDT, PASSOS e BEZERRA (2015) entre as principais funções da etapa de pós-processamento estão: a elaboração e a organização do conhecimento obtido, podendo incluir a simplificação de gráficos, diagramas, ou relatórios demonstrativos; além da conversão da forma de representação do conhecimento obtido, por exemplo.

Os objetivos da extração do conhecimento são definidos pela aplicação em que terá o conhecimento adquirido. Podemos enumerar dois objetivos:

- **Validação de hipóteses:** o objetivo consiste em validar hipóteses apresentadas por um especialista da área em que o KDD será aplicado.
- **Descoberta:** o objetivo consiste na descoberta de novos padrões, pode ser **preditiva**, onde são encontrados padrões para prever o futuro de entidades específicas. Ou **descritiva**, onde são encontrados padrões de forma que facilitem o entendimento para o usuário (FAYYAD; PIATESKY-SHAPIRO e SMYTH, 1996).

Figura 2. Pré-processamento e transformação de dados

(HAN e KAMBER, 2006)

2.2. Tarefas e técnicas de mineração de dados

A fase de mineração de dados consiste na fase de descoberta de padrões e do conhecimento, e pode ser dividida entre descoberta e validação de hipóteses (FAYYAD; PIATESKY-SHAPIO e SMYTH, 1996).

Nessa etapa, são definidas as técnicas e algoritmos a serem utilizados no problema em questão. Essa escolha depende do tipo de tarefa de KDD a ser realizada (GOLDSCHMIDT; PASSOS e BEZERRA, 2015).

2.2.1. Detecção de anomalias

Essa técnica, concentra-se na descoberta das alterações mais significativas nos dados a partir de valores previamente medidos (FAYYAD; PIATESKY-SHAPIO e SMYTH, 1996), ou seja, identificar registros no conjunto dados cujas características não atendam aos padrões considerados normais no contexto (GOLDSCHMIDT; PASSOS e BEZERRA, 2015).

Como exemplo, é possível citar o caso de uma compra atípica, ou seja, que divirja do perfil do cliente, utilizando um cartão de crédito. Nesse caso, é possível identificar que essa compra não pertence ao perfil do cliente utilizando a técnica de detecção de anomalias.

2.2.2. Regras de associação

Compreende a busca por ocorrências frequentes e simultâneas entre elementos de um contexto (SILVA; PERES e BOSCAROLI, 2016), ou seja, encontrar um modelo que descreva dependências significativas entre variáveis (FAYYAD; PIATESKY-SHAPIRO e SMYTH, 1996).

Como exemplo, é possível citar o caso de uma grande rede de mercados norte-americana. Como conhecimento extraído, utilizando a técnica de regras de associação, essa rede de mercados percebeu que um número razoável de compradores de fralda, também compravam cerveja em vésperas de finais de semana que teriam jogos transmitidos pela televisão. Por meio de uma análise feita por especialistas de domínio, foi constatado que esses clientes eram homens, e ao comprarem fraldas para seus filhos, aproveitavam para comprar cerveja para acompanhar os jogos enquanto cuidavam dos filhos. Nesse exemplo, a associação foi feita sobre fraldas e cerveja, e a rede utilizou esse conhecimento para aproximar as gôndolas dos dois produtos e incrementar a venda dos dois produtos (GOLDSCHMIDT; PASSOS e BEZERRA, 2015).

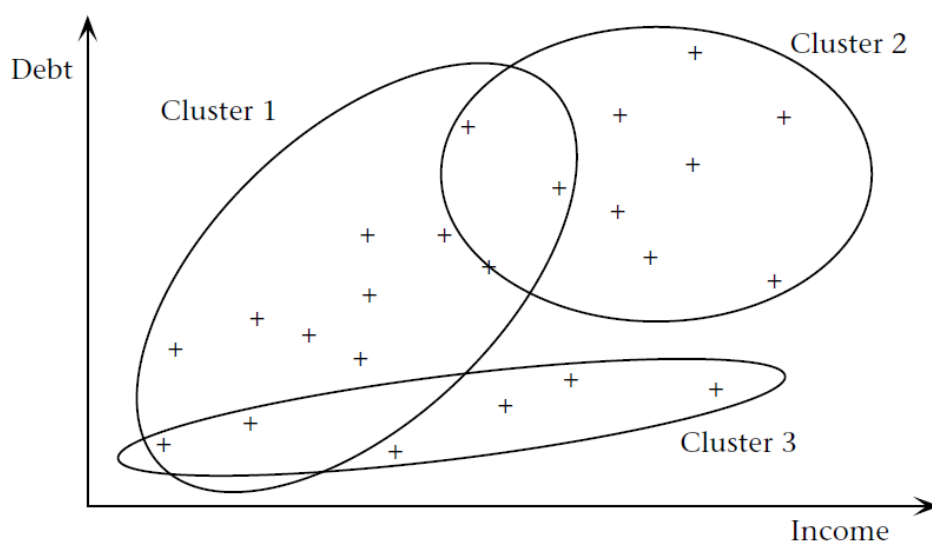
2.2.3. Agrupamento (*Clustering*) ⁶

Na técnica de agrupamento, os registros não possuem classes, ou rótulos, e se possuem os mesmos são ignorados. Nessa técnica não existe a necessidade do uso dessa informação. O objetivo é descobrir relações entre os dados por meios de similaridades e fornecer, como resposta, a indicação de quais dados são similares entre si, oferecendo um modelo de agrupamento ou perfis para grupos de dados (SILVA; PERES e BOSCAROLI, 2016).

A Figura 3 demonstra um agrupamento de um conjunto de dados de empréstimo em três *clusters*. É possível observar que os *clusters* se sobrepõem, permitindo que alguns registros pertençam a mais de um *cluster*.

⁶ Clusterização

Figura 3. Agrupamento em uma base de dados de empréstimo



(FAYYAD; PIATESKY-SHAPIO e SMYTH, 1996, 45)

2.2.4. Classificação

Consiste na tarefa de prever a qual classe pertence qualquer exemplar de um domínio sob análise, com base em um conjunto de dados já classificado (SILVA; PERES e BOSCAROLI, 2016).

Ou ainda, pode ser descrita como a busca por uma função que permita associar corretamente cada registro de um conjunto de dados a um único rótulo categórico, utilizando um conjunto de dados já classificado (GOLDSCHMIDT; PASSOS e BEZERRA, 2015).

Os atributos da base de dados são divididos em dois grupos. Um grupo contendo apenas o atributo-alvo, que se trata de um atributo categórico (seu domínio é composto por categorias / classes). E o outro, contendo os atributos que serão utilizados na predição do atributo-alvo. (GOLDSCHMIDT; PASSOS e BEZERRA, 2015).

Com o modelo definido, o mesmo pode ser aplicado a novos registros para prever as classes a que esses registros pertencem. Um exemplo pode ser visto na Figura 4, que representa uma nova entrada, sendo classificada com base em um conjunto de dados já classificado.

2.2.5. Regressão

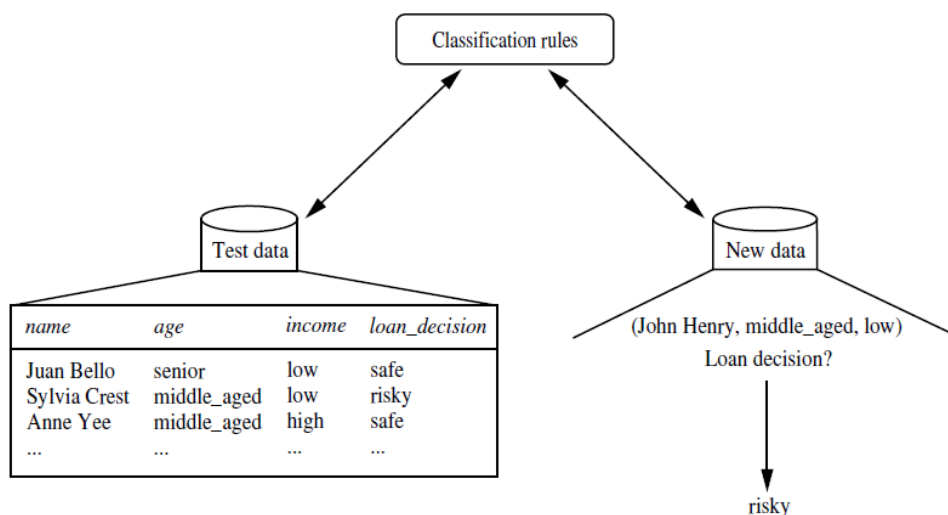
Consiste na tarefa de busca por funções, lineares ou não, que possam mapear as entradas de um conjunto de dados em valores reais. Ou seja, é uma tarefa similar à classificação, porém, o rótulo (classe) nesse caso é numérico (GOLDSCHMIDT; PASSOS e BEZERRA, 2015).

A tarefa de regressão é utilizada para estimar valores a partir de um conjunto de dados históricos, como por exemplo, prever os próximos valores de indicadores econômicos

analisando as entradas (atributos) de um conjunto de dados histórico (SILVA; PERES e BOSCAROLI, 2016).

Na Figura 5 é possível visualizar a reta traçada pela função linear, em que a variável *Debt*⁷ varia em função da variável *Income*⁸.

Figura 4. Modelo de Classificação



(HAN e KAMBER, 2006)

2.2.6. Sumarização

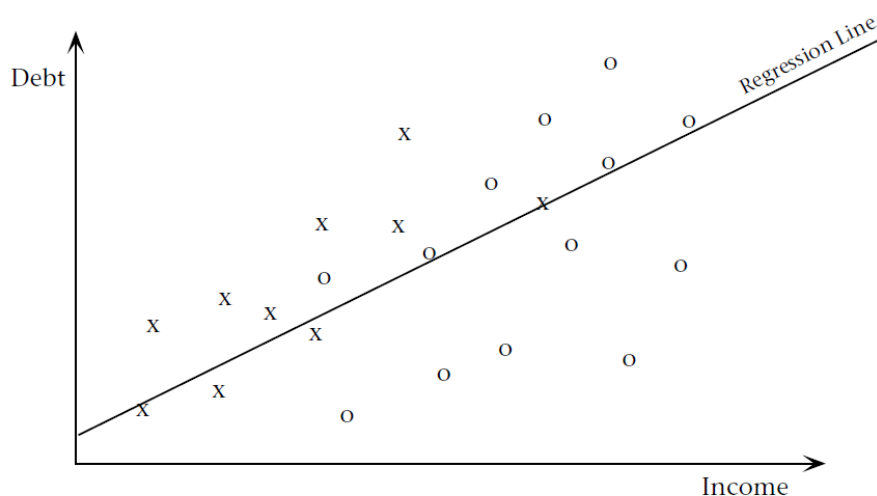
Consiste na tarefa de identificar e indicar características comuns entre registros do conjunto de dados.

Como exemplo, pode ser utilizada para identificar o perfil de assinantes de uma revista (faixa etária, sexo, escolaridade, profissão) para direcionar a publicidade ao público predominante (GOLDSCHMIDT; PASSOS e BEZERRA, 2015).

⁷ Dívida.

⁸ Renda.

Figura 5. Gráfico representando a tarefa de Regressão Linear



(FAYYAD; PIATESKY-SHAPIRO e SMYTH, 1996, 44)

2.3. Aprendizagem de máquina

Aprendizagem de máquina é uma subárea de inteligência artificial, que se concentra em algoritmos, ou modelos, que têm a capacidade de aprender, melhorando o seu desempenho nas tarefas futuras de aprendizagem após fazer observações sobre o mundo (RUSSEL e NORVIG, 2013).

A aprendizagem se define em três principais tipos:

Na **aprendizagem não supervisionada**, o modelo aprende com exemplos que possuem apenas entradas, e não possuem uma saída. A tarefa mais comum nesse caso é o **agrupamento**, como descrito anteriormente (RUSSEL e NORVIG, 2013).

Na **aprendizagem por reforço**, o modelo aprende a partir de recompensas ou punições. Como exemplo, em uma partida de xadrez o modelo pode ser recompensado ao ganhar uma partida, ou punido ao perder. E então caberá ao modelo decidir qual das ações anteriores ao reforço foram as maiores responsáveis por isso (RUSSEL e NORVIG, 2013).

Na **aprendizagem supervisionada**, o modelo aprende com exemplos que possuem entradas e saídas, e aprende uma função que faz o mapeamento da entrada para a saída. Quando a saída for categórica, o problema de aprendizagem será de **classificação**. Quando a saída for numérica, o problema de aprendizagem será de **regressão**. Os dois problemas foram abordados anteriormente (RUSSEL e NORVIG, 2013).

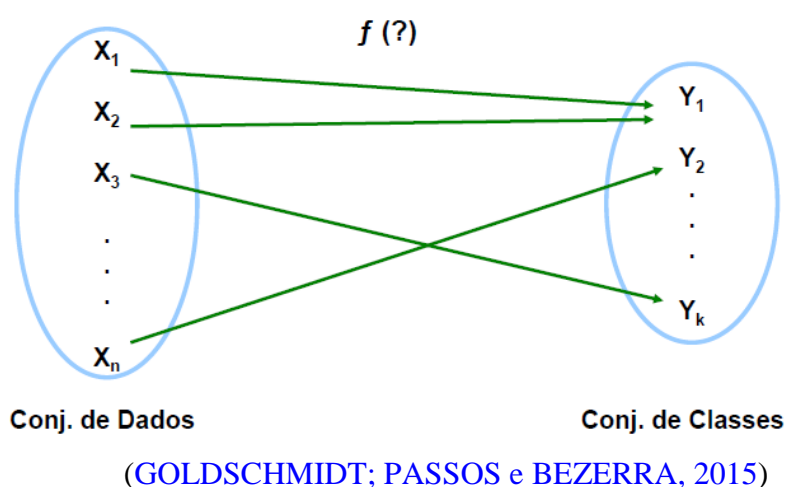
Esse trabalho se baseia em um algoritmo de **aprendizagem supervisionada**, envolvendo problemas de **classificação**, portanto o problema de classificação, assim como seus principais tipos, será abordado com mais profundidade.

2.4. Classificação

Para realizar a classificação, é necessário que seja encontrada uma função que associe uma nova entrada, contendo apenas atributos, a uma única saída (seu rótulo ou classe).

Essa função pode ser descrita como **hipótese $h(\mathbf{x})$** , que é a função que mais se aproxima da função original **$f(\mathbf{x})$** , do conjunto de dados que associa cada registro de atributos \mathbf{x} , a uma única saída y (GOLDSCHMIDT; PASSOS e BEZERRA, 2015), conforme a Figura 6.

Figura 6. Representação de um classificador



Para medir a precisão de uma **hipótese**, utilizamos um **conjunto de dados de testes** com exemplos distintos do **conjunto de dados de treinamento**, que é o conjunto de dados já conhecido e classificado (RUSSEL e NORVIG, 2013).

Caso o conjunto de dados de testes não seja suficientemente representativo, a hipótese pode ter um bom desempenho no teste, mas pode classificar erroneamente novas entradas, de dados reais. Nesse caso, a hipótese ajustou-se em excesso ao conjunto de dados de testes, o que é chamado de *overfitting*⁹. Por outro lado, quando as parametrizações do algoritmo de aprendizado são inadequadas, a hipótese pode se ajustar pouco ao conjunto de dados de testes, o que é chamado de *underfitting*¹⁰ (GOLDSCHMIDT; PASSOS e BEZERRA, 2015).

Nessa etapa, pode-se utilizar uma das diversas técnicas que serão exploradas posteriormente: *Holdout*, *Cross-Validation*, *Leave-One-Out* e *Bootstrap* (GOLDSCHMIDT; PASSOS e BEZERRA, 2015).

A seguir, serão abordados os tipos de **classificadores**.

⁹ Sobreajuste.

¹⁰ Subajuste.

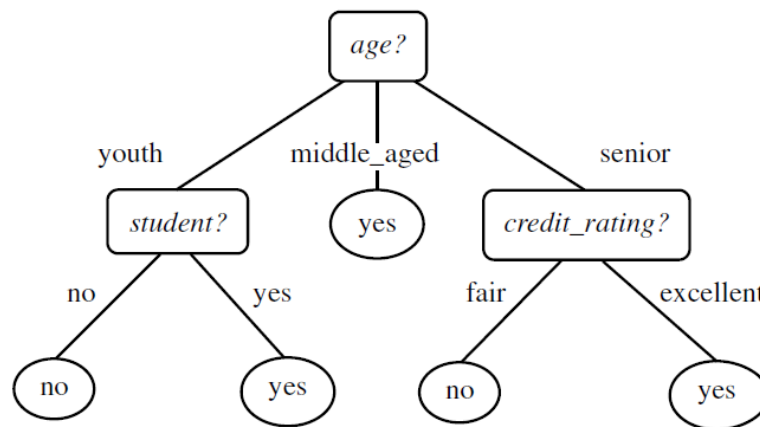
2.4.1. Árvores de decisão

Uma **árvore de decisão** representa uma função que recebe como entrada uma nova entrada, contendo apenas atributos, e associa a uma única **decisão**. A **árvore de decisão** é formada por **nós**, **ramificações** e **folhas**.

Para associar os atributos a uma única decisão, os atributos passam por uma série de testes até o algoritmo retornar essa **decisão**. Cada **nó**, corresponde a um teste para um dos atributos de entrada, e as **ramificações** são classificadas com os valores possíveis para o atributo. Cada **folha** corresponde à **decisão** a que será retornada pela função (RUSSEL e NORVIG, 2013).

A Figura 7, mostra uma árvore de decisão em que são classificados prováveis compradores de computadores. Os nós correspondem aos testes, as ramificações correspondem aos valores de cada atributo e as folhas correspondem às decisões (HAN e KAMBER, 2006).

Figura 7. Árvore de decisão de prováveis compradores de computadores

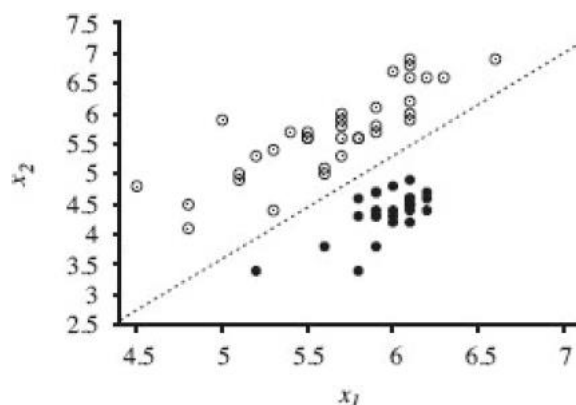


(HAN e KAMBER, 2006)

2.4.2. Modelos lineares

O **modelo linear**, é um modelo de **regressão linear** aplicado na **classificação**. Aplicando esse modelo, o algoritmo aprende uma função hipótese que retorna 1 quando o registro pertence a uma classe e 0 quando não pertence.

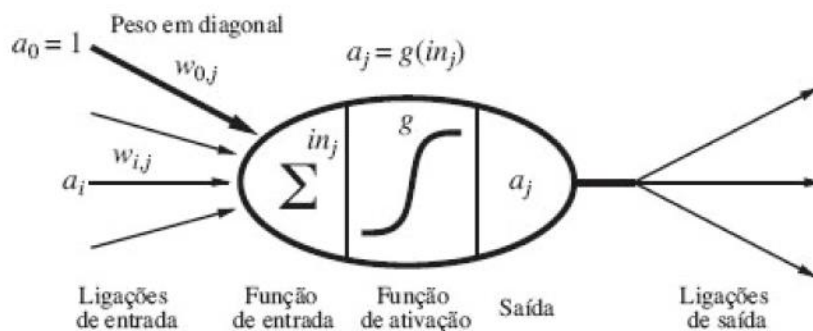
A Figura 8 demonstra como é aplicado o modelo linear. Os pontos no gráfico pertencem a duas classes: terremotos e explosões subterrâneas. Os atributos x_1 e x_2 , correspondem a magnitudes de onda do corpo e da superfície, respectivamente, calculados a partir do sinal sísmico. A tarefa é aprender uma função hipótese, que classificará novas entradas: 0 para terremotos e 1 para explosões (RUSSEL e NORVIG, 2013).

Figura 8. Classificador Linear

(RUSSEL e NORVIG, 2013)

2.4.3. Redes neurais

De forma muito simplificada, existe uma hipótese de que a atividade mental se resume basicamente na atividade eletroquímica em redes de células cerebrais chamadas **neurônios**. Baseados nessa hipótese, McCulloch e Pitts, desenvolveram um modelo matemático simples de um neurônio em 1943 (RUSSEL e NORVIG, 2013), conforme Figura 9.

Figura 9. Modelo matemático simples de um neurônio

(RUSSEL e NORVIG, 2013)

As redes neurais são compostas por **nós** ou **unidades**, representadas na Figura 9, conectadas por ligações direcionadas. Uma ligação de entrada, para a ligação de saída, serve para propagar a **função de ativação**, da função de entrada para a saída. Cada ligação tem um **peso** numérico (RUSSEL e NORVIG, 2013).

A fase de treinamento é utilizada para atribuir os pesos numéricos entre as ligações, ou conexões, e as redes neurais podem ser treinadas tanto como **aprendizado supervisionado**, como **não supervisionado**.

Em **aprendizagem supervisionada**, o algoritmo pode estimar o erro após o treinamento, comparando a classe desejada e a classe prevista, ajustando assim os pesos das conexões.

Utilizando **aprendizagem não supervisionada**, não há informações relativas à classificação desejada, portanto, as classes são definidas conforme a similaridade entre os dados de entrada, de forma parecida com a técnica de agrupamento (GOLDSCHMIDT; PASSOS e BEZERRA, 2015).

2.4.4. Modelos não paramétricos

Modelos **paramétricos**, são modelo em que é definida uma **função hipótese** na etapa de treinamento, e a partir desse ponto, o conjunto de dados de treinamento **não é mais utilizado**. As novas entradas são classificadas apenas a partir da função hipótese. Como exemplo, **regressão linear** e **redes neurais**.

Em contrapartida, um modelo **não paramétrico**, ao gerar a função hipótese, armazena o conjunto de dados de treinamento e o utiliza para classificar as novas entradas. Além disso, a cada nova classificação, esse modelo armazena os novos dados classificados, aumentando assim, o conjunto de dados de treinamento (RUSSEL e NORVIG, 2013). É conhecido também como *Instance-Based Learning*¹¹ (SILVA, PERES e BOSCARIOLI, 2016).

Também conhecido como: **aprendizado baseado em exemplos** ou **aprendizado baseado em memória** (RUSSEL e NORVIG, 2013). O exemplo mais comum, é o algoritmo que será estudado no próximo capítulo: K-vizinhos mais próximos.

2.4.5. Máquinas de vetores de suporte

Esse modelo é similar ao **modelo linear**, porém esse modelo consegue analisar conjuntos de dados que possuam mais de dois atributos através de um **hiperplano** (GOLDSCHMIDT, PASSOS e BEZERRA, 2015). Segundo GOLDSCHMIDT, PASSOS e BEZERRA (2015), **hiperplano** é a generalização do conceito de plano com duas ou mais dimensões.

O modelo **SVM**¹² constrói um **separador de margem máxima**, que é um limite de decisão com a maior distância possível dos pontos de exemplos. Além de também ser um método não paramétrico, ou seja, mantém os exemplos de treinamento. Assim, esse modelo combina as vantagens de um modelo paramétrico e não paramétrico, sendo resistente ao *overfitting* (RUSSEL e NORVIG, 2013).

¹¹ Aprendizado baseado em instâncias

¹² *Support vector machines* ou máquinas de vetores de suporte

2.5. Técnicas de treinamento

Conforme descrito anteriormente, para construir um classificador, é necessário encontrar a função hipótese. Para medir a precisão dessa função é utilizada uma das técnicas a seguir, e métricas que serão exploradas no próximo capítulo.

2.5.1. *Holdout*

Essa técnica consiste na divisão do conjunto de treinamento em dois subconjuntos distintos: um subconjunto para o **treinamento** e outro para realizar os **testes** (GOLDSCHMIDT; PASSOS e BEZERRA, 2015). Embora não exista um valor ótimo, geralmente os subconjuntos podem ser divididos em 70% e 30%, ou 60% e 40%, ambos para **treinamento** e **testes** respectivamente. Os exemplares que serão alocados em cada um dos subconjuntos devem escolhidos aleatoriamente (SILVA; PERES e BOSCAROLI, 2016).

Alternativamente, pode ser considerado o uso de um subconjunto de **validação**. Nesse caso, o conjunto de dados é dividido em três subconjuntos, em que após a construção do modelo preditivo utilizando os subconjuntos de **treinamento** e de **testes**, a precisão do modelo é avaliada utilizando o conjunto de **validação** (SILVA; PERES e BOSCAROLI, 2016).

Algo importante a ressaltar, é que utilizando essa técnica, o subconjunto de treinamento sempre deve conter mais de 50% dos exemplares do conjunto de dados (GOLDSCHMIDT; PASSOS e BEZERRA, 2015).

2.5.2. *K-Fold Cross-validation* ¹³

Nessa técnica, o conjunto de dados é dividido em K subconjuntos distintos. Considerando que o conjunto de dados possui N exemplares, e que esses exemplares serão escolhidos aleatoriamente, o conjunto de dados será dividido de forma que cada subconjunto possuirá a quantidade de exemplares aproximadamente igual (N/K) (GOLDSCHMIDT; PASSOS e BEZERRA, 2015).

Dentre o K subconjuntos, um subconjunto será usado como conjunto de **testes** e os outros $K-1$ subconjuntos serão utilizados como o conjunto de **treinamento**, esse processo será repetido K vezes e a cada iteração o subconjunto de **teste** será diferente, de forma que ao final todos os K subconjuntos tenham sido utilizados uma vez como o conjunto de **testes** (SILVA; PERES e BOSCAROLI, 2016).

Por fim, a avaliação pode ser feita de duas maneiras distintas: aplicando uma técnica estatística com os resultados de cada subconjunto (média, variância, desvio-padrão), ou somando-se os resultados do desempenho de cada subconjunto e dividindo pelo número de exemplares do conjunto de dados original (SILVA; PERES e BOSCAROLI, 2016).

¹³ Validação cruzada com K Conjuntos

2.5.3. *Leave-one-out*

Essa técnica é uma variação da técnica ***K-Fold Cross-validation***, em que K é igual à quantidade de registros (N) (GOLDSCHMIDT; PASSOS e BEZERRA, 2015). Ou seja, cada um dos subconjuntos possuirá exatamente um exemplar (N/K). A cada iteração será utilizado um único exemplar como conjunto de **testes** enquanto o restante dos registros será utilizado como o conjunto de **treinamento**.

Por se tratar de um método computacionalmente dispendioso, sua utilização é recomendada apenas para conjuntos de dados pequenos (GOLDSCHMIDT; PASSOS e BEZERRA, 2015).

2.5.4. *Bootstrap*

Essa técnica é similar ao ***holdout***, onde conjunto de dados é subdividido em dois subconjuntos: **testes** e **treinamento**, com a diferença da **reposição** nesse caso. São realizados N sorteios, para um conjunto de dados com N exemplares, em que cada exemplar sorteado é adicionado ao subconjunto de **treinamento**. Caso um exemplar seja sorteado novamente, ele será adicionado ao subconjunto de **treinamento** novamente. Ao terminarem os N sorteios, os exemplares que não tenham sido sorteados são adicionados ao conjunto de **testes** (GOLDSCHMIDT; PASSOS e BEZERRA, 2015).

2.6. Métricas de avaliação de desempenho

A métrica mais comum utilizada para mensurar o **desempenho** de um modelo preditivo, ou uma função hipótese, é a **acurácia**. A **matriz de confusão** fornece um detalhamento desse **desempenho**. Ambas serão abordadas a seguir.

2.6.1. Acurácia

A **acurácia** pode ser entendida como **taxa de acerto** do modelo preditivo. É obtida pela equação:

$$Acc(h) = 1 - Err(h)$$

Onde $Err(h)$ é a denominada **taxa de erro**, que é obtida pela equação:

$$Err(h) = \frac{1}{n} \sum_{i=1}^n \|y_i \neq h(i)\|$$

A **taxa de erro** é dada pelo somatório do comparativo entre todos os elementos classificados (GOLDSCHMIDT; PASSOS e BEZERRA, 2015).

Onde:

- Caso a expressão $\|y_i \neq h(i)\|$ seja verdadeira irá retornar 1, caso contrário 0;
- y_i é a classe real do i -ésimo elemento;

- $h(i)$ é a classe prevista pelo classificador para o i -ésimo elemento;
- n é o número de exemplos.

2.6.2. Matriz de confusão

É uma matriz de dimensões $C \times C$, em que C representa o número de classes envolvidas no problema de **classificação** (SILVA, PERES e BOSCAROLI, 2016).

A Figura 10 representa uma matriz de confusão, em que as colunas representam as classes preditas, e as linhas representam a classe a que o exemplo realmente pertence. Caso o conjunto de dados possua mais de duas classes, a matriz pode ser analisada como “classe positiva” (pertencente a classe analisada) ou “classe negativa” (demais classes) (SILVA, PERES e BOSCAROLI, 2016).

Cada campo da matriz será preenchido de acordo com o resultado do classificador, de forma a comparar cada classe predita (coluna) com sua classe real (linha).

Figura 10. Matriz de confusão

		Predicted class	
		C_1	C_2
Actual class	C_1	true positives	false negatives
	C_2	false positives	true negatives

(HAN e KAMBER, 2006)

Cada campo possui um significado, segundo SILVA, PERES e BOSCAROLI (2016):

- **VP** - Verdadeiro positivo (*true positives*): o exemplar é pertencente à **classe positiva**, e o classificador o classificou como pertencente à **classe positiva**, ou seja, **classificação correta na classe positiva**;
- **FP** - Falso positivo (*false positives*): o exemplar não é pertencente à **classe positiva**, e o classificador o classificou como pertencente à **classe positiva**, ou seja, **classificação incorreta na classe positiva**;
- **VN** - Verdadeiro negativo (*true negatives*): o exemplar é pertencente à **classe negativa**, e o classificador o classificou como pertencente à **classe negativa**, ou seja, **classificação correta na classe negativa**;
- **FN** - Falso negativo (*false negatives*): o exemplar não é pertencente à **classe negativa**, e o classificador o classificou como pertencente à **classe negativa**, ou seja, **classificação incorreta na classe negativa**.

Uma outra forma de calcular a **acurácia**, é utilizando a matriz de confusão, conforme a seguir.

$$\frac{VP + VN}{VP + VN + FP + FN}$$

Para diferentes formas de avaliar o classificador, existem diversas métricas, e algumas serão abordadas a seguir.

2.6.3. Sensibilidade ou revocação

Taxa de exemplos classificados como **verdadeiros positivos**, onde era esperada a **classe positiva** (SILVA, PERES e BOSCAROLI, 2016).

$$\frac{VP}{VP + FN}$$

2.6.4. Precisão

Taxa de **acertos** de exemplos classificados como **verdadeiros** ou **falsos positivos** (SILVA, PERES e BOSCAROLI, 2016).

$$\frac{VP}{VP + FP}$$

2.6.5. F-Score

Média harmônica entre **precisão** e **sensibilidade**. Onde, S é **sensibilidade** e P é **precisão** (HAN e KAMBER, 2006).

$$\frac{S * P}{\frac{(S + P)}{2}}$$

2.6.6. G-Score

Média geométrica entre **precisão** e **sensibilidade**. Onde, S é **sensibilidade** e P é **precisão** (MENON, AGARWAL e CHAWLA, 2013).

$$\sqrt{P * S}$$

2.7. Similaridade e Distância

Uma vez que um conjunto de dados pode ser interpretado como um conjunto de pontos em um espaço n -dimensional, **similaridade** pode ser definida como a **distância** entre dois pontos (GOLDSCHMIDT, PASSOS e BEZERRA, 2015). Quanto menor a **distância** entre dois pontos, mais **similares** eles serão.

Distância pode ser formalizada como uma função, que associa um valor real a cada par de pontos contidos no conjunto de dados (GOLDSCHMIDT, PASSOS e BEZERRA, 2015). Como função, deve atender às seguintes restrições:

- (i) $d(x, x) = 0$
- (ii) $d(x, y) = d(y, x)$
- (iii) $d(x, y) \leq d(x, z) + d(z, y)$

Onde d é distância, e x , y e z pontos quaisquer do conjunto de dados.

Para cada fórmula de cálculo de distância que será apresentada a seguir, considere x_i e y_i como o i -ésimo registro do conjunto de dados, e n como a quantidade de exemplares do conjunto de dados.

2.7.1. Bray Curtis

Também conhecida como distância de Sorensen, a distância **Bray Curtis** foi proposta em 1948 por Thorvald Sorensen (DEZA e DEZA, 2006).

$$d(x, y) = \frac{\sum_{i=1}^n |x_i - y_i|}{\sum_{i=1}^n (x_i + y_i)}$$

2.7.2. Canberra

A distância **Canberra** foi proposta em 1967 por William T. Williams e Godfrey Lance (DEZA e DEZA, 2006).

$$d(x, y) = \sum_{i=1}^n \frac{|x_i - y_i|}{|x_i| + |y_i|}$$

2.7.3. Chebyshev

A distância **Chebyshev** foi proposta por Patnuty Chebyshev (DEZA e DEZA, 2006).

$$d(x, y) = \max_i |x_i - y_i|$$

2.7.4. Euclidiana

A distância **Euclidiana** é também conhecida por distância de **Pitágoras** (DEZA e DEZA, 2006).

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

2.7.5. Manhattan

A distância de **Manhattan** é também conhecida como distância **City-Block** (DEZA e DEZA, 2006).

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

2.7.6. Minkowski

A distância de **Minkowski** utiliza um argumento p . Se $p = 1$, o resultado será o mesmo da distância de **Manhattan**, se $p = 2$, o resultado será o mesmo da distância **Euclidiana** (GOLDSCHMIDT, PASSOS e BEZERRA, 2015).

$$d(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}$$

2.7.7. Hamming

Proposta em 1950, por Richard Hamming, (DEZA e DEZA, 2006) essa distância também pode ser utilizada para calcular a distância entre atributos categóricos. Caso os atributos sejam iguais, a distância entre eles é 0, caso contrário é 1.

$$d(x, y) = \sum_{i=1}^n |x_i \neq y_i|$$

2.7.8. Correlacional

Também conhecido como **Coefficiente de Correlação de Pearson**, a **Similaridade Correlacional** é uma métrica que retorna um valor de -1 a 1, indicando o quão similar são os exemplares x e y (DEZA e DEZA, 2006).

$$s(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x}) (y_i - \bar{y})}{\sqrt{(\sum_{i=1}^n (x_i - \bar{x})^2) (\sum_{i=1}^n (y_i - \bar{y})^2)}}$$

Para calcular a distância entre os exemplares, é utilizada a fórmula a seguir.

$$d(x, y) = \sqrt{2 (1 - s)}$$

2.7.9. Cosseno

Também conhecida como **Similaridade de Orchini**, a **Similaridade de Cosseno** pode ser calculada de duas formas, a primeira para atributos numéricos (DEZA e DEZA, 2006).

$$\cos \theta = \frac{\sum_{i=1}^n (x_i * y_i)}{\sqrt{\sum_{i=1}^n (x_i)} \sqrt{\sum_{i=1}^n (y_i)}}$$

A segunda forma é utilizada para calcular a similaridade entre atributos categóricos. Onde n representa o tamanho de cada exemplar.

$$\cos \theta = \frac{x \cap y}{\sqrt{n(x)} \sqrt{n(y)}}$$

Para calcular a distância entre os exemplares, é utilizada a fórmula a seguir.

$$d(x, y) = 1 - \cos \theta$$

2.7.10. Jaccard

Também conhecida como **Similaridade de Tanimoto**, a **Similaridade de Jaccard** foi proposta por Paul Jaccard, em 1908 (DEZA e DEZA, 2006). A similaridade é obtida pela divisão da intersecção entre dois exemplares x e y pela união entre os exemplares x e y .

$$j(x, y) = \frac{x \cap y}{x \cup y}$$

Para calcular a distância entre os exemplares, é utilizada a fórmula a seguir.

$$d(x, y) = 1 - j$$

2.8. Estatística

No decorrer do trabalho serão utilizadas as fórmulas que serão descritas abaixo.

2.8.1. Média

Segundo LARSON e FARBER (2010) a **média** de um conjunto de dados é a soma das entradas de dados dividida pelo número de entradas. É obtida pela fórmula:

$$\bar{m} = \frac{\sum x}{n}$$

Onde n é a quantidade de registros da amostra.

2.8.2. Variância

A variância é uma medida de dispersão de uma distribuição ao redor de sua média. A interpretação dessa métrica, é de que valores maiores significam que a dispersão ao redor da média é maior. Valores menores significam que a dispersão está mais próxima a média, ou seja, x é pouco variável (CASELLA e BERGER, 2016). É dada pela fórmula:

$$\delta^2 = \frac{\sum (x - \bar{m})^2}{n}$$

(LARSON e FARBER, 2010)

2.8.3. Desvio padrão

Similar à variância, é interpretado da mesma forma. Porém, a interpretação se torna mais fácil, uma vez que a unidade de medida é a mesma da distribuição, ao passo que a variância usa a unidade de medida original elevada ao quadrado (CASELLA e BERGER, 2016). É dada pela fórmula:

$$\sigma = \sqrt{\delta^2}$$

([LARSON e FARBER, 2010](#))

2.9. Matemática

2.9.1. Logaritmo

Logaritmo é uma função dada por:

$$f(x) = \log_a x$$

Sejam $a > 0$, $a \neq 1$, e $x > 0$, dois números reais quaisquer, então existe um único y real, tal que ([GUIDORIZZI, 2001](#)):

$$a^y = x$$

Capítulo 3.

O ALGORITMO K-VIZINHOS MAIS PRÓXIMOS

Nesse capítulo será abordado o algoritmo KNN ¹⁴. O KNN é um algoritmo utilizado na fase de **Mineração de Dados**, quando o objetivo é a **descoberta preditiva**.

3.1. Definição

O algoritmo KNN é um algoritmo de **aprendizagem de máquina supervisionada**, e se trata de um **modelo não-paramétrico**, que pode ser utilizado para resolver problemas de **classificação** ou de **regressão**.

Nesse trabalho, o algoritmo será abordado com ênfase na **classificação**.

3.1.1. Algoritmo

O algoritmo KNN foi descrito pela primeira vez no início dos anos 50. Porém, como se trata de **modelo não-paramétrico**, só ganhou popularidade na década de 60, junto com o aumento do poder computacional. Desde então, tem se tornado bastante popular na área de classificação (HAN e KAMBER, 2006).

O KNN é um **modelo não-paramétrico**, portanto armazena seu **conjunto de treinamento** para realizar a classificação. Cada exemplar desse **conjunto de treinamento** é composto por n atributos, que representa um ponto no espaço n -dimensional (HAN e KAMBER, 2006), e cada atributo representa uma dimensão nesse espaço.

Para realizar a classificação de um novo exemplar, o algoritmo KNN funciona da seguinte forma (GOLDSCHMIDT, PASSOS e BEZERRA, 2015):

- Cálculo da distância, ou similaridade, do novo exemplar para os exemplares pertencentes ao **conjunto de treinamento**;
- Identificação dos k exemplares do **conjunto de treinamento** que apresentarem a menor distância para o novo exemplar;

¹⁴ K-Vizinhos Mais Próximos, do termo em inglês *K-Nearest Neighbor*

- Predição da classe do novo exemplar, que será a classe que tem mais votos, ou seja, a mais frequente entre esses k exemplares;
- Inserção do novo exemplar no **conjunto de treinamento**.

A seguir o pseudocódigo do algoritmo KNN. Considere x como conjunto de treinamento, y o exemplar a ter sua classe predita e k o número de vizinhos.

Entradas: x, y, k

Saída: rótulo predito pelo algoritmo para o novo exemplar

início

para $i \leftarrow \text{até fim}(x)$ **faça**

 Calcular distância $d[i] = (x_i, y)$

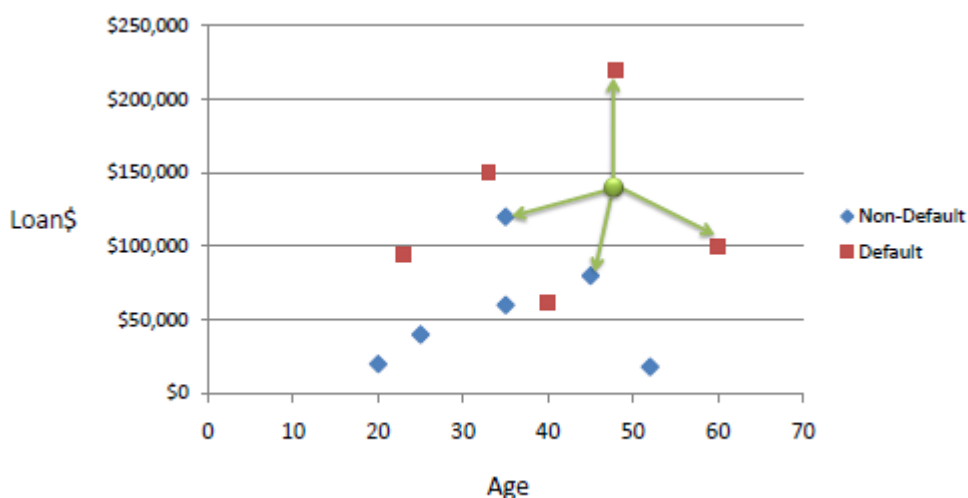
fim-para

identifique os k exemplares com a menor distância contida em d

retorne a classe com mais votos, entre os k exemplares

fim

Figura 11. Representação da aplicação do algoritmo KNN



(SAYAD, 2017)

Na [Figura 11](#), é representada a aplicação do algoritmo KNN em um problema de **classificação**, em que pessoas são classificadas, com os atributos *loan*¹⁵ e *age*¹⁶. As classes possíveis são as classes *default*¹⁷ e *non-default*¹⁸.

¹⁵ Empréstimo

¹⁶ Idade

¹⁷ Padrão ou comum

¹⁸ Incomum

O ponto verde é o exemplar que será classificado, e no caso do exemplo, os 4 exemplares mais próximos provocam um empate. Se fossem olhados 3 ou 5 exemplares, o empate não ocorreria, pois o novo exemplar seria classificado como *non-default* e *default*, respectivamente.

3.2. Implementação

A seguir, será abordada a implementação do algoritmo que foi utilizada nesse trabalho. Não foi utilizada nenhuma biblioteca, o algoritmo foi inteiramente construído e pode ser observado com mais detalhes no [Apêndice – Código Fonte](#).

3.2.1. Pré-processamento e transformação

Para a etapa de **pré-processamento**, foi implementada uma técnica para retirar os registros repetidos.

Para a etapa de **transformação**, foi implementada uma técnica para normalizar os dados. Utilizando essa técnica, os atributos de todos os exemplares do conjunto de treinamento ficam na mesma escala, entre 0 e 1. Do contrário, atributos que estejam em uma escala de valores maior influenciam os resultados de forma tendenciosa ([GOLDSCHMIDT, PASSOS e BEZERRA, 2015](#)). A técnica implementada foi a **normalização linear**, a seguir sua fórmula.

$$a' = \frac{a - \min}{\max - \min}$$

Onde:

- a' : valor normalizado;
- a : valor original;
- \min : *min* de a ;
- \max : *max* de a .

3.2.2. Treinamento

Para a etapa de treinamento, foi implementada a técnica **holdout**, em que o conjunto de dados é dividido em dois subconjuntos: **treinamento** e **teste**. Para aumentar a eficiência nessa etapa, e diminuir o *overfitting*, foram propostas duas técnicas.

A primeira, é **embalhar o conjunto de dados** antes de aplicar a técnica **holdout**, assim a cada rodada os dados nunca serão os mesmos.

Para diminuir as chances de **desbalancear** as classes na divisão, ou seja, diminuir as chances de um subconjunto possuir mais registros de determinada classe que o outro, foi implementada uma solução para manter o **balanceamento**. Ou seja, no momento da aplicação da técnica **holdout**, é mantida a proporção das classes entre os subconjuntos.

3.2.3. Classificação

Para a classificação propriamente dita, foram implementadas as **distâncias** descritas no capítulo anterior, com algumas propostas.

A primeira proposta é para a **classificação de conjuntos de dados** com atributos mistos, ou seja, atributos **numéricos** e **categóricos**, a proposta é que para os atributos **numéricos**, a distância seja calculada conforme a técnica escolhida, porém para atributos **categóricos**, seja utilizada a distância de **Hamming**, e no fim as duas distâncias sejam somadas.

Outra proposta é o tratamento para conjuntos de dados com **atributos faltantes**. Para esse assunto, foram propostas duas soluções.

Na primeira solução o algoritmo trata de forma similar a classificação com atributos mistos, onde também é utilizada a distância de **Hamming**. Ao calcular a distância para um exemplar com atributos faltantes, o cálculo dos atributos presentes é feito normalmente. Já os atributos faltantes são comparados, e no caso de os atributos nos dois exemplares estarem faltantes, o resultado dessa comparação será 1, em caso contrário será 0.

Na segunda solução é feito do preenchimento de atributos com **medidas estatísticas**, o atributo faltante é preenchido com a **média** daquele atributo no conjunto de dados.

Outra proposta é para solucionar o problema de empate na classificação. Conforme o exemplo abordado, a classificação pode empatar no caso de não haver uma classe com mais votos entre os k -vizinhos mais próximos, nesse caso, foi proposto que seja calculada a **média** ou o **desvio-padrão** entre as **distâncias** entre o exemplar que está sendo classificado e os exemplares pertencentes ao **conjunto de treinamento**.

O desvio-padrão foi proposto, pois mesmo com uma probabilidade muito baixa, a **média** pode empatar, já o **desvio-padrão** possui uma probabilidade muito menor de empate, já que leva em conta o quanto as **distâncias** entre os elementos variaram.

O treinamento em si é feito após a aplicação do método **holdout**, em que o subconjunto de **testes** é classificado, e o algoritmo usa o subconjunto de **treinamento** para realizar as classificações. Após as classificações, são retornadas as métricas de avaliação, a matriz de confusão, para que seja analisado o desempenho do algoritmo.

3.2.4. Escolha do k

No algoritmo original, o valor de k é passado como parâmetro. A proposta foi realizar a classificação dos dados usando um intervalo de valores para k , e ao fim, avalia-se qual valor k apresentou o melhor resultado, **acurácia** nesse caso. Assim, é possível descobrir qual é o melhor valor de k para posteriormente classificar dados do mundo real.

Capítulo 4.

BASES DE DADOS UTILIZADAS

As bases utilizadas nesse trabalho foram retiradas do *site* KEEL ¹⁹([KEEL, 2017](#)). O KEEL é um *site* bem conhecido e utilizado na comunidade de estudantes e pesquisadores de mineração de dados e aprendizagem de máquina. O KEEL fornece uma ferramenta de código aberto que pode ser utilizada para mineração de dados, utilizando diferentes algoritmos de aprendizado, modelos híbridos ou metodologias estatísticas. Além disso, o *site* KEEL fornece um repositório de base de dados para ser utilizado em pesquisas.

Todas as bases utilizadas são bases classificadas, ou seja, todos os registros possuem rótulos, já que esse estudo se baseia em um algoritmo de aprendizagem supervisionada. As bases foram escolhidas levando-se em conta: quantidade de classes, quantidade de instâncias, quantidade e tipos de atributos. Mais detalhes podem ser vistos no [Apêndice B – Conjunto de Dados](#).

4.1. Bases padrão

Tratam-se de bases de dados comuns, sem peculiaridades.

4.1.1. *Heart*

O objetivo dessa base de dados se resume a obtenção de um modelo para classificar pacientes com uma determinada doença cardíaca.

- Classes: 2
- Instâncias: 270
- Atributos: 13 (1 real e 12 inteiros)

¹⁹ *Knowledge Extraction based on Evolutionary Learning*

4.1.2. *Mammographic Mass*

O objetivo nessa base é classificar a severidade da massa mamária identificada em um exame de mamografia.

- Classes: 2
- Instâncias: 830
- Atributos: 5 (inteiros)

4.1.3. *Red Wine Quality*

O objetivo dessa base de dados é classificar a qualidade de Vinhos Tintos Verdes Portugueses.

- Classes: 11
- Instâncias: 1599
- Atributos: 11 (reais)

4.1.4. *Titanic*

O objetivo dessa base de dados é classificar se tripulantes a bordo do Titanic sobreviveram ou não.

- Classes: 2
- Instâncias: 2201
- Atributos: 3 (reais)

4.1.5. *Vehicle Silhouettes*

O objetivo dessa base de dados é classificar o tipo de um determinado carro.

- Classes: 4
- Instâncias: 846
- Atributos: 18 (inteiros)

4.2. Bases padrão com atributos categóricos

Tratam-se de bases de dados comuns, porém com atributos numéricos e categóricos.

4.2.1. *Automobile*

O objetivo dessa base de dados é classificar o risco de um carro em relação a seu preço, de acordo com suas características, classificação de risco de acordo com a companhia de seguro e a desvalorização em comparação com outros automóveis.

- Classes: 6
- Instâncias: 150
- Atributos: 25 (15 reais e 10 categóricos)

4.2.2. *German Credit*

O objetivo nessa base é classificar se clientes são bons ou ruins, com base em suas características e dados bancários.

- Classes: 2
- Instâncias: 1000
- Atributos: 20 (7 inteiros e 13 categóricos)

4.2.3. *Lymphography*

O objetivo dessa base de dados é classificar se o paciente possui um linfoma (tipo de câncer) e em caso positivo, qual o seu estágio.

- Classes: 4
- Instâncias: 148
- Atributos: 18 (3 inteiros e 15 categóricos)

4.2.4. *South African Heart*

O objetivo dessa base de dados é classificar se paciente possui ou não uma doença cardíaca. Os dados foram colhidos de pacientes homens, residentes de uma região com alto risco de doença cardíaca no Cabo Ocidental, África do Sul.

- Classes: 2
- Instâncias: 462
- Atributos: 9 (5 reais, 3 inteiros e 1 categórico)

4.3. Bases desbalanceadas

Tratam-se de bases de dados que possuem uma ou mais classes muito menores que as outras ([FERNÁNDEZ, GARCÍA, et al., 2009](#));

4.3.1. *Ecoli*

O objetivo dessa base de dados é prever a localização de proteínas. Como se trata de uma adaptação para classes desbalanceadas, possui a classe positiva (pertencente à classe original “pp”), com 15.48% das entradas e a classe negativa (pertencente às demais classes da base de dados original) com 84.52% das entradas. A proporção de desbalanceamento é de 5.46.

- Classes: 2
- Instâncias: 336
- Atributos: 7 (reais)

4.3.2. Iris

O objetivo dessa base de dados é prever a espécie de flores do gênero Íris, como se trata de uma adaptação para classes desbalanceadas, possui a classe positiva (pertencente à classe original “Iris-Setosa”), com 33.33% das entradas e a classe negativa (pertencente as demais classes da base de dados original) com 66.67% das entradas. A proporção de desbalanceamento é de 2.00.

- Classes: 2
- Instâncias: 150
- Atributos: 4 (reais)

4.3.3. Pima

O objetivo dessa base de dados é prever se um paciente possui diabetes ou não. É composta por registros de mulheres com menos de 21 anos de origem Pima. Possui a classe positiva, com 34.84% das entradas e a classe negativa com 66.16% das entradas. A proporção de desbalanceamento é de 1.90.

- Classes: 2
- Instâncias: 768
- Atributos: 8 (reais)

4.3.4. *Vehicle Silhouettes*

O objetivo dessa base de dados é classificar o tipo do carro. Possui a classe positiva (pertencente à classe original “Saab”), com 28.37% das entradas e a classe negativa (pertencente às demais classes da base de dados original) com 71.63% das entradas. A proporção de desbalanceamento é de 2.52.

- Classes: 2
- Instâncias: 846
- Atributos: 18 (inteiros)

4.3.5. Wisconsin

O objetivo dessa base de dados é classificar se pacientes que foram submetidos a cirurgia para câncer de mama possuem um tumor benigno ou maligno. Possui a classe benigno, com

65% das entradas e a classe maligno com 35% das entradas. A proporção de desbalanceamento é de 1.86.

- Classes: 2
- Instâncias: 683
- Atributos: 9 (inteiros)

4.4. Bases com atributos faltantes

Bases de dados em que algumas entradas possuem atributos que não possuem valores.

4.4.1. *Cylinder Bands*

O objetivo dessa base de dados é classificar, em um processo de impressão rotogravura, se determinada peça é um cilindro ou não.

- Classes: 2
- Instâncias: 539
- Atributos: 19 (13 reais e 6 inteiros)

4.4.2. *Heart Disease (Cleveland)*

O objetivo dessa base de dados é detectar se um paciente, morador de Cleveland, possui ou não uma doença cardíaca.

- Classes: 5
- Instâncias: 303
- Atributos: 13 (reais)

4.4.3. *Hepatitis*

O objetivo dessa base de dados é prever se pacientes portadores de hepatite irão sobreviver ou morrer.

- Classes: 2
- Instâncias: 155
- Atributos: 19 (2 reais e 17 inteiros)

4.4.4. *Mammographic Mass*

O objetivo nessa base é classificar a severidade da massa mamária identificada em um exame de mamografia.

- Classes: 2
- Instâncias: 961

- Atributos: 5 (5 inteiros)

4.5. Bases com atributos categóricos

Bases de dados que só possuem entradas com atributos categóricos.

4.5.1. *Car Evaluation*

O objetivo dessa base de dados é classificar a aceitação de um modelo de carro.

- Classes: 4
- Instâncias: 1728
- Atributos: 16 (categóricos)

4.5.2. *Tic-Tac-Toe Endgame*

O objetivo dessa base de dados é prever se um jogador ganhou ou perdeu uma partida de jogo da velha.

- Classes: 2
- Instâncias: 958
- Atributos: 9 (categóricos)

4.6. Bases com ruído de atributos

Tratam-se de bases que possuem ruído de atributos, ou seja, atributos errados que não pertencem a sua respectiva classe.

Todas as bases dessa categoria foram utilizadas 4 vezes, com 5%, 10%, 15% e 20% de ruídos.

4.6.1. *Contraceptive Method Choice*

O objetivo dessa base de dados é prever o método contraceptivo escolhido por uma mulher, com base em suas características (demográficas e socioeconômicas). Os dados foram colhidos na Indonésia.

- Classes: 3
- Instâncias: 1473
- Atributos: 9 (inteiros)

4.6.2. *Heart*

O objetivo dessa base de dados é prever se um paciente possui ou não uma doença cardíaca.

- Classes: 2
- Instâncias: 270
- Atributos: 13 (1 real e 12 inteiros)

4.6.3. Iris

O objetivo dessa base de dados é prever a espécie de flores do gênero Íris.

- Classes: 3
- Instâncias: 150
- Atributos: 4 (reais)

4.6.4. Pima

O objetivo dessa base de dados é prever se um paciente possui diabetes ou não. A base é composta por registros de mulheres com menos de 21 anos de origem Pima.

- Classes: 2
- Instâncias: 768
- Atributos: 8 (reais)

4.6.5. Sonar, Mines vs. Rocks

O objetivo dessa base de dados é prever o objeto analisado, através de um Sonar ²⁰, é uma rocha ou um cilindro metálico.

- Classes: 2
- Instâncias: 208
- Atributos: 60 (reais)

4.7. Bases com ruído de classes

Tratam-se de bases de dados com ruído de classes, ou seja, dados que estão inseridos na base de treinamento estão erroneamente classificados.

As bases utilizadas nessa categoria são as mesmas da categoria anterior, porém com ruído de classes ao invés de ruído de atributos, e todas as bases dessa categoria foram utilizadas 4 vezes, com 5%, 10%, 15% e 20% de ruídos.

²⁰ do inglês Sound Navigation and Ranging ou Navegação e Determinação da Distância pelo Som.

Capítulo 5.

EXPERIMENTOS

Para realizar os experimentos, o algoritmo K-Vizinhos Mais Próximos foi implementado com a linguagem C#, utilizando processamento *multithreading* ²¹ no cálculo de distâncias.

Os experimentos foram realizados em um computador equipado com um processador Intel Core i7-5500U com 2.40GHz de velocidade.

De forma geral, para cada base de dados foram comparadas as formas distintas de cálculo de distância e desempate. Os testes com balanceamento de classes, normalização de dados e retirada de elementos repetidos foram feitos com todas essas métricas juntas, ou em sua total ausência.

De forma específica, para bases de dados com atributos faltantes, serão comparadas também as métricas de substituir valores faltantes pela média, ou de utilizar a distância de Hamming nos atributos faltantes.

Nas bases de dados categóricas, foram utilizadas métricas de cálculo de distância específicas para atributos categóricos.

Em todos os testes, os resultados foram sobre o melhor K , encontrado pelo algoritmo conforme foi proposto. O K mínimo possível é o 3, para evitar o *overfitting* (RUSSEL e NORVIG, 2013) e para o K máximo foi utilizado \log para que o K máximo não cresça muito, evitando o *underfitting*. Para calcular o K máximo foi utilizado $\log_2 n$, onde n é a quantidade de exemplares da base de dados.

Para cada base de dados, foram feitos 25 testes para cada configuração de parâmetros.

Nesse capítulo serão analisados os resultados com o método de desempate desvio padrão, pois a diferença com a média se mostrou irrelevante. Para mais resultados e tabelas completas dos resultados analisados, consultar o [Apêndice C – Resultados](#).

²¹ Processamento paralelo, ou simultâneo

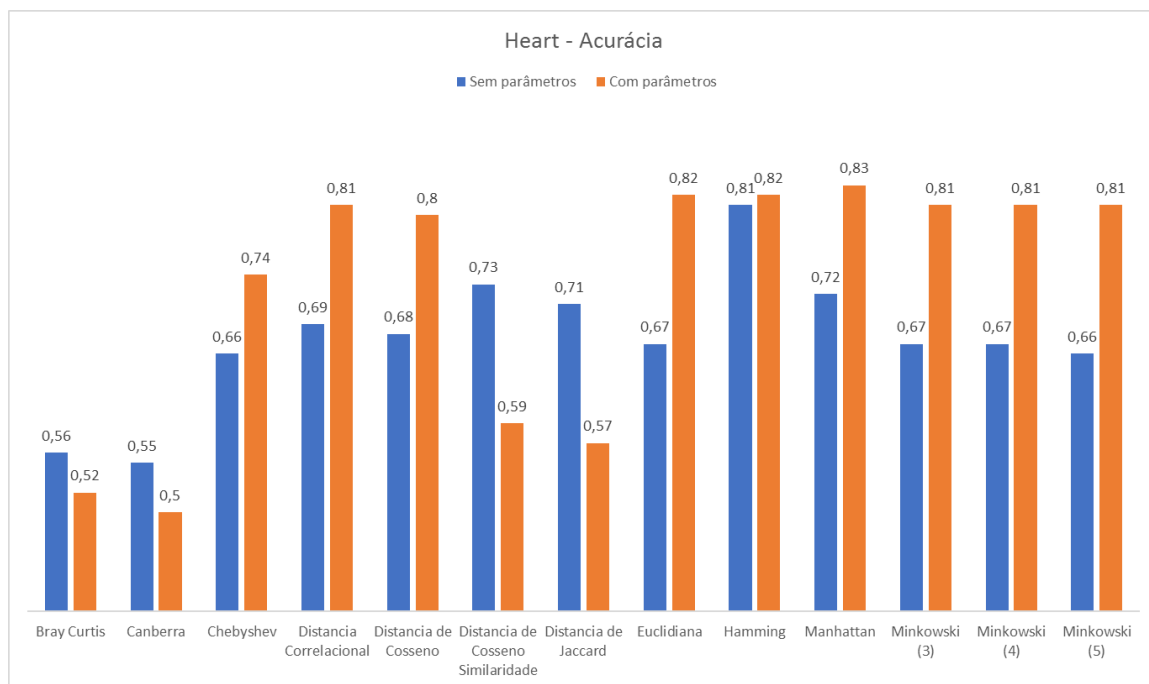
5.1. Bases padrão

5.1.1. *Heart*

Conforme a [Figura 12](#), os melhores resultados ficam para as distâncias Manhattan, Hamming e Euclidiana, por se tratar de uma base de dados padrão, com atributos inteiros. As diferenças entre o desempenho utilizando os parâmetros e não utilizando-os, é bastante considerável na maioria dos casos, apenas na distância de Hamming que não oferece uma diferença considerável por se tratar de uma métrica que apenas realiza comparação entre os dados.

Os piores resultados foram gerados pelas distâncias Bray Curtis e Canberra.

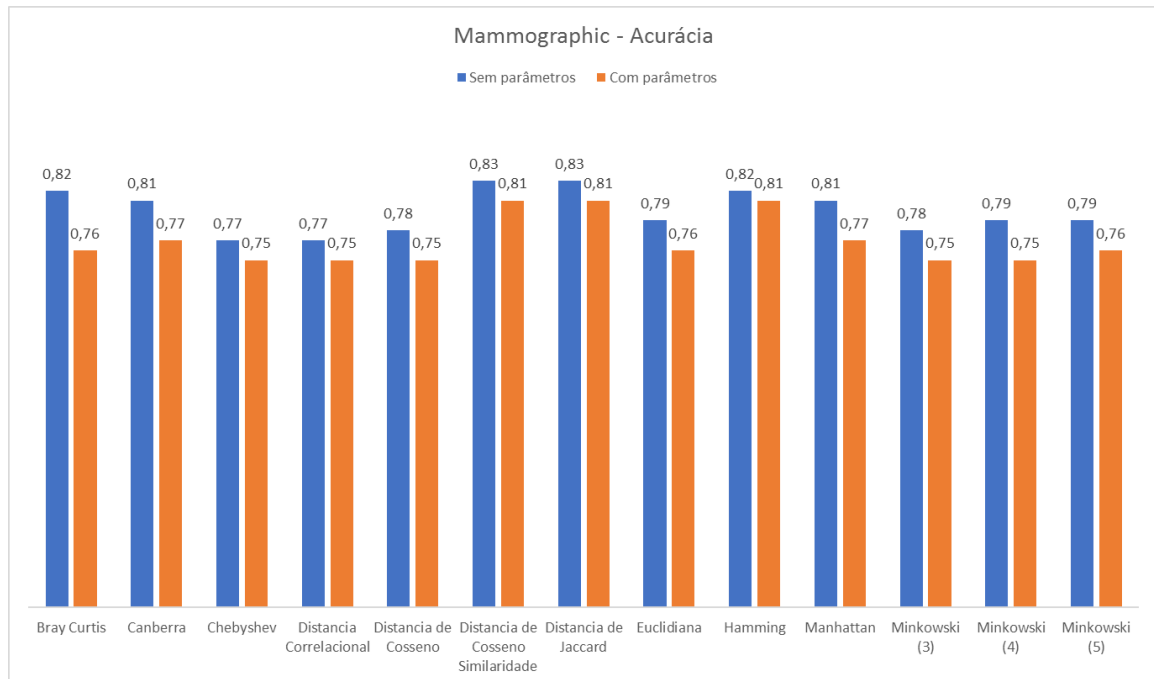
Figura 12. Resultados de classificação – Bases padrão – Heart Disease



5.1.2. *Mammographic Mass*

Conforme a [Figura 13](#), como essa base de dados possui atributos inteiros, e com um intervalo pequeno, os parâmetros se mostraram como uma opção ruim, uma vez que o algoritmo mostrou um desempenho superior ao não os utilizar. Os melhores resultados foram obtidos pelas distâncias de Cosseno Correlacional e Jaccard, pois a base de dados possui uma alta similaridade entre seus exemplares.

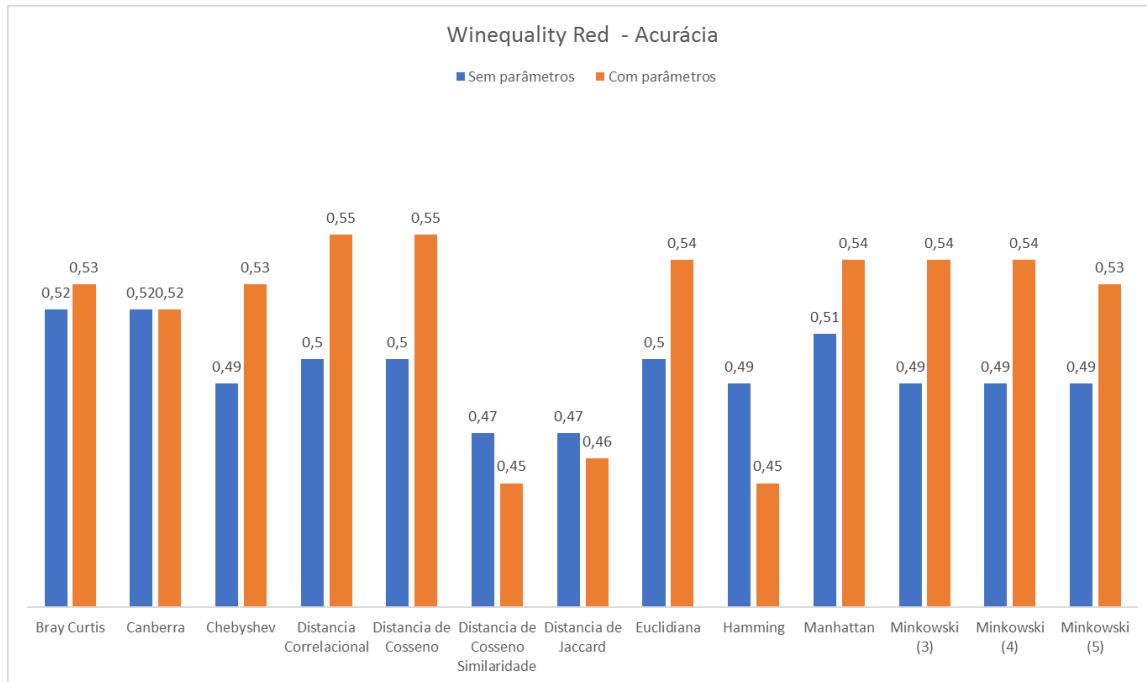
Em compensação, mesmo os piores resultados, Chebyshev e Distância Correlacional, não demonstraram uma diferença significativa de desempenho.

Figura 13. Resultados de classificação – Bases padrão – Mammographic Mass

5.1.3. Red Wine Quality

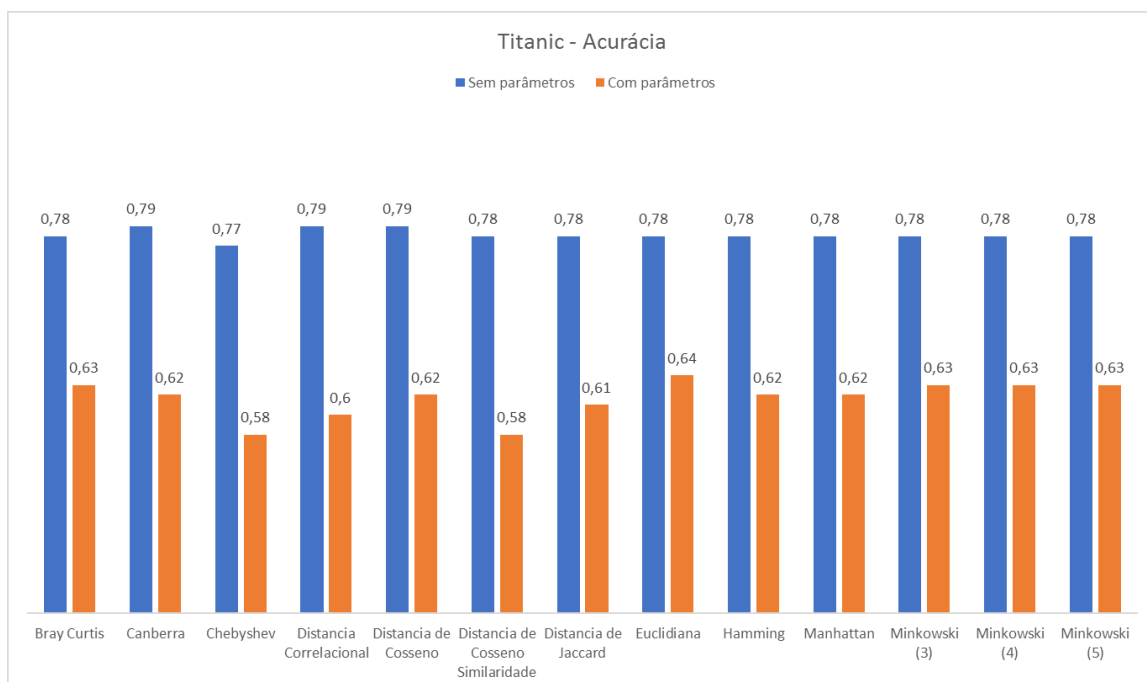
Essa base de dados apresenta atributos com intervalos de valores bastante diferentes, portanto, o desempenho geral do algoritmo foi melhor utilizando os parâmetros propostos. Em geral, o desempenho foi bastante parecido, entre 53% e 55% de acurácia (vide [Figura 13](#)).

Os piores resultados aconteceram utilizando as distâncias de Jaccard, Cosseno Similaridade e Hamming, por se tratar de uma base de dados com baixa similaridade.

Figura 14. Resultados de classificação – Bases padrão – Winequality Red

5.1.4. *Titanic*

Apesar dessa base de dados possuir atributos reais, eles se referem a classes, e não a valores numéricos. Portanto, utilizar os parâmetros propostos não apresenta uma boa solução. Todas as métricas apresentam o desempenho parecido, que varia entre 77% e 79% de acurácia (vide [Figura 15](#)).

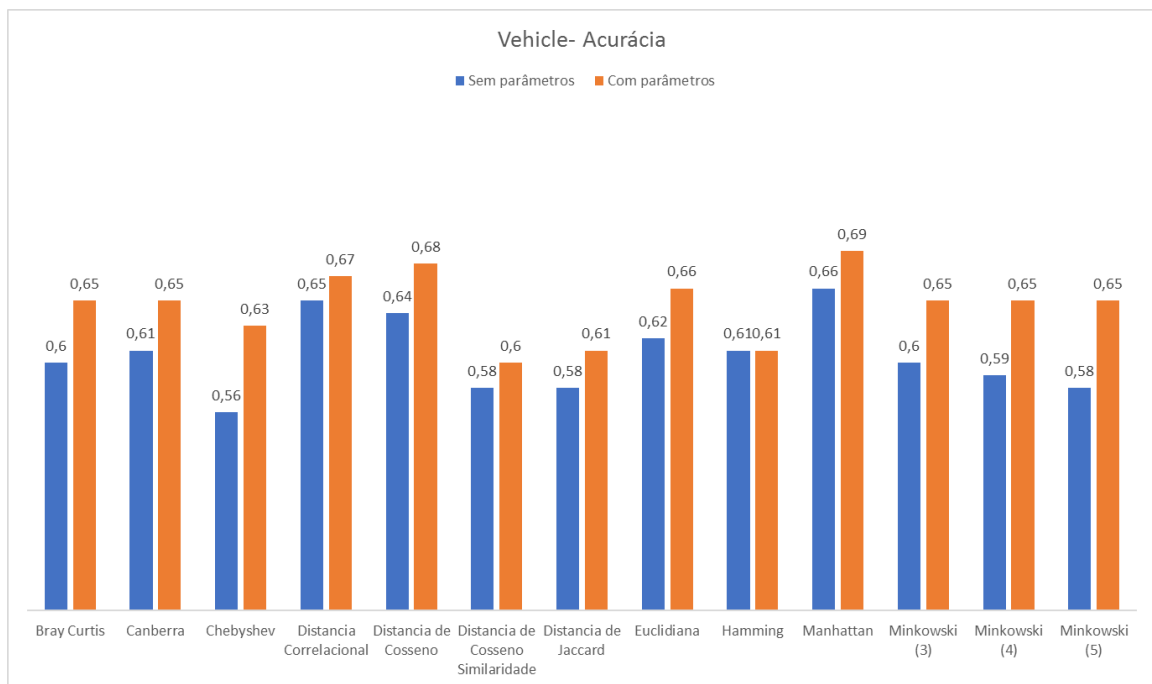
Figura 15. Resultados de classificação – Bases padrão – Titanic

Nesse caso, a melhor distância a ser utilizada é a distância que apresentar um menor tempo de execução, conforme será apresentado ao final dos experimentos.

5.1.5. *Vehicle Silhouettes*

Essa base de dados apresenta atributos inteiros, em diferentes escalas de valores, o que ocasiona em melhores resultados utilizando os parâmetros propostos, como se observa na Figura 16.

Figura 16. Resultados de classificação – Bases padrão – Vehicle Silhouettes



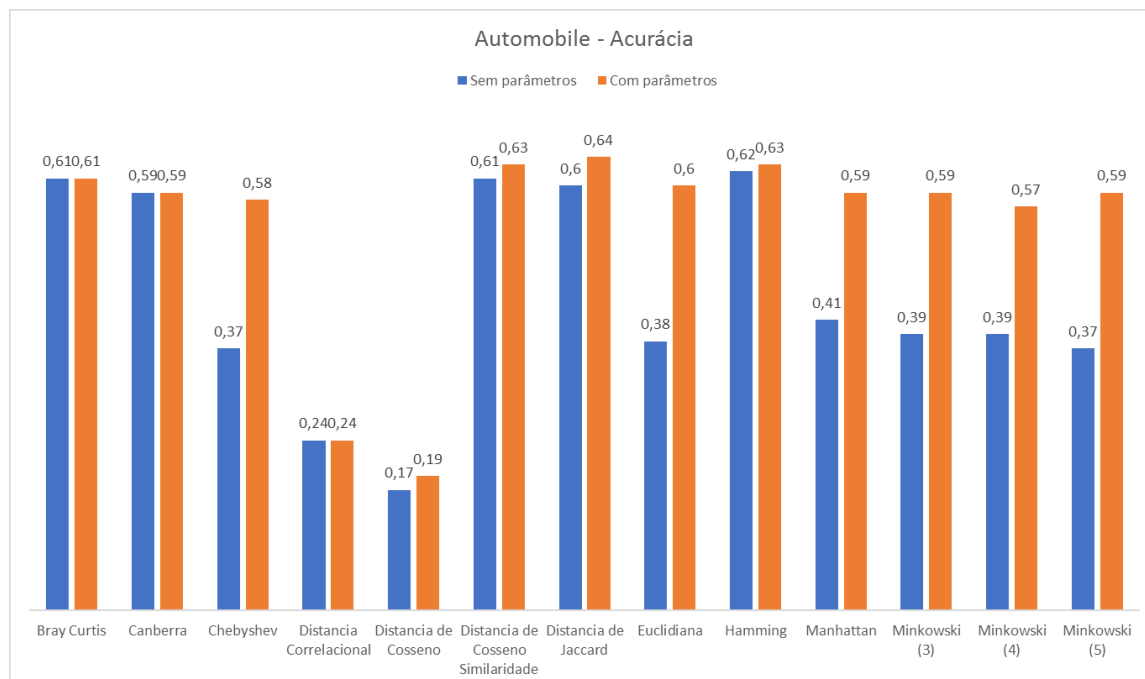
Os melhores resultados foram obtidos pelas distâncias de Manhattan e Distância de Cosseno. Os piores resultados foram obtidos pelas distâncias de Cosseno Similaridade, Jaccard e Hamming, por se tratar de uma base de dados com baixa similaridade.

5.2. Bases padrão com atributos categóricos

5.2.1. *Automobile*

Os melhores resultados foram obtidos pelas distâncias que utilizam a similaridade dos dados, como Cosseno Similaridade, Jaccard e Hamming, com uma pequena vantagem utilizando os parâmetros propostos (vide Figura 17).

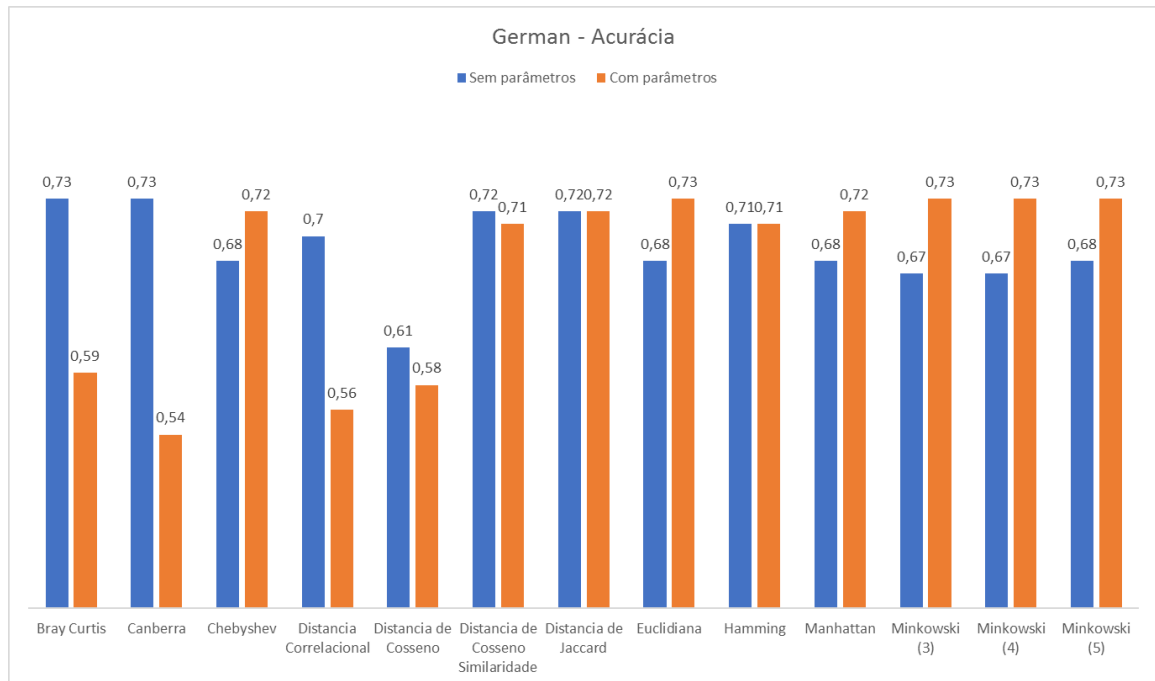
Os piores resultados foram obtidos pelas distâncias Correlacional e de Cosseno, pois ambas utilizam atributos numéricos para compor o divisor e retornar o coeficiente de similaridade, como essa base possui muitos atributos categóricos, o coeficiente não é calculado de forma correta.

Figura 17. Resultados de classificação – Bases padrão c/ atr. categóricos – Automobile

5.2.2. German Credit

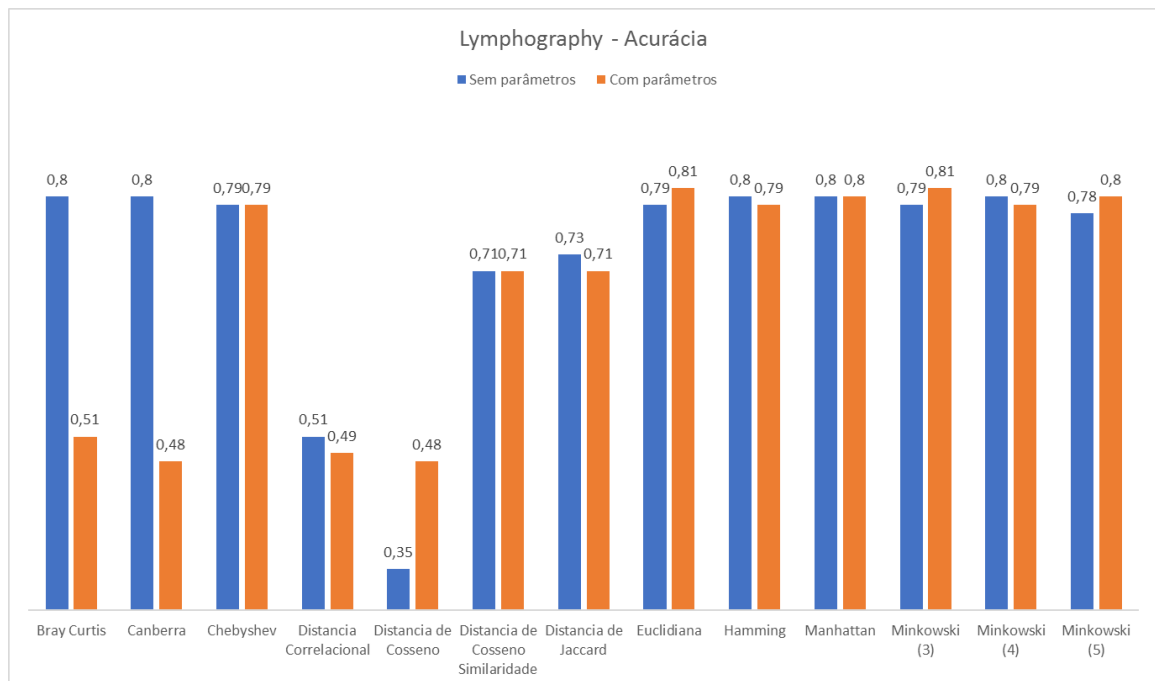
Ao utilizar os parâmetros propostos, as distâncias Euclidiana e Minkowski apresentaram os melhores resultados, porém empataram com as distâncias Bray Curtis e Canberra, quando utilizadas sem os parâmetros propostos, como mostra a [Figura 18](#). Esse desempenho se deve ao fato que o intervalo de dados dos atributos numéricos é pequeno.

Os piores resultados foram obtidos pela distância de Cosseno, pelo mesmo motivo do experimento anterior.

Figura 18. Resultados de classificação – Bases padrão c/ atr. categóricos – German Credit

5.2.3. Lymphography

Essa base de dados possui apenas três atributos numéricos, com um intervalo pequeno. Os melhores resultados foram obtidos pela distância Euclidiana e Minkowski, com o parâmetro 3, e utilizando os parâmetros propostos, conforme se observa na [Figura 19](#).

Figura 19. Resultados de classificação – Bases padrão c/ atr. categóricos – Lymphography

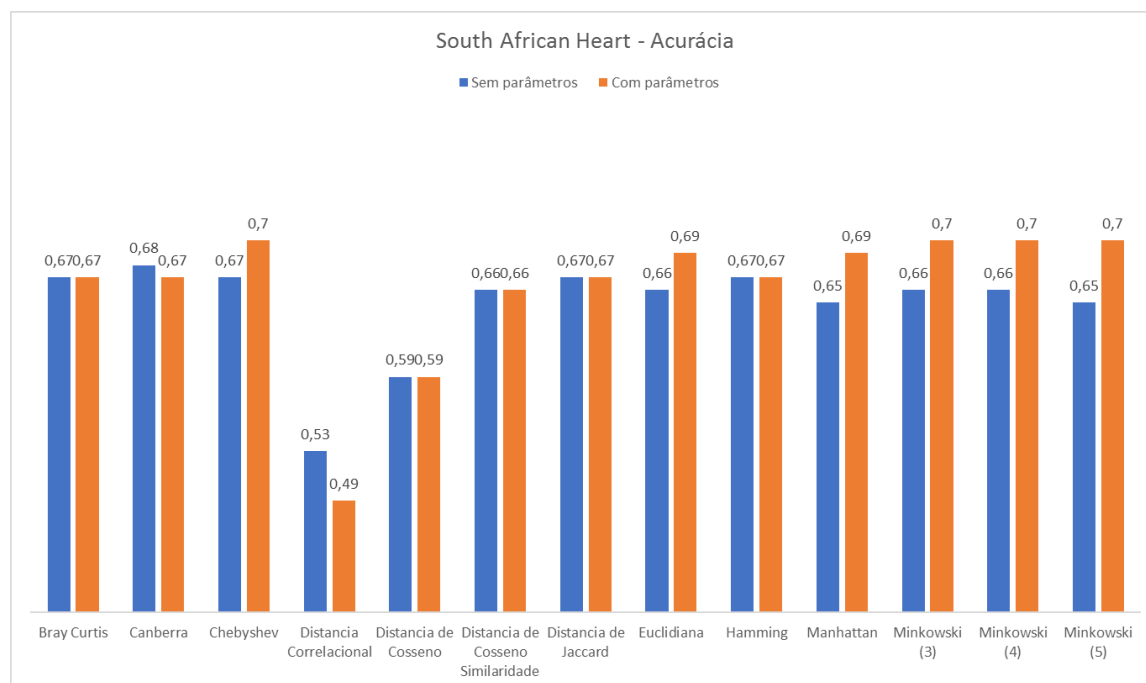
Porém, sem utilizar os parâmetros, com uma diferença de apenas 1% de acurácia, as distâncias Bray Curtis, Canberra, Hamming e Manhattan obtiveram bons resultados também.

Os piores resultados foram obtidos pelas distâncias Correlacional e de Cosseno, pelo mesmo motivo dos experimentos anteriores.

5.2.4. *South African Heart*

Essa base de dados possui apenas um atributo categórico. E os melhores resultados foram obtidos pelas distâncias Chebyshev, Minkowski e com uma diferença de 1% de acurácia, Euclidiana e Manhattan (vide [Figura 20](#)).

Figura 20. Resultados de classificação – Bases padrão c/ atr. categóricos – South African Heart



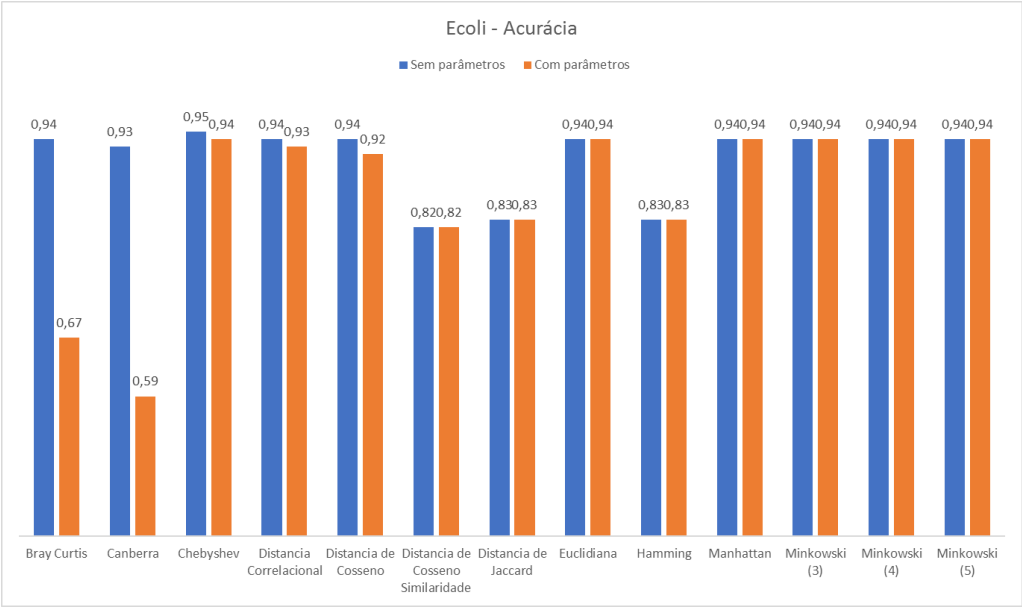
Os piores resultados foram obtidos pelas distâncias Correlacional e Cosseno, pelo mesmo motivo dos experimentos anteriores.

5.3. Bases desbalanceadas

5.3.1. *Ecoli*

Para essa base de dados o uso dos parâmetros não fez diferença, e em alguns casos até diminuiu o desempenho, pois os atributos dessa base já estão originalmente em um intervalo de 0-1.0. Os resultados são bastante parecidos, apenas as distâncias de Cosseno Similaridade, Jaccard e Hamming apresentaram um desempenho ruim, devido à baixa similaridade entre os dados (vide [Figura 21](#)).

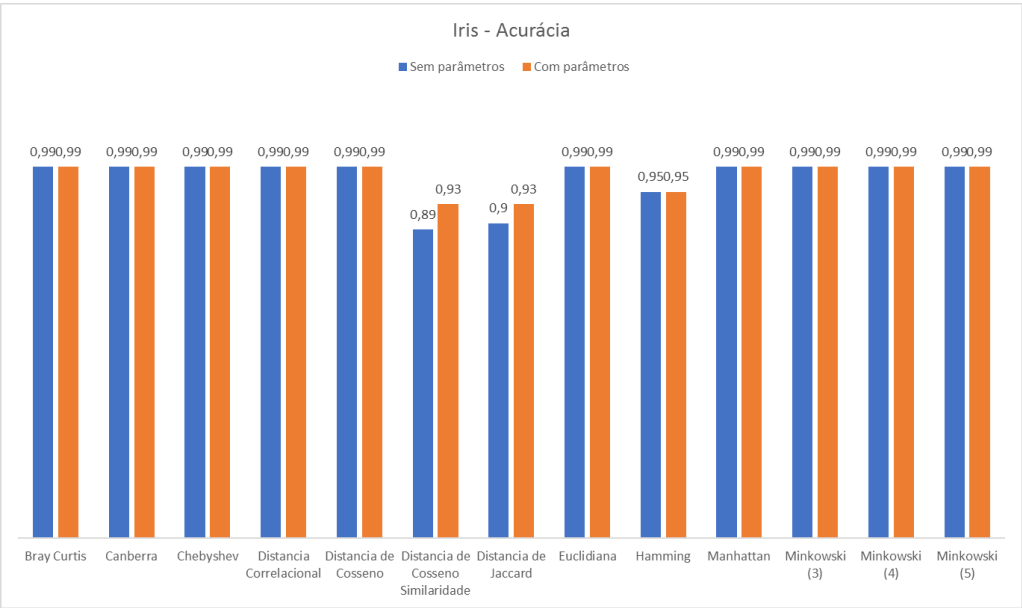
Figura 21. Resultados de classificação – Bases desbalanceadas – Ecoli



5.3.2. Iris

Apesar de não estar originalmente normalizada, praticamente todos os métodos de cálculo apresentaram os mesmos resultados, que se mostram muito bons, como se observa na [Figura 22](#). Os piores resultados foram obtidos pelas distâncias Cosseno de Similaridade, Jaccard e Hamming, pelos mesmos motivos da base de dados anterior.

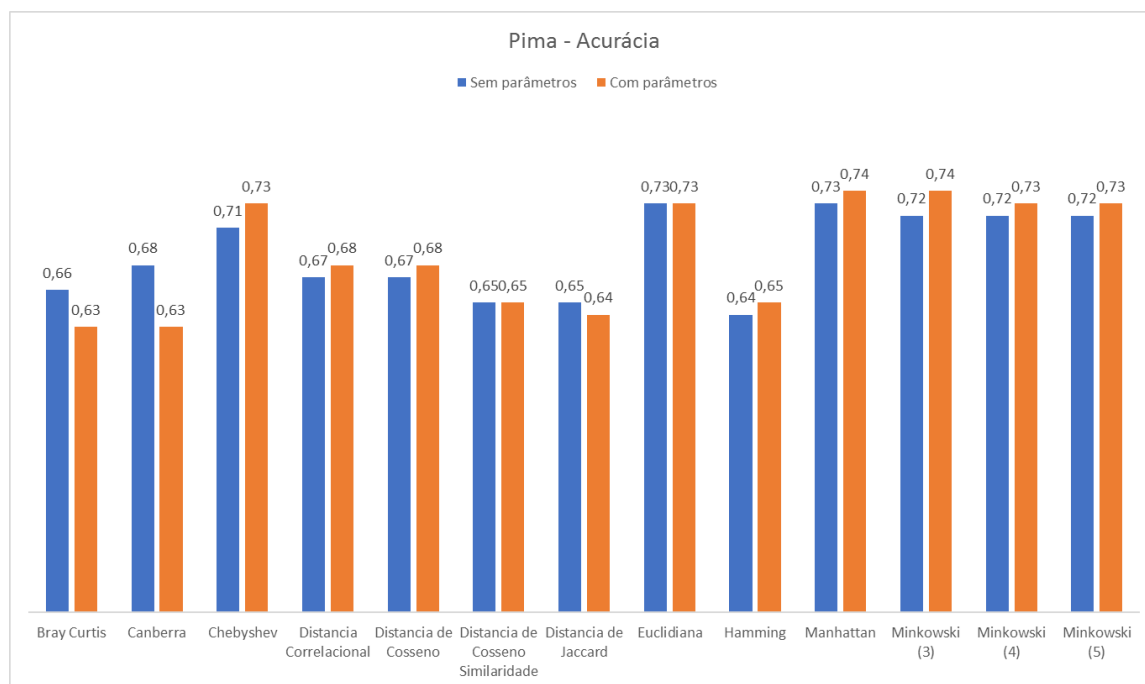
Figura 22. Resultados de classificação – Bases desbalanceadas – Iris



5.3.3. Pima

Essa base de dados apresenta um resultado muito similar, usando ou não os parâmetros propostos. As distâncias Manhattan e Eucliana apresentam os melhores resultados, porém a diferença para as demais é de apenas 1% (vide [Figura 23](#)).

Figura 23. Resultados de classificação – Bases desbalanceadas – Pima

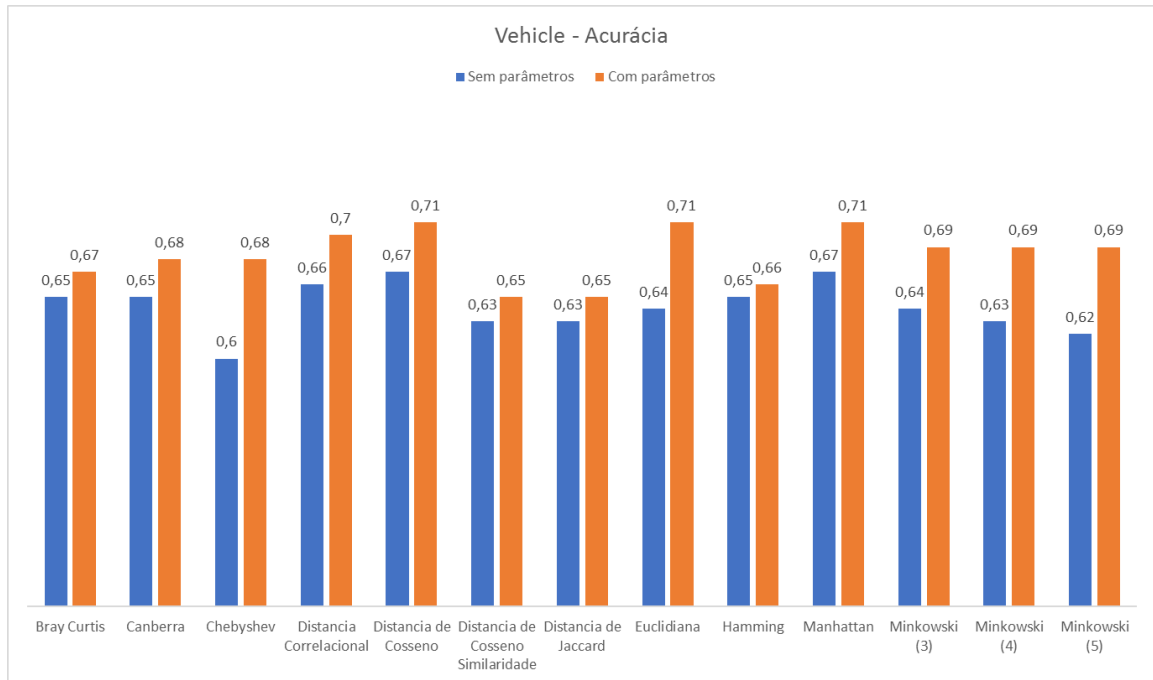


Os piores resultados são obtidos pelas distâncias de Cosseno Similaridade, Jaccard e Hamming.

5.3.4. Vehicle Silhouettes

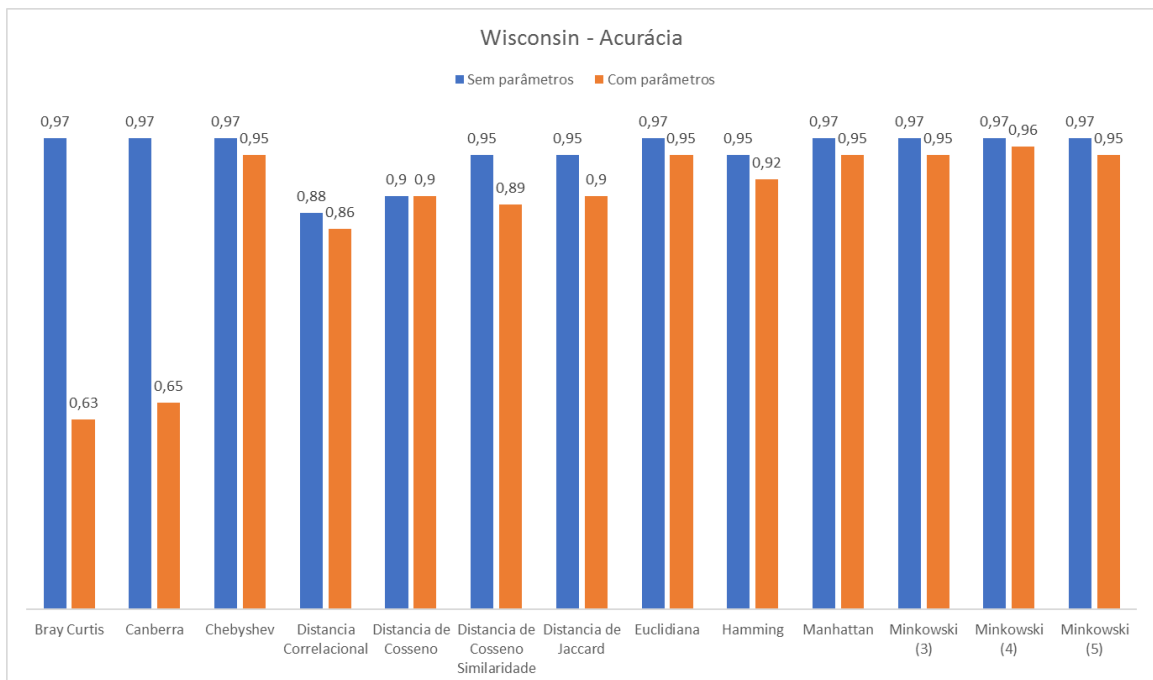
Como mostra a [Figura 24](#), como essa base de dados apresenta atributos em escalas diferentes, o algoritmo obteve um desempenho maior utilizando os parâmetros propostos. Os melhores resultados foram obtidos pelas distâncias de Cosseno, Euclidiana e Manhattan.

Os piores resultados foram obtidos pelas distâncias de Cosseno Similaridade, Jaccard e Hamming.

Figura 24. Resultados de classificação – Bases desbalanceadas – Vehicle

5.3.5. Wisconsin

Essa base de dados possui os dados em uma escala 0-10, portanto, o algoritmo apresentou um desempenho mais baixo ao utilizar os parâmetros propostos. Os resultados foram praticamente iguais, com exceção das distâncias Correlacional e de Cosseno que apresentaram um desempenho mais baixo (vide [Figura 25](#)).

Figura 25. Resultados de classificação – Bases desbalanceadas – Wisconsin

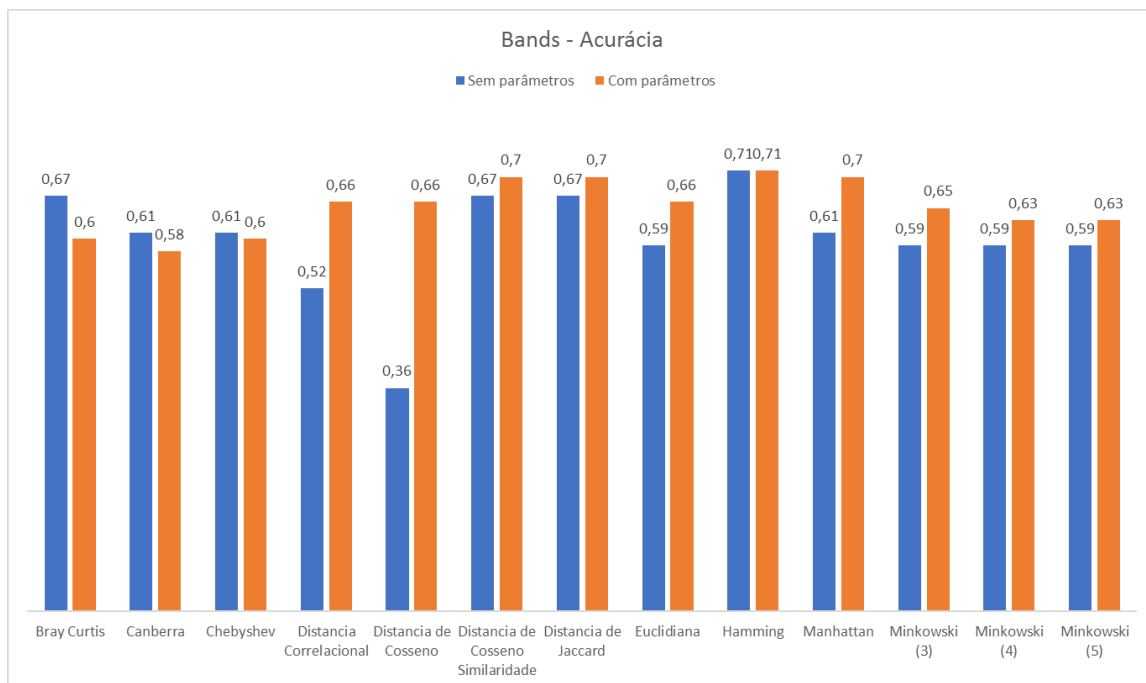
5.4. Bases com atributos faltantes

Para esse tipo de base de dados específico, “com parâmetros” inclui também a substituição de valores faltantes.

5.4.1. *Cylinder Bands*

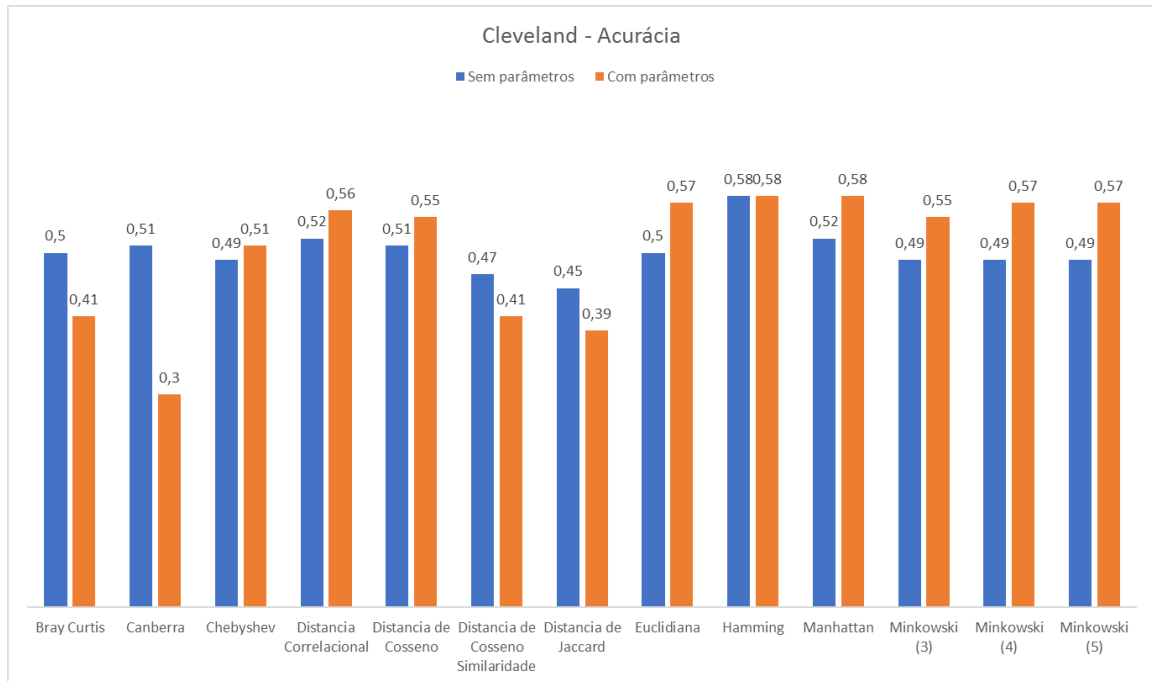
O melhor desempenho foi obtido pela distância de Hamming, porém as distâncias de Cosseno Similaridade e Jaccard também obtiveram o desempenho parecido, devido à alta similaridade entre a base de dados. Canberra e Chebyshev obtiveram o pior desempenho (vide [Figura 26](#)).

Figura 26. Resultados de classificação – Bases c/ atributos faltantes – Bands



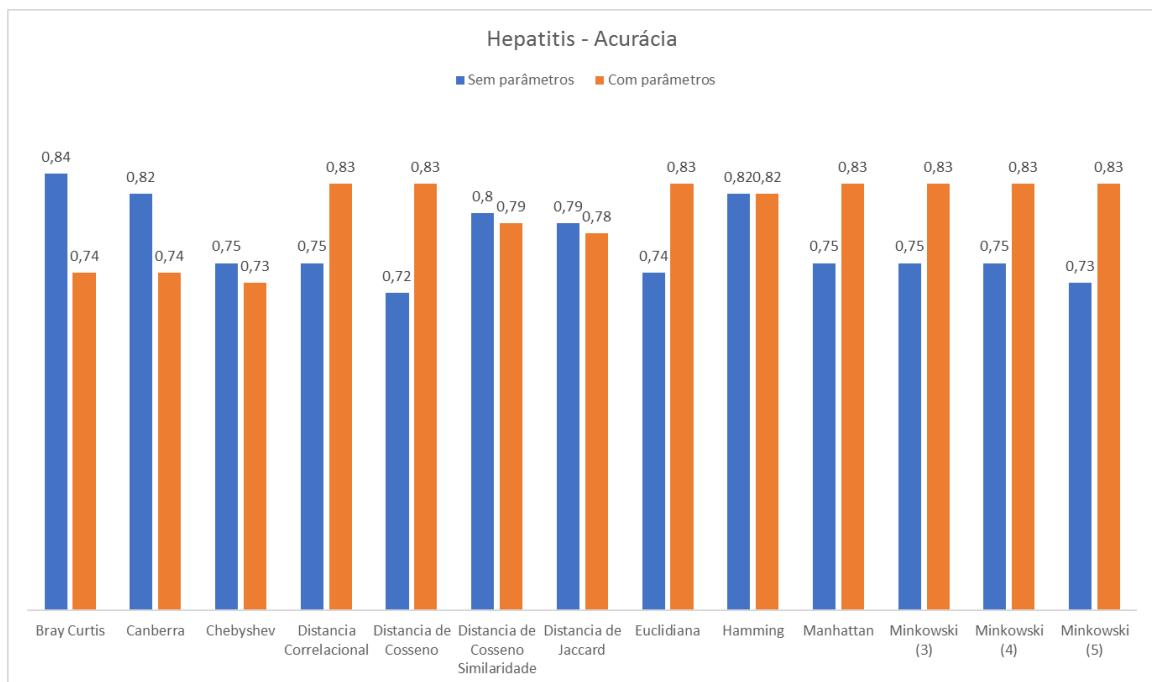
5.4.2. *Heart Disease (Cleveland)*

Com essa base, o maior desempenho do algoritmo foi utilizando a distância de Hamming. Os demais resultados foram melhores utilizando os parâmetros propostos, enquanto os piores resultados foram obtidos com as distâncias Cosseno de Similaridade e Jaccard, por se tratar de uma base com baixa similaridade (vide [Figura 27](#)).

Figura 27. Resultados de classificação – Bases c/ atributos faltantes – Cleveland

5.4.3. Hepatitis

Como pode se observa na [Figura 28](#), apesar de apresentar um desempenho melhor no geral, utilizando os parâmetros propostos, a distância que obteve o melhor desempenho foi a distância de Bray Curtis, porém sem utilizar os parâmetros propostos.

Figura 28. Resultados de classificação – Bases c/ atributos faltantes – Hepatitis

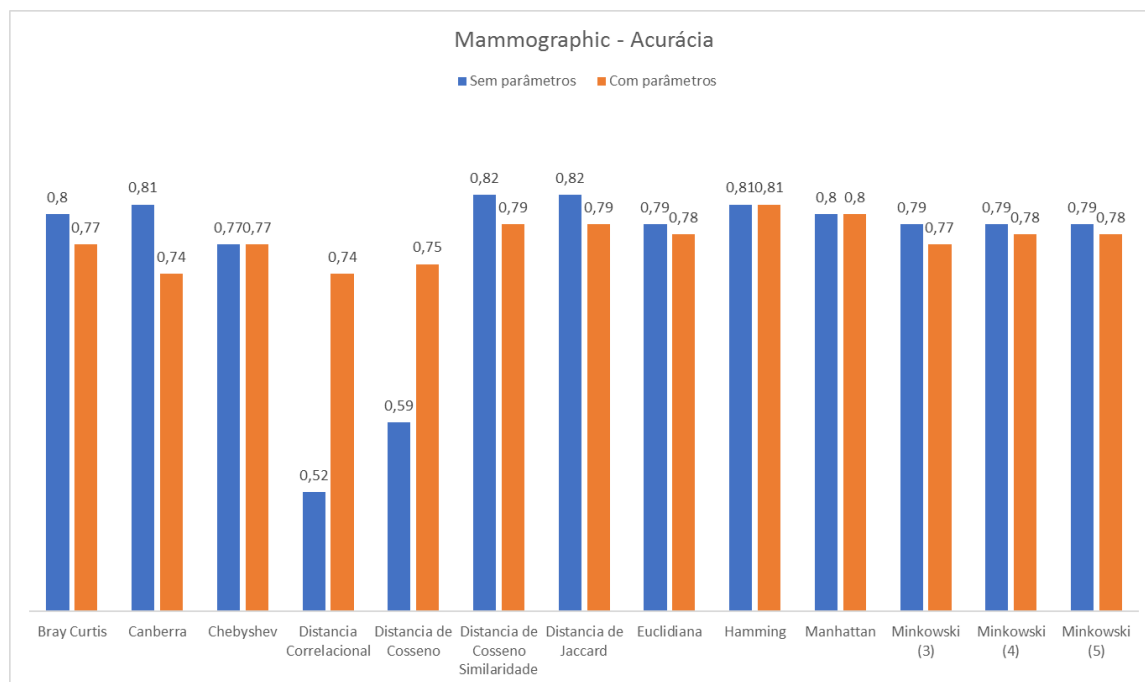
Isso se deve ao fato dos intervalos de dados serem estreitos, e o pior desempenho foi obtido pela distância de Chebyshev.

5.4.4. *Mammographic Mass*

Além de possuir intervalos de dados estreitos, essa base de dados possui uma similaridade alta, portanto as distâncias de Cosseno Similaridade e Jaccard obtiveram os melhores desempenhos, como se observa na [Figura 29](#). Em geral todas as distâncias obtiveram um resultado parecido, e nesse caso utilizar os parâmetros propostos não se mostrou uma boa escolha, já que o algoritmo obteve um desempenho igual ou pior, do ao não utilizá-los.

Os piores resultados foram obtidos pelas distâncias Correlacional e Cosseno.

Figura 29. Resultados de classificação – Bases c/ atributos faltantes – Mammographic



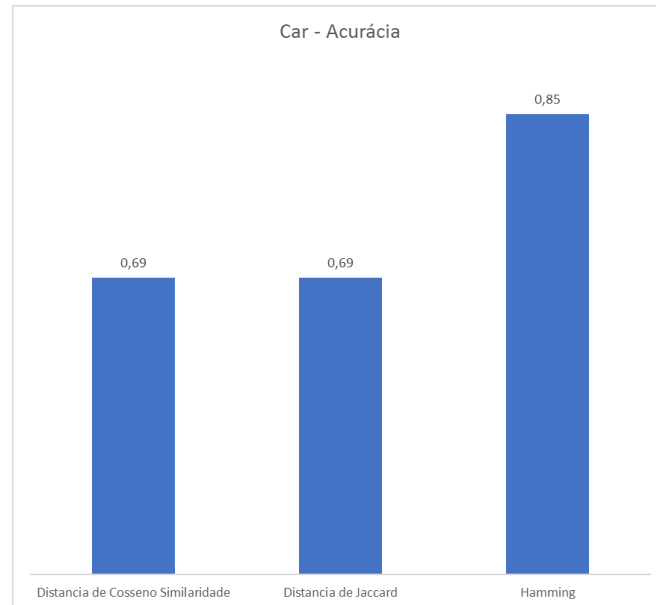
5.5. Bases com atributos categóricos

Para bases de dados com atributos categóricos, foram comparadas as distâncias de Hamming, Jaccard e Cosseno Similaridade. Por se tratar apenas de atributos categóricos, não serão comparados os resultados com parâmetros propostos, pois os mesmo se aplicam apenas a atributos numéricos.

5.5.1. *Car Evaluation*

Utilizando as distâncias de Cosseno Similaridade e Jaccard, o algoritmo obteve um desempenho mais baixo do que ao utilizar a distância de Hamming. O motivo mais provável, é que essa base de dados apresenta uma baixa similaridade entre os dados (vide [Figura 30](#)).

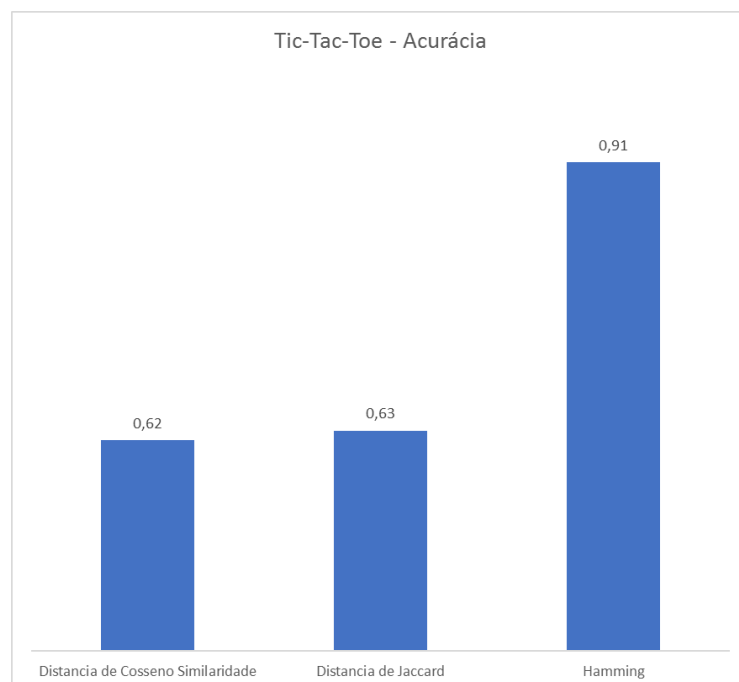
Figura 30. Resultados de classificação – Bases c/ atributos categóricos – Car



5.5.2. *Tic-Tac-Toe Endgame*

O melhor desempenho foi obtido utilizando a distância de Hamming, pelo mesmo motivo da base de dados anterior (vide [Figura 31](#)).

Figura 31. Resultados de classificação – Bases c/ atributos categóricos – Tic-Tac-Toe



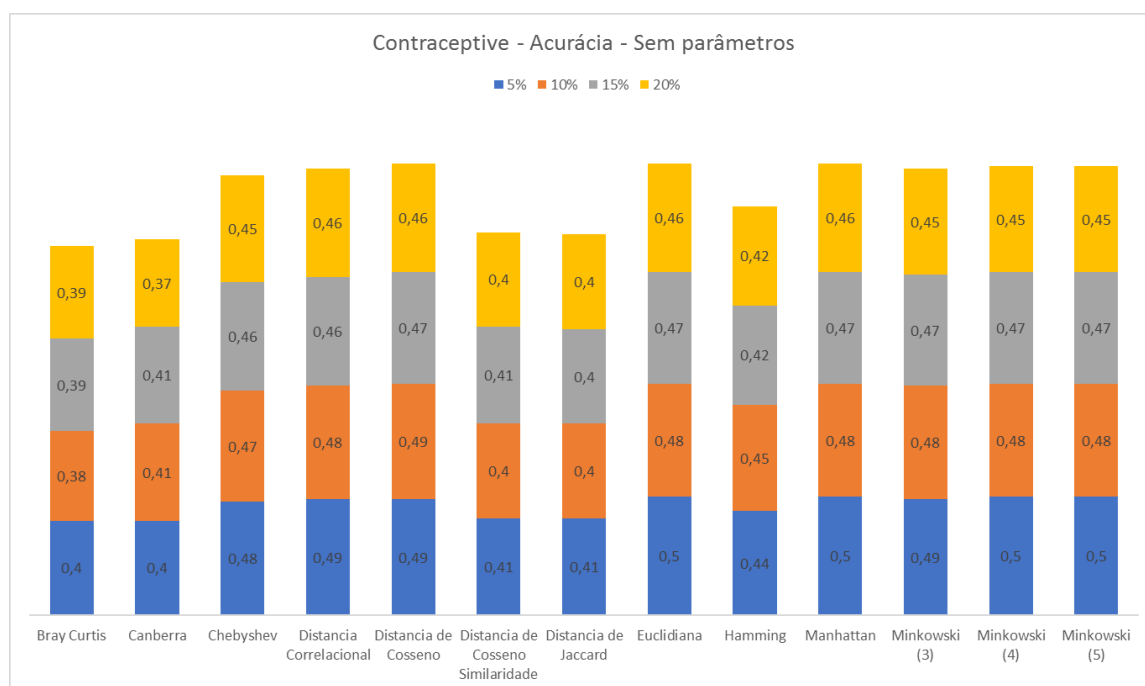
5.6. Bases com ruído de atributos

As análises a seguir, foram feitas utilizando apenas os melhores resultados obtidos, ou com os parâmetros propostos ou em sua ausência. Os resultados completos podem ser consultados no [Apêndice C – Resultados](#).

5.6.1. Contraceptive Method Choice

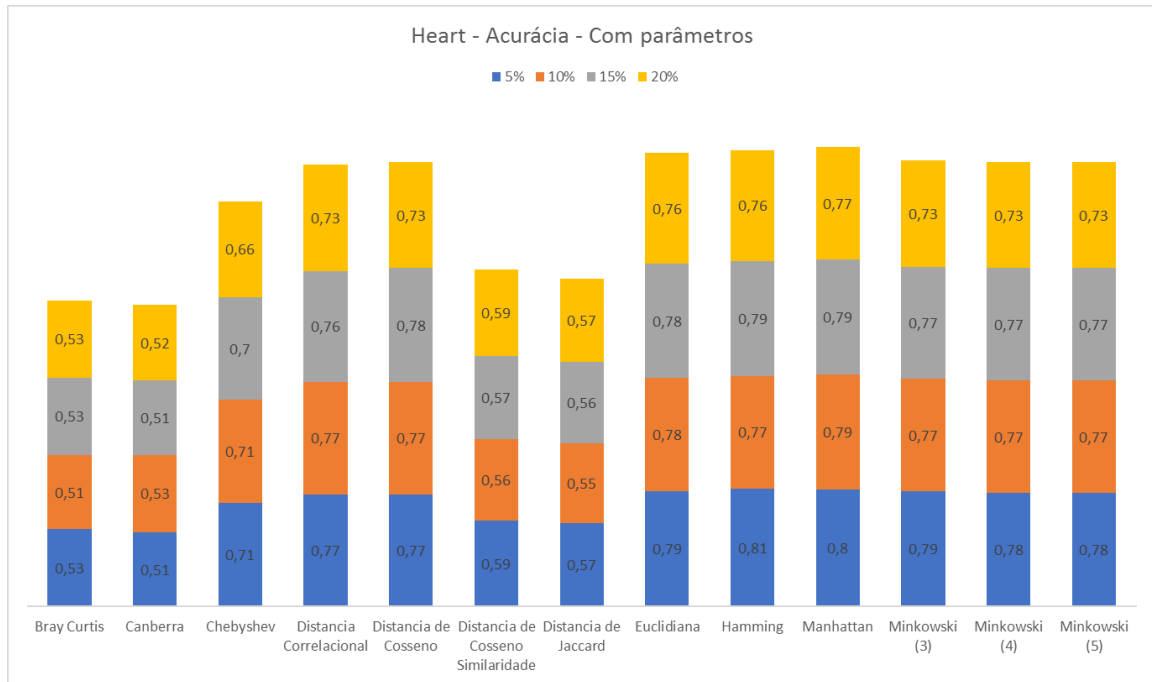
Como se observa na [Figura 32](#), o algoritmo apresentou um desempenho melhor ao não utilizar os parâmetros propostos. Apesar dos melhores resultados terem sido alcançados pela distância Euclidiana, os mesmos se mostraram insatisfatórios. Uma vez que a acurácia ficou abaixo de 50%.

Figura 32. Resultados de classificação – Bases c/ ruído de atributos – Contraceptive



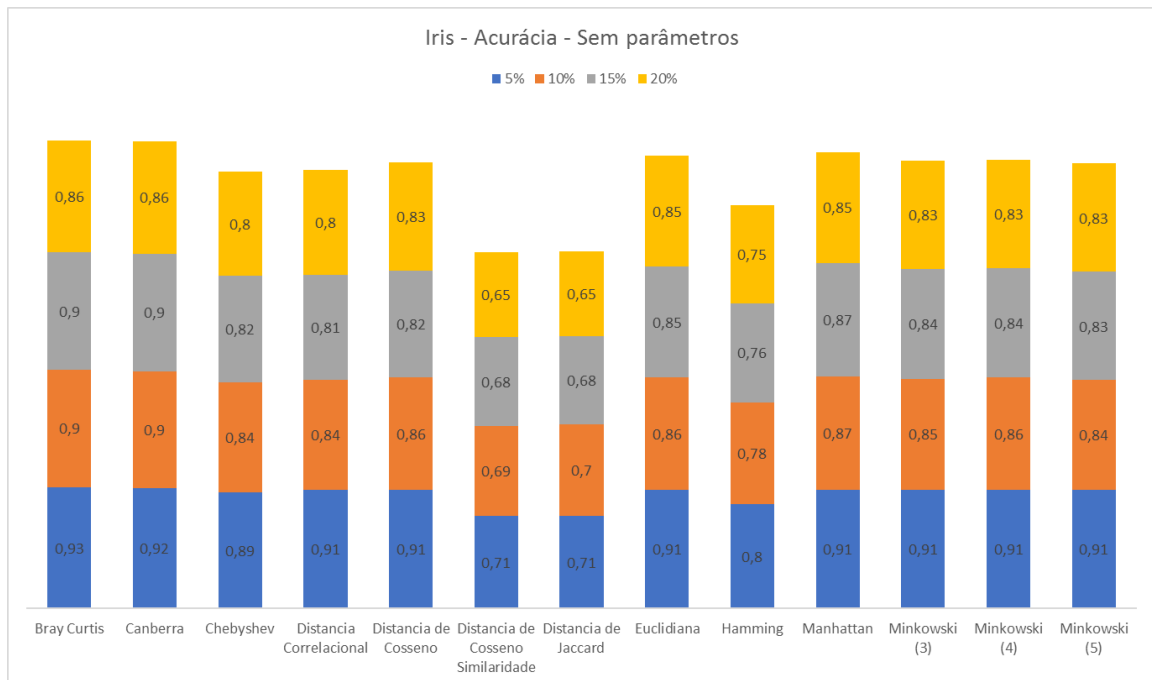
5.6.2. Heart

Nesse caso, o algoritmo obteve um desempenho melhor ao utilizar os parâmetros propostos. As distâncias que obteve melhores resultados foram as distâncias Euclidiana, Hamming e Manhattan. E os piores desempenhos foram obtidos pelas distâncias de Bray Curtis, Canberra, Cosseno Similaridade e Jaccard, pois alguns atributos não estão em uma faixa estreita de valores e a base possui uma baixa similaridade (vide [Figura 33](#)).

Figura 33. Resultados de classificação – Bases c/ ruído de atributos – Heart

5.6.3. Iris

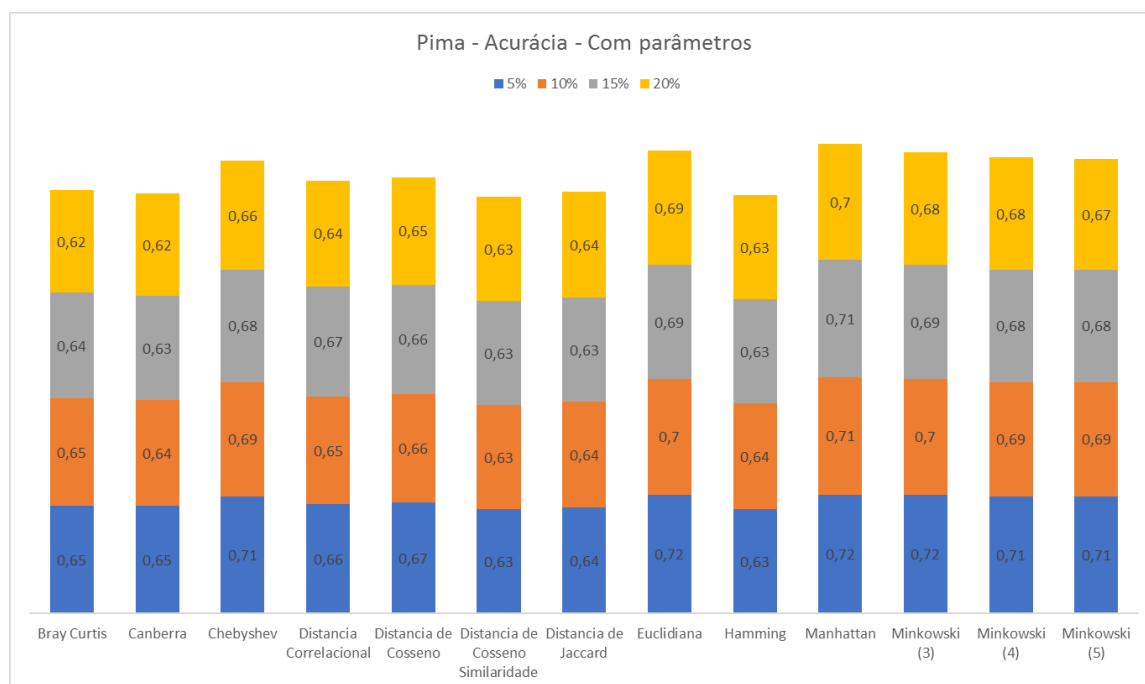
Como se observa na [Figura 34](#), os melhores resultados foram obtidos pelas distâncias Bray Curtis e Canberra, pois os atributos possuem uma faixa estreita de valores. Os piores resultados foram obtidos pelas distâncias de Cosseno Similaridade e Jaccard, pois a base de dados apresenta similaridade baixa.

Figura 34. Resultados de classificação – Bases c/ ruído de atributos – Iris

5.6.4. Pima

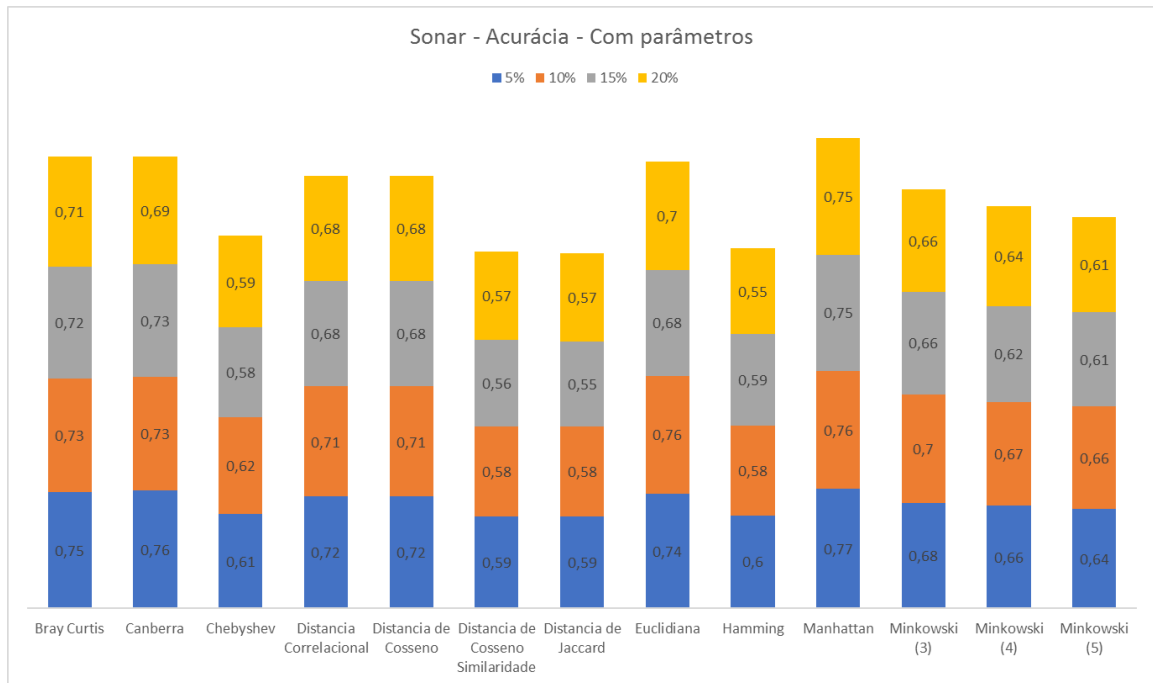
No caso dessa base de dados, o algoritmo obteve melhores resultados utilizando os parâmetros propostos. O melhor desempenho foi obtido pela distância Manhattan, porém o restante das distâncias obteve um desempenho parecido (vide [Figura 35](#)).

Figura 35. Resultados de classificação – Bases c/ ruído de atributos – Pima



5.6.5. Sonar, Mines vs. Rocks

Como se observa na [Figura 36](#), o algoritmo obteve um desempenho melhor utilizando os parâmetros propostos. O melhor desempenho foi obtido utilizando a distância de Manhattan. Os piores, foram obtidos utilizando as distâncias de Cosseno Similaridade, Jaccard e Hamming. O motivo é a baixa similaridade da base de dados.

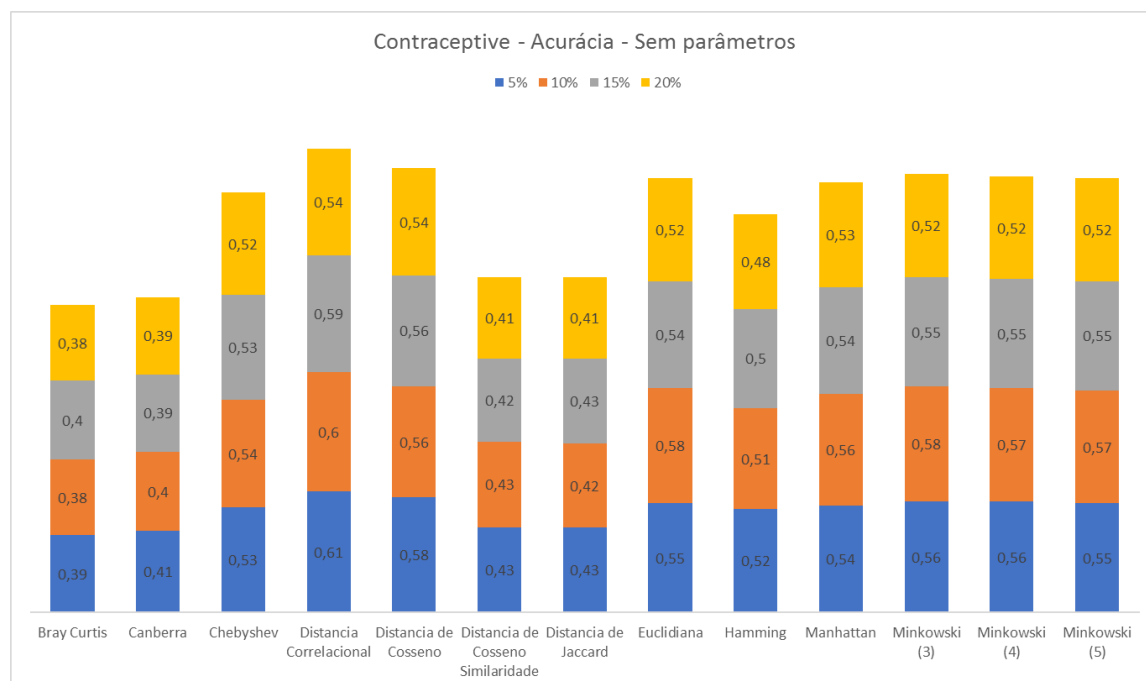
Figura 36. Resultados de classificação – Bases c/ ruído de atributos – Sonar

5.7. Bases com ruído de classes

As análises a seguir, foram feitas utilizando apenas os melhores resultados obtidos, ou com os parâmetros propostos ou em sua ausência. Os resultados completos podem ser consultados no [Apêndice C – Resultados](#).

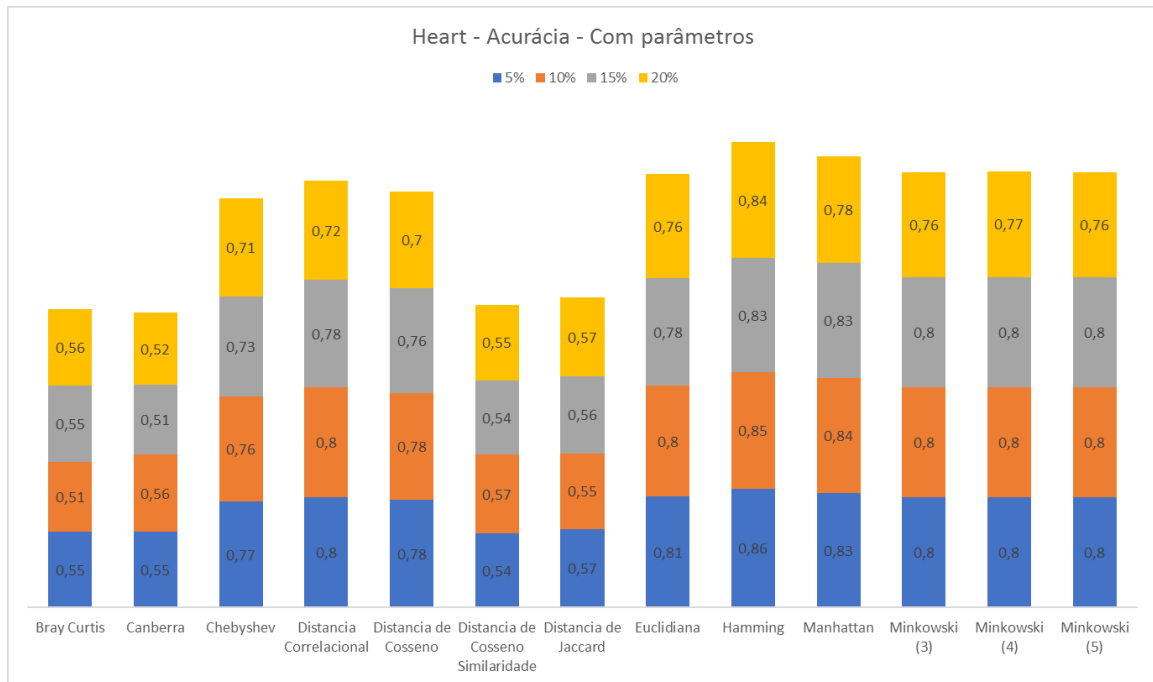
5.7.1. Contraceptive Method Choice

Como se observa na [Figura 37](#), o algoritmo obteve um melhor desempenho não utilizando os parâmetros propostos. O melhor desempenho foi obtido utilizando a distância Correlacional, enquanto os piores foram obtidos pelas distâncias BrayCurtis e Canberra. Isso se dá devido ao fato de nem todos os atributos estarem em uma faixa de valores estreita.

Figura 37. Resultados de classificação – Bases c/ ruído de classes – Contraceptive

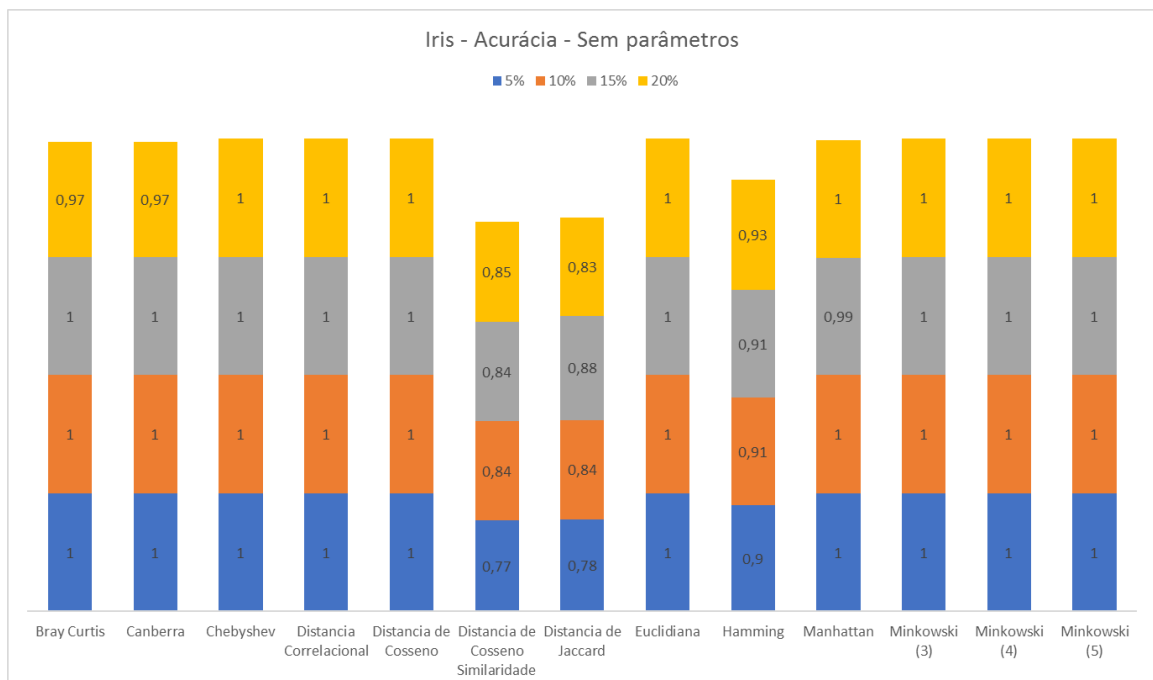
5.7.2. Heart

Como se observa na [Figura 38](#), o algoritmo obteve um melhor desempenho utilizando os parâmetros propostos. O melhor desempenho foi obtido utilizando a distância de Hamming, porém a distância de Manhattan obteve um desempenho bem próximo. Os piores desempenhos foram obtidos pelas distâncias BrayCurtis, Canberra, Cosseno similaridade e Canberra. O motivo é que alguns atributos não estão em uma faixa estreita de valores e a base possui uma baixa similaridade

Figura 38. Resultados de classificação – Bases c/ ruído de classes – Heart

5.7.3. Iris

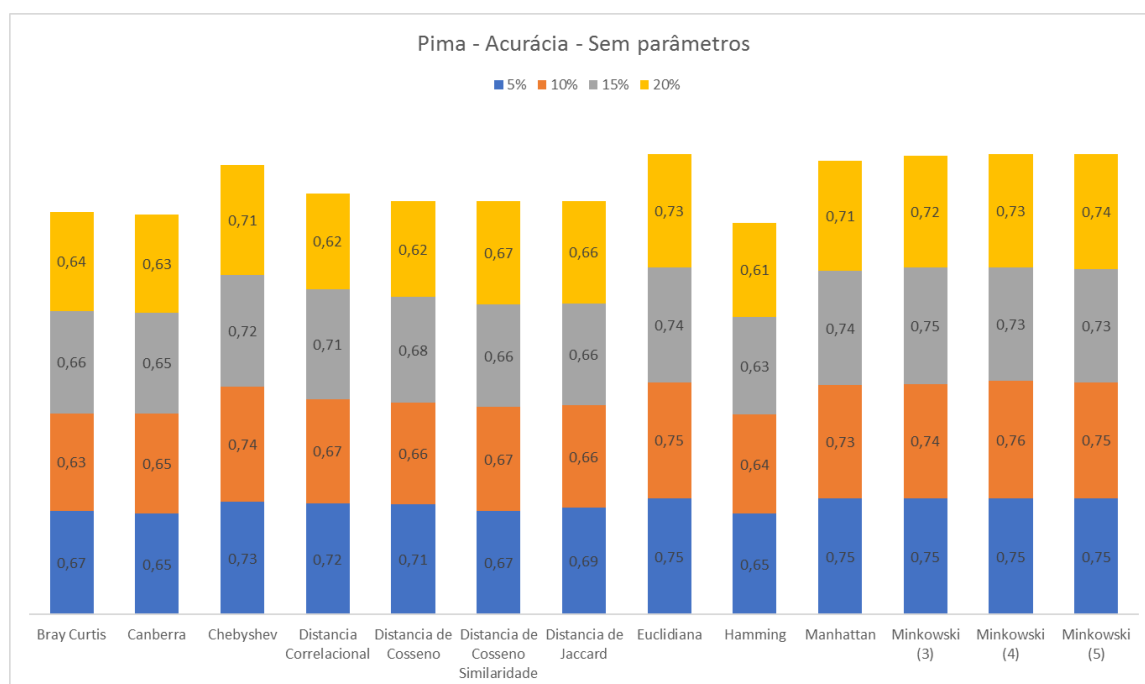
O algoritmo obteve resultados muito satisfatórios com praticamente todas as distâncias, sem utilizar os parâmetros propostos. As distâncias Cosseno Similaridade, Jaccard e Hamming obtiveram os piores resultados (vide [Figura 39](#)).

Figura 39. Resultados de classificação – Bases c/ ruído de classes – Iris

5.7.4. Pima

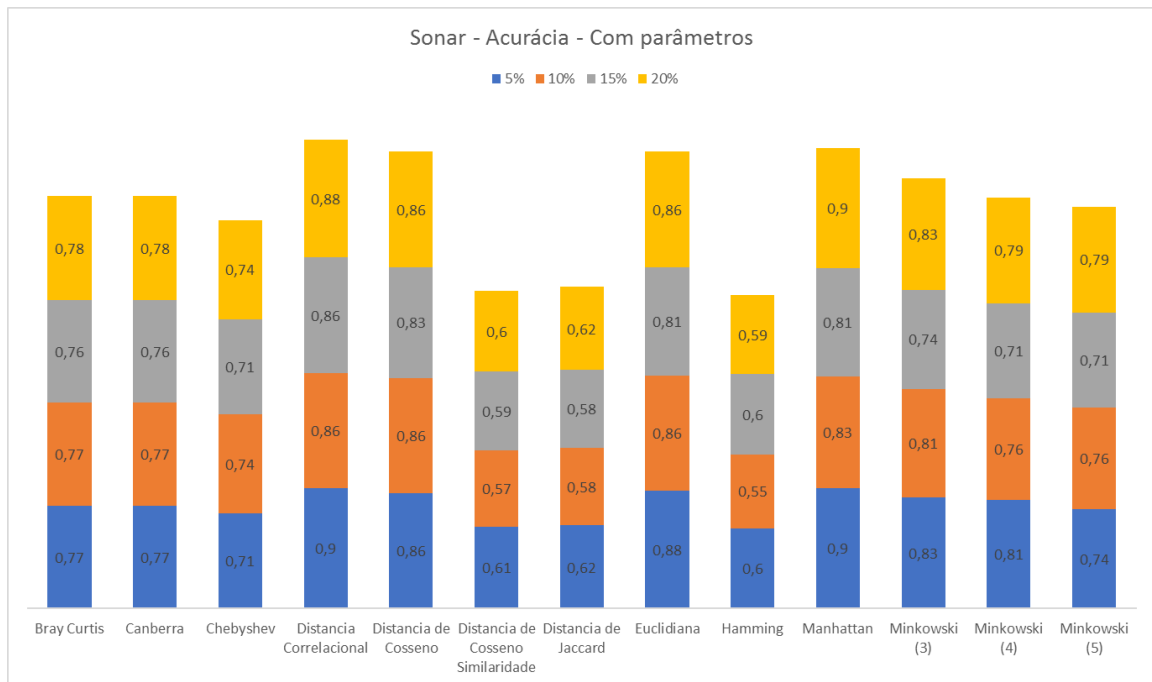
Como se observa na [Figura 40](#), o algoritmo obteve um desempenho melhor sem utilizar os parâmetros propostos. As distâncias que obtiveram os melhores desempenhos foram as distâncias Euclidiana, Manhattan e Minkowski. Os piores desempenhos foram obtidos pelas distâncias de Bray Curtis, Canberra e Hamming, devido a faixa de valores dos atributos não ser estreito.

Figura 40. Resultados de classificação – Bases c/ ruído de classes – Pima



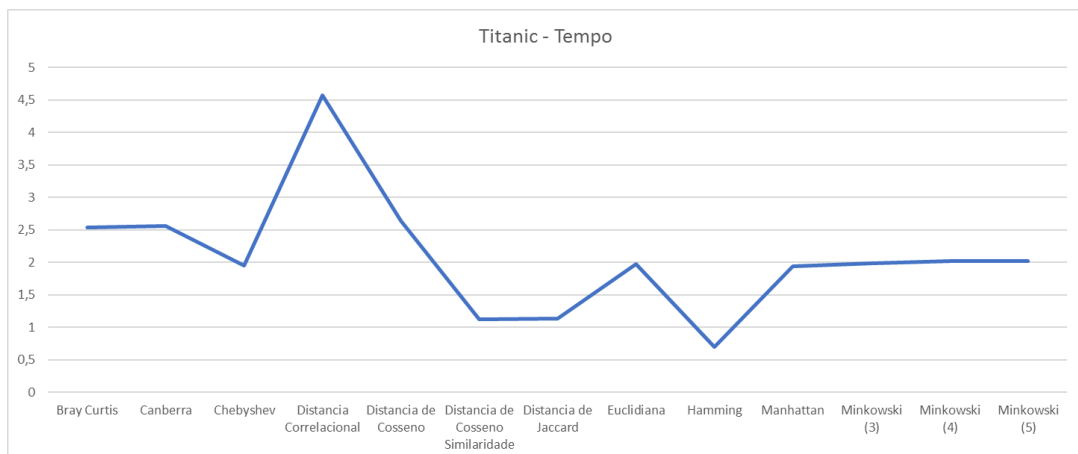
5.7.5. Sonar, Mines vs. Rocks

Como se observa na [Figura 41](#), o algoritmo obteve melhores resultados utilizando os parâmetros propostos. Os melhores desempenhos foram obtidos pelas distâncias Correlacional, Cosseno, Euclidiana e Manhattan. Os piores desempenhos foram obtidos pelas distâncias Cosseno Similaridade, Jaccard e Hamming. O motivo é a baixa similaridade entre os dados.

Figura 41. Resultados de classificação – Bases c/ ruído de classes – Sonar

5.8. Tempo de execução

Foi observado que, apesar dos intervalos de tempo variarem de acordo com o tamanho da base de dados, o gráfico é sempre igual. Portanto, muitas vezes o ganho em acuraria não compensa a diferença de tempo de execução. A [Figura 42](#) apresenta um gráfico de uma base de dados analisada, *Titanic*, todos os resultados podem consultados no [Apêndice C – Resultados](#).

Figura 42. Resultados de tempo de execução – Bases padrão – Titanic

Capítulo 6.

CONCLUSÕES

6.1. Principais contribuições

Com base nessa pesquisa, foi concluído que as métricas fazem uma diferença considerável na classificação, porém, dependendo da base de dados, que deve ser avaliada antes da escolha do método de cálculo e das métricas utilizadas com o algoritmo.

Outras fases do KDD se mostraram mais importantes dependendo da base de dados, como retirar registros repetidos, normalizar os dados e balancear as classes, porém em alguns casos o algoritmo teve um desempenho melhor sem utilizá-los.

Como método de desempate, o desvio padrão possui uma probabilidade menor de empate, em relação a média das distâncias, pois leva em conta a variação dessas distâncias. Porém não demonstraram uma diferença relevante nos resultados.

A escolha do melhor K de acordo com a base de dados se mostrou uma boa opção, entre $3 \leq K \leq \log_2 n$, pois além de encontrar o K que possui o melhor desempenho, evita o *overfitting* e *underfitting*.

Para bases de dados categóricas, a distância de Hamming se mostrou uma opção melhor que as distâncias de Cosseno Similaridade e Jaccard.

Em geral, para bases com alta similaridade entre os dados, as distâncias de Cosseno Similaridade, Jaccard e Hamming obtiveram um bom desempenho, e no caso da distância de Hamming um tempo mais baixo de execução que as outras distâncias.

Bray Curtis e Canberra obtiveram um bom desempenho em bases de dados com um intervalo de valores estreito.

Para bases de dados com atributos numéricos reais, as distâncias Euclidiana, Manhattan e Minkowski obtiveram um bom desempenho.

Para bases de dados com atributos faltantes, a proposta de comparar com a distância de Hamming o atributo faltante se mostrou igual, e alguns caso melhor do que substituir esses valores pela média dos atributos.

Portanto, antes de escolher as métricas e os métodos de cálculo, é necessário conhecer a base de dados e escolher a melhor configuração do algoritmo para aquela base de dados em específico.

6.2. Trabalhos futuros

Para esse trabalho, o algoritmo K-Vizinhos mais próximos foi aplicado apenas em problemas de classificação, outro trabalho a ser considerado é sua aplicação em problemas de regressão.

Outra sugestão é a pesquisa de propostas para o tratamento de base de dados com ruído de atributos e classes, pois nesse trabalho não foi considerada nenhuma.

Apesar de apresentar um bom tempo de execução, por ter sido implementado o processamento paralelo em um processador com quatro núcleos, uma sugestão é usar *GPU's* para realizar o processamento.

A comparação dos resultados de classificação com outros algoritmos de classificação é também um assunto interessante para ser investigado em trabalhos futuros.

Referências Bibliográficas

CASELLA, George.; BERGER, Roger L. **Inferência Estatística**. 2ª edição. São Paulo, Cengage Learning, 2016. ISBN: 978-85-221-0894-7.

DEZA, Michel Marie; DEZA, Elena. **Encyclopedia of Distances**. New York, Elsevier, 2006. ISBN: 978-0-444-52087-6.

FAYYAD, Usama; PIATESKY-SHAPIO, Gregory; SMYTH, Padhraic. **From data mining to knowledge discovery in databases**. AI MAGAZINE - American Association for Artificial Intelligence, p. 37–54, 1996.

FERNÁNDEZ, A.; GARCÍA, S.; DEL JESUS, M. J.; HERRERA, F. **A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets**. *Fuzzy Sets and Systems*. International Journal of Approximate Reasoning, p. 566-577, 2009.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel; BEZERRA, Eduardo. **Data Mining: Conceitos, técnicas, algoritmos, orientações e aplicações**. 2ª edição. Rio de Janeiro, Elsevier, 2015. ISBN: 978-85-352-7822-4.

GUIDORIZZI, Hamilton Luiz. **Um curso de cálculo: vol. 1**. 5ª edição. São Paulo, LTC, 2001.

HAN, Jiawei; KAMBER, Micheline. **Data Mining: Concepts and techniques**. EUA, Elsevier, 2006. ISBN: 978-1-55860-901-3.

KEEL. **A Software Tool to Assess Evolutionary Algorithms for Data Mining Problems**. Disponível em: <<http://www.keel.es/>>. Acesso em: 22 abr. 2017.

LARSON, Ron; FARBER, Betsy. **Estatística Aplicada**. 4ª edição. São Paulo, Pearson, 2010. ISBN: 978-85-7605-372-9.

MEIRA JUNIOR, Wagner; ZAKI, Mohammed J. **Data Mining and Analysis: Fundamental concepts and algorithms**. EUA, Cambridge Press, 2014. ISBN: 978-0-521-76633-3.

MENON, A.K.; AGARWAL, H.N.S.; CHAWLA, S. **On the Statistical Consistency of Algorithms for Binary Classification under Class Imbalance**. Proceedings of the 30th International Conference on Machine Learning, Atlanta, 16- 21 June 2013, p. 603-611.

RUSSEL, Stuart J.; NORVIG, Peter. **Inteligência Artificial**. 3ª edição. Rio de Janeiro, Elsevier, 2013. ISBN: 978-85-353-3701-6.

SAYAD, Sayad. **An Introduction to Data Mining**. Disponível em: <<http://www.saedsayad.com/>>. Acesso em: 20 maio 2017.

SILVA, Leandro Augusto da; PERES, Sarajane Marques; BOSCARIOLI, Clodis. **Introdução à mineração de dados: Com aplicações em R**. Rio de Janeiro, Elsevier, 2016. ISBN: 978-85-352-8447-8.

APÊNDICE A.

CÓDIGO FONTE

Para o desenvolvimento do trabalho, foi programada uma biblioteca na linguagem C#, que está disponível ²² em uma página *web*. Inicialmente seria disponibilizada nesse apêndice, porém o código ficou maior do que o esperado.

²² <https://github.com/netososilva/k-nearest-neighbor>

APÊNDICE B.

CONJUNTOS DE DADOS

B.1. Automobile

Normalized-losses	[65.0, 256.0]
Make	{ alfa-romero, audi, bmw, chevrolet, dodge, honda, isuzu, jaguar, mazda, mercedes-benz, mercury, mitsubishi, nissan, peugot, plymouth, porsche, renault, saab, subaru, toyota, volkswagen, volvo }
Fuel-type	{ diesel, gas }
Aspiration	{ std, turbo }
Num-of-doors	{ four, two }
Body-style	{ hardtop, wagon, sedan, hatchback, convertible }
Drive-wheels	{ 4wd, fwd, rwd }
Engine-location	{ front, rear }
Wheel-base	[86.6, 120.9]
Length	[141.1, 208.1]
Width	[60.3, 72.3]
Height	[47.8, 59.8]
Curb-weight	[1488.0, 4066.0]
Engine-type	{ dohc, dohcvt, l, ohc, ohcf, ohcv, rotor }
Num-of-cylinders	{ eight, five, four, six, three, twelve, two }
Engine-size	[61.0, 326.0]
Fuel-system	{ 1bbl, 2bbl, 4bbl, idi, mfi, mpfi, spdi, spfi }
Bore	[2.54, 3.94]
Stroke	[2.07, 4.17]
Compression-ratio	[7.0, 23.0]
Horsepower	[48.0, 288.0]
Peak-rpm	[4150.0, 6600.0]
City-mpg	[13.0, 49.0]
Highway-mpg	[16.0, 54.0]
Price	[5118.0, 45400.0]

Symboling	{-2, -1, 0, 1, 2, 3}
------------------	----------------------

B.2. Car Evaluation

Buying	{ vhigh,high,med,low }
Maint	{ vhigh,high,med,low }
Doors	{ 2,3,4,5more }
Persons	{ 2,4,more }
Lug_boot	{ small,med,big }
Safety	{ low,med,high }
Acceptability	{ unacc,acc,vgood,good }

B.3. Contraceptive Method Choice

Wife_age	[16,49]
Wife_education	[1,4]
Husband_education	[1,4]
Children	[0,16]
Wife_religion	[0,1]
Wife_working	[0,1]
Husband_occupation	[1,4]
Standard-of-living	[1,4]
Media_exposure	[0,1]
Contraceptive_method	{ 1,2,3 }

B.4. Cylinder Bands

Proof_cut	[25.0,72.5]
Viscosity	[35,72]
Caliper	[0.133,0.533]
Ink_temperature	[11.2,24.5]
Humidity	[57,105]
Roughness	[0.056,1.25]
Blade_pressure	[16,70]
Varnish_pct	[0.0,35.8]
Press_speed	[0,2600]
Ink_pct	[41.0,76.9]
Solvent_pct	[22.0,53.4]
Esa_voltage	[0.0,16.0]
ESA_amperage	[0.0,6.0]

Wax	[0.0,3.1]
Hardener	[0.0,3.0]
Roller_durometer	[28.0,60.0]
Density	[30,45]
Anode_ratio	[83.33,117.86]
Chrome_content	[90,100]
Band_type	{band,noband}

B.5. Ecoli

Mcg	[0.0, 0.89]
Gvh	[0.16, 1.0]
Lip	[0.48, 1.0]
Chg	[0.5, 1.0]
Aac	[0.0, 0.88]
Alm1	[0.03, 1.0]
Alm2	[0.0, 0.99]
Class	{positive, negative}

B.6. German Credit

StatusAccount	{A11, A12, A14, A13}
DurationMonth	[4, 72]
CreditHistory	{A34, A32, A33, A30, A31}
Purpose	{A43, A46, A42, A40, A41, A49, A44, A45, A410, A48}
CreditAmount	[250, 18424]
SavingsAccount	{A65, A61, A63, A64, A62}
EmploymentSince	{A75, A73, A74, A71, A72}
InstallmentRate	[1, 4]
StatusAndSex	{A93, A92, A91, A94}
Guarantors	{A101, A103, A102}
ResidenceSince	[1, 4]
Property	{A121, A122, A124, A123}
Age	[19, 75]
InstallmentPlans	{A143, A141, A142}
Housing	{A152, A153, A151}
NCredits	[1, 4]
Job	{A173, A172, A174, A171}
NPeopleMain	[1, 2]
Telephone	{A192, A191}
ForeignWorker	{A201, A202}
Customer	{1,2}

B.7. Hepatitis

Age	[7, 78]
Sex	[1, 2]
Steroid	[1, 2]
Antivirals	[1, 2]
Fatigue	[1, 2]
Malaise	[1, 2]
Anorexia	[1, 2]
LiverBig	[1, 2]
LiverFirm	[1, 2]
SpleenPalpable	[1, 2]
Spiders	[1, 2]
Ascites	[1, 2]
Varices	[1, 2]
Bilirubin	[0.3, 8.0]
AlkPhosphate	[26, 295]
Sgot	[14, 648]
AlbuMin	[2.1, 6.4]
ProTime	[0, 100]
Histology	[1, 2]
Class	{1, 2}

B.8. Iris

SepalLength	[4.3, 7.9]
SepalWidth	[2.0, 4.4]
PetalLength	[1.0, 6.9]
PetalWidth	[0.1, 2.5]
Class	{positive, negative}

B.9. Heart

Age	[29, 77]
Sex	[0, 1]
ChestPainType	[1, 4]
RestBloodPressure	[94, 200]
SerumCholestorl	[126, 564]
FastingBloodSugar	[0, 1]
ResElectrocardiographic	[0, 2]
MaxHeartRate	[71, 202]
ExerciseInduced	[0, 1]
Oldpeak	[0.0, 62.0]

Slope	[1, 3]
MajorVessels	[0, 3]
Thal	[3, 7]
Class	{1, 2}

B.10. Heart Disease (Cleveland)

Age	[29.0, 77.0]
Sex	[0.0, 1.0]
Cp	[1.0, 4.0]
Trestbps	[94.0, 200.0]
Chol	[126.0, 564.0]
Fbs	[0.0, 1.0]
Restecg	[0.0, 2.0]
Thalach	[71.0, 202.0]
Exang	[0.0, 1.0]
Oldpeak	[0.0, 6.2]
Slope	[1.0, 3.0]
Ca	[0.0, 3.0]
Thal	[3.0, 7.0]
Num	{0,1,2,3,4}

B.11. Lymphography

Lymphatics	{normal,arched,deformed,displaced}
Block_of_affere	{no,yes}
Bl_of_lymph_c	{no,yes}
Bl_of_lymph_s	{no,yes}
By_pass	{no,yes}
Extravasates	{no,yes}
Regeneration_of	{no,yes}
Early_uptake_in	{no,yes}
Lym_nodes_dimin	[0,3]
Lym_nodes_enlar	[1,4]
Changes_in_lym	{bean,oval,round}
Defect_in_node	{no,lacunar,lac_margin,lac_central}
Changes_in_node	{no,lacunar,lac_margin,lac_central}
Changes_in_stru	{no,grainy,drop_like,coarse,diluted,reticular,stripped,faint}
Special_forms	{no,chalices,vesicles}
Dislocation_of	{no,yes}
Exclusion_of_no	{no,yes}
No_of_nodes_in	[1,8]

Class	{normal,metastases,malign_lymph,fibrosis}
--------------	-------------------------------------------

B.12. Mammographic Mass

BI-RADS	[0,6]
Age	[18,96]
Shape	[1,4]
Margin	[1,5]
Density	[1,4]
Severity	{0, 1}

B.13. Pima

Preg	[0.0, 17.0]
Plas	[0.0, 199.0]
Pres	[0.0, 122.0]
Skin	[0.0, 99.0]
Insu	[0.0, 846.0]
Mass	[0.0, 67.1]
Pedi	[0.078, 2.42]
Age	[21.0, 81.0]
Class	{positive,negative}

B.14. Red Wine Quality

FixedAcidity	[4.6,15.9]
VolatileAcidity	[0.12,1.58]
CitricAcid	[0.0,1.0]
ResidualSugar	[0.9,15.5]
Chlorides	[0.012,0.611]
FreeSulfurDioxide	[1.0,72.0]
TotalSulfurDioxide	[6.0,289.0]
Density	[0.99007,1.00369]
PH	[2.74,4.01]
Sulphates	[0.33,2.0]
Alcohol	[8.4,14.9]
Quality	{0,1,2,3,4,5,6,7,8,9,10}

B.15. Sonar, Mines vs. Rocks

Band1	[0.0020,0.137]
Band2	[0.0010,0.234]
Band3	[0.0020,0.306]
Band4	[0.0060,0.426]
Band5	[0.0070,0.401]
Band6	[0.01,0.382]
Band7	[0.0030,0.373]
Band8	[0.0060,0.459]
Band9	[0.0080,0.683]
Band10	[0.011,0.711]
Band11	[0.029,0.734]
Band12	[0.024,0.706]
Band13	[0.018,0.713]
Band14	[0.027,0.997]
Band15	[0.0030,1.0]
Band16	[0.016,0.999]
Band17	[0.035,1.0]
Band18	[0.038,1.0]
Band19	[0.049,1.0]
Band20	[0.066,1.0]
Band21	[0.051,1.0]
Band22	[0.022,1.0]
Band23	[0.056,1.0]
Band24	[0.024,1.0]
Band25	[0.024,1.0]
Band26	[0.092,1.0]
Band27	[0.048,1.0]
Band28	[0.028,1.0]
Band29	[0.014,1.0]
Band30	[0.061,1.0]
Band31	[0.048,0.966]
Band32	[0.04,0.931]
Band33	[0.048,1.0]
Band34	[0.021,0.965]
Band35	[0.022,1.0]
Band36	[0.0080,1.0]
Band37	[0.035,0.95]
Band38	[0.038,1.0]
Band39	[0.037,0.986]
Band40	[0.012,0.93]

Band41	[0.036,0.9]
Band42	[0.0060,0.825]
Band43	[0.0,0.773]
Band44	[0.0,0.776]
Band45	[0.0,0.703]
Band46	[0.0,0.729]
Band47	[0.0,0.552]
Band48	[0.0,0.334]
Band49	[0.0,0.198]
Band50	[0.0,0.082]
Band51	[0.0,0.1]
Band52	[0.0010,0.071]
Band53	[0.0,0.039]
Band54	[0.0010,0.035]
Band55	[0.0010,0.045]
Band56	[0.0,0.039]
Band57	[0.0,0.036]
Band58	[0.0,0.044]
Band59	[0.0,0.036]
Band60	[0.0010,0.044]
Type	{R,M}

B.16. South African Heart

Sbp	[101, 218]
Tobacco	[0.0, 31.2]
Ldl	[0.98, 15.33]
Adiposity	[6.74, 42.49]
Famhist	{Present, Absent}
Typea	[13, 78]
Obesity	[14.7, 46.58]
Alcohol	[0.0, 147.19]
Age	[15, 64]
Chd	{0, 1}

B.17. Tic-Tac-Toe Endgame

TopLeft	{x, o, b}
TopMiddle	{x, o, b}
TopRight	{x, o, b}
MiddleLeft	{x, o, b}
MiddleMiddle	{o, b, x}
MiddleRight	{o, b, x}
BottomLeft	{x, o, b}
BottomMiddle	{o, x, b}
BottomRight	{o, x, b}
Class	{positive, negative}

B.18. Titanic

Class	[-1.87,0.965]
Age	[-0.228,4.38]
Sex	[-1.92,0.521]
Survived	{-1.0,1.0}

B.19. Vehicle Silhouettes

Compactness	[73, 119]
Circularity	[33, 59]
Distance_circularity	[40, 112]
Radius_ratio	[104, 333]
Praxis_aspect_ratio	[47, 138]
Max_length_aspect_ratio	[2, 55]
Scatter_ratio	[112, 265]
Elongatedness	[26, 61]
Praxis_rectangular	[17, 29]
Length_rectangular	[118, 188]
Major_variance	[130, 320]
Minor_variance	[184, 1018]
Gyration_radius	[109, 268]
Major_skewness	[59, 135]
Minor_skewness	[0, 22]
Minor_kurtosis	[0, 41]
Major_kurtosis	[176, 206]
Hollows_ratio	[181, 211]
Class	{van, saab, bus, opel}

B.20. Wisconsin

ClumpThickness	[1, 10]
CellSize	[1, 10]
CellShape	[1, 10]
MarginalAdhesion	[1, 10]
EpithelialSize	[1, 10]
BareNuclei	[1, 10]
BlandChromatin	[1, 10]
NormalNucleoli	[1, 10]
Mitoses	[1, 10]
Class	{negative,positive}

APÊNDICE C.

RESULTADOS

C.1 Bases padrão

C.1.1. Heart

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,55	0,08	0,52	0,06	0,56	0,07	0,52	0,06
Canberra	0,56	0,08	0,54	0,07	0,55	0,06	0,5	0,07
Chebyshev	0,67	0,03	0,73	0,04	0,66	0,03	0,74	0,02
Distancia Correlacional	0,67	0,04	0,81	0,03	0,69	0,02	0,81	0,02
Distancia de Cosseno	0,67	0,03	0,8	0,02	0,68	0,03	0,8	0,02
Distancia de Cosseno Similaridade	0,72	0,04	0,59	0,04	0,73	0,03	0,59	0,03
Distancia de Jaccard	0,72	0,05	0,58	0,04	0,71	0,04	0,57	0,05
Euclidiana	0,68	0,02	0,81	0,02	0,67	0,03	0,82	0,02
Hamming	0,81	0,02	0,81	0,02	0,81	0,02	0,82	0,02
Manhattan	0,71	0,02	0,82	0,02	0,72	0,02	0,83	0,02
Minkowski (3)	0,67	0,03	0,81	0,01	0,67	0,03	0,81	0,01

Minkowski (4)	0,67	0,03	0,81	0,02	0,67	0,02	0,81	0,02
Minkowski (5)	0,68	0,02	0,8	0,02	0,66	0,03	0,81	0,02

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	Sem parâmetros		Com parâmetros		Sem parâmetros		Com parâmetros	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,201	0,013	0,217	0,007	0,198	0,009	0,218	0,006
Canberra	0,201	0,014	0,215	0,005	0,196	0,008	0,22	0,008
Chebyshev	0,117	0,008	0,132	0,012	0,116	0,007	0,128	0,005
Distancia Correlacional	0,261	0,016	0,298	0,015	0,26	0,02	0,296	0,013
Distancia de Cosseno	0,162	0,015	0,177	0,006	0,155	0,007	0,179	0,008
Distancia de Cosseno Similaridade	0,075	0,006	0,072	0,004	0,074	0,007	0,073	0,007
Distancia de Jaccard	0,072	0,005	0,073	0,004	0,077	0,008	0,073	0,005
Euclidiana	0,119	0,006	0,137	0,006	0,12	0,006	0,137	0,006
Hamming	0,046	0,004	0,045	0,002	0,046	0,003	0,046	0,003
Manhattan	0,116	0,005	0,133	0,007	0,118	0,007	0,133	0,006
Minkowski (3)	0,121	0,007	0,141	0,011	0,125	0,011	0,138	0,008
Minkowski (4)	0,121	0,007	0,136	0,006	0,124	0,009	0,14	0,008
Minkowski (5)	0,121	0,006	0,136	0,005	0,12	0,006	0,138	0,006

C.1.2. Mammographic Mass

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,81	0,01	0,77	0,06	0,82	0,01	0,76	0,07
Canberra	0,81	0,01	0,77	0,04	0,81	0,01	0,77	0,05
Chebyshev	0,77	0,01	0,75	0,02	0,77	0,01	0,75	0,02
Distancia Correlacional	0,77	0,02	0,74	0,03	0,77	0,02	0,75	0,02
Distancia de Cosseno	0,78	0,01	0,75	0,02	0,78	0,01	0,75	0,02
Distancia de Cosseno Similaridade	0,83	0,01	0,82	0,01	0,83	0,01	0,81	0,03
Distancia de Jaccard	0,83	0,01	0,81	0,02	0,83	0,01	0,81	0,02
Euclidiana	0,8	0,01	0,76	0,02	0,79	0,01	0,76	0,02
Hamming	0,82	0,01	0,81	0,01	0,82	0,01	0,81	0,02
Manhattan	0,8	0,01	0,76	0,02	0,81	0,01	0,77	0,02
Minkowski (3)	0,79	0,01	0,76	0,02	0,78	0,01	0,75	0,03
Minkowski (4)	0,79	0,01	0,76	0,02	0,79	0,01	0,75	0,02
Minkowski (5)	0,79	0,01	0,75	0,03	0,79	0,01	0,76	0,02

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,547	0,018	0,333	0,009	0,548	0,018	0,34	0,015
Canberra	0,548	0,037	0,34	0,014	0,543	0,014	0,336	0,013
Chebyshev	0,427	0,011	0,24	0,007	0,434	0,034	0,246	0,014
Distancia Correlacional	0,953	0,023	0,51	0,012	0,956	0,02	0,511	0,014
Distancia de Cosseno	0,567	0,015	0,321	0,018	0,575	0,05	0,324	0,02
Distancia de Cosseno Similaridade	0,281	0,006	0,13	0,005	0,281	0,014	0,13	0,006
Distancia de Jaccard	0,281	0,013	0,13	0,007	0,283	0,009	0,128	0,003
Euclidiana	0,443	0,016	0,247	0,008	0,446	0,019	0,251	0,01
Hamming	0,186	0,006	0,093	0,004	0,184	0,003	0,093	0,004
Manhattan	0,424	0,006	0,244	0,011	0,428	0,012	0,245	0,012
Minkowski (3)	0,447	0,011	0,254	0,012	0,449	0,017	0,254	0,011
Minkowski (4)	0,454	0,011	0,253	0,013	0,446	0,01	0,255	0,012
Minkowski (5)	0,454	0,021	0,256	0,016	0,451	0,014	0,257	0,02

C.1.3. Red Wine Quality

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,52	0,01	0,53	0,01	0,52	0,01	0,53	0,01
Canberra	0,52	0,01	0,53	0,01	0,52	0,01	0,52	0,02
Chebyshev	0,49	0,01	0,54	0,01	0,49	0,01	0,53	0,01
Distancia Correlacional	0,5	0,01	0,55	0,01	0,5	0,01	0,55	0,01
Distancia de Cosseno	0,5	0,01	0,55	0,01	0,5	0,01	0,55	0,01
Distancia de Cosseno Similaridade	0,47	0,01	0,47	0,01	0,47	0,01	0,45	0,01
Distancia de Jaccard	0,47	0,01	0,47	0,01	0,47	0,01	0,46	0,01
Euclidiana	0,5	0,01	0,55	0,01	0,5	0,01	0,54	0,01
Hamming	0,5	0,01	0,47	0,01	0,49	0,01	0,45	0,02
Manhattan	0,52	0,01	0,55	0,01	0,51	0,01	0,54	0,01
Minkowski (3)	0,5	0,01	0,54	0,01	0,49	0,01	0,54	0,01
Minkowski (4)	0,5	0,01	0,54	0,01	0,49	0,01	0,54	0,01
Minkowski (5)	0,49	0,01	0,54	0,01	0,49	0,01	0,53	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	Sem parâmetros		Com parâmetros		Sem parâmetros		Com parâmetros	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	4,07	0,119	4,122	0,031	4,075	0,124	4,124	0,041
Canberra	4,068	0,157	4,144	0,04	4,078	0,134	4,142	0,035
Chebyshev	2,922	0,071	2,918	0,05	2,902	0,082	2,919	0,032
Distancia Correlacional	7,499	0,161	7,324	0,046	7,504	0,158	7,311	0,045
Distancia de Cosseno	4,274	0,118	4,272	0,108	4,285	0,135	4,249	0,092
Distancia de Cosseno Similaridade	1,194	0,059	0,847	0,026	1,223	0,081	0,838	0,017
Distancia de Jaccard	1,218	0,058	0,842	0,019	1,212	0,051	0,842	0,019
Euclidiana	3,072	0,097	3,004	0,037	3,108	0,155	3,003	0,049
Hamming	0,638	0,037	0,453	0,021	0,631	0,037	0,448	0,009
Manhattan	2,95	0,083	2,934	0,08	2,982	0,101	2,933	0,048
Minkowski (3)	3,036	0,088	3,019	0,083	3,034	0,075	3,001	0,048
Minkowski (4)	3,05	0,095	3	0,054	3,034	0,105	2,995	0,045
Minkowski (5)	3,044	0,099	2,996	0,045	3,048	0,086	3,005	0,057

C.1.4. Titanic

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,78	0,03	0,55	0,17	0,78	0,01	0,6	0,16
Canberra	0,79	0	0,6	0,15	0,79	0,01	0,57	0,16
Chebyshev	0,78	0,04	0,68	0,13	0,77	0,05	0,67	0,12
Distancia Correlacional	0,79	0,01	0,57	0,16	0,79	0,01	0,61	0,12
Distancia de Cosseno	0,78	0,04	0,46	0,19	0,79	0,01	0,46	0,19
Distancia de Cosseno Similaridade	0,78	0,01	0,62	0,12	0,78	0,04	0,61	0,13
Distancia de Jaccard	0,79	0,01	0,63	0,1	0,78	0,01	0,64	0,09
Euclidiana	0,78	0,05	0,64	0,12	0,78	0,05	0,7	0,11
Hamming	0,78	0,05	0,67	0,14	0,78	0,01	0,67	0,16
Manhattan	0,79	0,01	0,69	0,14	0,78	0,03	0,7	0,12
Minkowski (3)	0,78	0,01	0,7	0,1	0,78	0,05	0,69	0,14
Minkowski (4)	0,79	0,01	0,68	0,09	0,78	0,04	0,68	0,12
Minkowski (5)	0,79	0,01	0,7	0,11	0,78	0,02	0,67	0,12

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	2,546	0,08	1,162	1,175	2,535	0,085	1,159	1,169
Canberra	2,52	0,082	1,17	1,183	2,564	0,133	1,184	1,196
Chebyshev	1,942	0,066	0,873	0,883	1,948	0,085	0,881	0,89
Distancia Correlacional	4,55	0,122	2,096	2,117	4,574	0,101	2,099	2,12
Distancia de Cosseno	2,64	0,081	1,208	1,221	2,635	0,065	1,194	1,207
Distancia de Cosseno Similaridade	1,147	0,042	0,577	0,583	1,127	0,037	0,583	0,589
Distancia de Jaccard	1,126	0,051	0,584	0,59	1,133	0,05	0,58	0,586
Euclidiana	1,986	0,072	0,916	0,928	1,97	0,086	0,907	0,92
Hamming	0,69	0,028	0,341	0,344	0,698	0,037	0,353	0,358
Manhattan	1,928	0,075	0,898	0,908	1,941	0,071	0,889	0,899
Minkowski (3)	2,026	0,084	0,919	0,928	1,989	0,071	0,924	0,935
Minkowski (4)	2,018	0,083	0,923	0,933	2,022	0,07	0,915	0,925
Minkowski (5)	1,995	0,076	0,91	0,92	2,019	0,064	0,919	0,929

C.1.5. Vehicle Silhouettes

	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
<u>Distâncias</u>	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,6	0,02	0,65	0,01	0,6	0,01	0,65	0,02
Canberra	0,6	0,02	0,65	0,02	0,61	0,01	0,65	0,02
Chebyshev	0,56	0,01	0,64	0,02	0,56	0,01	0,63	0,02
Distancia Correlacional	0,64	0,01	0,67	0,01	0,65	0,02	0,67	0,02
Distancia de Cosseno	0,64	0,02	0,68	0,02	0,64	0,02	0,68	0,01
Distancia de Cosseno Similaridade	0,59	0,02	0,61	0,02	0,58	0,02	0,6	0,02
Distancia de Jaccard	0,59	0,02	0,61	0,02	0,58	0,02	0,61	0,02
Euclidiana	0,62	0,02	0,67	0,02	0,62	0,02	0,66	0,02
Hamming	0,63	0,02	0,62	0,02	0,61	0,01	0,61	0,02
Manhattan	0,66	0,02	0,69	0,02	0,66	0,02	0,69	0,02
Minkowski (3)	0,6	0,02	0,66	0,02	0,6	0,02	0,65	0,02
Minkowski (4)	0,58	0,01	0,65	0,02	0,59	0,01	0,65	0,02
Minkowski (5)	0,58	0,02	0,64	0,01	0,58	0,02	0,65	0,02

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	1,893	0,057	2,472	0,137	1,853	0,08	2,431	0,081
Canberra	1,859	0,061	2,491	0,125	1,86	0,05	2,468	0,096
Chebyshev	1,309	0,051	1,747	0,052	1,342	0,075	1,77	0,09
Distancia Correlacional	3,411	0,104	4,445	0,162	3,366	0,092	4,393	0,156
Distancia de Cosseno	1,904	0,065	2,536	0,096	1,903	0,065	2,543	0,09
Distancia de Cosseno Similaridade	0,663	0,039	0,692	0,043	0,673	0,048	0,684	0,04
Distancia de Jaccard	0,665	0,033	0,69	0,039	0,676	0,038	0,675	0,035
Euclidiana	1,421	0,07	1,828	0,06	1,413	0,043	1,848	0,075
Hamming	0,33	0,027	0,324	0,027	0,334	0,033	0,315	0,018
Manhattan	1,355	0,071	1,751	0,079	1,309	0,04	1,739	0,063
Minkowski (3)	1,386	0,06	1,838	0,067	1,393	0,056	1,808	0,06
Minkowski (4)	1,381	0,05	1,84	0,07	1,38	0,053	1,84	0,076
Minkowski (5)	1,401	0,045	1,844	0,078	1,394	0,058	1,827	0,064

C.2. Bases padrão com atributos categóricos

C.2.1. Automobile

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio</u>			
					<u>Padrão</u>			
	Sem parâmetros	Com parâmetros	Sem parâmetros	Com parâmetros	Sem parâmetros	Com parâmetros	Sem parâmetros	Com parâmetros
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,56	0,05	0,6	0,04	0,61	0,04	0,61	0,05
Canberra	0,59	0,06	0,6	0,04	0,59	0,05	0,59	0,05
Chebyshev	0,38	0,03	0,59	0,04	0,37	0,03	0,58	0,05
Distancia Correlacional	0,26	0,06	0,24	0,07	0,24	0,07	0,24	0,07
Distancia de Cosseno	0,16	0,08	0,17	0,06	0,17	0,06	0,19	0,06
Distancia de Cosseno Similaridade	0,6	0,06	0,65	0,04	0,61	0,05	0,63	0,04
Distancia de Jaccard	0,6	0,06	0,63	0,06	0,6	0,05	0,64	0,04
Euclidiana	0,37	0,03	0,59	0,03	0,38	0,03	0,6	0,05
Hamming	0,62	0,06	0,64	0,05	0,62	0,04	0,63	0,04
Manhattan	0,41	0,03	0,6	0,05	0,41	0,03	0,59	0,05
Minkowski (3)	0,38	0,03	0,59	0,04	0,39	0,03	0,59	0,04
Minkowski (4)	0,37	0,03	0,59	0,04	0,39	0,04	0,57	0,05
Minkowski (5)	0,38	0,04	0,58	0,03	0,37	0,04	0,59	0,04

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,074	0,007	0,098	0,009	0,072	0,004	0,098	0,014
Canberra	0,076	0,007	0,098	0,006	0,075	0,007	0,098	0,007
Chebyshev	0,06	0,006	0,072	0,005	0,059	0,005	0,071	0,003
Distancia Correlacional	0,127	0,007	0,167	0,008	0,128	0,007	0,168	0,016
Distancia de Cosseno	0,075	0,005	0,101	0,009	0,077	0,006	0,1	0,006
Distancia de Cosseno Similaridade	0,042	0,002	0,043	0,005	0,043	0,004	0,042	0,004
Distancia de Jaccard	0,044	0,004	0,043	0,005	0,045	0,008	0,044	0,006
Euclidiana	0,06	0,006	0,077	0,011	0,062	0,007	0,078	0,009
Hamming	0,022	0,002	0,022	0,002	0,021	0,001	0,021	0,001
Manhattan	0,061	0,007	0,073	0,006	0,06	0,009	0,074	0,007
Minkowski (3)	0,064	0,007	0,074	0,005	0,06	0,004	0,074	0,004
Minkowski (4)	0,063	0,007	0,074	0,006	0,061	0,009	0,073	0,004
Minkowski (5)	0,062	0,005	0,08	0,011	0,064	0,009	0,076	0,006

C.2.2. German Credit

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,73	0,01	0,6	0,17	0,73	0,01	0,59	0,17
Canberra	0,72	0,01	0,53	0,19	0,73	0,01	0,54	0,18
Chebyshev	0,67	0,01	0,72	0,01	0,68	0,01	0,72	0,01
Distancia Correlacional	0,68	0,08	0,56	0,2	0,7	0,01	0,56	0,2
Distancia de Cosseno	0,59	0,05	0,56	0,04	0,61	0,05	0,58	0,05
Distancia de Cosseno Similaridade	0,72	0,01	0,72	0,01	0,72	0,01	0,71	0,01
Distancia de Jaccard	0,71	0,01	0,72	0,01	0,72	0,01	0,72	0,01
Euclidiana	0,67	0,01	0,73	0,01	0,68	0,01	0,73	0,01
Hamming	0,71	0,01	0,72	0,01	0,71	0,01	0,71	0,01
Manhattan	0,68	0,01	0,73	0,01	0,68	0,01	0,72	0,01
Minkowski (3)	0,67	0,01	0,73	0,01	0,67	0,01	0,73	0,01
Minkowski (4)	0,68	0,01	0,72	0,01	0,67	0,01	0,73	0,01
Minkowski (5)	0,67	0,02	0,73	0,01	0,68	0,01	0,73	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	1,4	0,036	1,728	0,063	1,398	0,055	1,756	0,073
Canberra	1,432	0,064	1,748	0,062	1,421	0,052	1,726	0,057
Chebyshev	1,172	0,063	1,42	0,054	1,161	0,036	1,39	0,058
Distancia Correlacional	2,819	0,176	3,388	0,121	2,77	0,09	3,453	0,088
Distancia de Cosseno	1,454	0,057	1,778	0,067	1,426	0,045	1,787	0,056
Distancia de Cosseno Similaridade	1,08	0,044	1,084	0,048	1,094	0,033	1,088	0,037
Distancia de Jaccard	1,076	0,043	1,074	0,045	1,086	0,036	1,103	0,041
Euclidiana	1,208	0,055	1,42	0,043	1,172	0,039	1,404	0,045
Hamming	0,4	0,04	0,39	0,027	0,384	0,019	0,393	0,028
Manhattan	1,156	0,052	1,381	0,044	1,161	0,051	1,374	0,054
Minkowski (3)	1,198	0,047	1,403	0,067	1,178	0,046	1,427	0,068
Minkowski (4)	1,192	0,046	1,442	0,059	1,18	0,039	1,429	0,052
Minkowski (5)	1,18	0,047	1,421	0,059	1,197	0,038	1,424	0,045

C.2.3. Lymphography

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,8	0,03	0,5	0,06	0,8	0,03	0,51	0,06
Canberra	0,79	0,03	0,5	0,07	0,8	0,03	0,48	0,06
Chebyshev	0,8	0,03	0,8	0,03	0,79	0,03	0,79	0,03
Distancia Correlacional	0,53	0,06	0,49	0,06	0,51	0,07	0,49	0,06
Distancia de Cosseno	0,33	0,06	0,49	0,06	0,35	0,05	0,48	0,08
Distancia de Cosseno Similaridade	0,72	0,05	0,72	0,04	0,71	0,04	0,71	0,04
Distancia de Jaccard	0,72	0,04	0,72	0,03	0,73	0,04	0,71	0,04
Euclidiana	0,79	0,04	0,8	0,03	0,79	0,03	0,81	0,03
Hamming	0,78	0,04	0,79	0,04	0,8	0,04	0,79	0,03
Manhattan	0,8	0,02	0,8	0,03	0,8	0,02	0,8	0,03
Minkowski (3)	0,8	0,03	0,79	0,03	0,79	0,03	0,81	0,03
Minkowski (4)	0,8	0,03	0,79	0,04	0,8	0,03	0,79	0,03
Minkowski (5)	0,8	0,03	0,8	0,03	0,78	0,04	0,8	0,03

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,031	0,004	0,032	0,002	0,031	0,004	0,034	0,003
Canberra	0,03	0,003	0,036	0,005	0,029	0,003	0,035	0,005
Chebyshev	0,027	0,003	0,031	0,005	0,028	0,003	0,03	0,003
Distancia Correlacional	0,057	0,009	0,062	0,01	0,054	0,009	0,06	0,01
Distancia de Cosseno	0,032	0,004	0,034	0,003	0,033	0,005	0,036	0,005
Distancia de Cosseno Similaridade	0,032	0,006	0,032	0,004	0,033	0,007	0,031	0,004
Distancia de Jaccard	0,029	0,002	0,029	0,002	0,029	0,002	0,03	0,003
Euclidiana	0,028	0,004	0,03	0,003	0,028	0,003	0,031	0,004
Hamming	0,017	0,001	0,017	0,001	0,017	0,002	0,017	0,001
Manhattan	0,028	0,004	0,031	0,005	0,026	0,001	0,029	0,002
Minkowski (3)	0,03	0,004	0,03	0,003	0,029	0,005	0,032	0,005
Minkowski (4)	0,028	0,003	0,03	0,003	0,029	0,004	0,029	0,003
Minkowski (5)	0,03	0,004	0,031	0,003	0,028	0,003	0,03	0,003

C.2.4. South African Heart

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,66	0,06	0,66	0,05	0,67	0,02	0,67	0,03
Canberra	0,68	0,02	0,67	0,05	0,68	0,03	0,67	0,03
Chebyshev	0,66	0,02	0,69	0,02	0,67	0,02	0,7	0,02
Distancia Correlacional	0,54	0,15	0,51	0,15	0,53	0,15	0,49	0,15
Distancia de Cosseno	0,59	0,02	0,6	0,02	0,59	0,03	0,59	0,03
Distancia de Cosseno Similaridade	0,66	0,02	0,66	0,02	0,66	0,02	0,66	0,02
Distancia de Jaccard	0,66	0,02	0,66	0,03	0,67	0,02	0,67	0,02
Euclidiana	0,66	0,02	0,7	0,02	0,66	0,02	0,69	0,03
Hamming	0,67	0,02	0,67	0,02	0,67	0,02	0,67	0,03
Manhattan	0,66	0,02	0,69	0,02	0,65	0,02	0,69	0,02
Minkowski (3)	0,66	0,02	0,7	0,02	0,66	0,02	0,7	0,02
Minkowski (4)	0,66	0,02	0,7	0,02	0,66	0,02	0,7	0,02
Minkowski (5)	0,66	0,02	0,69	0,02	0,65	0,02	0,7	0,02

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,281	0,02	0,434	0,044	0,286	0,022	0,417	0,033
Canberra	0,278	0,015	0,43	0,044	0,288	0,022	0,434	0,041
Chebyshev	0,214	0,015	0,307	0,025	0,219	0,02	0,306	0,026
Distancia Correlacional	0,534	0,037	0,734	0,039	0,522	0,04	0,719	0,042
Distancia de Cosseno	0,3	0,037	0,44	0,042	0,312	0,033	0,439	0,043
Distancia de Cosseno Similaridade	0,11	0,01	0,111	0,005	0,114	0,012	0,115	0,007
Distancia de Jaccard	0,113	0,012	0,11	0,005	0,111	0,009	0,112	0,005
Euclidiana	0,222	0,011	0,315	0,023	0,23	0,023	0,313	0,023
Hamming	0,074	0,007	0,073	0,006	0,073	0,004	0,073	0,005
Manhattan	0,21	0,012	0,303	0,018	0,212	0,015	0,31	0,028
Minkowski (3)	0,232	0,026	0,328	0,034	0,22	0,008	0,334	0,041
Minkowski (4)	0,224	0,015	0,313	0,026	0,221	0,01	0,318	0,022
Minkowski (5)	0,225	0,015	0,323	0,03	0,224	0,016	0,325	0,04

C.3. Bases desbalanceadas

C.3.1. Ecoli

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	Sem parâmetros		Com parâmetros		Sem parâmetros		Com parâmetros	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,94	0,01	0,7	0,28	0,94	0,02	0,67	0,3
Canberra	0,93	0,02	0,7	0,28	0,93	0,02	0,59	0,33
Chebyshev	0,94	0,02	0,94	0,01	0,95	0,01	0,94	0,02
Distancia Correlacional	0,94	0,01	0,93	0,01	0,94	0,01	0,93	0,01
Distancia de Cosseno	0,94	0,01	0,93	0,01	0,94	0,01	0,92	0,01
Distancia de Cosseno Similaridade	0,83	0,01	0,83	0,02	0,82	0,02	0,82	0,01
Distancia de Jaccard	0,82	0,01	0,83	0,02	0,83	0,02	0,83	0,01
Euclidiana	0,94	0,01	0,94	0,01	0,94	0,01	0,94	0,01
Hamming	0,84	0,01	0,83	0,02	0,83	0,01	0,83	0,02
Manhattan	0,94	0,01	0,94	0,01	0,94	0,01	0,94	0,01
Minkowski (3)	0,94	0,01	0,94	0,01	0,94	0,01	0,94	0,01
Minkowski (4)	0,94	0,01	0,94	0,01	0,94	0,01	0,94	0,01
Minkowski (5)	0,93	0,01	0,94	0,01	0,94	0,01	0,94	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,144	0,008	0,179	0,018	0,142	0,005	0,178	0,015
Canberra	0,143	0,006	0,184	0,029	0,142	0,007	0,18	0,021
Chebyshev	0,111	0,006	0,133	0,01	0,11	0,006	0,134	0,013
Distancia Correlacional	0,238	0,007	0,293	0,017	0,243	0,019	0,292	0,012
Distancia de Cosseno	0,155	0,015	0,179	0,016	0,152	0,015	0,179	0,016
Distancia de Cosseno Similaridade	0,062	0,004	0,06	0,003	0,063	0,004	0,063	0,005
Distancia de Jaccard	0,064	0,005	0,064	0,006	0,068	0,007	0,062	0,004
Euclidiana	0,114	0,004	0,136	0,005	0,112	0,004	0,136	0,005
Hamming	0,042	0,002	0,042	0,003	0,043	0,003	0,043	0,004
Manhattan	0,11	0,004	0,134	0,009	0,113	0,006	0,135	0,011
Minkowski (3)	0,115	0,004	0,137	0,006	0,115	0,005	0,137	0,007
Minkowski (4)	0,115	0,004	0,137	0,008	0,116	0,005	0,134	0,005
Minkowski (5)	0,114	0,004	0,135	0,006	0,114	0,004	0,137	0,005

C.3.2. Iris

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,99	0	0,99	0	0,99	0	0,99	0
Canberra	0,99	0	0,99	0	0,99	0	0,99	0
Chebyshev	0,99	0	0,99	0	0,99	0,01	0,99	0
Distancia Correlacional	0,99	0,01	0,99	0	0,99	0	0,99	0
Distancia de Cosseno	0,99	0	0,99	0	0,99	0	0,99	0
Distancia de Cosseno Similaridade	0,89	0,04	0,93	0,04	0,89	0,05	0,93	0,03
Distancia de Jaccard	0,9	0,02	0,93	0,04	0,9	0,03	0,93	0,04
Euclidiana	0,99	0	0,99	0	0,99	0	0,99	0
Hamming	0,96	0,02	0,95	0,02	0,95	0,02	0,95	0,03
Manhattan	0,99	0,01	0,99	0	0,99	0	0,99	0
Minkowski (3)	0,99	0	0,99	0	0,99	0	0,99	0
Minkowski (4)	1	0,01	0,99	0	0,99	0	0,99	0
Minkowski (5)	0,99	0	0,99	0	0,99	0,01	0,99	0

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,026	0,003	0,031	0,002	0,027	0,005	0,031	0,002
Canberra	0,026	0,003	0,031	0,002	0,025	0,002	0,03	0,002
Chebyshev	0,022	0,002	0,025	0,003	0,022	0,002	0,025	0,002
Distancia Correlacional	0,038	0,004	0,048	0,003	0,038	0,004	0,048	0,007
Distancia de Cosseno	0,026	0,002	0,031	0,001	0,027	0,003	0,031	0,002
Distancia de Cosseno Similaridade	0,016	0,001	0,016	0,001	0,016	0,001	0,015	0,001
Distancia de Jaccard	0,016	0,001	0,015	0,001	0,016	0,001	0,016	0,001
Euclidiana	0,023	0,002	0,026	0,002	0,023	0,002	0,026	0,002
Hamming	0,012	0,002	0,012	0,001	0,012	0,001	0,011	0,001
Manhattan	0,022	0,001	0,025	0,002	0,022	0,002	0,026	0,002
Minkowski (3)	0,023	0,004	0,026	0,002	0,022	0,002	0,026	0,002
Minkowski (4)	0,023	0,002	0,026	0,002	0,023	0,003	0,026	0,003
Minkowski (5)	0,023	0,002	0,026	0,002	0,023	0,002	0,026	0,002

C.3.3. Pima

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,67	0,04	0,63	0,07	0,66	0,05	0,63	0,07
Canberra	0,67	0,04	0,63	0,07	0,68	0,03	0,63	0,07
Chebyshev	0,72	0,02	0,73	0,01	0,71	0,01	0,73	0,01
Distancia Correlacional	0,67	0,02	0,68	0,01	0,67	0,02	0,68	0,01
Distancia de Cosseno	0,67	0,02	0,68	0,01	0,67	0,01	0,68	0,02
Distancia de Cosseno Similaridade	0,65	0,02	0,64	0,02	0,65	0,02	0,65	0,02
Distancia de Jaccard	0,66	0,02	0,65	0,02	0,65	0,02	0,64	0,02
Euclidiana	0,73	0,02	0,73	0,01	0,73	0,02	0,73	0,02
Hamming	0,64	0,02	0,64	0,02	0,64	0,01	0,65	0,01
Manhattan	0,73	0,01	0,73	0,01	0,73	0,02	0,74	0,01
Minkowski (3)	0,72	0,01	0,73	0,01	0,72	0,01	0,74	0,02
Minkowski (4)	0,72	0,02	0,74	0,01	0,72	0,01	0,73	0,01
Minkowski (5)	0,72	0,02	0,72	0,01	0,72	0,01	0,73	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,727	0,064	1,024	0,048	0,72	0,033	1,033	0,052
Canberra	0,712	0,035	1,028	0,046	0,715	0,077	1,029	0,06
Chebyshev	0,523	0,029	0,758	0,049	0,544	0,029	0,753	0,046
Distancia Correlacional	1,265	0,062	1,775	0,087	1,275	0,061	1,798	0,088
Distancia de Cosseno	0,744	0,029	1,048	0,052	0,74	0,036	1,069	0,056
Distancia de Cosseno Similaridade	0,25	0,009	0,255	0,016	0,244	0,007	0,251	0,008
Distancia de Jaccard	0,246	0,006	0,247	0,004	0,247	0,009	0,247	0,005
Euclidiana	0,56	0,037	0,776	0,044	0,555	0,037	0,789	0,048
Hamming	0,18	0,015	0,163	0,01	0,168	0,014	0,159	0,006
Manhattan	0,519	0,029	0,747	0,039	0,546	0,026	0,748	0,04
Minkowski (3)	0,58	0,029	0,778	0,046	0,564	0,038	0,765	0,046
Minkowski (4)	0,566	0,034	0,783	0,051	0,577	0,035	0,777	0,045
Minkowski (5)	0,564	0,038	0,781	0,043	0,575	0,027	0,783	0,044

C.3.4. Vehicle Silhouettes

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,65	0,02	0,68	0,02	0,65	0,01	0,67	0,02
Canberra	0,64	0,02	0,69	0,01	0,65	0,02	0,68	0,02
Chebyshev	0,62	0,02	0,68	0,02	0,6	0,01	0,68	0,01
Distancia Correlacional	0,66	0,01	0,7	0,02	0,66	0,01	0,7	0,01
Distancia de Cosseno	0,66	0,02	0,7	0,02	0,67	0,02	0,71	0,02
Distancia de Cosseno Similaridade	0,64	0,01	0,66	0,01	0,63	0,01	0,65	0,01
Distancia de Jaccard	0,64	0,01	0,66	0,01	0,63	0,02	0,65	0,01
Euclidiana	0,65	0,01	0,7	0,01	0,64	0,01	0,71	0,02
Hamming	0,66	0,02	0,66	0,02	0,65	0,02	0,66	0,01
Manhattan	0,68	0,01	0,71	0,02	0,67	0,02	0,71	0,01
Minkowski (3)	0,63	0,02	0,69	0,02	0,64	0,02	0,69	0,02
Minkowski (4)	0,63	0,02	0,7	0,01	0,63	0,02	0,69	0,02
Minkowski (5)	0,62	0,01	0,69	0,01	0,62	0,01	0,69	0,02

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	1,71	0,091	2,763	0,105	1,716	0,068	2,743	0,109
Canberra	1,74	0,092	2,763	0,112	1,715	0,047	2,741	0,103
Chebyshev	1,221	0,044	1,914	0,068	1,247	0,045	1,916	0,081
Distancia Correlacional	3,208	0,113	4,85	0,116	3,168	0,104	4,835	0,118
Distancia de Cosseno	1,794	0,061	2,827	0,096	1,809	0,079	2,825	0,104
Distancia de Cosseno Similaridade	0,605	0,038	0,591	0,043	0,586	0,033	0,608	0,04
Distancia de Jaccard	0,608	0,041	0,607	0,049	0,595	0,033	0,597	0,038
Euclidiana	1,295	0,049	1,942	0,056	1,281	0,041	1,933	0,052
Hamming	0,262	0,01	0,265	0,017	0,265	0,018	0,259	0,007
Manhattan	1,237	0,059	1,885	0,057	1,239	0,047	1,882	0,055
Minkowski (3)	1,297	0,042	1,941	0,035	1,279	0,048	1,954	0,058
Minkowski (4)	1,304	0,051	1,944	0,045	1,302	0,047	1,936	0,039
Minkowski (5)	1,304	0,045	1,947	0,043	1,298	0,053	1,939	0,045

C.3.5. Wisconsin

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,97	0	0,64	0,05	0,97	0,01	0,63	0,06
Canberra	0,97	0	0,64	0,05	0,97	0	0,65	0,01
Chebyshev	0,96	0,01	0,95	0,01	0,97	0	0,95	0,01
Distancia Correlacional	0,87	0,03	0,85	0,02	0,88	0,05	0,86	0,02
Distancia de Cosseno	0,9	0,01	0,9	0,01	0,9	0,01	0,9	0,01
Distancia de Cosseno Similaridade	0,96	0,01	0,9	0,03	0,95	0,03	0,89	0,05
Distancia de Jaccard	0,95	0,01	0,9	0,03	0,95	0,01	0,9	0,07
Euclidiana	0,97	0,01	0,96	0,01	0,97	0,01	0,95	0,01
Hamming	0,95	0,01	0,92	0,01	0,95	0,01	0,92	0,02
Manhattan	0,97	0,01	0,95	0,01	0,97	0,01	0,95	0,01
Minkowski (3)	0,97	0,01	0,96	0,01	0,97	0,01	0,95	0,01
Minkowski (4)	0,97	0,01	0,95	0,01	0,97	0,01	0,96	0,01
Minkowski (5)	0,97	0,01	0,95	0,01	0,97	0,01	0,95	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,605	0,041	0,355	0,023	0,62	0,041	0,352	0,029
Canberra	0,594	0,039	0,35	0,016	0,588	0,032	0,354	0,021
Chebyshev	0,442	0,036	0,252	0,005	0,449	0,041	0,267	0,022
Distancia Correlacional	1,055	0,04	0,665	0,024	1,049	0,031	0,649	0,036
Distancia de Cosseno	0,615	0,045	0,355	0,017	0,626	0,042	0,383	0,025
Distancia de Cosseno Similaridade	0,252	0,016	0,124	0,005	0,247	0,01	0,123	0,006
Distancia de Jaccard	0,249	0,011	0,124	0,016	0,247	0,011	0,123	0,006
Euclidiana	0,459	0,037	0,261	0,008	0,463	0,041	0,266	0,015
Hamming	0,145	0,012	0,069	0,003	0,139	0,009	0,07	0,004
Manhattan	0,439	0,023	0,255	0,012	0,426	0,027	0,266	0,015
Minkowski (3)	0,478	0,045	0,259	0,006	0,459	0,04	0,26	0,008
Minkowski (4)	0,46	0,027	0,264	0,015	0,467	0,036	0,259	0,006
Minkowski (5)	0,474	0,038	0,26	0,007	0,459	0,024	0,262	0,015

C.4. Bases com atributos faltantes

C.4.1. Cylinder Bands

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	Sem parâmetros		Com parâmetros		Sem parâmetros		Com parâmetros	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,66	0,08	0,59	0,13	0,67	0,05	0,6	0,12
Canberra	0,68	0,02	0,57	0,13	0,61	0,12	0,58	0,13
Chebyshev	0,61	0,02	0,61	0,02	0,61	0,02	0,6	0,02
Distancia Correlacional	0,5	0,09	0,65	0,02	0,52	0,08	0,66	0,02
Distancia de Cosseno	0,36	0,04	0,65	0,02	0,36	0,04	0,66	0,02
Distancia de Cosseno Similaridade	0,66	0,02	0,7	0,02	0,67	0,03	0,7	0,02
Distancia de Jaccard	0,66	0,02	0,69	0,02	0,67	0,02	0,7	0,02
Euclidiana	0,6	0,02	0,66	0,02	0,59	0,02	0,66	0,02
Hamming	0,7	0,02	0,72	0,02	0,71	0,02	0,71	0,02
Manhattan	0,61	0,02	0,69	0,02	0,61	0,02	0,7	0,02
Minkowski (3)	0,59	0,02	0,64	0,02	0,59	0,02	0,65	0,02
Minkowski (4)	0,58	0,02	0,63	0,02	0,59	0,02	0,63	0,02
Minkowski (5)	0,58	0,02	0,62	0,02	0,59	0,02	0,63	0,02

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,748	0,034	1,09	0,054	0,754	0,041	1,096	0,053
Canberra	0,736	0,035	1,094	0,069	0,737	0,04	1,092	0,067
Chebyshev	0,54	0,035	0,789	0,045	0,544	0,035	0,784	0,045
Distancia Correlacional	1,35	0,049	1,918	0,063	1,341	0,046	1,949	0,086
Distancia de Cosseno	0,774	0,044	1,131	0,057	0,765	0,036	1,12	0,05
Distancia de Cosseno Similaridade	0,283	0,023	0,278	0,01	0,284	0,022	0,289	0,024
Distancia de Jaccard	0,278	0,018	0,279	0,016	0,278	0,016	0,28	0,022
Euclidiana	0,577	0,048	0,792	0,046	0,58	0,034	0,797	0,041
Hamming	0,128	0,008	0,133	0,012	0,128	0,008	0,127	0,008
Manhattan	0,537	0,032	0,78	0,036	0,538	0,028	0,768	0,034
Minkowski (3)	0,572	0,039	0,781	0,045	0,572	0,042	0,79	0,043
Minkowski (4)	0,573	0,04	0,798	0,052	0,574	0,035	0,796	0,035
Minkowski (5)	0,573	0,036	0,802	0,052	0,564	0,029	0,813	0,075

C.4.2. Heart Disease (Cleveland)

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	Sem parâmetros		Com parâmetros		Sem parâmetros		Com parâmetros	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,5	0,13	0,34	0,2	0,25	0,18	0,41	0,19
Canberra	0,51	0,13	0,37	0,2	0,33	0,21	0,3	0,2
Chebyshev	0,49	0,03	0,52	0,04	0,51	0,04	0,51	0,04
Distancia Correlacional	0,52	0,04	0,56	0,02	0,54	0,02	0,56	0,02
Distancia de Cosseno	0,51	0,04	0,56	0,02	0,55	0,02	0,55	0,03
Distancia de Cosseno Similaridade	0,47	0,04	0,42	0,04	0,4	0,05	0,41	0,07
Distancia de Jaccard	0,45	0,03	0,43	0,05	0,42	0,05	0,39	0,06
Euclidiana	0,5	0,02	0,57	0,02	0,57	0,02	0,57	0,01
Hamming	0,58	0,02	0,57	0,02	0,58	0,02	0,58	0,02
Manhattan	0,52	0,02	0,58	0,02	0,58	0,02	0,58	0,02
Minkowski (3)	0,49	0,03	0,56	0,02	0,57	0,02	0,55	0,02
Minkowski (4)	0,49	0,02	0,57	0,02	0,55	0,03	0,57	0,02
Minkowski (5)	0,49	0,02	0,56	0,02	0,56	0,02	0,57	0,02

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,214	0,017	0,214	0,019	0,2	0,029	0,222	0,026
Canberra	0,204	0,006	0,208	0,012	0,202	0,029	0,207	0,012
Chebyshev	0,158	0,018	0,152	0,012	0,139	0,009	0,154	0,011
Distancia Correlacional	0,369	0,028	0,374	0,027	0,316	0,024	0,381	0,033
Distancia de Cosseno	0,218	0,016	0,211	0,009	0,197	0,024	0,233	0,025
Distancia de Cosseno Similaridade	0,081	0,006	0,081	0,007	0,08	0,005	0,077	0,004
Distancia de Jaccard	0,079	0,009	0,078	0,005	0,081	0,006	0,076	0,006
Euclidiana	0,17	0,02	0,171	0,025	0,139	0,007	0,16	0,022
Hamming	0,044	0,002	0,046	0,003	0,046	0,004	0,046	0,004
Manhattan	0,152	0,009	0,157	0,023	0,142	0,015	0,157	0,015
Minkowski (3)	0,164	0,021	0,167	0,021	0,141	0,006	0,16	0,014
Minkowski (4)	0,168	0,017	0,164	0,018	0,145	0,011	0,164	0,023
Minkowski (5)	0,159	0,019	0,159	0,016	0,141	0,008	0,158	0,011

C.4.3. Hepatitis

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,82	0,02	0,74	0,16	0,84	0,03	0,74	0,16
Canberra	0,84	0,03	0,68	0,23	0,82	0,02	0,74	0,16
Chebyshev	0,75	0,04	0,73	0,14	0,75	0,04	0,73	0,13
Distancia Correlacional	0,77	0,06	0,83	0,02	0,75	0,14	0,83	0,02
Distancia de Cosseno	0,73	0,15	0,83	0,02	0,72	0,15	0,83	0,02
Distancia de Cosseno Similaridade	0,79	0,02	0,78	0,02	0,8	0,03	0,79	0,03
Distancia de Jaccard	0,79	0,02	0,79	0,02	0,79	0,02	0,78	0,02
Euclidiana	0,74	0,04	0,83	0,02	0,74	0,07	0,83	0,03
Hamming	0,82	0,02	0,82	0,02	0,82	0,04	0,82	0,02
Manhattan	0,75	0,05	0,83	0,02	0,75	0,06	0,83	0,02
Minkowski (3)	0,76	0,05	0,82	0,02	0,74	0,06	0,83	0,02
Minkowski (4)	0,76	0,04	0,83	0,02	0,75	0,05	0,83	0,02
Minkowski (5)	0,73	0,06	0,83	0,02	0,73	0,06	0,83	0,02

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,077	0,012	0,09	0,014	0,06	0,006	0,08	0,009
Canberra	0,06	0,004	0,079	0,007	0,064	0,009	0,08	0,007
Chebyshev	0,047	0,005	0,068	0,013	0,05	0,008	0,066	0,011
Distancia Correlacional	0,108	0,006	0,135	0,008	0,103	0,004	0,134	0,009
Distancia de Cosseno	0,08	0,016	0,089	0,015	0,07	0,013	0,081	0,007
Distancia de Cosseno Similaridade	0,031	0,003	0,031	0,002	0,033	0,004	0,032	0,003
Distancia de Jaccard	0,032	0,002	0,035	0,005	0,032	0,005	0,033	0,004
Euclidiana	0,051	0,009	0,06	0,008	0,057	0,01	0,066	0,012
Hamming	0,019	0,001	0,018	0,001	0,019	0,001	0,019	0,001
Manhattan	0,055	0,01	0,065	0,014	0,046	0,003	0,056	0,007
Minkowski (3)	0,048	0,006	0,058	0,004	0,05	0,007	0,072	0,01
Minkowski (4)	0,061	0,009	0,062	0,01	0,048	0,007	0,062	0,011
Minkowski (5)	0,049	0,006	0,061	0,007	0,05	0,007	0,063	0,011

C.4.4. Mammographic Mass

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,8	0,02	0,78	0,02	0,8	0,01	0,77	0,06
Canberra	0,8	0,01	0,77	0,03	0,81	0,01	0,74	0,07
Chebyshev	0,77	0,01	0,77	0,02	0,77	0,01	0,77	0,02
Distancia Correlacional	0,52	0,02	0,74	0,01	0,52	0,03	0,74	0,02
Distancia de Cosseno	0,6	0,07	0,75	0,02	0,59	0,07	0,75	0,02
Distancia de Cosseno Similaridade	0,82	0,01	0,79	0,03	0,82	0,01	0,79	0,03
Distancia de Jaccard	0,82	0,01	0,79	0,04	0,82	0,01	0,79	0,05
Euclidiana	0,79	0,01	0,79	0,01	0,79	0,01	0,78	0,02
Hamming	0,82	0,01	0,81	0,01	0,81	0,01	0,81	0,02
Manhattan	0,8	0,01	0,8	0,01	0,8	0,01	0,8	0,01
Minkowski (3)	0,79	0,01	0,78	0,02	0,79	0,01	0,77	0,01
Minkowski (4)	0,79	0,01	0,78	0,02	0,79	0,01	0,78	0,02
Minkowski (5)	0,79	0,01	0,78	0,01	0,79	0,02	0,78	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,666	0,038	0,457	0,027	0,671	0,038	0,468	0,037
Canberra	0,678	0,04	0,456	0,033	0,664	0,037	0,451	0,032
Chebyshev	0,535	0,035	0,363	0,029	0,524	0,037	0,352	0,03
Distancia Correlacional	1,183	0,037	0,767	0,036	1,172	0,036	0,777	0,049
Distancia de Cosseno	0,7	0,036	0,472	0,035	0,692	0,037	0,458	0,032
Distancia de Cosseno Similaridade	0,336	0,032	0,172	0,013	0,326	0,025	0,167	0,014
Distancia de Jaccard	0,335	0,035	0,169	0,013	0,331	0,027	0,17	0,013
Euclidiana	0,54	0,035	0,348	0,026	0,537	0,035	0,355	0,027
Hamming	0,199	0,01	0,107	0,009	0,201	0,012	0,109	0,006
Manhattan	0,547	0,047	0,348	0,028	0,516	0,035	0,346	0,031
Minkowski (3)	0,537	0,036	0,352	0,023	0,546	0,023	0,36	0,032
Minkowski (4)	0,549	0,029	0,364	0,031	0,543	0,029	0,352	0,032
Minkowski (5)	0,551	0,042	0,367	0,038	0,536	0,028	0,359	0,029

C.5. Bases com atributos categóricos

C.5.1. Car Evaluation

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	Sem parâmetros		Com parâmetros		Sem parâmetros		Com parâmetros	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Distancia de Cosseno Similaridade	0,7	0,01	0,69	0,02	0,69	0,01	0,69	0,02
Distancia de Jaccard	0,7	0,01	0,69	0,02	0,69	0,01	0,69	0,02
Hamming	0,86	0,01	0,86	0,01	0,85	0,01	0,85	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	Sem parâmetros		Com parâmetros		Sem parâmetros		Com parâmetros	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Distancia de Cosseno Similaridade	1,139	0,045	1,14	0,033	1,147	0,046	1,132	0,041
Distancia de Jaccard	1,137	0,03	1,119	0,032	1,125	0,036	1,123	0,041
Hamming	0,705	0,095	0,677	0,026	0,673	0,03	0,671	0,025

C.5.2. Tic-Tac-Toe Endgame

	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	Sem parâmetros		Com parâmetros		Sem parâmetros		Com parâmetros	
<u>Distâncias</u>	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Distancia de Cosseno Similaridade	0,62	0,06	0,58	0,09	0,62	0,06	0,58	0,11
Distancia de Jaccard	0,63	0,04	0,59	0,1	0,63	0,04	0,6	0,1
Hamming	0,91	0,02	0,91	0,01	0,91	0,02	0,91	0,01

	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	Sem parâmetros		Com parâmetros		Sem parâmetros		Com parâmetros	
<u>Distâncias</u>	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Distancia de Cosseno Similaridade	0,58	0,023	0,59	0,031	0,575	0,025	0,578	0,027
Distancia de Jaccard	0,582	0,032	0,595	0,029	0,58	0,021	0,576	0,038
Hamming	0,288	0,019	0,286	0,015	0,289	0,021	0,29	0,016

C.6. Bases com ruído de atributos

C.6.1.1. Contraceptive Method Choice - 05%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio</u>			
					<u>Padrão</u>			
	Sem parâmetros	Com parâmetros	Sem parâmetros	Com parâmetros	Sem parâmetros	Com parâmetros	Sem parâmetros	Com parâmetros
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,41	0,02	0,36	0,07	0,4	0,04	0,34	0,07
Canberra	0,4	0,02	0,35	0,08	0,4	0,03	0,32	0,07
Chebyshev	0,49	0,01	0,43	0,01	0,48	0,02	0,43	0,01
Distancia Correlacional	0,5	0,01	0,45	0,02	0,49	0,01	0,45	0,01
Distancia de Cosseno	0,5	0,01	0,45	0,01	0,49	0,01	0,45	0,01
Distancia de Cosseno Similaridade	0,4	0,02	0,43	0,03	0,41	0,01	0,41	0,02
Distancia de Jaccard	0,41	0,02	0,42	0,02	0,41	0,02	0,42	0,02
Euclidiana	0,5	0,01	0,45	0,01	0,5	0,01	0,45	0,01
Hamming	0,45	0,01	0,45	0,01	0,44	0,01	0,44	0,01
Manhattan	0,5	0,01	0,46	0,01	0,5	0,01	0,45	0,01
Minkowski (3)	0,5	0,01	0,45	0,01	0,49	0,01	0,45	0,02
Minkowski (4)	0,5	0,01	0,45	0,01	0,5	0,01	0,45	0,01
Minkowski (5)	0,5	0,01	0,46	0,01	0,5	0,01	0,45	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	2,497	0,087	2,759	0,104	2,548	0,085	2,761	0,12
Canberra	2,523	0,096	2,77	0,122	2,526	0,109	2,773	0,116
Chebyshev	1,8	0,066	1,959	0,086	1,818	0,069	1,971	0,088
Distancia Correlacional	4,569	0,147	5,025	0,178	4,67	0,124	4,956	0,137
Distancia de Cosseno	2,552	0,082	2,825	0,12	2,579	0,08	2,879	0,172
Distancia de Cosseno Similaridade	1,118	0,04	0,958	0,036	1,096	0,034	0,976	0,04
Distancia de Jaccard	1,109	0,052	0,962	0,044	1,117	0,044	0,996	0,097
Euclidiana	1,825	0,05	2,041	0,086	1,863	0,067	2,024	0,081
Hamming	0,522	0,041	0,463	0,026	0,532	0,039	0,472	0,029
Manhattan	1,784	0,063	1,991	0,088	1,81	0,075	1,973	0,074
Minkowski (3)	1,846	0,057	2,083	0,089	1,858	0,076	2,036	0,1
Minkowski (4)	1,869	0,065	2,071	0,084	1,823	0,041	2,064	0,075
Minkowski (5)	1,841	0,059	2,065	0,078	1,841	0,051	2,054	0,091

C.6.1.2. Contraceptive Method Choice - 10%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,4	0,04	0,36	0,07	0,38	0,05	0,33	0,07
Canberra	0,4	0,04	0,35	0,07	0,41	0,02	0,34	0,07
Chebyshev	0,48	0,02	0,42	0,02	0,47	0,01	0,42	0,01
Distancia Correlacional	0,48	0,01	0,45	0,01	0,48	0,01	0,45	0,01
Distancia de Cosseno	0,48	0,01	0,45	0,02	0,49	0,01	0,45	0,01
Distancia de Cosseno Similaridade	0,4	0,02	0,42	0,02	0,4	0,01	0,42	0,02
Distancia de Jaccard	0,4	0,02	0,42	0,03	0,4	0,01	0,41	0,02
Euclidiana	0,49	0,01	0,44	0,01	0,48	0,01	0,45	0,01
Hamming	0,45	0,01	0,45	0,01	0,45	0,01	0,45	0,01
Manhattan	0,48	0,01	0,45	0,01	0,48	0,01	0,45	0,01
Minkowski (3)	0,48	0,01	0,45	0,02	0,48	0,01	0,44	0,01
Minkowski (4)	0,49	0,01	0,44	0,01	0,48	0,01	0,44	0,01
Minkowski (5)	0,49	0,01	0,45	0,01	0,48	0,01	0,44	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	2,495	0,078	2,936	0,08	2,588	0,066	2,924	0,086
Canberra	2,561	0,154	2,969	0,109	2,541	0,125	2,911	0,084
Chebyshev	1,805	0,093	2,049	0,077	1,831	0,106	2,074	0,075
Distancia Correlacional	4,585	0,12	5,214	0,102	4,651	0,123	5,211	0,113
Distancia de Cosseno	2,592	0,089	3,025	0,083	2,606	0,092	2,985	0,07
Distancia de Cosseno Similaridade	1,107	0,031	1,006	0,039	1,121	0,043	1,022	0,046
Distancia de Jaccard	1,13	0,047	1,006	0,043	1,122	0,034	1,016	0,04
Euclidiana	1,85	0,068	2,135	0,08	1,891	0,076	2,159	0,067
Hamming	0,528	0,04	0,481	0,033	0,522	0,036	0,508	0,047
Manhattan	1,8	0,062	2,1	0,08	1,826	0,068	2,081	0,07
Minkowski (3)	1,846	0,046	2,209	0,118	1,858	0,066	2,174	0,105
Minkowski (4)	1,841	0,056	2,19	0,098	1,836	0,054	2,216	0,121
Minkowski (5)	1,848	0,057	2,17	0,083	1,856	0,087	2,181	0,076

C.6.1.3. Contraceptive Method Choice - 15%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,4	0,03	0,35	0,08	0,39	0,04	0,38	0,07
Canberra	0,4	0,04	0,34	0,08	0,41	0,04	0,36	0,07
Chebyshev	0,46	0,02	0,41	0,01	0,46	0,01	0,4	0,01
Distancia Correlacional	0,47	0,01	0,43	0,01	0,46	0,01	0,42	0,01
Distancia de Cosseno	0,48	0,01	0,43	0,02	0,47	0,01	0,42	0,01
Distancia de Cosseno Similaridade	0,4	0,02	0,42	0,02	0,41	0,02	0,41	0,02
Distancia de Jaccard	0,4	0,02	0,42	0,02	0,4	0,01	0,4	0,02
Euclidiana	0,47	0,01	0,43	0,01	0,47	0,01	0,43	0,01
Hamming	0,43	0,01	0,43	0,01	0,42	0,01	0,42	0,01
Manhattan	0,47	0,01	0,43	0,01	0,47	0,01	0,43	0,01
Minkowski (3)	0,47	0,01	0,43	0,01	0,47	0,01	0,42	0,01
Minkowski (4)	0,47	0,01	0,42	0,01	0,47	0,01	0,42	0,01
Minkowski (5)	0,47	0,01	0,43	0,01	0,47	0,01	0,42	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	2,537	0,076	3,08	0,143	2,621	0,125	3,072	0,111
Canberra	2,571	0,087	3,054	0,067	2,519	0,083	3,086	0,109
Chebyshev	1,85	0,066	2,142	0,073	1,847	0,068	2,14	0,072
Distancia Correlacional	4,619	0,193	5,4	0,11	4,686	0,149	5,416	0,151
Distancia de Cosseno	2,617	0,098	3,131	0,083	2,606	0,119	3,064	0,088
Distancia de Cosseno Similaridade	1,112	0,039	1,039	0,038	1,113	0,033	1,061	0,049
Distancia de Jaccard	1,124	0,045	1,027	0,04	1,119	0,05	1,063	0,055
Euclidiana	1,888	0,067	2,27	0,092	1,909	0,104	2,257	0,08
Hamming	0,528	0,037	0,5	0,035	0,518	0,03	0,501	0,037
Manhattan	1,833	0,089	2,21	0,073	1,829	0,083	2,227	0,076
Minkowski (3)	1,868	0,069	2,286	0,088	1,863	0,074	2,288	0,077
Minkowski (4)	1,881	0,104	2,277	0,086	1,863	0,06	2,316	0,094
Minkowski (5)	1,879	0,106	2,288	0,069	1,903	0,127	2,298	0,079

C.6.1.4. Contraceptive Method Choice - 20%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,4	0,03	0,36	0,06	0,39	0,05	0,37	0,06
Canberra	0,4	0,04	0,37	0,07	0,37	0,07	0,36	0,08
Chebyshev	0,45	0,01	0,41	0,01	0,45	0,01	0,4	0,01
Distancia Correlacional	0,46	0,01	0,43	0,01	0,46	0,01	0,43	0,01
Distancia de Cosseno	0,47	0,01	0,43	0,01	0,46	0,01	0,43	0,01
Distancia de Cosseno Similaridade	0,4	0,01	0,42	0,02	0,4	0,02	0,41	0,02
Distancia de Jaccard	0,4	0,01	0,42	0,02	0,4	0,01	0,41	0,02
Euclidiana	0,46	0,01	0,42	0,01	0,46	0,01	0,42	0,01
Hamming	0,43	0,01	0,43	0,01	0,42	0,01	0,43	0,01
Manhattan	0,46	0,01	0,43	0,01	0,46	0,01	0,43	0,01
Minkowski (3)	0,46	0,01	0,42	0,01	0,45	0,01	0,42	0,01
Minkowski (4)	0,45	0,01	0,42	0,01	0,45	0,01	0,42	0,01
Minkowski (5)	0,45	0,01	0,42	0,01	0,45	0,01	0,42	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	2,588	0,096	3,112	0,097	2,598	0,089	3,066	0,088
Canberra	2,575	0,11	3,111	0,074	2,548	0,114	3,097	0,092
Chebyshev	1,831	0,048	2,158	0,076	1,864	0,066	2,169	0,077
Distancia Correlacional	4,652	0,129	5,475	0,132	4,773	0,107	5,476	0,122
Distancia de Cosseno	2,642	0,101	3,185	0,074	2,669	0,085	3,118	0,088
Distancia de Cosseno Similaridade	1,135	0,057	1,057	0,05	1,148	0,053	1,065	0,046
Distancia de Jaccard	1,132	0,05	1,063	0,048	1,13	0,041	1,07	0,037
Euclidiana	1,881	0,064	2,303	0,087	1,902	0,065	2,311	0,101
Hamming	0,524	0,036	0,516	0,033	0,539	0,038	0,512	0,039
Manhattan	1,842	0,063	2,246	0,083	1,872	0,079	2,23	0,099
Minkowski (3)	1,886	0,074	2,307	0,084	1,882	0,062	2,3	0,071
Minkowski (4)	1,905	0,07	2,291	0,076	1,857	0,041	2,302	0,084
Minkowski (5)	1,881	0,077	2,288	0,085	1,874	0,055	2,299	0,091

C.6.2.1. Heart - 5%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,51	0,06	0,58	0,08	0,56	0,08	0,53	0,08
Canberra	0,52	0,07	0,54	0,07	0,55	0,08	0,51	0,07
Chebyshev	0,7	0,03	0,66	0,02	0,66	0,02	0,71	0,03
Distancia Correlacional	0,77	0,02	0,67	0,02	0,67	0,03	0,77	0,03
Distancia de Cosseno	0,78	0,03	0,68	0,02	0,69	0,03	0,77	0,02
Distancia de Cosseno Similaridade	0,58	0,04	0,71	0,04	0,71	0,04	0,59	0,03
Distancia de Jaccard	0,56	0,05	0,72	0,03	0,72	0,03	0,57	0,03
Euclidiana	0,8	0,02	0,66	0,02	0,66	0,03	0,79	0,02
Hamming	0,81	0,02	0,81	0,02	0,82	0,02	0,81	0,02
Manhattan	0,81	0,02	0,7	0,01	0,7	0,03	0,8	0,02
Minkowski (3)	0,79	0,02	0,66	0,03	0,66	0,02	0,79	0,03
Minkowski (4)	0,79	0,02	0,65	0,02	0,65	0,03	0,78	0,02
Minkowski (5)	0,78	0,02	0,67	0,03	0,66	0,02	0,78	0,02

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,169	0,015	0,138	0,009	0,169	0,019	0,136	0,007
Canberra	0,183	0,027	0,138	0,012	0,178	0,027	0,141	0,015
Chebyshev	0,121	0,007	0,107	0,01	0,125	0,011	0,103	0,006
Distancia Correlacional	0,287	0,014	0,248	0,028	0,283	0,014	0,245	0,019
Distancia de Cosseno	0,189	0,032	0,145	0,01	0,177	0,021	0,143	0,015
Distancia de Cosseno Similaridade	0,062	0,004	0,063	0,006	0,064	0,006	0,065	0,006
Distancia de Jaccard	0,064	0,006	0,063	0,006	0,07	0,007	0,069	0,01
Euclidiana	0,132	0,015	0,108	0,01	0,128	0,007	0,108	0,008
Hamming	0,038	0,003	0,039	0,002	0,041	0,004	0,039	0,002
Manhattan	0,127	0,013	0,104	0,009	0,123	0,006	0,106	0,015
Minkowski (3)	0,129	0,01	0,111	0,014	0,126	0,007	0,109	0,008
Minkowski (4)	0,129	0,008	0,107	0,006	0,127	0,008	0,108	0,007
Minkowski (5)	0,127	0,008	0,108	0,008	0,127	0,008	0,109	0,015

C.6.2.2. Heart - 10%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,53	0,07	0,5	0,06	0,55	0,08	0,51	0,07
Canberra	0,55	0,09	0,52	0,07	0,53	0,09	0,53	0,06
Chebyshev	0,63	0,02	0,7	0,03	0,65	0,03	0,71	0,02
Distancia Correlacional	0,64	0,03	0,76	0,02	0,64	0,03	0,77	0,02
Distancia de Cosseno	0,64	0,03	0,77	0,02	0,64	0,03	0,77	0,02
Distancia de Cosseno Similaridade	0,7	0,03	0,56	0,04	0,7	0,03	0,56	0,03
Distancia de Jaccard	0,69	0,04	0,57	0,04	0,69	0,04	0,55	0,04
Euclidiana	0,65	0,03	0,78	0,02	0,64	0,02	0,78	0,02
Hamming	0,77	0,02	0,77	0,02	0,78	0,02	0,77	0,02
Manhattan	0,68	0,02	0,79	0,02	0,67	0,03	0,79	0,02
Minkowski (3)	0,63	0,02	0,78	0,02	0,64	0,03	0,77	0,02
Minkowski (4)	0,64	0,03	0,78	0,03	0,63	0,03	0,77	0,02
Minkowski (5)	0,63	0,03	0,77	0,03	0,64	0,03	0,77	0,03

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,136	0,008	0,18	0,03	0,14	0,009	0,168	0,01
Canberra	0,138	0,008	0,19	0,027	0,139	0,01	0,188	0,03
Chebyshev	0,102	0,005	0,123	0,008	0,107	0,009	0,126	0,009
Distancia Correlacional	0,24	0,015	0,306	0,031	0,252	0,025	0,291	0,024
Distancia de Cosseno	0,156	0,025	0,198	0,03	0,147	0,016	0,179	0,02
Distancia de Cosseno Similaridade	0,061	0,004	0,062	0,005	0,064	0,007	0,062	0,005
Distancia de Jaccard	0,064	0,006	0,062	0,005	0,069	0,007	0,071	0,007
Euclidiana	0,111	0,008	0,133	0,013	0,105	0,006	0,13	0,015
Hamming	0,04	0,003	0,037	0,002	0,039	0,003	0,042	0,004
Manhattan	0,107	0,009	0,125	0,007	0,104	0,011	0,127	0,012
Minkowski (3)	0,109	0,009	0,134	0,02	0,107	0,008	0,127	0,008
Minkowski (4)	0,108	0,006	0,127	0,01	0,109	0,007	0,126	0,008
Minkowski (5)	0,108	0,01	0,129	0,007	0,108	0,006	0,131	0,015

C.6.2.3. Heart - 15%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,57	0,08	0,53	0,06	0,55	0,08	0,53	0,07
Canberra	0,58	0,08	0,52	0,07	0,57	0,07	0,51	0,07
Chebyshev	0,62	0,03	0,69	0,03	0,63	0,03	0,7	0,03
Distancia Correlacional	0,64	0,03	0,76	0,02	0,62	0,03	0,76	0,02
Distancia de Cosseno	0,64	0,03	0,76	0,02	0,65	0,02	0,78	0,02
Distancia de Cosseno Similaridade	0,7	0,03	0,57	0,04	0,69	0,03	0,57	0,04
Distancia de Jaccard	0,69	0,03	0,58	0,04	0,69	0,05	0,56	0,04
Euclidiana	0,62	0,03	0,77	0,03	0,62	0,03	0,78	0,02
Hamming	0,78	0,02	0,79	0,02	0,79	0,02	0,79	0,02
Manhattan	0,67	0,02	0,78	0,02	0,66	0,03	0,79	0,02
Minkowski (3)	0,62	0,02	0,77	0,02	0,62	0,02	0,77	0,02
Minkowski (4)	0,62	0,02	0,77	0,02	0,62	0,02	0,77	0,02
Minkowski (5)	0,62	0,02	0,77	0,02	0,62	0,02	0,77	0,02

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,143	0,018	0,201	0,033	0,138	0,011	0,2	0,033
Canberra	0,139	0,008	0,206	0,024	0,137	0,008	0,206	0,035
Chebyshev	0,109	0,015	0,141	0,022	0,104	0,008	0,138	0,013
Distancia Correlacional	0,244	0,02	0,315	0,018	0,255	0,028	0,325	0,038
Distancia de Cosseno	0,152	0,018	0,205	0,031	0,145	0,016	0,187	0,022
Distancia de Cosseno Similaridade	0,063	0,005	0,068	0,007	0,066	0,008	0,066	0,006
Distancia de Jaccard	0,065	0,007	0,066	0,006	0,068	0,008	0,069	0,005
Euclidiana	0,107	0,004	0,143	0,013	0,11	0,007	0,138	0,012
Hamming	0,039	0,003	0,039	0,003	0,039	0,002	0,04	0,004
Manhattan	0,108	0,016	0,138	0,016	0,103	0,005	0,137	0,017
Minkowski (3)	0,107	0,005	0,143	0,018	0,111	0,015	0,143	0,013
Minkowski (4)	0,108	0,007	0,143	0,019	0,11	0,011	0,143	0,021
Minkowski (5)	0,11	0,008	0,14	0,012	0,111	0,011	0,142	0,015

C.6.2.4. Heart - 20%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,55	0,09	0,51	0,07	0,58	0,06	0,53	0,07
Canberra	0,57	0,06	0,5	0,08	0,55	0,06	0,52	0,07
Chebyshev	0,62	0,03	0,66	0,02	0,62	0,02	0,66	0,03
Distancia Correlacional	0,62	0,03	0,72	0,02	0,61	0,03	0,73	0,02
Distancia de Cosseno	0,62	0,03	0,72	0,02	0,62	0,02	0,73	0,02
Distancia de Cosseno Similaridade	0,68	0,03	0,58	0,03	0,68	0,03	0,59	0,04
Distancia de Jaccard	0,67	0,03	0,59	0,04	0,67	0,03	0,57	0,03
Euclidiana	0,63	0,02	0,75	0,03	0,63	0,03	0,76	0,03
Hamming	0,75	0,03	0,76	0,02	0,75	0,03	0,76	0,02
Manhattan	0,63	0,02	0,77	0,02	0,63	0,03	0,77	0,02
Minkowski (3)	0,61	0,02	0,73	0,02	0,63	0,03	0,73	0,03
Minkowski (4)	0,62	0,02	0,73	0,03	0,62	0,02	0,73	0,02
Minkowski (5)	0,62	0,03	0,74	0,03	0,61	0,02	0,73	0,03

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,146	0,018	0,184	0,028	0,143	0,011	0,183	0,026
Canberra	0,142	0,013	0,189	0,028	0,146	0,016	0,191	0,032
Chebyshev	0,108	0,012	0,132	0,013	0,108	0,01	0,13	0,02
Distancia Correlacional	0,25	0,021	0,303	0,027	0,267	0,037	0,306	0,031
Distancia de Cosseno	0,158	0,027	0,199	0,032	0,158	0,025	0,186	0,027
Distancia de Cosseno Similaridade	0,065	0,007	0,066	0,009	0,065	0,006	0,066	0,01
Distancia de Jaccard	0,065	0,006	0,064	0,005	0,07	0,01	0,067	0,007
Euclidiana	0,111	0,008	0,134	0,01	0,11	0,008	0,131	0,011
Hamming	0,039	0,003	0,038	0,002	0,039	0,004	0,041	0,004
Manhattan	0,111	0,012	0,13	0,016	0,112	0,019	0,129	0,01
Minkowski (3)	0,115	0,012	0,132	0,009	0,111	0,012	0,137	0,016
Minkowski (4)	0,112	0,012	0,131	0,01	0,111	0,009	0,134	0,012
Minkowski (5)	0,11	0,008	0,135	0,014	0,115	0,012	0,133	0,011

C.6.3.1. Iris - 05%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,92	0,02	0,91	0,02	0,93	0,01	0,91	0,02
Canberra	0,92	0,02	0,91	0,02	0,92	0,03	0,91	0,02
Chebyshev	0,9	0,02	0,88	0,03	0,89	0,02	0,88	0,02
Distancia Correlacional	0,9	0,02	0,76	0,03	0,91	0,02	0,76	0,03
Distancia de Cosseno	0,91	0,02	0,76	0,04	0,91	0,02	0,76	0,03
Distancia de Cosseno Similaridade	0,71	0,05	0,75	0,05	0,71	0,05	0,73	0,03
Distancia de Jaccard	0,73	0,03	0,74	0,05	0,71	0,04	0,73	0,05
Euclidiana	0,91	0,02	0,9	0,02	0,91	0,02	0,91	0,02
Hamming	0,81	0,04	0,8	0,04	0,8	0,04	0,79	0,03
Manhattan	0,91	0,02	0,93	0,02	0,91	0,02	0,93	0,02
Minkowski (3)	0,9	0,03	0,89	0,02	0,91	0,02	0,89	0,02
Minkowski (4)	0,9	0,02	0,89	0,02	0,91	0,02	0,88	0,02
Minkowski (5)	0,9	0,03	0,89	0,02	0,91	0,01	0,9	0,02

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	Sem parâmetros		Com parâmetros		Sem parâmetros		Com parâmetros	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,025	0,002	0,034	0,006	0,025	0,003	0,034	0,004
Canberra	0,025	0,003	0,034	0,006	0,025	0,002	0,034	0,005
Chebyshev	0,022	0,002	0,026	0,003	0,021	0,002	0,026	0,002
Distancia Correlacional	0,042	0,007	0,056	0,01	0,038	0,005	0,053	0,009
Distancia de Cosseno	0,026	0,003	0,035	0,004	0,028	0,005	0,035	0,004
Distancia de Cosseno Similaridade	0,015	0,001	0,015	0,001	0,016	0,003	0,015	0,001
Distancia de Jaccard	0,015	0,001	0,015	0,001	0,015	0,001	0,015	0,001
Euclidiana	0,022	0,002	0,027	0,003	0,022	0,002	0,026	0,003
Hamming	0,012	0,001	0,012	0,001	0,012	0,002	0,012	0,001
Manhattan	0,022	0,002	0,027	0,003	0,021	0,002	0,025	0,003
Minkowski (3)	0,022	0,002	0,027	0,003	0,022	0,002	0,028	0,005
Minkowski (4)	0,022	0,002	0,028	0,004	0,021	0,001	0,029	0,005
Minkowski (5)	0,023	0,003	0,027	0,003	0,021	0,002	0,026	0,003

C.6.3.2. Iris - 10%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,89	0,02	0,89	0,03	0,9	0,02	0,89	0,02
Canberra	0,89	0,02	0,88	0,03	0,9	0,02	0,89	0,02
Chebyshev	0,84	0,03	0,84	0,03	0,84	0,02	0,84	0,02
Distancia Correlacional	0,83	0,02	0,69	0,03	0,84	0,02	0,68	0,03
Distancia de Cosseno	0,85	0,03	0,72	0,04	0,86	0,02	0,72	0,04
Distancia de Cosseno Similaridade	0,69	0,05	0,73	0,05	0,69	0,05	0,73	0,05
Distancia de Jaccard	0,68	0,06	0,73	0,05	0,7	0,05	0,74	0,05
Euclidiana	0,87	0,03	0,87	0,02	0,86	0,02	0,88	0,03
Hamming	0,79	0,05	0,79	0,03	0,78	0,05	0,79	0,04
Manhattan	0,87	0,02	0,91	0,02	0,87	0,02	0,9	0,02
Minkowski (3)	0,87	0,03	0,86	0,03	0,85	0,02	0,86	0,03
Minkowski (4)	0,86	0,03	0,85	0,03	0,86	0,02	0,85	0,02
Minkowski (5)	0,84	0,02	0,84	0,03	0,84	0,03	0,85	0,02

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,026	0,003	0,033	0,004	0,026	0,003	0,033	0,004
Canberra	0,025	0,002	0,032	0,003	0,026	0,003	0,031	0,002
Chebyshev	0,022	0,002	0,027	0,003	0,023	0,003	0,027	0,006
Distancia Correlacional	0,041	0,005	0,05	0,006	0,041	0,006	0,049	0,005
Distancia de Cosseno	0,028	0,004	0,036	0,005	0,026	0,003	0,034	0,004
Distancia de Cosseno Similaridade	0,015	0,001	0,015	0	0,015	0,001	0,015	0,001
Distancia de Jaccard	0,015	0,001	0,015	0,001	0,015	0,001	0,015	0,001
Euclidiana	0,022	0,002	0,028	0,005	0,022	0,003	0,026	0,002
Hamming	0,012	0,001	0,012	0,001	0,012	0,001	0,012	0,002
Manhattan	0,021	0,001	0,026	0,003	0,021	0,002	0,025	0,002
Minkowski (3)	0,022	0,002	0,026	0,003	0,022	0,002	0,026	0,002
Minkowski (4)	0,022	0,002	0,027	0,003	0,022	0,002	0,029	0,004
Minkowski (5)	0,022	0,002	0,025	0,002	0,022	0,002	0,026	0,003

C.6.3.3. Iris - 15%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,89	0,02	0,89	0,03	0,9	0,02	0,89	0,02
Canberra	0,9	0,02	0,89	0,03	0,9	0,02	0,9	0,03
Chebyshev	0,83	0,04	0,83	0,03	0,82	0,03	0,84	0,03
Distancia Correlacional	0,82	0,03	0,69	0,04	0,81	0,03	0,7	0,03
Distancia de Cosseno	0,82	0,04	0,72	0,04	0,82	0,03	0,73	0,04
Distancia de Cosseno Similaridade	0,66	0,04	0,69	0,04	0,68	0,04	0,68	0,05
Distancia de Jaccard	0,68	0,05	0,7	0,03	0,68	0,04	0,69	0,06
Euclidiana	0,85	0,03	0,87	0,03	0,85	0,04	0,87	0,02
Hamming	0,78	0,05	0,76	0,04	0,76	0,04	0,76	0,05
Manhattan	0,88	0,03	0,89	0,02	0,87	0,03	0,89	0,03
Minkowski (3)	0,85	0,03	0,86	0,03	0,84	0,04	0,86	0,03
Minkowski (4)	0,83	0,03	0,85	0,03	0,84	0,04	0,86	0,03
Minkowski (5)	0,83	0,03	0,85	0,03	0,83	0,03	0,84	0,03

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,026	0,004	0,031	0,003	0,026	0,002	0,033	0,004
Canberra	0,026	0,004	0,031	0,003	0,027	0,004	0,031	0,003
Chebyshev	0,023	0,002	0,027	0,004	0,022	0,002	0,025	0,003
Distancia Correlacional	0,04	0,005	0,05	0,006	0,041	0,005	0,051	0,006
Distancia de Cosseno	0,028	0,004	0,035	0,006	0,028	0,004	0,035	0,005
Distancia de Cosseno Similaridade	0,015	0,001	0,015	0,001	0,016	0,001	0,015	0,001
Distancia de Jaccard	0,015	0,001	0,015	0,001	0,015	0,001	0,015	0,001
Euclidiana	0,023	0,002	0,027	0,006	0,022	0,001	0,025	0,002
Hamming	0,011	0,002	0,011	0,001	0,011	0,001	0,012	0,002
Manhattan	0,022	0,002	0,026	0,003	0,023	0,003	0,025	0,003
Minkowski (3)	0,022	0,001	0,025	0,002	0,023	0,004	0,025	0,002
Minkowski (4)	0,023	0,003	0,025	0,002	0,023	0,003	0,029	0,005
Minkowski (5)	0,022	0,002	0,026	0,003	0,023	0,002	0,027	0,004

C.6.3.4. Iris - 20%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,86	0,03	0,85	0,02	0,86	0,03	0,86	0,03
Canberra	0,86	0,02	0,86	0,03	0,86	0,03	0,86	0,02
Chebyshev	0,81	0,03	0,8	0,03	0,8	0,03	0,79	0,03
Distancia Correlacional	0,8	0,02	0,68	0,03	0,8	0,03	0,67	0,02
Distancia de Cosseno	0,84	0,03	0,69	0,03	0,83	0,02	0,68	0,03
Distancia de Cosseno Similaridade	0,67	0,04	0,7	0,04	0,65	0,06	0,69	0,04
Distancia de Jaccard	0,67	0,05	0,68	0,04	0,65	0,04	0,69	0,05
Euclidiana	0,84	0,03	0,84	0,03	0,85	0,03	0,84	0,04
Hamming	0,76	0,04	0,74	0,04	0,75	0,05	0,74	0,04
Manhattan	0,86	0,03	0,88	0,02	0,85	0,04	0,87	0,03
Minkowski (3)	0,84	0,04	0,84	0,03	0,83	0,04	0,82	0,03
Minkowski (4)	0,82	0,03	0,83	0,03	0,83	0,03	0,82	0,03
Minkowski (5)	0,83	0,03	0,83	0,03	0,83	0,03	0,82	0,03

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,027	0,003	0,034	0,004	0,027	0,003	0,032	0,003
Canberra	0,027	0,002	0,031	0,002	0,027	0,003	0,032	0,003
Chebyshev	0,024	0,002	0,03	0,006	0,023	0,002	0,026	0,003
Distancia Correlacional	0,043	0,006	0,05	0,007	0,042	0,005	0,054	0,009
Distancia de Cosseno	0,027	0,003	0,037	0,006	0,03	0,006	0,034	0,004
Distancia de Cosseno Similaridade	0,015	0,001	0,016	0,001	0,015	0,001	0,015	0,001
Distancia de Jaccard	0,015	0,001	0,015	0,001	0,016	0,002	0,015	0,001
Euclidiana	0,024	0,003	0,028	0,005	0,023	0,002	0,026	0,002
Hamming	0,012	0,001	0,011	0,001	0,012	0,001	0,012	0,001
Manhattan	0,022	0,002	0,026	0,002	0,022	0,002	0,026	0,002
Minkowski (3)	0,023	0,001	0,026	0,002	0,024	0,002	0,027	0,003
Minkowski (4)	0,023	0,002	0,026	0,003	0,024	0,002	0,03	0,003
Minkowski (5)	0,023	0,002	0,027	0,003	0,023	0,001	0,028	0,003

C.6.4.1. Pima - 05%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,66	0,05	0,65	0,07	0,68	0,03	0,65	0,06
Canberra	0,66	0,05	0,63	0,07	0,68	0,03	0,65	0,06
Chebyshev	0,71	0,01	0,7	0,01	0,7	0,02	0,71	0,01
Distancia Correlacional	0,66	0,02	0,66	0,01	0,65	0,02	0,66	0,02
Distancia de Cosseno	0,65	0,01	0,66	0,02	0,66	0,02	0,67	0,01
Distancia de Cosseno Similaridade	0,65	0,01	0,64	0,02	0,65	0,01	0,63	0,03
Distancia de Jaccard	0,65	0,01	0,63	0,02	0,65	0,02	0,64	0,02
Euclidiana	0,71	0,02	0,72	0,01	0,71	0,01	0,72	0,01
Hamming	0,64	0,02	0,63	0,02	0,64	0,02	0,63	0,02
Manhattan	0,72	0,02	0,72	0,01	0,72	0,02	0,72	0,01
Minkowski (3)	0,71	0,02	0,71	0,02	0,71	0,01	0,72	0,02
Minkowski (4)	0,71	0,02	0,71	0,02	0,7	0,02	0,71	0,01
Minkowski (5)	0,7	0,02	0,71	0,01	0,71	0,02	0,71	0,02

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,8	0,044	1,071	0,061	0,798	0,051	1,054	0,044
Canberra	0,778	0,045	1,066	0,038	0,78	0,043	1,042	0,044
Chebyshev	0,61	0,044	0,776	0,044	0,587	0,048	0,766	0,038
Distancia Correlacional	1,408	0,048	1,804	0,068	1,384	0,043	1,821	0,074
Distancia de Cosseno	0,806	0,044	1,061	0,044	0,786	0,045	1,108	0,027
Distancia de Cosseno Similaridade	0,248	0,01	0,249	0,013	0,246	0,009	0,252	0,015
Distancia de Jaccard	0,255	0,017	0,247	0,008	0,245	0,008	0,249	0,013
Euclidiana	0,61	0,045	0,795	0,039	0,604	0,053	0,817	0,034
Hamming	0,171	0,016	0,164	0,019	0,167	0,014	0,161	0,013
Manhattan	0,617	0,048	0,771	0,049	0,587	0,031	0,774	0,039
Minkowski (3)	0,616	0,043	0,808	0,039	0,604	0,045	0,805	0,051
Minkowski (4)	0,608	0,044	0,803	0,035	0,604	0,049	0,801	0,041
Minkowski (5)	0,603	0,052	0,801	0,046	0,61	0,041	0,788	0,047

C.6.4.2. Pima - 10%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,66	0,02	0,63	0,06	0,66	0,03	0,65	0,06
Canberra	0,65	0,05	0,64	0,07	0,66	0,03	0,64	0,06
Chebyshev	0,69	0,02	0,68	0,02	0,69	0,01	0,69	0,01
Distancia Correlacional	0,65	0,02	0,65	0,01	0,65	0,02	0,65	0,01
Distancia de Cosseno	0,65	0,01	0,65	0,02	0,65	0,02	0,66	0,01
Distancia de Cosseno Similaridade	0,64	0,02	0,64	0,02	0,64	0,01	0,63	0,02
Distancia de Jaccard	0,64	0,02	0,63	0,03	0,64	0,02	0,64	0,02
Euclidiana	0,69	0,01	0,7	0,01	0,69	0,02	0,7	0,01
Hamming	0,64	0,01	0,64	0,02	0,64	0,02	0,64	0,02
Manhattan	0,7	0,02	0,71	0,01	0,7	0,02	0,71	0,01
Minkowski (3)	0,7	0,02	0,7	0,01	0,69	0,01	0,7	0,01
Minkowski (4)	0,69	0,02	0,69	0,02	0,69	0,02	0,69	0,01
Minkowski (5)	0,69	0,01	0,69	0,01	0,68	0,02	0,69	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	Sem parâmetros		Com parâmetros		Sem parâmetros		Com parâmetros	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,806	0,042	1,047	0,051	0,813	0,044	1,065	0,043
Canberra	0,795	0,049	1,067	0,094	0,816	0,035	1,064	0,048
Chebyshev	0,599	0,051	0,779	0,057	0,609	0,044	0,795	0,056
Distancia Correlacional	1,455	0,048	1,856	0,111	1,392	0,048	1,854	0,089
Distancia de Cosseno	0,832	0,045	1,08	0,08	0,805	0,043	1,112	0,052
Distancia de Cosseno Similaridade	0,25	0,014	0,253	0,019	0,253	0,021	0,262	0,026
Distancia de Jaccard	0,245	0,009	0,248	0,01	0,251	0,017	0,254	0,017
Euclidiana	0,632	0,048	0,797	0,033	0,62	0,047	0,816	0,041
Hamming	0,167	0,014	0,164	0,014	0,166	0,017	0,16	0,01
Manhattan	0,613	0,04	0,789	0,039	0,6	0,043	0,787	0,042
Minkowski (3)	0,627	0,045	0,806	0,044	0,606	0,04	0,807	0,055
Minkowski (4)	0,639	0,043	0,779	0,04	0,638	0,045	0,801	0,047
Minkowski (5)	0,629	0,054	0,793	0,048	0,638	0,051	0,793	0,044

C.6.4.3. Pima - 15%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,64	0,05	0,61	0,06	0,65	0,04	0,64	0,06
Canberra	0,64	0,04	0,62	0,06	0,66	0,03	0,63	0,07
Chebyshev	0,67	0,02	0,68	0,02	0,67	0,02	0,68	0,01
Distancia Correlacional	0,65	0,02	0,66	0,01	0,65	0,02	0,67	0,02
Distancia de Cosseno	0,65	0,01	0,66	0,01	0,65	0,01	0,66	0,01
Distancia de Cosseno Similaridade	0,65	0,01	0,63	0,02	0,65	0,02	0,63	0,03
Distancia de Jaccard	0,65	0,02	0,63	0,03	0,64	0,01	0,63	0,03
Euclidiana	0,68	0,01	0,7	0,02	0,69	0,01	0,69	0,02
Hamming	0,63	0,02	0,63	0,03	0,63	0,02	0,63	0,02
Manhattan	0,69	0,01	0,7	0,02	0,69	0,01	0,71	0,01
Minkowski (3)	0,68	0,01	0,68	0,01	0,68	0,01	0,69	0,01
Minkowski (4)	0,67	0,01	0,68	0,01	0,67	0,02	0,68	0,02
Minkowski (5)	0,67	0,02	0,68	0,02	0,67	0,01	0,68	0,02

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,826	0,049	1,06	0,051	0,827	0,053	1,074	0,046
Canberra	0,829	0,044	1,072	0,048	0,826	0,045	1,081	0,037
Chebyshev	0,604	0,042	0,794	0,046	0,617	0,045	0,796	0,04
Distancia Correlacional	1,469	0,072	1,842	0,053	1,455	0,055	1,832	0,075
Distancia de Cosseno	0,855	0,04	1,072	0,042	0,857	0,044	1,126	0,051
Distancia de Cosseno Similaridade	0,249	0,013	0,255	0,023	0,255	0,021	0,247	0,008
Distancia de Jaccard	0,245	0,012	0,256	0,022	0,251	0,016	0,251	0,013
Euclidiana	0,647	0,047	0,812	0,052	0,65	0,05	0,813	0,045
Hamming	0,168	0,016	0,168	0,016	0,169	0,011	0,165	0,014
Manhattan	0,612	0,043	0,793	0,041	0,6	0,047	0,789	0,058
Minkowski (3)	0,626	0,035	0,802	0,051	0,64	0,038	0,815	0,061
Minkowski (4)	0,631	0,044	0,806	0,054	0,626	0,048	0,835	0,089
Minkowski (5)	0,64	0,048	0,801	0,04	0,636	0,046	0,786	0,051

C.6.4.4. Pima - 20%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,65	0,04	0,63	0,05	0,66	0,04	0,62	0,06
Canberra	0,65	0,05	0,62	0,05	0,64	0,05	0,62	0,06
Chebyshev	0,67	0,02	0,67	0,02	0,67	0,01	0,66	0,02
Distancia Correlacional	0,63	0,01	0,65	0,01	0,64	0,02	0,64	0,01
Distancia de Cosseno	0,64	0,02	0,65	0,01	0,64	0,01	0,65	0,01
Distancia de Cosseno Similaridade	0,65	0,01	0,63	0,02	0,64	0,01	0,63	0,02
Distancia de Jaccard	0,64	0,02	0,63	0,03	0,65	0,02	0,64	0,02
Euclidiana	0,68	0,01	0,69	0,02	0,69	0,02	0,69	0,01
Hamming	0,63	0,02	0,63	0,03	0,64	0,01	0,63	0,02
Manhattan	0,69	0,01	0,7	0,01	0,68	0,01	0,7	0,01
Minkowski (3)	0,67	0,02	0,68	0,01	0,68	0,02	0,68	0,01
Minkowski (4)	0,67	0,01	0,67	0,01	0,68	0,02	0,68	0,01
Minkowski (5)	0,67	0,02	0,67	0,01	0,67	0,02	0,67	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,854	0,042	1,111	0,05	0,866	0,048	1,113	0,059
Canberra	0,867	0,05	1,107	0,06	0,871	0,054	1,085	0,046
Chebyshev	0,639	0,05	0,825	0,042	0,638	0,044	0,805	0,04
Distancia Correlacional	1,542	0,072	1,872	0,061	1,516	0,066	1,897	0,084
Distancia de Cosseno	0,887	0,046	1,122	0,053	0,882	0,049	1,167	0,052
Distancia de Cosseno Similaridade	0,252	0,015	0,254	0,018	0,244	0,01	0,253	0,015
Distancia de Jaccard	0,253	0,013	0,25	0,01	0,247	0,011	0,252	0,017
Euclidiana	0,661	0,046	0,81	0,043	0,667	0,051	0,837	0,048
Hamming	0,172	0,019	0,163	0,011	0,173	0,014	0,165	0,015
Manhattan	0,63	0,036	0,806	0,036	0,635	0,053	0,812	0,041
Minkowski (3)	0,665	0,039	0,816	0,05	0,649	0,041	0,831	0,044
Minkowski (4)	0,668	0,048	0,812	0,047	0,649	0,045	0,829	0,042
Minkowski (5)	0,644	0,045	0,807	0,043	0,655	0,043	0,824	0,042

C.6.5.1. Sonar, Mines vs. Rocks - 05%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,76	0,03	0,75	0,03	0,77	0,03	0,75	0,03
Canberra	0,75	0,04	0,75	0,03	0,74	0,04	0,76	0,03
Chebyshev	0,57	0,04	0,61	0,03	0,57	0,04	0,61	0,03
Distancia Correlacional	0,67	0,04	0,72	0,05	0,68	0,04	0,72	0,04
Distancia de Cosseno	0,68	0,03	0,72	0,03	0,7	0,04	0,72	0,03
Distancia de Cosseno Similaridade	0,56	0,03	0,58	0,03	0,56	0,04	0,59	0,04
Distancia de Jaccard	0,56	0,03	0,59	0,04	0,57	0,05	0,59	0,03
Euclidiana	0,69	0,03	0,74	0,04	0,7	0,04	0,74	0,03
Hamming	0,59	0,03	0,59	0,04	0,59	0,04	0,6	0,03
Manhattan	0,73	0,05	0,78	0,04	0,73	0,04	0,77	0,03
Minkowski (3)	0,65	0,04	0,68	0,03	0,66	0,04	0,68	0,04
Minkowski (4)	0,64	0,03	0,66	0,04	0,64	0,04	0,66	0,03
Minkowski (5)	0,61	0,04	0,65	0,02	0,63	0,04	0,64	0,03

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,388	0,058	0,592	0,038	0,396	0,045	0,605	0,058
Canberra	0,394	0,05	0,606	0,057	0,39	0,036	0,612	0,062
Chebyshev	0,281	0,034	0,395	0,034	0,287	0,03	0,385	0,031
Distancia Correlacional	0,776	0,054	1,061	0,05	0,794	0,058	1,091	0,05
Distancia de Cosseno	0,414	0,056	0,602	0,062	0,407	0,046	0,608	0,04
Distancia de Cosseno Similaridade	0,203	0,029	0,188	0,022	0,19	0,017	0,185	0,027
Distancia de Jaccard	0,212	0,044	0,177	0,01	0,203	0,027	0,186	0,023
Euclidiana	0,302	0,039	0,4	0,036	0,3	0,033	0,388	0,021
Hamming	0,052	0,009	0,055	0,009	0,051	0,007	0,05	0,007
Manhattan	0,299	0,041	0,388	0,033	0,31	0,05	0,387	0,032
Minkowski (3)	0,289	0,029	0,406	0,046	0,294	0,048	0,418	0,047
Minkowski (4)	0,291	0,04	0,398	0,026	0,286	0,022	0,397	0,026
Minkowski (5)	0,287	0,026	0,394	0,035	0,286	0,039	0,405	0,044

C.6.5.2. Sonar, Mines vs. Rocks - 10%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,75	0,03	0,73	0,04	0,76	0,04	0,73	0,03
Canberra	0,75	0,04	0,72	0,03	0,75	0,04	0,73	0,04
Chebyshev	0,59	0,04	0,61	0,04	0,6	0,03	0,62	0,04
Distancia Correlacional	0,68	0,05	0,71	0,03	0,69	0,03	0,71	0,04
Distancia de Cosseno	0,7	0,03	0,71	0,03	0,69	0,04	0,71	0,04
Distancia de Cosseno Similaridade	0,56	0,04	0,58	0,04	0,56	0,03	0,58	0,03
Distancia de Jaccard	0,56	0,03	0,58	0,04	0,56	0,03	0,58	0,03
Euclidiana	0,67	0,05	0,75	0,04	0,71	0,03	0,76	0,03
Hamming	0,59	0,03	0,58	0,03	0,57	0,04	0,58	0,03
Manhattan	0,73	0,03	0,76	0,04	0,74	0,03	0,76	0,04
Minkowski (3)	0,66	0,04	0,7	0,03	0,66	0,05	0,7	0,03
Minkowski (4)	0,62	0,03	0,67	0,04	0,64	0,04	0,67	0,03
Minkowski (5)	0,62	0,04	0,64	0,05	0,62	0,03	0,66	0,03

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,409	0,05	0,596	0,037	0,409	0,053	0,638	0,062
Canberra	0,413	0,055	0,62	0,056	0,413	0,06	0,627	0,057
Chebyshev	0,291	0,043	0,381	0,026	0,294	0,046	0,387	0,023
Distancia Correlacional	0,789	0,081	1,056	0,042	0,811	0,093	1,079	0,054
Distancia de Cosseno	0,419	0,058	0,618	0,051	0,422	0,049	0,616	0,045
Distancia de Cosseno Similaridade	0,217	0,029	0,189	0,024	0,198	0,025	0,188	0,025
Distancia de Jaccard	0,209	0,03	0,185	0,03	0,214	0,031	0,19	0,028
Euclidiana	0,299	0,026	0,406	0,041	0,318	0,03	0,397	0,024
Hamming	0,054	0,01	0,058	0,011	0,049	0,006	0,051	0,007
Manhattan	0,296	0,032	0,405	0,043	0,323	0,05	0,39	0,024
Minkowski (3)	0,287	0,024	0,396	0,018	0,29	0,028	0,416	0,046
Minkowski (4)	0,299	0,029	0,403	0,037	0,298	0,029	0,405	0,032
Minkowski (5)	0,293	0,032	0,402	0,029	0,298	0,031	0,398	0,019

C.6.5.3. Sonar, Mines vs. Rocks - 15%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,73	0,05	0,71	0,03	0,72	0,04	0,72	0,03
Canberra	0,72	0,04	0,71	0,04	0,72	0,05	0,73	0,03
Chebyshev	0,58	0,04	0,58	0,03	0,58	0,03	0,58	0,03
Distancia Correlacional	0,66	0,04	0,67	0,04	0,66	0,04	0,68	0,04
Distancia de Cosseno	0,64	0,04	0,67	0,03	0,65	0,04	0,68	0,04
Distancia de Cosseno Similaridade	0,52	0,03	0,56	0,03	0,52	0,05	0,56	0,03
Distancia de Jaccard	0,52	0,04	0,55	0,03	0,52	0,03	0,55	0,03
Euclidiana	0,67	0,04	0,69	0,03	0,68	0,04	0,68	0,05
Hamming	0,58	0,03	0,58	0,04	0,58	0,03	0,59	0,04
Manhattan	0,72	0,03	0,76	0,03	0,73	0,03	0,75	0,04
Minkowski (3)	0,63	0,03	0,65	0,05	0,63	0,04	0,66	0,04
Minkowski (4)	0,61	0,04	0,62	0,03	0,62	0,03	0,62	0,04
Minkowski (5)	0,6	0,04	0,61	0,03	0,6	0,04	0,61	0,04

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,423	0,044	0,605	0,058	0,417	0,029	0,643	0,047
Canberra	0,412	0,027	0,607	0,05	0,414	0,031	0,596	0,059
Chebyshev	0,3	0,027	0,388	0,029	0,3	0,028	0,397	0,029
Distancia Correlacional	0,802	0,049	1,048	0,067	0,82	0,063	1,066	0,064
Distancia de Cosseno	0,424	0,028	0,618	0,057	0,455	0,054	0,622	0,061
Distancia de Cosseno Similaridade	0,199	0,027	0,194	0,022	0,192	0,009	0,194	0,03
Distancia de Jaccard	0,199	0,024	0,18	0,013	0,201	0,021	0,189	0,024
Euclidiana	0,307	0,029	0,399	0,023	0,316	0,032	0,396	0,019
Hamming	0,053	0,007	0,055	0,009	0,049	0,005	0,049	0,005
Manhattan	0,295	0,025	0,4	0,04	0,336	0,06	0,386	0,022
Minkowski (3)	0,302	0,02	0,421	0,049	0,3	0,015	0,417	0,046
Minkowski (4)	0,303	0,032	0,401	0,022	0,299	0,02	0,399	0,018
Minkowski (5)	0,299	0,014	0,403	0,027	0,304	0,022	0,404	0,035

C.6.5.4. Sonar, Mines vs. Rocks - 20%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,73	0,03	0,7	0,04	0,72	0,04	0,71	0,03
Canberra	0,73	0,04	0,71	0,04	0,73	0,05	0,69	0,04
Chebyshev	0,56	0,04	0,59	0,04	0,56	0,04	0,59	0,02
Distancia Correlacional	0,64	0,03	0,67	0,03	0,65	0,03	0,68	0,03
Distancia de Cosseno	0,65	0,04	0,67	0,03	0,67	0,04	0,68	0,03
Distancia de Cosseno Similaridade	0,53	0,03	0,56	0,03	0,53	0,04	0,57	0,03
Distancia de Jaccard	0,52	0,04	0,57	0,03	0,53	0,03	0,57	0,05
Euclidiana	0,68	0,04	0,68	0,04	0,66	0,04	0,7	0,03
Hamming	0,56	0,05	0,56	0,04	0,56	0,03	0,55	0,03
Manhattan	0,73	0,05	0,75	0,04	0,72	0,04	0,75	0,03
Minkowski (3)	0,64	0,04	0,66	0,03	0,64	0,03	0,66	0,03
Minkowski (4)	0,61	0,03	0,64	0,02	0,61	0,04	0,64	0,04
Minkowski (5)	0,58	0,03	0,63	0,03	0,6	0,04	0,61	0,02

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,444	0,054	0,627	0,053	0,434	0,044	0,645	0,052
Canberra	0,435	0,04	0,627	0,065	0,433	0,038	0,62	0,046
Chebyshev	0,304	0,03	0,398	0,034	0,31	0,032	0,405	0,036
Distancia Correlacional	0,837	0,054	1,097	0,058	0,834	0,056	1,089	0,05
Distancia de Cosseno	0,439	0,027	0,621	0,054	0,47	0,052	0,618	0,044
Distancia de Cosseno Similaridade	0,192	0,021	0,187	0,02	0,193	0,02	0,186	0,02
Distancia de Jaccard	0,209	0,028	0,188	0,02	0,201	0,031	0,197	0,028
Euclidiana	0,313	0,022	0,401	0,026	0,321	0,037	0,413	0,035
Hamming	0,054	0,009	0,06	0,009	0,05	0,005	0,055	0,007
Manhattan	0,293	0,014	0,405	0,04	0,323	0,049	0,398	0,036
Minkowski (3)	0,317	0,035	0,423	0,054	0,309	0,031	0,41	0,032
Minkowski (4)	0,31	0,019	0,416	0,045	0,315	0,024	0,439	0,068
Minkowski (5)	0,306	0,014	0,42	0,038	0,309	0,024	0,417	0,038

C.7. Bases com ruído de classes

C.7.1.1. Contraceptive Method Choice - 05%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio</u>			
					<u>Padrão</u>			
	Sem parâmetros	Com parâmetros	Sem parâmetros	Com parâmetros	Sem parâmetros	Com parâmetros	Sem parâmetros	Com parâmetros
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,41	0,03	0,42	0,02	0,39	0,05	0,41	0,04
Canberra	0,39	0,05	0,41	0,04	0,41	0,02	0,41	0,02
Chebyshev	0,54	0,02	0,48	0,01	0,53	0,02	0,48	0,01
Distancia Correlacional	0,59	0	0,49	0,01	0,61	0,01	0,49	0,01
Distancia de Cosseno	0,57	0	0,5	0,01	0,58	0	0,5	0,01
Distancia de Cosseno Similaridade	0,42	0,02	0,45	0,02	0,43	0,02	0,44	0,02
Distancia de Jaccard	0,43	0,02	0,44	0,02	0,43	0,02	0,44	0,02
Euclidiana	0,55	0,01	0,52	0,01	0,55	0,01	0,51	0,01
Hamming	0,53	0,02	0,5	0,02	0,52	0,02	0,49	0,02
Manhattan	0,56	0,01	0,53	0,01	0,54	0,02	0,52	0,01
Minkowski (3)	0,56	0,01	0,52	0,01	0,56	0,01	0,5	0,01
Minkowski (4)	0,56	0,01	0,52	0,01	0,56	0,01	0,5	0,01
Minkowski (5)	0,56	0,01	0,52	0,01	0,55	0,01	0,5	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	1,72	0,065	1,975	0,06	1,74	0,077	1,991	0,09
Canberra	1,7	0,053	2,003	0,086	1,707	0,068	1,98	0,066
Chebyshev	1,277	0,055	1,406	0,045	1,245	0,047	1,408	0,057
Distancia Correlacional	3,215	0,093	3,611	0,084	3,251	0,1	3,624	0,079
Distancia de Cosseno	1,752	0,052	2,043	0,075	1,788	0,119	2,032	0,076
Distancia de Cosseno Similaridade	0,737	0,035	0,656	0,037	0,752	0,033	0,664	0,043
Distancia de Jaccard	0,747	0,037	0,65	0,034	0,736	0,044	0,656	0,048
Euclidiana	1,294	0,054	1,459	0,053	1,271	0,063	1,47	0,051
Hamming	0,32	0,026	0,289	0,043	0,318	0,028	0,281	0,022
Manhattan	1,257	0,057	1,426	0,052	1,225	0,037	1,429	0,062
Minkowski (3)	1,28	0,047	1,491	0,055	1,29	0,061	1,482	0,061
Minkowski (4)	1,283	0,049	1,479	0,037	1,275	0,051	1,485	0,065
Minkowski (5)	1,274	0,049	1,47	0,045	1,306	0,081	1,476	0,05

C.7.1.2. Contraceptive Method Choice - 10%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,4	0,03	0,4	0,04	0,38	0,05	0,4	0,06
Canberra	0,4	0,02	0,4	0,04	0,4	0,03	0,41	0,02
Chebyshev	0,54	0,02	0,48	0,01	0,54	0,03	0,48	0,01
Distancia Correlacional	0,58	0	0,51	0,01	0,6	0,01	0,5	0,01
Distancia de Cosseno	0,56	0,01	0,51	0,01	0,56	0,01	0,49	0,01
Distancia de Cosseno Similaridade	0,43	0,02	0,44	0,02	0,43	0,02	0,44	0,03
Distancia de Jaccard	0,43	0,02	0,45	0,01	0,42	0,02	0,44	0,02
Euclidiana	0,57	0,01	0,49	0,01	0,58	0,02	0,48	0,01
Hamming	0,52	0,02	0,5	0,02	0,51	0,01	0,49	0,02
Manhattan	0,57	0,01	0,51	0,01	0,56	0,02	0,5	0,01
Minkowski (3)	0,57	0,01	0,5	0,01	0,58	0,01	0,49	0,01
Minkowski (4)	0,57	0,01	0,5	0,01	0,57	0,01	0,49	0,01
Minkowski (5)	0,57	0,01	0,5	0,01	0,57	0,01	0,49	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	1,702	0,064	2,001	0,077	1,71	0,059	2,005	0,071
Canberra	1,705	0,067	2,015	0,076	1,729	0,124	1,999	0,084
Chebyshev	1,287	0,059	1,429	0,056	1,252	0,043	1,42	0,055
Distancia Correlacional	3,253	0,098	3,64	0,102	3,283	0,084	3,661	0,097
Distancia de Cosseno	1,758	0,063	2,072	0,087	1,757	0,066	2,069	0,078
Distancia de Cosseno Similaridade	0,754	0,046	0,679	0,048	0,762	0,046	0,671	0,051
Distancia de Jaccard	0,747	0,047	0,668	0,046	0,747	0,047	0,671	0,043
Euclidiana	1,322	0,097	1,464	0,055	1,274	0,041	1,476	0,082
Hamming	0,311	0,021	0,283	0,02	0,322	0,04	0,286	0,018
Manhattan	1,278	0,07	1,443	0,035	1,246	0,064	1,414	0,044
Minkowski (3)	1,322	0,116	1,509	0,061	1,33	0,081	1,493	0,05
Minkowski (4)	1,313	0,079	1,49	0,046	1,313	0,074	1,494	0,052
Minkowski (5)	1,292	0,069	1,508	0,059	1,317	0,107	1,484	0,062

C.7.1.3. Contraceptive Method Choice - 15%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,39	0,04	0,39	0,05	0,4	0,03	0,4	0,04
Canberra	0,39	0,05	0,36	0,07	0,39	0,04	0,39	0,04
Chebyshev	0,53	0,02	0,47	0,01	0,53	0,02	0,47	0,01
Distancia Correlacional	0,58	0	0,49	0,01	0,59	0,01	0,5	0,01
Distancia de Cosseno	0,57	0	0,5	0,01	0,56	0	0,5	0,01
Distancia de Cosseno Similaridade	0,43	0,02	0,45	0,02	0,42	0,02	0,44	0,02
Distancia de Jaccard	0,43	0,02	0,44	0,02	0,43	0,02	0,45	0,02
Euclidiana	0,55	0,01	0,5	0,01	0,54	0,01	0,49	0,01
Hamming	0,51	0,02	0,48	0,02	0,5	0,02	0,47	0,02
Manhattan	0,55	0,01	0,51	0,01	0,54	0,01	0,5	0,01
Minkowski (3)	0,55	0,01	0,5	0,01	0,55	0,01	0,49	0,01
Minkowski (4)	0,55	0,01	0,5	0,01	0,55	0,01	0,5	0,01
Minkowski (5)	0,55	0,01	0,49	0,01	0,55	0,01	0,49	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	1,705	0,063	2	0,061	1,708	0,065	2,009	0,071
Canberra	1,686	0,04	2,013	0,083	1,699	0,058	2,029	0,08
Chebyshev	1,292	0,041	1,426	0,067	1,26	0,042	1,416	0,055
Distancia Correlacional	3,27	0,099	3,64	0,088	3,286	0,095	3,674	0,185
Distancia de Cosseno	1,763	0,071	2,074	0,093	1,782	0,076	2,047	0,082
Distancia de Cosseno Similaridade	0,756	0,037	0,689	0,044	0,757	0,045	0,675	0,058
Distancia de Jaccard	0,76	0,033	0,668	0,066	0,742	0,047	0,679	0,05
Euclidiana	1,293	0,054	1,478	0,052	1,293	0,058	1,494	0,051
Hamming	0,317	0,032	0,289	0,025	0,308	0,017	0,295	0,027
Manhattan	1,288	0,097	1,436	0,046	1,238	0,047	1,405	0,03
Minkowski (3)	1,333	0,089	1,515	0,068	1,306	0,076	1,481	0,043
Minkowski (4)	1,307	0,049	1,507	0,056	1,314	0,057	1,502	0,048
Minkowski (5)	1,304	0,099	1,492	0,067	1,314	0,079	1,513	0,088

C.7.1.4. Contraceptive Method Choice - 20%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,37	0,05	0,4	0,04	0,38	0,04	0,4	0,04
Canberra	0,39	0,02	0,38	0,05	0,39	0,03	0,4	0,05
Chebyshev	0,52	0,02	0,46	0,01	0,52	0,02	0,46	0,01
Distancia Correlacional	0,54	0	0,47	0,01	0,54	0	0,47	0,01
Distancia de Cosseno	0,54	0	0,47	0,01	0,54	0,01	0,47	0,01
Distancia de Cosseno Similaridade	0,41	0,02	0,43	0,02	0,41	0,02	0,43	0,02
Distancia de Jaccard	0,42	0,02	0,43	0,02	0,41	0,02	0,43	0,02
Euclidiana	0,54	0,01	0,47	0,01	0,52	0,01	0,47	0,01
Hamming	0,49	0,01	0,47	0,01	0,48	0,02	0,46	0,02
Manhattan	0,54	0,01	0,48	0,01	0,53	0,01	0,49	0,01
Minkowski (3)	0,53	0,01	0,47	0,01	0,52	0,01	0,47	0,01
Minkowski (4)	0,53	0,01	0,46	0,01	0,52	0,01	0,46	0,01
Minkowski (5)	0,53	0,01	0,46	0,01	0,52	0,01	0,46	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	1,719	0,06	2,012	0,064	1,738	0,053	2,017	0,084
Canberra	1,709	0,06	2,019	0,089	1,709	0,053	2,013	0,072
Chebyshev	1,291	0,063	1,419	0,037	1,26	0,036	1,421	0,055
Distancia Correlacional	3,291	0,126	3,644	0,114	3,326	0,211	3,677	0,103
Distancia de Cosseno	1,763	0,08	2,11	0,091	1,794	0,098	2,073	0,096
Distancia de Cosseno Similaridade	0,762	0,051	0,668	0,048	0,769	0,055	0,663	0,05
Distancia de Jaccard	0,761	0,033	0,688	0,075	0,764	0,061	0,661	0,035
Euclidiana	1,3	0,065	1,502	0,059	1,28	0,045	1,498	0,063
Hamming	0,313	0,023	0,291	0,024	0,317	0,031	0,292	0,025
Manhattan	1,276	0,032	1,448	0,053	1,236	0,024	1,434	0,055
Minkowski (3)	1,309	0,048	1,501	0,07	1,302	0,046	1,498	0,058
Minkowski (4)	1,303	0,053	1,516	0,061	1,31	0,051	1,521	0,059
Minkowski (5)	1,315	0,046	1,5	0,049	1,313	0,037	1,5	0,048

C.7.2.1. Heart - 05%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,59	0,07	0,57	0,09	0,55	0,07	0,55	0,09
Canberra	0,57	0,08	0,54	0,08	0,56	0,09	0,55	0,08
Chebyshev	0,69	0,01	0,76	0,02	0,7	0,01	0,77	0,02
Distancia Correlacional	0,69	0	0,8	0	0,69	0	0,8	0
Distancia de Cosseno	0,72	0	0,78	0	0,74	0	0,78	0
Distancia de Cosseno Similaridade	0,72	0,02	0,56	0,06	0,71	0,03	0,54	0,06
Distancia de Jaccard	0,72	0,03	0,56	0,05	0,72	0,04	0,57	0,06
Euclidiana	0,69	0	0,8	0	0,69	0,01	0,81	0
Hamming	0,84	0,02	0,87	0,02	0,84	0,01	0,86	0,02
Manhattan	0,75	0,02	0,83	0	0,75	0,01	0,83	0
Minkowski (3)	0,72	0	0,8	0	0,72	0,01	0,8	0
Minkowski (4)	0,72	0,01	0,8	0	0,72	0	0,8	0
Minkowski (5)	0,71	0,01	0,8	0	0,7	0,01	0,8	0

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,09	0,01	0,108	0,006	0,085	0,004	0,111	0,015
Canberra	0,085	0,004	0,107	0,006	0,088	0,006	0,11	0,008
Chebyshev	0,069	0,009	0,079	0,007	0,065	0,007	0,076	0,007
Distancia Correlacional	0,177	0,024	0,212	0,025	0,16	0,017	0,204	0,02
Distancia de Cosseno	0,092	0,018	0,11	0,006	0,092	0,011	0,112	0,016
Distancia de Cosseno Similaridade	0,038	0,005	0,038	0,005	0,036	0,005	0,034	0,003
Distancia de Jaccard	0,036	0,004	0,037	0,008	0,036	0,005	0,038	0,007
Euclidiana	0,065	0,005	0,079	0,006	0,066	0,01	0,081	0,008
Hamming	0,018	0,002	0,018	0,004	0,018	0,003	0,018	0,003
Manhattan	0,067	0,01	0,08	0,008	0,069	0,009	0,076	0,004
Minkowski (3)	0,066	0,005	0,08	0,005	0,065	0,004	0,079	0,004
Minkowski (4)	0,066	0,004	0,084	0,008	0,065	0,005	0,081	0,006
Minkowski (5)	0,066	0,004	0,08	0,005	0,066	0,005	0,083	0,014

C.7.2.1. Heart - 10%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,6	0,07	0,54	0,09	0,58	0,08	0,51	0,09
Canberra	0,55	0,08	0,52	0,07	0,58	0,07	0,56	0,09
Chebyshev	0,71	0,02	0,76	0,02	0,7	0,02	0,76	0,03
Distancia Correlacional	0,67	0	0,8	0	0,69	0,01	0,8	0
Distancia de Cosseno	0,7	0	0,78	0	0,72	0	0,78	0
Distancia de Cosseno Similaridade	0,69	0,04	0,57	0,06	0,69	0,04	0,57	0,06
Distancia de Jaccard	0,71	0,04	0,57	0,05	0,71	0,03	0,55	0,07
Euclidiana	0,65	0,01	0,8	0,01	0,68	0,01	0,8	0
Hamming	0,83	0,02	0,85	0,02	0,83	0,01	0,85	0,02
Manhattan	0,74	0,01	0,83	0	0,74	0,01	0,84	0,01
Minkowski (3)	0,68	0,01	0,8	0	0,69	0	0,8	0
Minkowski (4)	0,68	0,01	0,8	0	0,68	0,01	0,8	0,01
Minkowski (5)	0,68	0,01	0,8	0	0,68	0,01	0,8	0

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,089	0,008	0,109	0,007	0,085	0,004	0,108	0,006
Canberra	0,087	0,006	0,112	0,019	0,087	0,006	0,11	0,009
Chebyshev	0,064	0,006	0,08	0,011	0,075	0,013	0,078	0,009
Distancia Correlacional	0,169	0,02	0,214	0,023	0,165	0,019	0,203	0,023
Distancia de Cosseno	0,09	0,007	0,109	0,006	0,091	0,008	0,109	0,006
Distancia de Cosseno Similaridade	0,035	0,003	0,036	0,007	0,039	0,005	0,035	0,004
Distancia de Jaccard	0,035	0,004	0,036	0,004	0,035	0,003	0,036	0,004
Euclidiana	0,067	0,009	0,079	0,006	0,078	0,014	0,08	0,007
Hamming	0,018	0,002	0,018	0,001	0,017	0,001	0,019	0,002
Manhattan	0,065	0,007	0,08	0,008	0,076	0,012	0,076	0,004
Minkowski (3)	0,066	0,006	0,083	0,006	0,066	0,006	0,081	0,005
Minkowski (4)	0,068	0,012	0,083	0,008	0,07	0,009	0,079	0,005
Minkowski (5)	0,067	0,009	0,084	0,011	0,068	0,007	0,085	0,013

C.7.2.1. Heart - 15%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,55	0,09	0,54	0,09	0,59	0,07	0,55	0,08
Canberra	0,55	0,1	0,55	0,09	0,57	0,09	0,51	0,09
Chebyshev	0,62	0,01	0,72	0,04	0,62	0,02	0,73	0,03
Distancia Correlacional	0,68	0,01	0,78	0	0,69	0,01	0,78	0
Distancia de Cosseno	0,71	0,01	0,76	0	0,72	0,01	0,76	0,01
Distancia de Cosseno Similaridade	0,7	0,04	0,55	0,06	0,7	0,03	0,54	0,05
Distancia de Jaccard	0,69	0,03	0,54	0,05	0,69	0,03	0,56	0,05
Euclidiana	0,61	0,01	0,78	0	0,61	0,01	0,78	0
Hamming	0,83	0,02	0,82	0,02	0,82	0,02	0,83	0,03
Manhattan	0,67	0,01	0,83	0	0,67	0,01	0,83	0,01
Minkowski (3)	0,59	0	0,8	0	0,63	0,01	0,8	0
Minkowski (4)	0,6	0,01	0,8	0	0,65	0	0,8	0
Minkowski (5)	0,61	0,01	0,8	0	0,65	0,01	0,8	0

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,087	0,007	0,108	0,007	0,088	0,007	0,109	0,015
Canberra	0,086	0,006	0,109	0,009	0,087	0,005	0,112	0,011
Chebyshev	0,071	0,012	0,079	0,014	0,063	0,004	0,079	0,007
Distancia Correlacional	0,166	0,025	0,212	0,025	0,17	0,019	0,21	0,02
Distancia de Cosseno	0,089	0,006	0,108	0,004	0,097	0,016	0,111	0,009
Distancia de Cosseno Similaridade	0,036	0,005	0,036	0,005	0,04	0,007	0,036	0,004
Distancia de Jaccard	0,036	0,003	0,037	0,009	0,035	0,004	0,038	0,006
Euclidiana	0,068	0,007	0,082	0,009	0,074	0,011	0,079	0,007
Hamming	0,018	0,002	0,018	0,002	0,018	0,001	0,018	0,002
Manhattan	0,064	0,007	0,081	0,006	0,077	0,014	0,08	0,008
Minkowski (3)	0,066	0,005	0,082	0,007	0,065	0,004	0,082	0,009
Minkowski (4)	0,065	0,003	0,08	0,004	0,066	0,004	0,081	0,008
Minkowski (5)	0,066	0,005	0,082	0,007	0,068	0,01	0,083	0,009

C.7.2.1. Heart - 20%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,54	0,08	0,53	0,09	0,53	0,09	0,56	0,09
Canberra	0,54	0,08	0,54	0,08	0,53	0,07	0,52	0,1
Chebyshev	0,6	0,01	0,71	0,03	0,58	0,02	0,71	0,04
Distancia Correlacional	0,59	0	0,7	0	0,59	0	0,72	0
Distancia de Cosseno	0,61	0	0,7	0	0,63	0,01	0,7	0
Distancia de Cosseno Similaridade	0,63	0,03	0,56	0,06	0,65	0,04	0,55	0,05
Distancia de Jaccard	0,64	0,03	0,56	0,05	0,64	0,03	0,57	0,05
Euclidiana	0,63	0,01	0,76	0	0,61	0,01	0,76	0
Hamming	0,85	0,03	0,83	0,04	0,85	0,02	0,84	0,03
Manhattan	0,68	0,01	0,78	0	0,68	0,01	0,78	0,01
Minkowski (3)	0,65	0	0,76	0,01	0,62	0,01	0,76	0
Minkowski (4)	0,61	0,01	0,78	0,01	0,61	0,01	0,77	0,01
Minkowski (5)	0,61	0,01	0,76	0	0,61	0,01	0,76	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,085	0,004	0,107	0,006	0,086	0,005	0,112	0,013
Canberra	0,09	0,007	0,107	0,008	0,087	0,006	0,109	0,006
Chebyshev	0,064	0,006	0,077	0,006	0,065	0,007	0,077	0,005
Distancia Correlacional	0,166	0,016	0,194	0,012	0,16	0,01	0,196	0,009
Distancia de Cosseno	0,089	0,006	0,111	0,009	0,088	0,006	0,111	0,01
Distancia de Cosseno Similaridade	0,035	0,003	0,036	0,006	0,038	0,005	0,038	0,008
Distancia de Jaccard	0,037	0,004	0,037	0,007	0,036	0,003	0,035	0,004
Euclidiana	0,07	0,01	0,079	0,006	0,073	0,011	0,08	0,009
Hamming	0,018	0,001	0,018	0,002	0,018	0,002	0,019	0,003
Manhattan	0,072	0,011	0,083	0,011	0,075	0,017	0,078	0,006
Minkowski (3)	0,066	0,006	0,084	0,011	0,067	0,006	0,081	0,006
Minkowski (4)	0,066	0,006	0,082	0,007	0,073	0,008	0,083	0,007
Minkowski (5)	0,066	0,009	0,081	0,006	0,066	0,005	0,083	0,007

C.7.3.1. Iris - 05%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	1	0	1	0	1	0	1	0
Canberra	1	0	1	0	1	0	1	0
Chebyshev	1	0	1	0	1	0	1	0
Distancia Correlacional	1	0	0,75	0,01	1	0	0,75	0,01
Distancia de Cosseno	1	0	0,72	0,01	1	0,01	0,71	0
Distancia de Cosseno Similaridade	0,79	0,04	0,86	0,08	0,77	0,04	0,87	0,06
Distancia de Jaccard	0,78	0,05	0,86	0,05	0,78	0,03	0,86	0,06
Euclidiana	1	0	1	0	1	0	1	0
Hamming	0,9	0,05	0,94	0,02	0,9	0,05	0,93	0,03
Manhattan	1	0	1	0	1	0	1	0
Minkowski (3)	1	0	0,97	0,01	1	0	0,96	0
Minkowski (4)	1	0	0,96	0	1	0	0,97	0,01
Minkowski (5)	1	0	0,97	0,01	1	0	1	0

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	Sem parâmetros		Com parâmetros		Sem parâmetros		Com parâmetros	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,012	0,001	0,016	0,002	0,013	0,001	0,016	0,002
Canberra	0,013	0,001	0,017	0,003	0,013	0,001	0,016	0,001
Chebyshev	0,011	0,001	0,012	0,001	0,011	0,001	0,013	0,001
Distancia Correlacional	0,02	0,001	0,028	0,004	0,021	0,002	0,025	0,002
Distancia de Cosseno	0,014	0,002	0,016	0,002	0,014	0,001	0,016	0,001
Distancia de Cosseno Similaridade	0,006	0,001	0,006	0,001	0,006	0,001	0,006	0,002
Distancia de Jaccard	0,006	0,001	0,006	0,001	0,007	0,001	0,006	0,001
Euclidiana	0,011	0,001	0,013	0,001	0,011	0,002	0,013	0,002
Hamming	0,005	0,001	0,004	0,001	0,005	0,001	0,005	0,001
Manhattan	0,011	0,001	0,012	0,002	0,011	0,002	0,013	0,001
Minkowski (3)	0,011	0,002	0,013	0,001	0,011	0,001	0,013	0,001
Minkowski (4)	0,011	0,001	0,013	0,001	0,011	0,002	0,013	0,002
Minkowski (5)	0,011	0,001	0,013	0,002	0,011	0,001	0,013	0,001

C.7.3.1. Iris - 10%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	1	0	1	0	1	0	1	0
Canberra	1	0	1	0	1	0	1	0
Chebyshev	1	0	1	0,01	1	0	1	0,01
Distancia Correlacional	1	0	0,75	0,01	1	0	0,75	0
Distancia de Cosseno	1	0,01	0,71	0	1	0,01	0,71	0
Distancia de Cosseno Similaridade	0,85	0,06	0,88	0,06	0,84	0,06	0,87	0,05
Distancia de Jaccard	0,85	0,04	0,87	0,05	0,84	0,06	0,86	0,05
Euclidiana	1	0	1	0	1	0	1	0,01
Hamming	0,92	0,03	0,93	0,03	0,91	0,04	0,93	0,04
Manhattan	1	0	1	0	1	0	1	0
Minkowski (3)	1	0	0,96	0	1	0	0,96	0
Minkowski (4)	1	0	0,96	0	1	0	0,97	0,01
Minkowski (5)	1	0	0,97	0,01	1	0	1	0

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,013	0,002	0,017	0,003	0,014	0,002	0,015	0,001
Canberra	0,014	0,002	0,016	0,003	0,013	0,001	0,016	0,001
Chebyshev	0,011	0,001	0,012	0	0,011	0,001	0,013	0,002
Distancia Correlacional	0,021	0,002	0,026	0,003	0,021	0,002	0,027	0,004
Distancia de Cosseno	0,014	0,001	0,017	0,002	0,013	0,001	0,017	0,002
Distancia de Cosseno Similaridade	0,007	0,001	0,006	0,001	0,006	0	0,006	0,001
Distancia de Jaccard	0,007	0,001	0,006	0,001	0,007	0,001	0,006	0,001
Euclidiana	0,011	0,001	0,012	0,001	0,011	0,001	0,013	0,002
Hamming	0,005	0,001	0,004	0,001	0,005	0,001	0,005	0,001
Manhattan	0,01	0,001	0,012	0	0,011	0,001	0,013	0,002
Minkowski (3)	0,011	0,001	0,013	0,002	0,011	0,001	0,013	0,001
Minkowski (4)	0,011	0,001	0,013	0,001	0,011	0,001	0,012	0,001
Minkowski (5)	0,011	0,001	0,013	0,001	0,011	0,001	0,013	0,002

C.7.3.1. Iris - 15%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	1	0	1	0	1	0	1	0,01
Canberra	1	0	1	0	1	0	1	0
Chebyshev	1	0	1	0	1	0,01	1	0
Distancia Correlacional	1	0	0,72	0,01	1	0	0,72	0,01
Distancia de Cosseno	1	0	0,65	0,01	1	0	0,68	0,01
Distancia de Cosseno Similaridade	0,85	0,06	0,85	0,05	0,84	0,04	0,86	0,06
Distancia de Jaccard	0,87	0,05	0,85	0,05	0,88	0,07	0,85	0,07
Euclidiana	0,97	0	0,96	0	1	0	0,96	0
Hamming	0,92	0,04	0,93	0,02	0,91	0,04	0,93	0,04
Manhattan	0,97	0	0,96	0	0,99	0,02	1	0,01
Minkowski (3)	0,97	0,01	0,96	0	1	0,01	0,96	0
Minkowski (4)	0,97	0,01	0,97	0,01	1	0,01	0,96	0
Minkowski (5)	0,97	0	0,97	0,01	1	0	1	0,01

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,013	0,001	0,016	0,002	0,014	0,002	0,016	0,002
Canberra	0,013	0,001	0,016	0,001	0,013	0,002	0,016	0,002
Chebyshev	0,011	0,001	0,012	0,001	0,011	0,002	0,012	0,001
Distancia Correlacional	0,022	0,004	0,03	0,005	0,021	0,002	0,025	0,002
Distancia de Cosseno	0,014	0,001	0,016	0,002	0,014	0,001	0,017	0,002
Distancia de Cosseno Similaridade	0,006	0	0,006	0,001	0,007	0,001	0,006	0,001
Distancia de Jaccard	0,006	0,001	0,006	0,001	0,007	0,001	0,006	0
Euclidiana	0,011	0,001	0,013	0,001	0,012	0,002	0,013	0,002
Hamming	0,005	0,001	0,004	0,001	0,005	0	0,005	0,001
Manhattan	0,012	0,001	0,012	0,001	0,011	0,001	0,013	0,001
Minkowski (3)	0,011	0,001	0,012	0,001	0,011	0,001	0,013	0,001
Minkowski (4)	0,011	0,001	0,012	0,001	0,011	0,001	0,013	0,002
Minkowski (5)	0,012	0,002	0,013	0,001	0,011	0,002	0,012	0,001

C.7.3.1. Iris - 20%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,97	0,01	0,96	0	0,97	0,01	0,96	0
Canberra	0,97	0,01	0,96	0	0,97	0	0,96	0
Chebyshev	1	0	1	0	1	0	0,96	0
Distancia Correlacional	1	0	0,68	0,01	1	0	0,68	0,01
Distancia de Cosseno	1	0	0,71	0	1	0	0,71	0
Distancia de Cosseno Similaridade	0,85	0,06	0,86	0,06	0,85	0,06	0,86	0,06
Distancia de Jaccard	0,87	0,05	0,88	0,04	0,83	0,05	0,86	0,05
Euclidiana	1	0	1	0,01	1	0	1	0,01
Hamming	0,92	0,04	0,94	0,02	0,93	0,03	0,94	0,03
Manhattan	1	0	1	0,01	1	0	1	0
Minkowski (3)	1	0	0,96	0	1	0	0,96	0
Minkowski (4)	1	0	0,96	0	1	0	0,96	0
Minkowski (5)	1	0	0,97	0,01	1	0	0,96	0

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,013	0,001	0,017	0,002	0,013	0,001	0,016	0,002
Canberra	0,013	0,001	0,016	0,002	0,013	0,002	0,016	0,003
Chebyshev	0,01	0,001	0,012	0,001	0,011	0,002	0,013	0,001
Distancia Correlacional	0,021	0,003	0,026	0,004	0,022	0,003	0,026	0,003
Distancia de Cosseno	0,014	0,001	0,017	0,002	0,014	0,002	0,017	0,002
Distancia de Cosseno Similaridade	0,007	0,001	0,006	0,001	0,006	0,001	0,006	0,001
Distancia de Jaccard	0,007	0,001	0,006	0,001	0,007	0,001	0,006	0,001
Euclidiana	0,011	0,001	0,012	0,001	0,011	0,001	0,013	0,001
Hamming	0,005	0,001	0,005	0,001	0,005	0	0,005	0,001
Manhattan	0,011	0,002	0,012	0	0,01	0,001	0,013	0,002
Minkowski (3)	0,011	0,001	0,014	0,002	0,011	0,001	0,013	0,001
Minkowski (4)	0,011	0,001	0,012	0,001	0,011	0,001	0,012	0,001
Minkowski (5)	0,011	0,001	0,013	0,002	0,011	0,002	0,014	0,002

C.7.4.1. Pima - 05%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,68	0,05	0,63	0,04	0,67	0,04	0,63	0,04
Canberra	0,67	0,05	0,63	0,04	0,65	0,06	0,65	0,01
Chebyshev	0,74	0,01	0,72	0,01	0,73	0,01	0,73	0,01
Distancia Correlacional	0,72	0	0,67	0	0,72	0	0,65	0
Distancia de Cosseno	0,72	0	0,69	0	0,71	0	0,69	0
Distancia de Cosseno Similaridade	0,68	0,02	0,59	0,1	0,67	0,02	0,63	0,08
Distancia de Jaccard	0,68	0,02	0,63	0,07	0,69	0,02	0,64	0,06
Euclidiana	0,75	0	0,71	0	0,75	0	0,72	0
Hamming	0,65	0,03	0,6	0,07	0,65	0,02	0,61	0,07
Manhattan	0,75	0	0,72	0	0,75	0	0,72	0
Minkowski (3)	0,75	0	0,72	0	0,75	0	0,73	0
Minkowski (4)	0,75	0	0,71	0	0,75	0	0,72	0
Minkowski (5)	0,75	0	0,73	0	0,75	0	0,73	0

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,524	0,036	0,73	0,055	0,537	0,04	0,706	0,039
Canberra	0,504	0,04	0,729	0,048	0,497	0,031	0,729	0,05
Chebyshev	0,372	0,027	0,522	0,046	0,373	0,034	0,528	0,058
Distancia Correlacional	0,901	0,033	1,301	0,068	0,948	0,032	1,278	0,068
Distancia de Cosseno	0,549	0,047	0,747	0,036	0,523	0,038	0,735	0,044
Distancia de Cosseno Similaridade	0,165	0,019	0,175	0,017	0,158	0,02	0,165	0,021
Distancia de Jaccard	0,166	0,02	0,164	0,02	0,169	0,025	0,171	0,018
Euclidiana	0,374	0,027	0,513	0,034	0,392	0,024	0,535	0,046
Hamming	0,084	0,006	0,082	0,013	0,084	0,006	0,08	0,006
Manhattan	0,375	0,033	0,516	0,035	0,383	0,033	0,517	0,033
Minkowski (3)	0,407	0,034	0,536	0,037	0,392	0,026	0,532	0,047
Minkowski (4)	0,399	0,037	0,529	0,035	0,389	0,029	0,543	0,049
Minkowski (5)	0,398	0,033	0,53	0,032	0,399	0,04	0,532	0,037

C.7.4.1. Pima - 10%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,61	0,06	0,6	0,06	0,63	0,05	0,62	0,05
Canberra	0,65	0,04	0,62	0,06	0,65	0,04	0,63	0,04
Chebyshev	0,74	0,01	0,71	0,01	0,74	0,01	0,71	0,01
Distancia Correlacional	0,64	0	0,62	0	0,67	0	0,62	0
Distancia de Cosseno	0,66	0,01	0,66	0	0,66	0	0,66	0
Distancia de Cosseno Similaridade	0,68	0,03	0,58	0,12	0,67	0,03	0,61	0,09
Distancia de Jaccard	0,67	0,02	0,57	0,12	0,66	0,02	0,6	0,1
Euclidiana	0,75	0	0,7	0	0,75	0	0,69	0
Hamming	0,64	0,02	0,6	0,05	0,64	0,02	0,6	0,08
Manhattan	0,73	0,01	0,73	0	0,73	0,01	0,72	0
Minkowski (3)	0,74	0	0,72	0	0,74	0	0,72	0
Minkowski (4)	0,76	0	0,74	0	0,76	0	0,74	0
Minkowski (5)	0,75	0	0,71	0	0,75	0	0,71	0

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,533	0,04	0,733	0,039	0,53	0,046	0,726	0,04
Canberra	0,505	0,044	0,728	0,047	0,501	0,029	0,726	0,041
Chebyshev	0,383	0,032	0,526	0,045	0,38	0,04	0,525	0,034
Distancia Correlacional	0,931	0,05	1,28	0,054	0,95	0,035	1,263	0,054
Distancia de Cosseno	0,555	0,044	0,76	0,036	0,561	0,049	0,741	0,045
Distancia de Cosseno Similaridade	0,176	0,021	0,167	0,024	0,161	0,019	0,161	0,023
Distancia de Jaccard	0,164	0,018	0,165	0,018	0,163	0,021	0,168	0,017
Euclidiana	0,382	0,036	0,517	0,036	0,385	0,021	0,533	0,026
Hamming	0,083	0,006	0,079	0,005	0,086	0,007	0,08	0,005
Manhattan	0,374	0,03	0,527	0,036	0,395	0,043	0,517	0,036
Minkowski (3)	0,396	0,026	0,524	0,029	0,403	0,042	0,524	0,024
Minkowski (4)	0,401	0,044	0,525	0,04	0,392	0,031	0,557	0,043
Minkowski (5)	0,395	0,03	0,54	0,038	0,395	0,036	0,527	0,031

C.7.4.1. Pima - 15%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,68	0,04	0,61	0,05	0,66	0,05	0,59	0,06
Canberra	0,66	0,06	0,62	0,05	0,65	0,06	0,64	0,05
Chebyshev	0,72	0,01	0,72	0	0,72	0,01	0,72	0
Distancia Correlacional	0,71	0	0,68	0	0,71	0	0,67	0
Distancia de Cosseno	0,68	0	0,7	0	0,68	0,01	0,7	0
Distancia de Cosseno Similaridade	0,66	0,03	0,62	0,08	0,66	0,03	0,59	0,12
Distancia de Jaccard	0,67	0,02	0,6	0,09	0,66	0,02	0,59	0,12
Euclidiana	0,74	0	0,72	0	0,74	0	0,73	0
Hamming	0,63	0,02	0,57	0,09	0,63	0,02	0,59	0,07
Manhattan	0,74	0	0,73	0	0,74	0	0,75	0
Minkowski (3)	0,75	0	0,74	0	0,75	0	0,74	0
Minkowski (4)	0,73	0	0,73	0	0,73	0	0,75	0
Minkowski (5)	0,73	0	0,74	0	0,73	0	0,75	0

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	Sem parâmetros		Com parâmetros		Sem parâmetros		Com parâmetros	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,523	0,033	0,723	0,032	0,543	0,051	0,74	0,044
Canberra	0,512	0,04	0,715	0,034	0,508	0,032	0,723	0,038
Chebyshev	0,372	0,025	0,533	0,042	0,387	0,041	0,52	0,034
Distancia Correlacional	0,917	0,038	1,277	0,044	0,948	0,027	1,276	0,06
Distancia de Cosseno	0,548	0,043	0,762	0,049	0,557	0,054	0,741	0,039
Distancia de Cosseno Similaridade	0,172	0,017	0,173	0,029	0,162	0,022	0,161	0,022
Distancia de Jaccard	0,171	0,02	0,172	0,018	0,173	0,018	0,168	0,018
Euclidiana	0,383	0,027	0,528	0,049	0,386	0,016	0,539	0,027
Hamming	0,083	0,005	0,079	0,003	0,084	0,006	0,08	0,006
Manhattan	0,377	0,029	0,528	0,038	0,376	0,025	0,512	0,034
Minkowski (3)	0,389	0,032	0,526	0,04	0,388	0,026	0,529	0,042
Minkowski (4)	0,399	0,036	0,54	0,037	0,391	0,034	0,561	0,037
Minkowski (5)	0,402	0,033	0,536	0,029	0,399	0,034	0,549	0,043

C.7.4.1. Pima - 20%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,66	0,05	0,61	0,06	0,64	0,06	0,62	0,06
Canberra	0,65	0,05	0,6	0,06	0,63	0,06	0,61	0,06
Chebyshev	0,7	0,01	0,72	0,01	0,71	0,01	0,72	0,01
Distancia Correlacional	0,62	0	0,67	0	0,62	0	0,69	0
Distancia de Cosseno	0,64	0	0,68	0	0,62	0	0,69	0
Distancia de Cosseno Similaridade	0,66	0,03	0,65	0,04	0,67	0,02	0,61	0,09
Distancia de Jaccard	0,66	0,02	0,62	0,07	0,66	0,03	0,58	0,1
Euclidiana	0,73	0	0,71	0	0,73	0	0,72	0
Hamming	0,6	0,02	0,58	0,07	0,61	0,03	0,58	0,07
Manhattan	0,71	0,01	0,74	0	0,71	0	0,74	0
Minkowski (3)	0,72	0	0,71	0	0,72	0	0,73	0
Minkowski (4)	0,73	0	0,74	0	0,73	0	0,75	0
Minkowski (5)	0,73	0	0,72	0	0,74	0	0,73	0

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,533	0,038	0,73	0,043	0,528	0,047	0,728	0,045
Canberra	0,514	0,037	0,726	0,043	0,519	0,036	0,743	0,044
Chebyshev	0,372	0,03	0,518	0,034	0,393	0,043	0,522	0,037
Distancia Correlacional	0,926	0,042	1,278	0,052	0,949	0,029	1,276	0,057
Distancia de Cosseno	0,543	0,038	0,763	0,041	0,538	0,048	0,743	0,042
Distancia de Cosseno Similaridade	0,171	0,022	0,165	0,026	0,166	0,019	0,16	0,021
Distancia de Jaccard	0,172	0,022	0,175	0,022	0,169	0,022	0,17	0,02
Euclidiana	0,39	0,03	0,536	0,034	0,38	0,017	0,542	0,037
Hamming	0,083	0,004	0,081	0,007	0,083	0,005	0,082	0,008
Manhattan	0,38	0,031	0,534	0,034	0,389	0,035	0,531	0,039
Minkowski (3)	0,405	0,04	0,529	0,041	0,385	0,025	0,545	0,034
Minkowski (4)	0,405	0,044	0,56	0,036	0,396	0,043	0,558	0,038
Minkowski (5)	0,4	0,039	0,551	0,038	0,391	0,036	0,548	0,034

C.7.5.1. Sonar, Mines vs. Rocks - 05%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,86	0	0,77	0,03	0,86	0	0,77	0,03
Canberra	0,86	0	0,77	0,02	0,86	0	0,77	0,02
Chebyshev	0,76	0	0,69	0	0,76	0	0,71	0
Distancia Correlacional	0,86	0	0,9	0	0,86	0	0,9	0
Distancia de Cosseno	0,86	0	0,86	0	0,86	0	0,86	0
Distancia de Cosseno Similaridade	0,62	0,03	0,61	0,05	0,61	0,03	0,61	0,04
Distancia de Jaccard	0,61	0,02	0,62	0,05	0,6	0,03	0,62	0,04
Euclidiana	0,86	0	0,88	0	0,86	0	0,88	0
Hamming	0,62	0,04	0,58	0,05	0,62	0,03	0,6	0,05
Manhattan	0,93	0,01	0,9	0	0,9	0	0,9	0
Minkowski (3)	0,81	0	0,83	0	0,81	0	0,83	0
Minkowski (4)	0,83	0,01	0,81	0	0,81	0	0,81	0
Minkowski (5)	0,83	0	0,74	0	0,81	0	0,74	0

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,278	0,026	0,389	0,03	0,278	0,029	0,377	0,02
Canberra	0,26	0,013	0,381	0,02	0,269	0,019	0,387	0,031
Chebyshev	0,18	0,01	0,28	0,03	0,188	0,022	0,267	0,019
Distancia Correlacional	0,513	0,035	0,735	0,041	0,535	0,046	0,721	0,059
Distancia de Cosseno	0,27	0,011	0,393	0,019	0,293	0,024	0,386	0,021
Distancia de Cosseno Similaridade	0,127	0,006	0,117	0,008	0,133	0,018	0,112	0,006
Distancia de Jaccard	0,129	0,007	0,113	0,007	0,129	0,008	0,113	0,007
Euclidiana	0,197	0,023	0,273	0,018	0,193	0,015	0,272	0,026
Hamming	0,027	0,002	0,027	0,004	0,03	0,006	0,026	0,002
Manhattan	0,18	0,01	0,291	0,03	0,175	0,006	0,284	0,026
Minkowski (3)	0,186	0,007	0,271	0,013	0,188	0,011	0,268	0,011
Minkowski (4)	0,183	0,008	0,264	0,008	0,189	0,016	0,27	0,011
Minkowski (5)	0,191	0,02	0,27	0,014	0,186	0,01	0,266	0,009

C.7.5.1. Sonar, Mines vs. Rocks - 10%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	<u>\bar{m}</u>	<u>$\sigma \pm$</u>	<u>\bar{m}</u>	<u>$\sigma \pm$</u>	<u>\bar{m}</u>	<u>$\sigma \pm$</u>	<u>\bar{m}</u>	<u>$\sigma \pm$</u>
Bray Curtis	0,86	0	0,77	0,02	0,86	0	0,77	0,02
Canberra	0,86	0	0,78	0,02	0,86	0	0,77	0,02
Chebyshev	0,69	0	0,71	0	0,69	0,01	0,74	0
Distancia Correlacional	0,86	0	0,86	0	0,86	0	0,86	0
Distancia de Cosseno	0,83	0	0,83	0	0,83	0	0,86	0
Distancia de Cosseno Similaridade	0,57	0,02	0,56	0,05	0,58	0,03	0,57	0,05
Distancia de Jaccard	0,57	0,02	0,58	0,05	0,57	0,03	0,58	0,05
Euclidiana	0,81	0	0,86	0	0,81	0	0,86	0
Hamming	0,63	0,03	0,55	0,06	0,63	0,04	0,55	0,05
Manhattan	0,88	0	0,83	0	0,88	0	0,83	0
Minkowski (3)	0,76	0	0,81	0	0,76	0	0,81	0
Minkowski (4)	0,76	0	0,76	0	0,76	0	0,76	0
Minkowski (5)	0,74	0	0,76	0	0,74	0	0,76	0

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,281	0,027	0,391	0,027	0,283	0,033	0,376	0,019
Canberra	0,269	0,017	0,377	0,019	0,269	0,025	0,402	0,034
Chebyshev	0,184	0,012	0,272	0,021	0,184	0,016	0,265	0,018
Distancia Correlacional	0,512	0,043	0,733	0,035	0,522	0,042	0,733	0,056
Distancia de Cosseno	0,274	0,018	0,387	0,013	0,293	0,029	0,395	0,029
Distancia de Cosseno Similaridade	0,134	0,009	0,115	0,007	0,137	0,007	0,114	0,009
Distancia de Jaccard	0,135	0,008	0,113	0,006	0,138	0,008	0,117	0,006
Euclidiana	0,193	0,02	0,273	0,012	0,19	0,013	0,273	0,023
Hamming	0,027	0,003	0,026	0,004	0,029	0,004	0,026	0,002
Manhattan	0,182	0,017	0,276	0,027	0,178	0,013	0,266	0,02
Minkowski (3)	0,186	0,008	0,276	0,019	0,192	0,014	0,274	0,017
Minkowski (4)	0,189	0,014	0,266	0,009	0,195	0,019	0,269	0,013
Minkowski (5)	0,184	0,007	0,268	0,01	0,191	0,008	0,268	0,008

C.7.5.1. Sonar, Mines vs. Rocks - 15%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,83	0	0,77	0,03	0,86	0	0,76	0,02
Canberra	0,83	0	0,76	0,02	0,86	0	0,76	0,02
Chebyshev	0,69	0	0,71	0	0,69	0	0,71	0
Distancia Correlacional	0,86	0	0,86	0	0,86	0	0,86	0
Distancia de Cosseno	0,83	0	0,83	0,01	0,83	0	0,83	0
Distancia de Cosseno Similaridade	0,59	0,03	0,6	0,05	0,59	0,02	0,59	0,05
Distancia de Jaccard	0,59	0,03	0,6	0,04	0,58	0,02	0,58	0,05
Euclidiana	0,81	0	0,83	0,01	0,81	0	0,81	0
Hamming	0,59	0,03	0,61	0,06	0,58	0,03	0,6	0,05
Manhattan	0,83	0	0,81	0	0,83	0	0,81	0
Minkowski (3)	0,76	0	0,76	0	0,74	0	0,74	0
Minkowski (4)	0,74	0	0,71	0	0,74	0	0,71	0
Minkowski (5)	0,79	0	0,71	0	0,79	0	0,71	0

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,276	0,034	0,38	0,024	0,276	0,032	0,375	0,019
Canberra	0,262	0,019	0,381	0,019	0,267	0,021	0,384	0,026
Chebyshev	0,186	0,013	0,263	0,016	0,186	0,015	0,282	0,024
Distancia Correlacional	0,507	0,038	0,726	0,038	0,518	0,043	0,755	0,032
Distancia de Cosseno	0,276	0,018	0,418	0,034	0,289	0,026	0,396	0,023
Distancia de Cosseno Similaridade	0,14	0,008	0,118	0,008	0,138	0,015	0,113	0,006
Distancia de Jaccard	0,139	0,008	0,113	0,006	0,134	0,006	0,115	0,006
Euclidiana	0,189	0,017	0,277	0,02	0,193	0,015	0,273	0,015
Hamming	0,028	0,004	0,026	0,002	0,029	0,004	0,027	0,006
Manhattan	0,191	0,017	0,274	0,024	0,183	0,022	0,283	0,022
Minkowski (3)	0,195	0,016	0,276	0,017	0,194	0,013	0,268	0,013
Minkowski (4)	0,186	0,012	0,266	0,008	0,186	0,006	0,288	0,027
Minkowski (5)	0,186	0,009	0,269	0,013	0,185	0,009	0,267	0,009

C.7.5.1. Sonar, Mines vs. Rocks - 20%

<u>Distâncias</u>	<u>Acurácia</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,83	0	0,79	0,02	0,83	0	0,78	0,02
Canberra	0,83	0	0,78	0,02	0,83	0	0,78	0,02
Chebyshev	0,76	0	0,74	0,02	0,76	0,01	0,74	0,02
Distancia Correlacional	0,79	0	0,88	0	0,79	0	0,88	0
Distancia de Cosseno	0,81	0	0,86	0	0,81	0	0,86	0
Distancia de Cosseno Similaridade	0,58	0,04	0,63	0,05	0,58	0,03	0,6	0,05
Distancia de Jaccard	0,57	0,04	0,63	0,04	0,58	0,03	0,62	0,05
Euclidiana	0,79	0	0,88	0	0,79	0	0,86	0
Hamming	0,62	0,04	0,6	0,04	0,64	0,03	0,59	0,06
Manhattan	0,86	0	0,9	0	0,86	0	0,9	0
Minkowski (3)	0,81	0	0,83	0	0,81	0	0,83	0
Minkowski (4)	0,76	0	0,79	0	0,76	0	0,79	0
Minkowski (5)	0,74	0	0,79	0	0,72	0,01	0,79	0

<u>Distâncias</u>	<u>Tempo de execução</u>							
	<u>Método de desempate Média</u>				<u>Método de desempate Desvio Padrão</u>			
	<u>Sem parâmetros</u>		<u>Com parâmetros</u>		<u>Sem parâmetros</u>		<u>Com parâmetros</u>	
	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$	\bar{m}	$\sigma \pm$
Bray Curtis	0,269	0,024	0,392	0,031	0,274	0,029	0,39	0,032
Canberra	0,263	0,021	0,381	0,017	0,266	0,018	0,387	0,034
Chebyshev	0,182	0,012	0,27	0,023	0,188	0,023	0,263	0,014
Distancia Correlacional	0,495	0,038	0,73	0,042	0,503	0,039	0,747	0,043
Distancia de Cosseno	0,277	0,018	0,399	0,018	0,288	0,022	0,401	0,021
Distancia de Cosseno Similaridade	0,139	0,009	0,115	0,007	0,136	0,009	0,114	0,007
Distancia de Jaccard	0,138	0,007	0,114	0,007	0,136	0,011	0,116	0,008
Euclidiana	0,191	0,013	0,27	0,011	0,198	0,024	0,282	0,027
Hamming	0,027	0,004	0,028	0,006	0,032	0,005	0,026	0,003
Manhattan	0,182	0,012	0,269	0,022	0,176	0,01	0,276	0,025
Minkowski (3)	0,186	0,012	0,273	0,014	0,186	0,008	0,277	0,017
Minkowski (4)	0,193	0,019	0,269	0,011	0,192	0,022	0,271	0,012
Minkowski (5)	0,19	0,013	0,266	0,01	0,187	0,009	0,271	0,011