ESCOLA NAVAL

talant de bien faire

Catarina Isabel Ramos de Pádua Santos

Acústica Submarina

Classificação de navios aplicando algoritmos de aprendizagem supervisionada *(data mining)*

Dissertação para obtenção do grau de Mestre em Ciências Militares Navais,

na especialidade de Marinha

Alfeite

2015

ESCOLA NAVAL

Catarina Isabel Ramos de Pádua Santos

Acústica Submarina

Classificação de navios aplicando algoritmos de aprendizagem supervisionada (data mining)

Dissertação para obtenção do grau de Mestre em Ciências Militares Navais,

na especialidade de Marinha

Orientação de:

Co-orientação de:

O Aluno Mestrando	O Co-orientador
Catarina Pádua Santos	Doutor Victor Lobo

O Orientador

1TEN TSN-AMB Quaresma dos Santos

Alfeite

2015

Epígrafe

"Se parares o teu barco e mergulhares um longo tubo na água irás escutar na sua extremidade navios a uma grande distância."

Leonardo Da Vinci, 1490

Dedicatória

Dedico à minha família, pelo apoio inestimável ao longo de todo o meu percurso académico.

Agradecimentos

Considero relevante destacar aqueles que mais contribuíram para a realização da presente dissertação. Agradeço:

- Ao 1TEN TSN-AMB Quaresma dos Santos, meu orientador, pelos ensinamentos passados e apoio prestado na realização da dissertação;
- Ao meu coorientador Professor Doutor Victor Lobo, pela motivação e colaboração na realização do trabalho;
- Ao Professor Doutor Mário Gatta, pelo interesse demonstrado pelo tema e pelas sugestões efetuadas;
- À 1TEN EN-AEL Mendes Vieira, pela disponibilidade e pelas correções realizadas;
- À divisão de Oceanografia do Instituto Hidrográfico, pela forma como me recebeu nas suas instalações, durante o desenvolvimento da dissertação;
- Ao Tiago Teles, pelo incansável apoio, pelas sugestões e pela colaboração, ao longo do período de elaboração da tese;
- À secção de fontes acústicas do Centro de Gestão e Análise de Dados Operacionais pela visita ao centro e ajuda neste trabalho;
- Ao Comandante do NRP *Almirante Gago Coutinho*, pela compreensão e flexibilidade durante o estágio de embarque;
- Ao 2TEN STH Teixeira Carvalho pelo apoio e recomendações efetuadas;
- À minha família pelo apoio constante e preocupação;

Resumo

O oceano é atualmente explorado a vários níveis. A nível político e económico, recorde-se que 90% do comércio mundial é feito por mar. A nível de investigação, muitos são os que se dedicam ao estudo das características do leito do mar e do próprio meio subaquático, bem como da fauna e da flora marinhas. Deste crescente interesse pelo mundo submarino, surge a necessidade de o compreender enquanto meio de transmissão e receção de sinais e a melhor forma de o fazer é através da acústica submarina, por ser a onda do tipo sonora a que se propaga mais facilmente na água. A presente dissertação recorre a várias técnicas de *data mining* com o objetivo final de identificar uma fonte sonora, mais particularmente um navio, partindo de registos acústicos de navios já conhecidos. É realizado o enquadramento teórico e são apresentados os conceitos considerados pertinentes na área da acústica submarina e processamento de sinal, bem como descritos os métodos de classificação utilizados para identificar um novo registo acústico de navio. Os dados utilizados para treinar e testar os classificadores foram fornecidos pelo Instituto Hidrográfico, tendo sido todos obtidos no exercício naval *Recognized Environmental Picture* em 2013 (REP 13). Na fase de testes, após o pré-processamento dos dados, os resultados apontam, de uma forma geral, para uma elevada taxa de sucesso dos classificadores na identificação de um novo dado. Deve-se no entanto considerar uma limitação ao estudo, o facto dos registos acústicos terem sido recolhidos todos no mesmo local, num pequeno intervalo temporal. Futuramente fará sentido aplicar este método a uma escala maior, para que o sinal acústico de um navio por classificar seja facilmente associado ao navio a que pertence desde que conste numa base de dados. Esta ferramenta poderá ser utilizada pela Marinha com diversos propósito, entre os quais monotorização costeira e fiscalização.

Palavras-chave:

Acústica submarina, *data mining*, pré-processamento, classificadores, monotorização.

Abstract

Today, oceans are persistently explored at different levels. On a political and economic level, around 90% of the world trade is conducted through sea. On a scientific level, both aquatic fauna and flora, as well as the characteristics of the seabed and the aquatic medium itself are all important objects of research. Being so, it is important to understand the sea as a mean of transmission and reception of signals, and the best way to do so is through submarine acoustics, as sound wave is the one that best propagates underwater. This thesis uses several data mining techniques to ultimately identify a determined sound source (more particularly a vessel), through the acoustic recordings of already known vessels. We present the theoric fundamentals in submarine acoustics and signal processing considered pertinent and describe the classification methods used to identify a new acoustic recording of a vessel. The data used to train and test the classifiers was provided by the Hydrographic Institute and recorded during the Recognized Environmental Picture naval exercise, in 2013 (REP 13). After preprocessing the data gathered in the test phase, the results indicate a high success rate in the classification of new data. However, it should be considered as a limitation, the fact that all the acoustic records have been collected in the same place, in a small period of time. In the future, it is important to implement this method on a larger scale, in order to easily associate the acoustic recording of an unidentified ship to it's recording previously registered in a vessel acoustic database. This tool can be useful to the Navy for several purposes, including coastal monitoring and surveillance.

Keywords:

Submarine acoustics, data mining, preprocessing, classifiers, monitoring

Índice

Índice de figuras

Índice de tabelas

Índice de algoritmos

Abreviaturas, siglas e acrónimos

ACINT	*Acoustics Intelligence*
CADOP	Centro de Análise e Gestão de Dados Operacionais
CIRA	*Centre d'Interprétation et de Reconnaissance Acoustique*
CRISP-DM	*Cross-Industry Standard Process for Data Mining*
CSV	*Comma-separated values*
DEMON	*Detection of Envelope Modulation on Noise*
DFT	*Discrete Fourier Transform*
FFT	*Fast Fourrier Transform*
FT	*Fourier Transform*
IPqM	Instituto de Pesquisas da Marinha do Brasil
LDG	Lancha de Desembarque Grande
LOFAR	*Low Frequency Analysis and Recording*
MATLAB	*Matrix Laboratory Software*
ONI	*Office of Naval Intelligence*
REP	*Recognized Environmental Picture*
RPM	Rotações por Minuto
SDAC	Sistema de Deteção, Acompanhamento e Classificação de Contatos
SI	Sistema Internacional de unidades
SICLA	Sistema de Clasificación Acústica
WEKA	*Waikato Environment for Knowledge Analysis*

Capítulo 1 - Introdução

A acústica submarina estuda a propagação do som na água, bem como os fenómenos associadas à interação das ondas sonoras com o meio. As ondas sonoras são ondas mecânicas, que se propagam através de pequenas oscilações das partículas do meio. Como a onda sonora se transmite por compressão e descompressão, a sua velocidade depende basicamente da elasticidade do meio. Assim, sendo a água um meio mais elástico, a velocidade, e o alcance, de propagação da onda é maior na água do que no ar (OGP, 2008, p. 4). Esta característica permite aos animais marinhos, especialmente os mamíferos, comunicar a grandes distâncias através da emissão de sons que se propagam pelo oceano (Urick, 1984, pp. 7-1,7-7).

A ciência da acústica submarina é relativamente recente, podendo atribuir-se a Leonardo Da Vinci[1] a autoria de um dos primeiros protótipos de escuta focado na deteção de navios através do ruído submarino (Lurton, 2002, p. 5). Desde então, muitos foram os que deram o seu contributo nesta área. O físico suíço Daniel Colladon e o matemático francês Charles Sturm, realizaram a primeira medição quantitativa da velocidade do som na água, no Lago Léman. O desenvolvimento de sonares ativos e passivos foi acelerado com a produção em larga escala de submarinos durante a Primeira Guerra Mundial. Os transdutores piezoelétricos foram utilizados pela primeira vez durante a Segunda Grande Guerra (Silva, 2014).

Atualmente a acústica submarina é utilizada para os mais variados fins, tais como monitorização ambiental, monitorização de parâmetros oceanográficos, exploração de recursos submarinos, prevenção de desastres, comunicação e controlo de robótica submarina, bem como na área da defesa e em diferentes operações militares.

A observação e análise do ruído submarino é por excelência uma capacidade militar que há muito preocupa as marinhas de guerra, e em particular a portuguesa. A utilização de submarinos há mais de 100 anos levou ao desenvolvimento de competências que permitiram durante este período operar em segurança e explorar da melhor maneira esta arma.

[1] Inventor italiano, 1490 (Lurton, 2002).

NÃO CLASSIFICADO

Assim sendo, este trabalho pretende explorar este domínio da ciência, integrando atividades de investigação em curso no Instituto Hidrográfico e na Escola Naval. Os dados sobre os quais se desenvolve foram adquiridos no exercício naval REP13, que teve por objetivo avaliar o desempenho de novos sensores e plataformas, utilizadas em missões de reconhecimento GEOINTEL e em levantamento das condições ambientais (METOC) num teatro de operações costeiras e estuarinas.

O Instituto Hidrográfico implementou para o efeito um sistema de vigilância da barra do porto de Setúbal, que registou, durante 6 dias, o ruído submarino, as anomalias magnéticas e fotografou os navios que praticaram o canal desta barra.

A presente tese visa explorar os resultados alcançados nesta série, focando a análise dos dados acústicos e a aplicação de classificadores para a identificação de navios e de padrões de atividade marítima. Como resultado, pretende-se demonstrar como esta informação pode ser analisada e processada no sentido de organizá-la numa base de dados focada na classificação por tipo de navio, numa perspetiva militar a ser utilizada pela Marinha.

1.1. Motivação

A onda sonora emitida por uma fonte em meio submarino está naturalmente sujeita a alterações subjacentes às características do meio onde se propaga, nomeadamente devido à atenuação causada por fenómenos de espalhamento e de absorção na água; à redução da sua velocidade de propagação, quando comparada com a velocidade de propagação das ondas eletromagnéticas no ar; ao fenómeno de refração, que consiste no desvio no sentido de propagação, devido às variações da velocidade do som e à sua reflexão por diferentes interfaces; a deformações do sinal emitido, devido à heterogeneidade do meio, tais como alterações de frequência devido ao efeito de Doppler e ao ruído presente no oceano, que pode por si só camuflar o sinal acústico antropogénico, oriundo de atividade humana (Lurton, 2002, p. 11). Pode-se portanto concluir que estamos perante um processo complexo e heterogéneo no espaço e no tempo. O estudo da transmissão e da receção de informação através da água são de extrema importância para o exercício da defesa militar. Adequando ao atual contexto de atuação da Marinha, este saber torna-se crucial na capacidade de conhecimento situacional marítimo, podendo

NÃO CLASSIFICADO

reconhecer, localizar e identificar um determinado navio, recorrendo à sua escuta e análise por comparação com registos de uma base de dados acústica.

1.2. Estrutura do documento

A atual tese encontra-se dividida em seis capítulos principais:

1. Introdução, em que se contextualiza o tema e se define a motivação e o objetivo do trabalho;

2. Aplicações da acústica submarina na área da defesa, que se refere de uma forma geral ao que já existe feito, quer a nível nacional quer a nível internacional, no mesmo âmbito;

3. Enquadramento Teórico, revisitando a física por detrás dos processos acústicos e as ferramentas matemáticas habitualmente utilizadas na sua análise. É feita uma abordagem às ondas acústicas, ao processamento de sinal e às características e fenómenos associados ao meio subaquático enquanto meio de propagação;

4. Arquitetura da Solução, define uma forma para resolver o problema inicial, neste caso, identificar um navio recorrendo a técnicas de *data mining* com aprendizagem supervisionada. Contém todos os conceitos relevantes para a compreensão e aplicação dos classificadores;

5. Descrição e pré-processamento dos dados, é feita a descrição dos dados, bem como dos processos de seleção e pré-processamento dos mesmos;

6. Testes e Resultados, este capítulo é composto pelos dois grupos de testes realizados, que diferem no conjunto de teste e no conjunto de treino. Para cada teste, são aplicados três classificadores, vizinho mais próximo, *naive de Bayes* e árvore de decisão J48, cada um com três formas diferentes de testar os dados, *holdhout*, validação cruzada e com introdução de um terceiro navio para teste. Apresentam-se os resultados obtidos, para os diferentes casos.

7. Conclusões, por fim é feita a análise dos resultados obtidos face ao objetivo inicial, são mencionadas algumas sugestões para trabalhos futuros e referidas as limitações percecionadas.

NÃO CLASSIFICADO

Capítulo 2 - Aplicações da acústica submarina na área da defesa

Desde a II Guerra Mundial que a análise acústica tem vindo a ser explorada e aperfeiçoada. Esta realidade sente-se principalmente ao nível das Forças Armadas, nomeadamente nas Marinhas, onde existe um grande interesse em compreender e dominar os sinais que se propagam pelo oceano. No seguimento desta necessidade, muitas são as entidades públicas e privadas no mundo que se dedicam a este estudo. Apresentam-se de seguida alguns exemplos de entidades ligadas a esta área.

O Grupo de Sonar do Instituto de Pesquisas da Marinha do Brasil (IPqM) desenvolveu o Sistema de Deteção, Acompanhamento e Classificação de Contatos (SDAC), que utiliza os sons detetados na análise do tipo *Low Frequency Analysis and Recording*[2] (LOFAR) e as características obtidas na análise *Detection of Envelope Modulation on Noise*[3] (DEMON) para identificar os parâmetros de um contacto e classificá-lo, utilizando uma base de dados (Filho, 2011).

O *Acoustics Intelligence Laboratory*, faz parte do *Office of Naval Intelligence* (ONI), nos Estados Unidos. Os especialistas em *Acoustics Intelligence* (ACINT), fazem uso de bases de dados de acústica para vários propósitos, nomeadamente identificar vulnerabilidades acústicas (Althage, 2004). O departamento *Signatures* do *Carderock Division's* do *Naval Surface Warfare Center* também nos Estados Unidos, é líder Mundial no desenvolvimento de tecnologias direcionadas para o controlo de assinaturas acústicas e promoção da discrição dos meios navais da Marinha Americana.

O *Centre d'Interprétation et de Reconnaissance Acoustique* (CIRA), em Toulon, França, forma os operadores da guerra acústica da Marinha Francesa. Esta unidade assegura a base de dados acústicos de referência da Marinha Nacional Francesa ("CIRA," 2014).

A Companhia inglesa *Drumgrange*, desenvolveu um sistema de monitorização, utilizado amplamente pela *Royal Navy* do Reino Unido para controlar as assinaturas

[2] Análise e gravação a baixa frequência.

[3] Demodulação de ruído.

NÃO CLASSIFICADO

acústicas de submarinos e navios de guerra de superfície ("Platform Signature Monitoring System ", 2014).

A Armada Espanhola utiliza a bordo dos submarinos o *Sistema de Clasificación Acústica* (SICLA), que permite a classificação dos contactos detetados. Trata-se de um sistema que se adapta a qualquer configuração de *software* e *hardware* e pode ser integrado em qualquer Sistema de Combate ou instalado num computador portátil, conectado com qualquer dispositivo acústico. Permite a análise dos contactos em tempo real ou diferido, em LOFAR, DEMON, espectro e áudio. ("SICLA - Sistema de Clasificación Acústica - SAES," 2014).

Na Marinha Portuguesa, a secção de acústica do Centro de Análise e Gestão de Dados Operacionais (CADOP), é responsável pela análise e classificação de dados acústicos. Esta informação, contudo, dado à sua importância não pode ser partilhada nem divulgada, tal como os sistemas de *software* de análise e classificação utilizados.

NÃO CLASSIFICADO

Capítulo 3 - Enquadramento Teórico

3.1. Ondas Acústicas

3.1.1. Natureza e Propagação

As ondas acústicas, vulgarmente designadas por som, são pulsos energéticos, originados por uma perturbação mecânica que se propaga através de um meio material (sólido, líquido ou gasoso). Esta propagação traduz-se em compressões e rarefações locais, que são transmitidas de um ponto para os pontos em seu redor, (Herman Medwin, 1997, p. 17) como traduz a Figura 1.

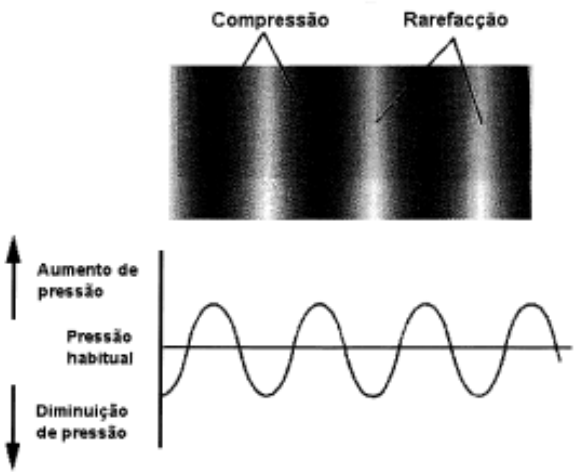

Figura 1 - Representação transversal de uma onda sonora ("PEEAS 43 (A)")

Este fenómeno deve-se às propriedades elásticas do meio. A matéria constituinte do meio elástico tende e preservar o seu comprimento, forma e volume quando sujeito a forças externas, regressando ao seu estado original quando estas forças desaparecem.

O som é assim uma onda de pressão do tipo longitudinal, uma vez que as partículas oscilam paralelamente em relação à direção de propagação da própria onda (Hodges, 2010, p. 1).

NÃO CLASSIFICADO

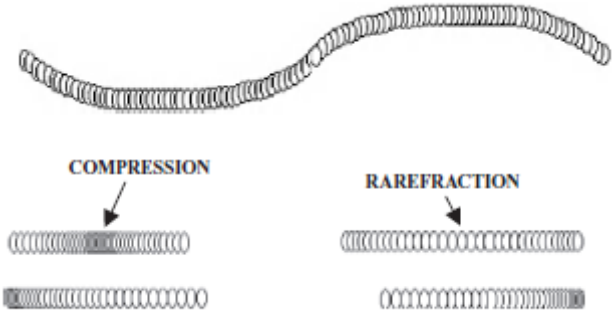

Figura 2 - Mola a representar ondas do tipo transversal (em cima) e longitudinal (em baixo) (Hodges, 2010, p. 3).

Todas as ondas sonoras podem ser decompostas em harmónicas simples. Um sinal harmónico simples é traduzido por uma onda sinusoidal periódica, de frequência e amplitude constantes. Ou seja, em teoria todos os sinais acústicos podem ser decompostos em conjuntos de harmónicas, cujas frequências são múltiplas das restantes. Portanto, é todo o sinal periódico descrito pelas funções sinusoidais:

$$Asen(\omega t)\ e\ Acos(\omega t) \tag{1}$$

Em que A é a amplitude, ω é a velocidade angular e t é o tempo.

3.1.2. Caracterização

As ondas acústicas podem-se caracterizar através de um conjunto de parâmetros, tornando-se essencial compreender a informação que estes traduzem.

- **Comprimento de Onda**

O comprimento de onda (λ) corresponde ao intervalo espacial entre dois pontos consecutivos que se encontrem na mesma fase de vibração. O comprimento de onda corresponde à distância que a onda avança durante um período do sinal (T) a uma velocidade c. Tem como unidades mais comuns m ou cm.

$$\lambda = cT = \frac{c}{f} \tag{2}$$

Em que f é a frequência do sinal.

NÃO CLASSIFICADO

- **Frequência**

A frequência (f) de um fenómeno periódico, como é o caso da onda sonora, traduz-se no número de vezes que esse fenómeno se verifica por unidade de tempo, ou seja, no número de vezes que a pressão oscila em torno da pressão média, por unidade de tempo. O número de oscilações ou ciclos por segundo de uma onda sinusoidal tem como unidade de medida o Hertz (Hz).

As principais frequências associadas ao ruído produzido por um navio encontram-se entre 1 Hz a 100 KHz (Collier, 1998, p. 409).

- **Período**

O Período (T) da onda corresponde ao intervalo de tempo entre a emissão de dois pulsos. Ou seja, traduz a duração de um ciclo completo de oscilação de uma onda. No Sistema Internacional de unidades (SI) o período é medido em segundos, s:

$$T = \frac{1}{f} \tag{3}$$

- **Amplitude**

A amplitude (A) representa o afastamento máximo da posição de equilíbrio. Quanto maior a amplitude de uma onda sonora, maior a intensidade do som. Pode expressar-se em db[4], m cm ou Pa.

- **Velocidade de Propagação**

A velocidade de propagação da onda acústica (c) traduz-se na rapidez com que a onda se propaga, isto é, a distância que percorre num determinado intervalo de tempo, as unidades SI são m/s:

$$c = \frac{\lambda}{T} \tag{4}$$

[4] O decibel (db) corresponde a um valor logarítmico que exprime uma razão entre dois níveis de grandezas (Rumsey & McCormick, 2012, p. 14).

NÃO CLASSIFICADO

A velocidade de propagação do som é uma caraterística do meio em que este se propaga. Na água do mar, a velocidade depende das condições de temperatura, pressão e salinidade. Pode-se portanto definir em função da densidade e do módulo da elasticidade (Lurton, 2002, p. 13):

$$c = \sqrt{\frac{E}{\rho}} \tag{5}$$

Sendo a água muito menos compressível do que o ar e com um módulo de elasticidade superior, a velocidade de propagação do som na água é mais elevada aqui do que no ar. A velocidade de propagação das ondas sonoras em meio aquático é muito maior do que no ar porque o aumento de densidade é compensado por um aumento de elasticidade, rondando os 1500 m/s, enquanto no ar, o valor típico é cerca de 340 m/s.

Como referido anteriormente, diferentes situações de temperatura, pressão e salinidade vão afetar a velocidade do som na água, isto porque influenciam a compressibilidade e a densidade do fluído. A densidade aumenta com a diminuição da temperatura e aumenta com o aumento da pressão e da salinidade. Por outro lado, a elasticidade aumenta com o aumento da pressão e salinidade e aumenta com a diminuição da temperatura. Cada um destes parâmetros afeta de forma diferente a velocidade, contudo a temperatura é o fator que induz maiores variações. O acréscimo de 1ºC gera um aumento de 4.6 m/s a 0ºC e 2.5 m/s a 21.1ºC, enquanto mais 100 m de profundidade aumentam em 1.7 m/s a velocidade do som, devido ao aumento da pressão hidrostática e por cada 1 ppm (partes por mil) somado, obtém-se um aumento de 1,4 m/s (Hodges, 2010, p. 5).

NÃO CLASSIFICADO

A Figura 3 representa um perfil climatológico da velocidade do som, em função das variações de temperatura, salinidade e pressão, para uma zona noroeste do Atlântico,

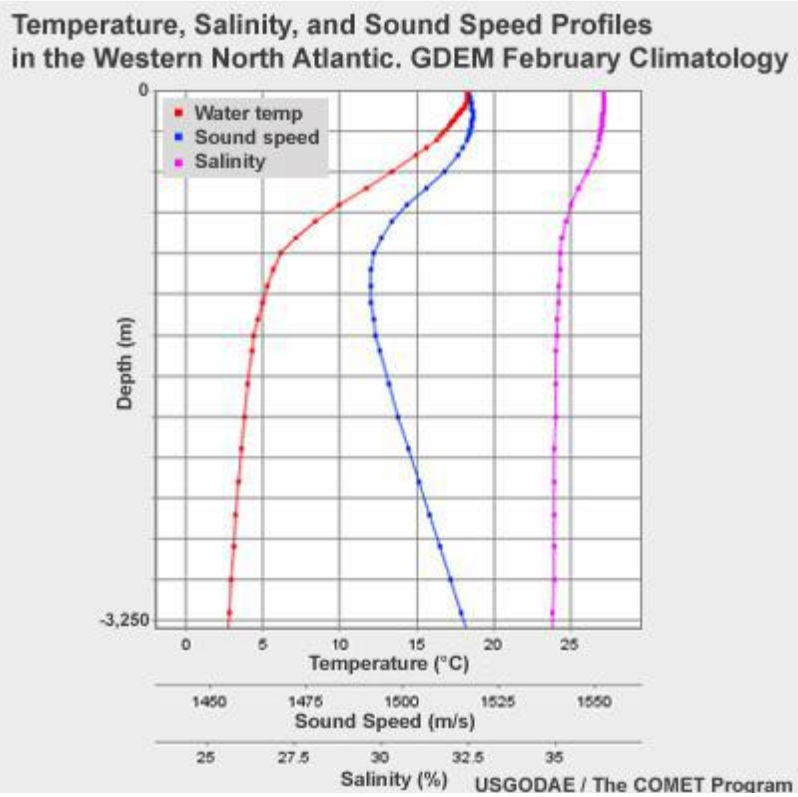

Figura 3 - Perfil da velocidade do som (MetEd, 2015b)

no mês de Fevereiro.

Pode-se concluir que *c* varia pouco com a salinidade e que o efeito da temperatura e da pressão são mais significativos. Até cerca de 1000 m, o efeito da temperatura prevalece uma vez que o seu gradiente é maior nesta região, levando à diminuição da velocidade do som com a diminuição da temperatura da água. A partir dos 1000 m, aproximadamente, a temperatura sofre pouca variação pelo que o efeito da pressão começa a ser dominante, resultando num aumento da velocidade do som com o aumento da profundidade.

O efeito da salinidade torna-se significativo em latitudes elevadas, em que a temperatura tem pouca variação e a salinidade mostra gradientes verticais importantes, diminuindo *c* para menores valores de salinidade e aumentando para maiores valores desta.

NÃO CLASSIFICADO

- **Intensidade e Potência**

A intensidade acústica I é o valor médio do fluxo de energia associado a um sinal acústico, por unidade de superfície e unidade de tempo, estando definida em W/m^2. Representa-se por:

$$I = \frac{p^2}{\rho c} \tag{6}$$

Em que p é a pressão eficaz[5], N/m^2, ρ a densidade do meio, kg/m^3, e c a velocidade do som, m/s.

A potência acústica P em Watts (W) é portanto dada por:

$$P = I \times Área \tag{7}$$

Ou seja, é a intensidade acústica (I) por unidade de área ($Área$) do meio de propagação.

3.2. Processamento de Sinal

3.2.1. Teorema da Amostragem

Para que seja possível processar o sinal, é necessário a sua conversão de analógico para digital. Neste processo, o sinal $x(t)$, é amostrado a uma frequência definida pelo registador, passando a ser representado por um conjunto discreto de números. Estas amostras representam valores da onda original para cada instante pré-definido pela frequência de amostragem. Ao intervalo de tempo, normalmente constante, entre as amostras dá-se o nome de intervalo de amostragem (Ta), sendo o seu inverso a frequência de amostragem (fa).

Para que o sinal analógico possa ser reconstituído com exatidão, as amostras deverão estar convenientemente espaçadas. É necessário amostrar em número suficiente para extrair a componente de maior frequência do sinal de interesse, contudo se houver

[5] A pressão eficaz, também denominada pressão *root mean square* (RMS) é a que interessa avaliar, no caso dos sons puros é calculada por: $p_{eficaz} = \frac{p_{máxima}}{\sqrt{2}}$ (Costa, 2013).

NÃO CLASSIFICADO

sobreamostragem, resultado da criação de mais amostras do que as necessárias, teremos a introdução de dados desnecessários, que ocupa espaço de armazenamento em vão (William J. Emery, 2001, p. 3).

Segundo o Teorema de Amostragem, a frequência de amostragem terá que ser, no mínimo, superior ao dobro da frequência máxima contida no sinal analógico, ou seja:

$$fa > 2fmax \tag{8}$$

Para um dado intervalo de amostragem (Δt), a frequência de amostragem mínima designa-se frequência de Nyquist (fn) e corresponde à maior frequência de um sinal original que pode estar contida na amostragem:

$$fn = \frac{1}{2\Delta t} = 2fmax \tag{9}$$

Se o teorema anterior for respeitado no registo de um sinal, não ocorrerão fenómenos de distorção, denominados por *aliasing*. Quando a amostragem é realizada abaixo da frequência de *Nyquist*, obtém- se uma representação incorreta do sinal, que pode levar à obtenção de um sinal com frequência inferior à do sinal original (Imperadeiro, 2010, p. 14). O *aliasing* impossibilita portanto a correta reconstrução do sinal de origem.

A Figura 4 representa três sinais com frequências diferentes, todos amostrados a 1 kHz. Como se pode constatar, apenas o primeiro sinal poderá ser reconstituído uma vez

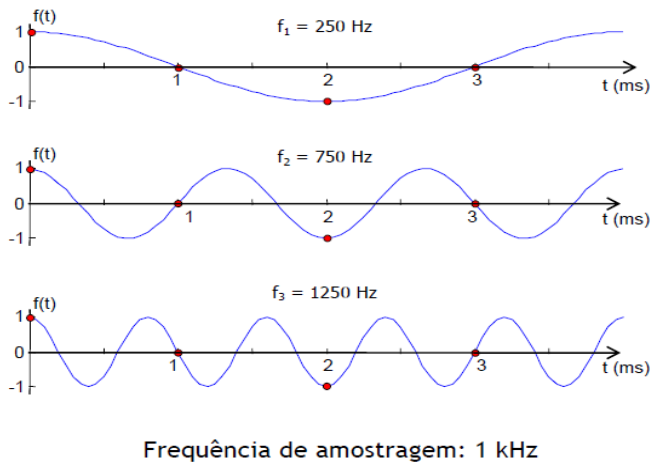

Figura 4 - Exemplo de amostragem para três frequências diferentes em que só a primeira obedece ao Teorema de Nyquist.

NÃO CLASSIFICADO

que é o único em que a frequência de amostragem obedece ao teorema de *Nyquist* ($1\ kHz > \ (2 \times 250)\ Hz$).

3.2.2. Duração do Sinal

No processo de amostragem é igualmente essencial garantir que a série de dados tem a duração suficiente para que seja representativo. Os dados devem ser recolhidos ao longo de um período que permita observar ciclos repetidos de um determinado fenómeno (William J. Emery, 2001, p. 5). É o comprimento total da série que permite extrair a sua frequência mais baixa, também designada frequência fundamental (f_0):

$$f_0 = \frac{1}{N\Delta t} = \frac{1}{Tt} = \Delta f \tag{10}$$

Em que Tt é o comprimento total da série de dados, N é o número de amostras obtido e Δf é a resolução em frequência ou diferença em frequência mínima.

Em teoria, na análise de um sinal, é possível distinguir toda a gama de frequências igual ou superior à frequência fundamental e menor à frequência de *Nyquist*.

Para extrair as frequências de interesse torna-se necessário gerir estes parâmetros. Por um lado, o tamanho da série a amostrar deve ser grande (T grande), para que f_0 cubra as frequências mais baixas do espectro e a resolução em frequência seja maior (Δf pequeno). Por outro lado, a amostragem deve ser rápida o suficiente para que fn se afaste de todas as frequências que pretendemos analisar, garantindo assim uma correta representação da sua intensidade espectral (Δt pequeno).

3.2.3. Transformada de Fourier

A Transformada de Fourier $X(\omega)$ é um dos métodos mais utilizados para identificar componentes periódicos em séries de tempo quase-estacionárias de dados oceanográficos (William J. Emery, 2001, p. 380).

Em 1807, o matemático francês, Joseph Fourier exercendo a função de administrador de Napoleão, estudou o problema da dissipação do calor através de blocos de metal, para melhorar a produção de canhões. Fourier defende que qualquer série de tempo de comprimento finito, repetida infinitamente, definida num intervalo [0, T], pode ser reproduzida através da soma linear de cossenos e senos ("séries de Fourier").

NÃO CLASSIFICADO

Trata-se de uma distribuição complexa que integra informação relativa às sinusoides que compõe o sinal $x(t)$. A transformada, decompõe o sinal original em somas de senos e cossenos e representa as suas amplitudes no domínio das suas frequências. Esta representação do sinal designa-se espectro e representa a variação de amplitude do sinal em função da frequência, ignorando as suas fases.

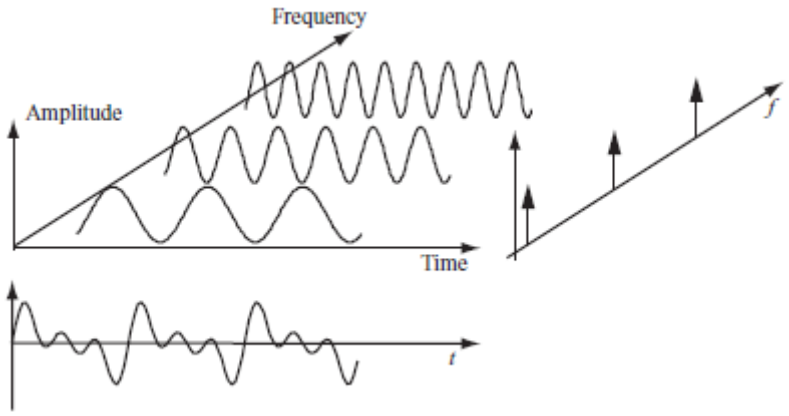

Figura 5 – Representação de um sinal, decomposto em três sinais com diferentes frequências e a mesma amplitude, vistos no domínio do tempo e da frequência (Brandt, 2011, p. 170).

Desta forma tem-se que a transformada de Fourier $X(\omega)$ de um sinal $x(t)$ é dada por (Li, 2012, pp. 41-43):

$$X(\omega) = \int_{-\infty}^{+\infty} x(t)\,[\cos(\omega t) - jsen(\omega t)], dt \tag{11}$$

Em que $\omega = \frac{2\pi}{T}$ é a frequência angular e $j = \sqrt{-1}$, referente à componente imaginária do sinal. Recorde-se que, pela fórmula de Euler, $e^{-j\omega t} = \cos(\omega t) - jsen(\omega t)$.

A *Discret Fourier Transform* (DFT) surge do advento dos computadores digitais e da necessidade de definir uma sequência discreta da transformada de Fourier (Lyons, 2010, p. 50):

$$X(m) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi nm/N} \tag{12}$$

Em que $X(m)$ é o valor da transformada discreta de Fourier, m é um valor entre $0\ e\ N-1$ correspondendo a frequências angulares entre $0\ e\ 2\pi$, $x(n)$ é a sequência de

NÃO CLASSIFICADO

valores de entrada, no domínio do tempo para n entre $0\ e\ N-1$. N é o número de amostras da sequência de entrada e o número de pontos no domínio da frequência da DFT de saída.

Sendo fn a maior frequência que se pode discriminar e f_0 o limite de resolução em frequência, então o número máximo de componentes que podemos estimar numa análise é (William J. Emery, 2001):

$$\frac{fn}{f_0} = \frac{\frac{1}{2\Delta t}}{\frac{1}{N\Delta t}} = \frac{N}{2} \tag{13}$$

Sinais periódicos podem ser representados através de séries de Fourier, como a soma de sinusoides cujas frequências são múltiplas da frequência fundamental do sinal. Da relação entre a frequência dos sons harmónicos (fh) e a frequência fundamental (f_0) , sendo h o número da harmónica, tem-se que:

$$fh = hf_0 \tag{14}$$

3.2.4. Potência Espectral

É possível obter-se a distribuição dos valores de potência em função da frequência. A potência espectral pode ser definida para o sinal completo de uma vez, dando origem a um *periodograma* ou dividindo o sinal em segmentos e calculando a média dos *periodograma* obtidos para esses segmentos, obtendo desta forma a densidade espectral de potência (Press, 1992).

- **Periodograma**

  Para a obtenção de um *periodograma* tem-se que:

$$Period. = \frac{|X(\omega)|^2}{N} \tag{15}$$

Sendo $X(\omega)$ a transformada de Fourier do sinal $x(n)$ e N o número de amostras do sinal.

NÃO CLASSIFICADO

- **Densidade Espectral de Potência**

A densidade espectral de potência trata-se da representação no domínio da frequência da potência por Hz em relação à frequência. Para sinais de longa duração, a média dos *periodogramas* dos segmentos do sinal, distribuí de forma mais precisa a potência pelas frequências corretas, para além de reduzir o ruído das amplitudes da potência. Por outro lado tem uma resolução em frequência inferior, limitada ao número de pontos utilizados para cada *Fast Fourrier Transform*[6] (FFT). A precisão deste método também aumenta com a dimensão de cada segmento. Para que não haja perda de informação, deve haver sobreposição dos referidos segmentos.

Existem diferenças significativas no nível médio de ruído para determinadas frequências e instantes no tempo, que devem ser sujeitas a um algoritmo de normalização. Os esquemas de normalização passam por recorrer à informação que rodeia a célula que se pretende representar e reduzir essas diferenças para a escala da sua representação.

3.2.5. Análise LOFAR E DEMON

A rotação dos hélices dos navios origina componentes de ruído irradiado de baixa frequência. No domínio da frequência caracterizam-se por ter uma amplitude maior do que a média e permite a deteção e identificação de navios.

A análise de baixas frequências, LOFAR e por demodulação de ruído, DEMON, (análise de banda-larga e de banda-estreita respetivamente), são considerados os métodos de análise mais efetivos em processamento digital de sinal e assumem que existem componentes singulares de frequências características no sinal detetado (Li, 2012, pp. 317-324).

Em LOFAR, a aplicação direta da análise de Fourier extrai os componentes de frequência do sinal de origem, abaixo dos 100 Hz, associados ao ruído das máquinas do navio e imprime frequências características. No processamento DEMON extraem-se as frequências mais baixas do sinal, vulgarmente designado por "som de fundo", após a eliminação do ruído ambiente (cavitação). A esta informação dá-se vulgarmente o nome

[6] Transformada rápida de Fourier, reduz o tempo de cálculo da DFT de N^2 passos para $(N/2)\log_2(N)$. O número de pontos terá que ser sempre uma potência de 2 (2^n pontos).

NÃO CLASSIFICADO

de “portadora” ou “envelope” do sinal, capaz de se propagar a grandes distancias. Esta análise, ao longo do tempo permite identificar o tipo de propulsão do navio, o que possibilita determinar o número de veios e o número de pás do contacto de interesse.

3.2.6. Espectro do Ruído Irradiado

O ruído irradiado por um navio dá origem a um espectro contínuo, sobreposto por componentes discretas de banda-estreita (conceito a desenvolver no ponto seguinte), os tonais. A intensidade da componente contínua e da discreta diminui com o aumento da frequência.

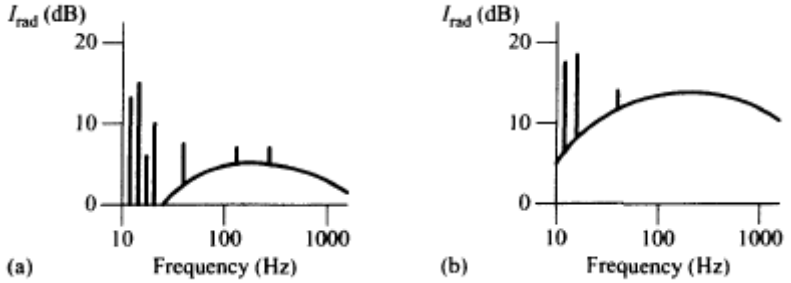

Figura 6 - Componente contínua e tonal do ruído irradiado para: (a) baixa velocidade e (b) velocidade elevada (Waite, 2002, p. 126)

As frequências mais baixas do espectro são dominadas pelos tonais produzidos pelas máquinas e pelos tonais resultantes do movimento das pás dos hélices. Com o aumento da frequência, estas linhas acabam por ficar integradas na parte contínua do espectro (a). Com o aumento da velocidade do navio, ou diminuição da profundidade, no caso dos submarinos, a componente contínua aumenta de intensidade e estende-se às frequências mais baixas do espectro, como ilustrado na Figura 6 em (b) (Waite, 2002, p. 126).

3.2.7. Banda-Estreita e Banda-Larga

Para interpretar o espectro de ruído irradiado utilizam-se técnicas de banda-larga e banda-estreita.

Os sonares de banda-larga (*broadband*), examinam a energia total numa grande banda de frequências. A sua performance aumenta com a largura de banda, desde que não exceda o espectro do ruído irradiado pelo alvo.

Os sonares de banda-estreita (*narrowband*), dividem a energia total em células de pequenos intervalos de frequência, a fim de detetarem linhas discretas de sinal irradiado.

NÃO CLASSIFICADO

A sua performance aumenta com a redução da largura de banda (Waite, 2002, pp. 129-131).

Existe ruído irradiado das duas categorias, banda-larga e banda-estreita, que dão origem a sinais que cobrem uma larga gama de frequências e a sinais compostos por tonais discretos, respetivamente.

Quanto às suas características temporais, os sinais podem ser classificados como contínuos, intermitentes ou transientes. Uma fonte de ruído intermitente, manifesta-se durante certos períodos com determinada regularidade, enquanto fontes transientes são ruídos inopinados e imprevisíveis (uma porta a bater, o movimento do leme na água).

Os sinais de banda-estreita têm necessariamente larguras de banda diferentes entre si. Um hélice, cuja velocidade em Rotações por Minuto (RPM) varia com o tempo, fazendo consequentemente variar a frequência de rotação das pás, dentro de uma gama de valores e os sistemas elétricos, normalmente bem regulados e com frequências muito estáveis, requerem necessariamente larguras de banda diferentes. Obteremos melhores resultados com filtros muito estreitos para frequências estáveis e filtros mais abrangentes para frequências menos estáveis (Hodges, 2010, pp. 183-184).

A Figura 7 representa o exemplo do espectro de um navio onde se podem identificar os componentes de banda-estreita e a componente contínua que corresponde a ruído de banda-larga.

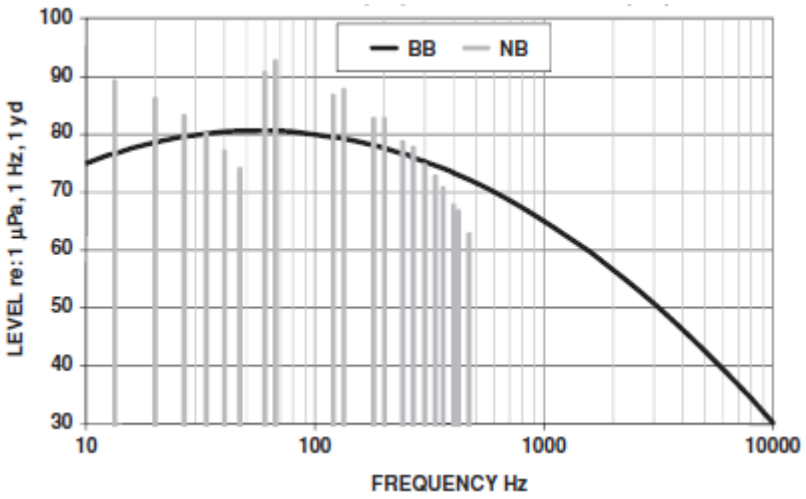

Figura 7 - Componentes linha de banda-estreita (NB) e banda-larga (BB) do ruído irradiado por um navio (Hodges, 2010, p. 184).

NÃO CLASSIFICADO

3.3. Meio Subaquático

3.3.1. Propagação do Som na Água

Certas propriedades da água fazem com que seja um excelente meio de propagação do som, como já foi referido, a sua elasticidade propicia a transmissão da vibração de umas partículas para as outras.

A velocidade de propagação do som na água varia com as características do meio, normalmente entre os 1450 e os 1550 m/s, dependendo das condições de temperatura, pressão e salinidade (Lurton, 2002, p. 13). Estes fatores combinam-se, para um determinado local, profundidade, estação do ano, localização geográfica e período do dia.

A propagação do som dentro de água pode fazer-se através de múltiplos caminhos. Uma mesma fonte emite som que se propaga por diferentes percursos, resultando em tempos de chegada e a intensidades distintas, a um mesmo recetor. Claro está que, quanto maior for a distância entre a fonte e o recetor, mais evidentes serão estas diferenças (Waite, 2002, p. 66).

3.3.2. Fontes de Ruído

Ao contrário do que se poderia pensar, o meio submarino é na realidade muito ruidoso, devido a uma série de diferentes causas (Urick, 1984, pp. 7-19). Há uma série de sons que se misturam com as ondas acústicas originadas pelos navios que se pretendem analisar. Para que seja possível distinguir e identificar um determinado navio, é essencial ter em conta a existência e influência de várias fontes de ruído (ver anexo 1). O designado ruído ambiente é o ruído de fundo do próprio oceano, não tendo origem ou direção particular e depende do nível de ruído associado a cada uma das fontes de ruído. Trata-se, em termos práticos, do ruído detetado por um sonar, na ausência de qualquer sinal ou ruído próprio do sistema (Lurton, 2002, p. 103). Respeitante às referidas fontes, podem-se considerar dois grupos principais, no que diz respeito à sua origem:

Fontes antropogénicas (ação humana);

Fontes ambientais (emitidos por agentes biológicos e por processos naturais).

NÃO CLASSIFICADO

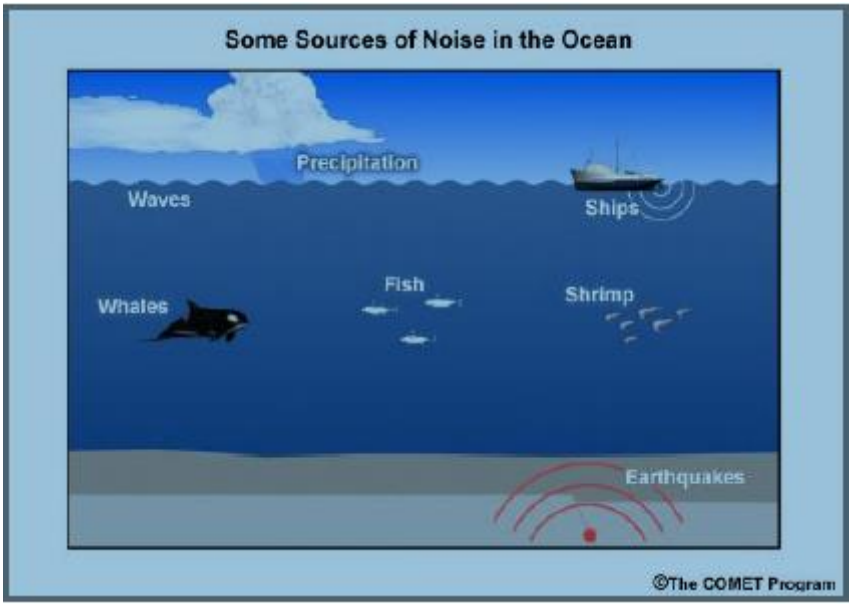

Figura 8 - Exemplos de fontes de ruído no Oceano (MetEd, 2015a).

- **Fontes Antropogénicas**

A atividade dos navios que se encontram a longas distâncias gera ondas acústicas que se propagam facilmente pelo oceano. Trata-se de um som de banda larga (conceito que será explorado posteriormente) que é detetado pelos sensores acústicos. As zonas portuárias, indústrias junto à costa, operações de dragagem e as próprias instalações costeiras, produzem diferentes ruídos de alguma intensidade. Estes ruídos são ainda destorcidos pelos fenómenos de propagação das ondas sonoras, a abordar posteriormente. Tratando-se frequentemente da principal causa de ruído acústico, corresponde a frequências entres os 10 Hz e 1 kHz (Richardson, Greene Jr, Malme, & Thomson, 2013, p. 92).

- **Fontes Ambientais**

Dentro dos fatores ambientais é possível ainda a distinção de diversas origens de ruído:

Agitação da superfície da água, a interação do vento atmosférico com as águas oceânicas geram uma série de fenómenos ruidosos, como vagas, bolhas de ar e "spray" (salpicos). A intensidade do ruído será proporcional à velocidade e continuidade do vento que se fizer sentir no local. Gera ruído desde alguns Hz até algumas dezenas de kHz (Lurton, 2002, pp. 107-108);

NÃO CLASSIFICADO

Biológicas, os próprios animais marinhos e plantas aquáticas que estão associados a este habitat geram sons muito distintos. Estes sons podem ser o resultado da comunicação entre mamíferos marinhos (como o caso dos Cetáceos) ou até mesmo da ecolocalização de presas e obstáculos. As características dos sinais emitidos variam de espécie para espécie. As baleias por exemplo, emitem vocalizações de baixa frequência capazes de se propagarem até centenas de quilómetros (entre 12 Hz e alguns milhares de Hz). Enquanto mamíferos de menores dimensões fazem soar assobios de alta frequência (superior a 1 kHz). Os cetáceos odontocetes, como é o caso dos golfinhos e toninhas recorrem à ecolocalização, produzindo "clicks" entre os 50 e os 200 kHz (Lurton, 2002, pp. 110-111).

Outros fenómenos, como a queda de chuva e neve, a recorrente atividade sísmica de baixa intensidade, o ruído do *stress* mecânico do gelo polar provocado por mudanças de temperatura, constituem fontes de ruído significativas (Lobo, 2002, pp. 190-191). A atividade sísmica e vulcânica está limitada a frequências muito baixas, na ordem das dezenas de Hz. Por outro lado, a chuva gera ruído de alta intensidade, devido ao impacto das gotas na superfície da água e à implosão das bolhas de ar geradas pelo embate. À semelhança das frequências de ruído originadas pela agitação da superfície da água por ação do vento, para a chuva, as frequências típicas identificadas estão entre 1 e 100 kHz.

3.3.3. Sons Gerados por Navios

Existem uma série de fontes associadas a cada navio que contribuem para a totalidade do ruído emitido pelas plataformas (Lobo, 2002, pp. 184-189).

O ruído irradiado é a fonte de sinais para sonares passivos. Consiste no ruído irradiado por um navio e medido a determinada distância (Waite, 2002, p. 93).

- **Ruído Irradiado pelos Hélices**

Os hélices, por se encontrarem em contacto direto com o meio de propagação, produzem ruído distinto do associado às máquinas no interior do navio. As depressões geradas pelo movimento das pás na água provoca a cavitação, que se caracteriza por ruído de banda-larga a elevadas frequências. Neste processo, a rotação acelerada das pás do hélice, origina uma redução significativa da pressão que dá origem a uma evaporação rápida das moléculas de água (geração de bolhas de ar) emitindo um som "explosivo". As

NÃO CLASSIFICADO

frequências geradas pelo movimento giratório dos hélices, definem linhas espectrais a frequências muito baixas, entre 0,1 e 10 Hz. As frequências emitidas dependem da velocidade de rotação das pás e da sua geometria. As frequências das séries harmónicas identificadas pode ser dada por (Hodges, 2010, p. 185):

$$fm = mnfr \tag{16}$$

Em que m é o número de harmónicas, n o número de pás do hélice e fr a frequência de rotação do hélice.

- **Ruído das Máquinas**

As vibrações originadas pelas explosões periódicas dos cilindros de um motor a *diesel*, as descontinuidades repetitivas das pás de uma turbina, os dentes das engrenagens e irregularidades das máquinas rotativas dão origem a componentes de ruído de banda-estreita. Estando estas fontes associadas ao sistema de propulsão principal do navio, as suas frequências e amplitudes aumentam com a velocidade. Por outro lado, a turbulência e cavitação do fluido dos sistemas hidráulicos do navio, bombas e válvulas, produzem sinais de banda-larga.

- **Ruído Hidrodinâmico**

O ruído hidrodinâmico está associado a uma série fatores. A circulação do fluxo de água ao longo da estrutura do navio gera ondas de frequência muito baixa. O próprio contacto da água com o casco do navio provoca turbulência, que se reflete em ruido de maior frequência. Em último lugar, pequenas irregularidades no casco, com a passagem da água, provocam cavitação, que origina ruído de frequências ainda mais elevadas (Lobo, 2002, p. 188).

- **Ruído da Atividade de Bordo**

A atividade do pessoal embarcado, especialmente tratando-se de um navio civil, é muito ruidosa. Os navios militares, principalmente submarinos, têm um cuidado redobrado com ruídos desta natureza, tendo políticas muito rígidas em relação ao ruído produzido pela guarnição. Algumas atividades realizadas no mar são particularmente sonorosas, nomeadamente as que envolvem perfuração do leito marinho, pesca de arrasto, reboque, lançamento de submersíveis, entres outras (Lurton, 2002, p. 113).

NÃO CLASSIFICADO

3.3.4. Perda ou Alteração na Propagação do Sinal

No seu trajeto ao longo do oceano, as ondas sonoras sofrem determinados fenómenos que alteram as suas características ou o seu trajeto. (Waite, 2002, pp. 43-49).

- **Divergência**

As perdas por divergência ou espalhamento estão ligadas ao facto de o sinal acústico enfraquecer progressivamente com a distância à fonte (devido à propagação radial do som). Trata-se do efeito de redução da densidade de energia. Sabendo que a potência total do sinal proveniente da fonte é dada por $P = Intensidade \times área$, se a área aumenta, com o aumento da distância à fonte, a intensidade terá que diminuir para que a potência total P permaneça constante.

- **Atenuação**

A atenuação inclui as perdas por absorção e por retro-dispersão (*scattering*). Estas perdas estão dependentes da frequência de transmissão, da temperatura da água e igualmente da distância.

As perdas por absorção ocorrem principalmente devido a dois fatores, a viscosidade e a relaxação molecular e causam uma perda de intensidade acústica ao longo do percurso. Tanto a água doce como salgada estão sujeitas a perdas devido à viscosidade, sendo essa perda proporcional ao quadrado da frequência. Perdas por relaxação molecular apenas ocorrem em água salgada. Este mecanismo consiste na redução das moléculas em iões, induzida pela pressão do som. Para frequências muito elevadas (maiores do que 500 kHz), as alterações de pressão são demasiado rápidas para que este fenómeno ocorra, logo não há absorção da energia do sinal. A relaxação do sulfato de magnésio é dominante entre os 2 e os 500 kHz e a do ácido bórico abaixo dos 2 kHz (Waite, 2002, p. 47).

Transmitindo-se dentro de água, o som está sujeito a múltiplos obstáculos. Ao entrar em contacto com a fauna marinha, partículas suspensas, bolhas de ar, ao atravessar a estrutura heterogénea do próprio oceano ou ainda quando sofre reflexão no leito do mar e na superfície da água (zonas de fronteira), o sinal sonoro acaba por dispersar em várias direções. Para sonares ativos, a porção de energia sonora incidente que é refletida de volta para a fonte designa-se *backscattering* e está associada ao fenómeno de reverberação.

NÃO CLASSIFICADO

- **Reflexão**

As perdas por reflexão relacionam-se com a rugosidade dos refletores e com a frequência do sinal. Ocorre quando o sinal embate na superfície e no fundo. Não sendo estas fronteiras refletores perfeitos, a energia sonora incidente é sempre superior à da onda sonora refletida.

- **Refração**

A refração é descrita pela lei de Snell, num meio em que a velocidade do som não é constante. Quando a velocidade do som varia continuamente com a profundidade, pode-se considerar que o meio de propagação se encontra dividido em finas camadas, com diferentes características. Segundo esta lei, a direção de propagação do som vai tender a desviar-se no sentido da camada de menor velocidade de propagação, como representado na Figura 9. Desta forma, acaba por descrever curvas (Waite, 2002, pp. 59-60). Sendo θ o ângulo de incidência da onda sonora na fronteira das diferentes camadas, e c a velocidade do som,verifica-se a seguinte relação:

$$\frac{\cos\theta 1}{c1} = \frac{cos\theta 2}{c2} = \frac{cos\theta 3}{c3} \tag{17}$$

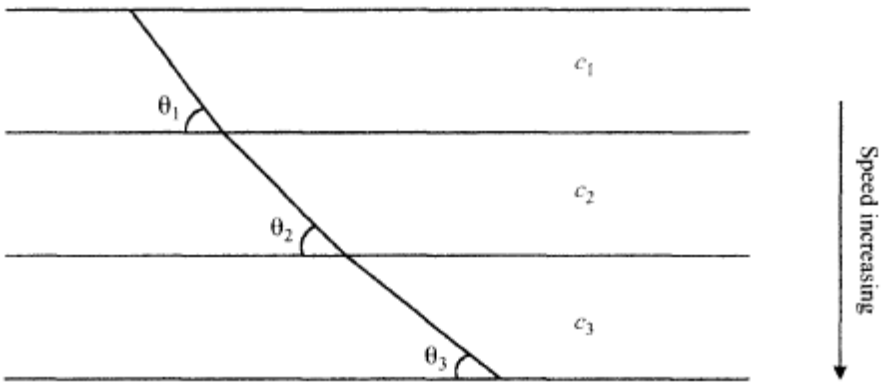

Figura 9- Ilustração da Lei de Snell (Waite, 2002, p. 59)

NÃO CLASSIFICADO

- **Efeito de Doppler**

Às alterações de frequências que resultam do movimento relativo entre a fonte sonora e o recetor, dá-se o nome de efeito de Doppler. Este efeito verifica-se devido à alteração da duração do percurso entre a fonte e o recetor durante o tempo de transmissão (Lurton, 2002, p. 34).

A variação em frequência é definida por (Hodges, 2010, p. 9):

$$\Delta f = fs\frac{vs - vr}{c - vs} \tag{18}$$

Em que fs é a frequência da fonte sonora em kHz, vs a velocidade relativa da fonte em nós (positiva quando há aproximação e negativa quando há afastamento), vr é a velocidade relativa do recetor em nós (positiva ou negativa à semelhança de vs) e c é a velocidade do som no meio (1500 m/s na água).

Como as velocidades são muito pequenas comparadas com a velocidade do som, é válido dizer que, no oceano:

$$\Delta f = fs\frac{\Delta v}{c} \cong 0{,}35fs\Delta v \tag{19}$$

Para o caso que se pretende estudar, em que o recetor está fixo e é apenas a fonte sonora que se aproxima ou se afasta em relação a ele tem-se que, no recetor, a frequência observada é dada por:

$$fr = fs\frac{c}{c - vs} \tag{20}$$

Se é o recetor que tem movimento relativo em relação à fonte tem-se:

$$fr = fs\frac{c - vr}{c} \tag{21}$$

Caso tanto o recetor como a fonte, tenham movimento relativo em relação ao outro, a frequência no recetor é dada por:

NÃO CLASSIFICADO

$$fr = fs\frac{c - vr}{c - vs} \tag{22}$$

3.3.5. Sonares

Existem dois grandes grupos de sonares, ativos e passivos. Ambos têm como objetivo detetar sinais do alvo. No caso dos sonares ativos, o som é emitido pelo próprio sonar e o sinal que se pretende analisar corresponde aos ecos dos alvos encontrados no caminho da sua propagação. No caso dos sonares passivos, o sinal analisado é o próprio som irradiado pelos alvos. Os sonares do tipo ativo, geram um pulso sonoro que se propaga através da água até um alvo, e regressa como um eco. Os alvos são detetados, localizados e classificados a partir dos ecos recebidos pelo sonar. Os sonares passivos foram concebido para detetar o ruído irradiado por um alvo inserido no ruído de fundo do meio (Waite, 2002, pp. 125-128). Na realização desta dissertação, exploraram-se registos acústicos, obtidos por escuta passiva, através de um hidrofone.

NÃO CLASSIFICADO

Capítulo 4 - Arquitetura da Solução: Aprendizagem Supervisionada

Partindo dos conceitos anteriormente expostos e tendo por base a tecnologia de que atualmente dispomos, considera-se ser possível edificar um sistema capaz de classificar cada navio, com base no ruído que emite.

Assim sendo, pretende-se utilizar dados previamente recolhidos, selecioná-los, pré processá-los e recorrer a técnicas de *data mining*[7], em particular, classificação com aprendizagem supervisionada, para identificar cada navio.

4.1. O *Data mining*

O *data mining* está associado à exploração de dados. Este processo emergente permite identificar correlações, padrões e tendências, processando grandes quantidades de dados, recorrendo a tecnologias de reconhecimento de padrões, estatística e técnicas matemáticas (Freitas, 2013, pp. 1-2).

A grande vantagem do *data mining* é a capacidade de transformar grandes blocos de dados e informação em conhecimento, superando a capacidade analítica do ser humano. Atualmente, o crescimento explosivo da recolha de dados, o armazenamento dos dados em *data warehouses*, o aumento da disponibilidade de acesso dos dados devido à internet, o desenvolvimento de *software* de *data mining* bem como o melhoramento da capacidade de computação a nível de processamento e armazenamento, têm contribuído para o significativo desenvolvimento desta área exploratória (Larose, 2005, p. 4).

[7] Prospeção ou processamento de dados.

NÃO CLASSIFICADO

4.1.1. Organização dos Dados

O *data mining* pode ser aplicado a quaisquer conjuntos de dados. As formas mais comuns de representar e organizar os dados são as bases de dados, os *data warehouses* e os dados transacionais. A Figura 10 ilustra os três exemplos referidos (Han, Kamber, & Pei, 2011, pp. 8-14).

Figura 10 - Formas de representar os dados

4.1.2. Fases de um Projeto *Data mining*

De acordo com a *Cross-Industry Standard Process for Data Mining* (CRISP-DM), um projeto de *data mining* compreende um ciclo de seis fases, como ilustra a Figura 11. Uma primeira fase dedicada à compreensão da investigação, em que se enunciam os objetivos e requisitos do projeto e traça-se uma estratégia para atingir esses objetivos. Na segunda fase recolhem-se os dados, avalia-se a qualidade dos mesmos e procede-se à sua seleção. Na fase seguinte, preparam-se os dados até se ter um data set[8] final que se irá usar nas fases seguintes. Na quarta fase selecionam-se e aplicam-se técnicas de modulação e calibram-se os modelos por forma a otimizar os resultados. A quinta fase foi concebida para avaliar os modelos e a sua qualidade em função dos objetivos definidos na primeira fase, antes de se implementarem no terreno. Por fim, na sexta fase aplicam-se os modelos criados (Larose, 2005, pp. 5-10).

[8] Conjunto de dados.

NÃO CLASSIFICADO

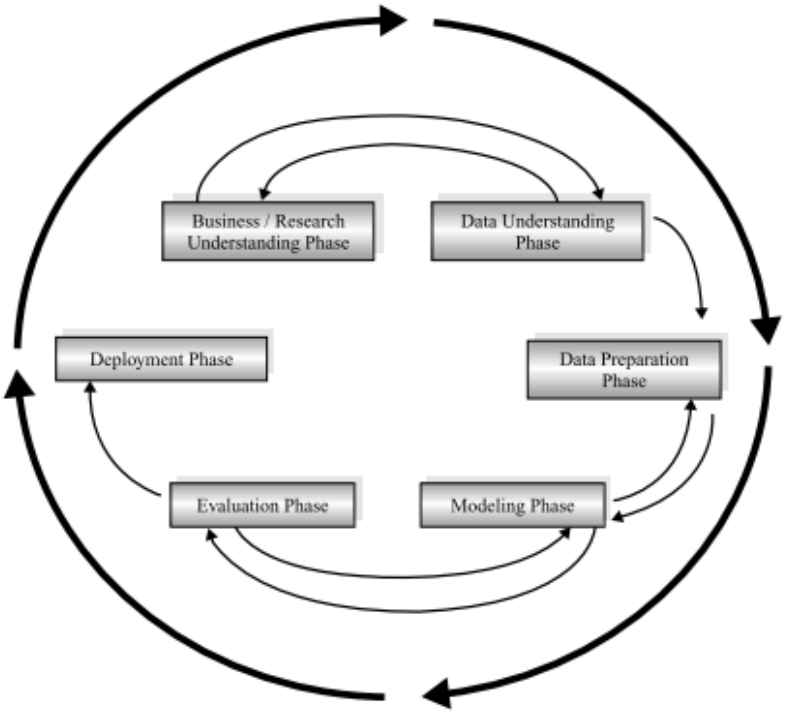

Figura 11 - Fases do processo de data mining (Larose, 2005, p. 6)

4.1.3. Objetivos do *Data mining*

O *data mining* por ser utilizado para vários fins, sendo os mais comuns a descrição, estima, predição, classificação, *clustering* e associação (Larose, 2005, pp. 11-17). Na descrição os investigadores pretendem encontrar formas de descrever padrões e tendências inerentes aos dados. Na estima efetua-se uma classificação dos dados, sendo a variável alvo sempre do tipo numérica. A predição assemelha-se à classificação e à estima, à exceção de que os resultados obtidos se aplicam apenas ao futuro. Na classificação utiliza-se uma variável alvo categórica, que pode ser particionada em várias categorias. Já o *clustering* consiste no agrupamento de registos, observações ou casos em classes de objetos semelhantes. O *data mining* pode ainda ser utilizado para efetuar associações, ou seja, procurar atributos que estejam relacionados (Hand, Mannila, & Smyth, 2001, pp. 12-15).

4.2. *Machine Learning* e Métodos de Aprendizagem

O machine learning é uma área do data mining que combina inteligência artificial com estatística e tem por intuito gerar algoritmos que aprendam ao explorar o espaço de n dimensões de um determinado conjunto de dados e encontrem generalizações válidas.

NÃO CLASSIFICADO

Uma das suas principais vertentes é o machine learning indutivo, em que a generalização é obtida a partir de um conjunto de amostras e formalizada recorrendo a diferentes técnicas e modelos. A aprendizagem indutiva pode ser definida como o processo que estima as dependências input-output desconhecidas ou a estrutura de um sistema, recorrendo a um número limitado de observações ou a medidas de inputs e outputs do sistema (Kantardzic, 2011, pp. 89-101). A Figura 12 representa o processo de aprendizagem descrito:

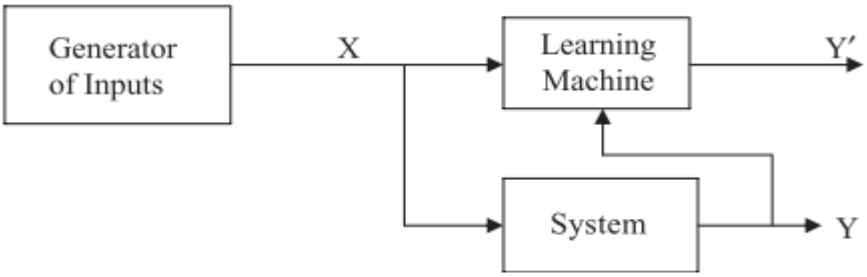

Figura 12 - Processo de machine learning utilizando as informações do sistema para gerar um output Y' para o input X.(Kantardzic, 2011, p. 89)

Como exemplos de aprendizagem indutiva destacam-se os problemas de interpolação, regressão, classificação, *clustering* e estima da densidade.

Ainda dentro da aprendizagem indutiva, existem dois tipos de métodos: aprendizagem supervisionada e aprendizagem não supervisionada. A aprendizagem supervisionada está associada ao processo de classificação, em que a supervisão do sistema passa por aprender com os exemplos identificados do conjunto de treino. A aprendizagem não supervisionada está basicamente associada ao clustering. Neste método, são descobertas classes de dados dentro do conjunto de treino, os exemplos não estão previamente classificados, assim sendo, não existe a supervisão do sistema (Han et al., 2011, pp. 24-25).

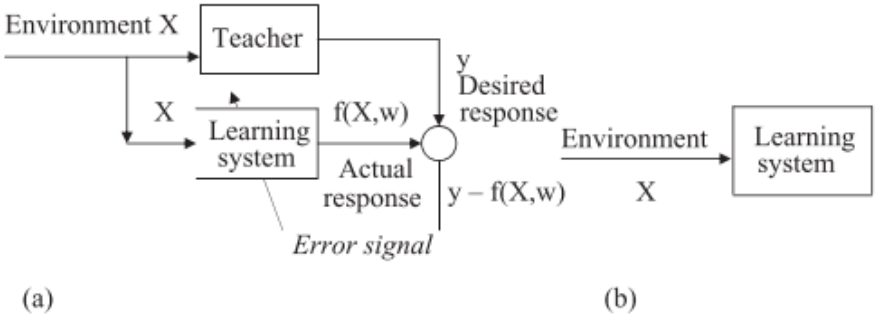

Figura 13 - Dois tipos de aprendizagem indutiva: Aprendizagem supervisionada a) e aprendizagem não supervisionada b) (Kantardzic, 2011, p. 100).

NÃO CLASSIFICADO

4.3. Aprendizagem Supervisionada

Neste método de aprendizagem indutiva, como referido anteriormente, o algoritmo utilizado aprende com os exemplos identificados do conjunto de treino e associa cado novo caso aos que já conhece (Witten & Frank, 2005, p. 43). No âmbito desta dissertação, foram selecionados três classificadores distintos baseados na aprendizagem supervisionada, que apesar da dificuldade que existe *a priori* em definir qual o algoritmo que tem melhores resultados para um dado conjunto de dados, se consideram eficazes, dado o seu já reconhecido bom desempenho (Gama & Brazdil, 1995). Assim sendo, os classificadores utilizados foram o vizinho mais próximo (IBK), o *naive de Bayes* e a árvore de decisão (C4.5).

4.3.1. Classificador Vizinho mais Próximo ou K-vizinhos

O classificador Vizinho mais Próximo ou k-vizinhos tem vindo a ser amplamente utilizado na área de reconhecimento de padrões, apesar de estar associado a alguma demora na fase de teste, daqui serem igualmente conhecidos por *lazzy*. De entre os vários métodos de aprendizagem supervisionada que existem, destaca-se pelo seu alto desempenho consistente (Freitas, 2013, pp. 34-35). O k-vizinhos insere-se no grupo da aprendizagem baseada em instâncias, em que o conjunto de treino é simplesmente armazenado e o processo de classificação passa por comparar um novo conjunto de dados (conjunto de teste) com os dados de treino devidamente classificados, não há a construção de um modelo, como acontece para outros classificadores (Han et al., 2011, p. 423).

As instâncias de treino são descritas por n atributos e cada uma representa um ponto num espaço de dimensão n. Para um dado conjunto de teste, este classificador identifica o padrão de distância para as k instâncias de treino consideradas mais próximas da instância desconhecida (Han et al., 2011, p. 423).

O grau de semelhança ou proximidade entre os dados determina-se com recurso a uma função de distância $d(x, y)$, com as seguintes propriedades (Larose, 2005, p. 99):

1. $d(x, y) \geq 0 \; e \; d(x, y) = 0 \; se \; e \; só \; se \; x = y$ (23)
2. $d(x, y) = d(y, x)$

NÃO CLASSIFICADO

3. $d(x,z) \leq d(x,y) + d(y,z)$

Algumas das distâncias mais utilizadas são as distâncias de *Minkowski (Zezula, Amato, Dohnal, & Batko, 2006, p. 10)*, em particular a de Manhattan[9] e a Euclidiana[10]:

$$dMinkowski(x,y) = \sqrt[p]{\sum_{i=1}^{n} |xi - yi|^p}\,, \tag{24}$$

$$dManhattan(x,y) = \sum_{i=1}^{n} |xi - yi| \tag{25}$$

$$dEuclidiana(x,y) = \sqrt{\sum_{i=1}^{n} (xi - yi)^2} \tag{26}$$

Sendo $x = x^1, x^2, \ldots, x^n$ *e* $y = y^1, y^2, \ldots, y^n$ os valores dos n atributos de dois conjuntos de dados que se pretendem comparar. Antes de proceder ao cálculo das distâncias, é importante normalizar[11] os dados, para que os atributos tenham igual peso na classificação (Han et al., 2011, p. 113).

Após normalizados, é então necessário prever qual a classe de uma determinada instância $(x, c(x))$. Para tal o algoritmo determina os k vizinhos, pertencentes ao conjunto de treino, que distam menos desta, classificando-a segundo a expressão:

$$\hat{c}(x) \leftarrow arg \max_{v \epsilon V} \sum_{i=1}^{k} \delta(v, c(y_i))\,, \tag{27}$$

[9] Distância de *Minkowski* de ordem 1.

[10] Distância de *Minkowski* de ordem 2. A distância Euclidiana torna-se menos discriminante com o aumento do número de atributos, devendo considerar-se a utilização de outra função para estes casos (Kantardzic, 2011, p. 120).

[11] Na normalização a escala dos dados é reduzida para valores entre -1,0 e 1,0 ou 0,0 a 1,0, por exemplo.

NÃO CLASSIFICADO

Em que $y_1, \dots, y_k$ correspondem aos k vizinhos mais próximos de x, pertencentes ao conjunto de treino, sendo V o conjunto de classes e,

$$\delta(x,y) = \begin{cases} 0, & x \neq y \\ 1, & x = y \end{cases}. \tag{28}$$

A escolha do parâmetro k não é um processo trivial. Normalmente é selecionado por experimentação ou partindo do conhecimento *à priori* do problema de classificação considerado. Para vários valores de k, o que tiver menor taxa de erro deve ser o escolhido. Por um lado, quanto maior o valor de k, maior será o tempo de processamento na fase de teste do algoritmo mas mais suave será a fronteira entre diferentes classes, reduzindo a influência do ruído. Contudo o tempo de teste é independente do número de classes, daqui a vantagem de utilização deste algoritmo em problemas com múltiplas classes. Por outro lado, um k mais pequeno irá preservar o comportamento local de interesse. De uma forma geral, são utilizados k na ordem das unidades e dezenas, mais do que nas centenas e milhares (Kantardzic, 2011, p. 120). Outra perspetiva a considerar na fase de seleção do k, é que este deverá ser preferencialmente impar, para que o classificador não tenha dúvidas em relação à classe dominante, evitando a situação de ter o mesmo número de vizinhos de duas classes distintas (Larose, 2005, p. 98).

O processo de classificação do vizinho mais próximo pode-se resumir nos seguintes passos:

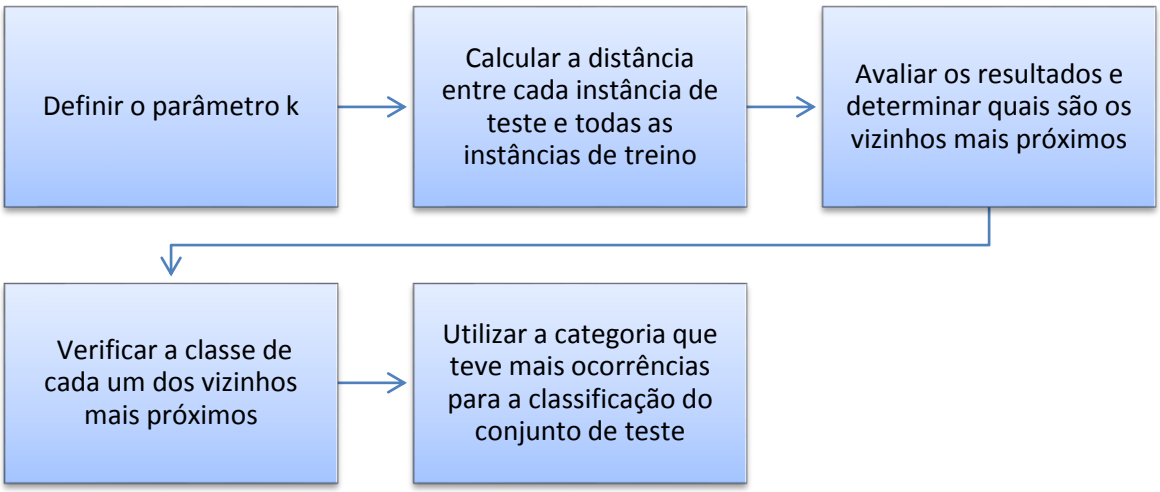

Figura 14 – Passos gerais do processo de classificação dos k-vizinhos

NÃO CLASSIFICADO

Aplicando este classificador ao reconhecimento do espectro de diferentes navios considere-se o seguinte exemplo, que expressa a importância da escolha acertada do parâmetro k. Vamos assumir que podemos representar cada som como um ponto (neste caso representado graficamente como um polígono) num plano, de tal forma que distâncias entre os pontos nesse plano são proporcionais às distâncias entre os sons representados (ver Figura 16). Cada polígono vermelho representa uma instância que sabemos ser da lancha de fiscalização Auriga, e cada polígono verde representa uma instância que sabemos ser da lancha de desembarque grande Bacamarte. O conjunto das referidas instâncias constitui o conjunto de treino. O polígono sem cor representa a instância que se pretende classificar (que neste caso pertence à LDG) e faz parte do conjunto de teste.

Para k=1, k=2 e k=3, o classificador identificaria corretamente a instância como pertencente à Bacamarte. Contudo, se aumentássemos os k-vizinhos para 7 ou mais, obteríamos uma classificação errada, uma vez que a maioria das instâncias passaria a ser da outra classe, pelo critério da maioria.

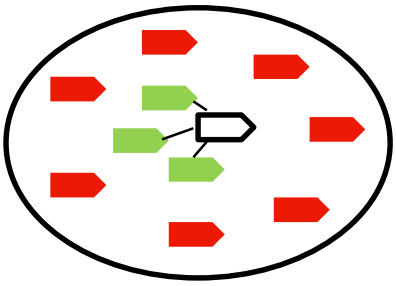

Figura 16 - Exemplo 1, classificador k-vizinhos com k=3.

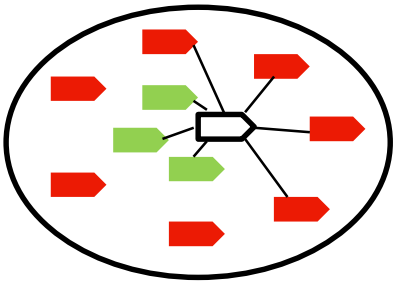

Figura 15 - Exemplo 1, classificador k vizinhos com k=7

4.3.2. Classificação *Bayesiana*

Os métodos de classificação *Bayesiana* consistem em métodos de predição de classes baseados no teorema de *Bayes*, que será explicado de seguida. Estudos nesta área revelam que o classificador *naive de Bayes*, tal como redes neuronais e árvores de decisão, tem bons desempenhos em termos de eficácia (baixo erro de classificação), e velocidade de processamento, quando aplicado em grandes quantidades de dados (Han et al., 2011, p. 350).

NÃO CLASSIFICADO

- **Teorema de *Bayes***

Para compreender o teorema de *Bayes*, considere-se uma instância B e uma hipótese A. Se estivermos perante um problema de classificação, aquilo que pretendemos calcular é $P(A|B)$, ou seja a probabilidade de A ocorrer, sabendo que ocorreu B.

Considere-se, a título de exemplo, que se pretende determinar se uma embarcação de pesca está ou não em infração consoante o tempo que passou desde a última fiscalização a que foi sujeita. A embarcação de pesca *Neptuno* é um arrastão e foi sujeita a fiscalização há seis anos. Para se determinar se está ou não em infração, calcula-se a probabilidade de estar em infração sabendo que foi fiscalizada há seis anos e é um arrastão. Desta forma tem-se que:

- $P(A)$ – Probabilidade da embarcação de pesca estar em infração;
- $P(B)$ – Probabilidade da embarcação de pesca ter sido fiscalizada há seis anos e ser um arrastão;
- $P(B|A)$ – Probabilidade da embarcação de pesca ter sido fiscalizada há seis anos e ser um arrastão, sabendo que está em infração.

Para calcular $P(A|B)$, partindo da regra de *La Place* (Hartigan, 1983, p. 1):

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \tag{29}$$

Multiplicando e dividindo por $P(A)$ obtém-se,

$$\frac{\mathrm{P(A|B)\ P(A)}}{\mathrm{P(A)}} = \frac{\mathrm{P(A \cap B)}}{\mathrm{P(B)}}, \tag{30}$$

Isolando $P(A|B)$,

$$P(A|B) = \frac{P(A \cap B)\ P(A)}{P(A)\ P(B)}, \tag{31}$$

Ao simplificar, obtém-se o teorema de *Bayes* (Hartigan, 1983, pp. 29-31):

NÃO CLASSIFICADO

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}. \quad (32)$$

Esta relação foi idealizado para tentar contornar o problema de não se poder saber a priori uma classe dado um dado. Assim, com esta expressão consegue-se saber a *priori* como são os dados de uma classe dada a própria classe.

- ***Naive de Bayes***

Considere-se um conjunto de treino T, composto por um conjunto de instâncias B_i com dimensão n, sendo que cada uma é representada pelo vetor de atributos $B_i = (b_i^1, b_i^2, \dots, b_i^j \dots b_{in}^{jn}, C_i \dots C_{in})$, associado a n atributos, $X^1, X^2, \dots, X^n$ e a uma classe C.

Temos então um conjunto de treino T_{ex}, composto por cinco instâncias com duas dimensões, em que $j = 1$ corresponde ao atributo tempo em anos e $j = 2$ ao tipo de embarcação de pesca. Assumindo que a instância B_1 é uma embarcação de pesca que foi fiscalizada há 1 ano, é uma cercadora e está legal ($C_1 = Legal$), a instância B_2 um arrastão que foi fiscalizado há 2 anos e está em infração ($C_2 = Infrator$), a instância B_3 um palangre fiscalizado há 3 anos e está legal ($C_3 = Legal$), a instância B_4 um polivalente, fiscalizado há 4 anos e está legal ($C_4 = Legal$), e a instância B_5, um arrastão fiscalizado há 5 anos e está em infração ($C_5 = Infrator$), tem-se: $B_1(1, CERCADORA, Legal)$, $B_2(2, ARRASTÃO, Infrator)$, $B_3(3, PALANGRE, Legal)$,$B_4(4, POLIVALENTE, Legal)$ e $B_5(5, ARRASTÃO, Infrator)$.

Desta forma, podemos dizer que o vetor atributo tempo, em anos, corresponde a $X^1(1, 2, 3{,}4{,}5)$ e o vetor atributo tipo de embarcação a X^2(CERCADORA, ARRASTÃO, PALANGRE,POLIVALENTE,ARRASTÃO). A seguinte matriz e a Figura 17 ilustram o conjunto de treino T^{ex}. As instâncias legais estão representadas a verde e as correspondentes a infração a vermelho.

NÃO CLASSIFICADO

	X_1	X_2	C_i
B_1	1	CERCADORA	Legal
B_2	2	ARRASTÃO	Infrator
B_3	3	PALANGRE	Legal
B_4	4	POLIVALENTE	Legal
B_5	5	ARRASTÃO	Infrator

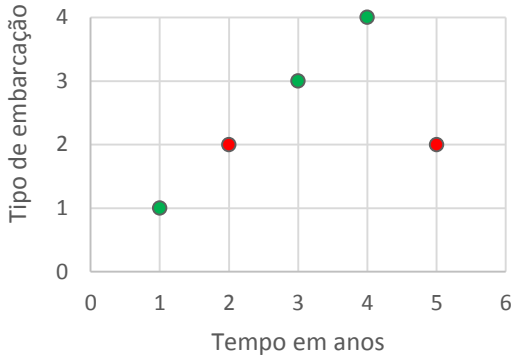

Figura 17 - Gráfico de dispersão do conjunto de treino T_{ex}

Note-se que para cada tipo de embarcação foi associado um valor numérico $(1, 2, 3\ e\ 4)$. Cada ponto representa uma instância e a cada eixo, os dois atributos considerados.

Existem m classes, $C^1, C^2, \ldots, C^m$ e o classificador irá identificar B como pertencente à classe que tenha a maior probabilidade condicionada em B, ou seja:

$$P(C^i|B) > P(C^j|B)\ para\ 1 \leq j \leq m, j \neq i. \tag{33}$$

Para tal é necessário maximizar $P(C^i|B)$, recorrendo ao teorema de *Bayes*,

$$P(C^i|B) = \frac{P(B|C^i)\,P(C^i)}{P(B)}. \tag{34}$$

Uma vez que $P(B)$ é constante para todas as classes, apenas é necessário maximizar $P(B|C^i)\,P(C^i)$. Sempre que a probabilidade de cada classe não é conhecida, assume-se que estas são equiprováveis, logo $P(C^i) = 1/m$, onde m é o número de classes.

Regressando ao exemplo, considere-se $P(C^1)$ a probabilidade da classe "Infrator" e $P(C^2)$ a probabilidade da classe "Legal". Desta forma, $P(C^1) = \#C^{1,T_{ex}}/\#T^{ex} = 2/5$ e $P(C^2) = \#C^{2,T_{ex}}/\#T^{ex} = 3/5$, onde $\#C^{1,T_{ex}}$ é o número de elementos de T^{ex} que têm a classe C^1.

Caso o conjunto T tenha muitos atributos, torna-se bastante difícil calcular $P(B|C^i)$, devido ao elevado número de probabilidades a calcular (Hand et al., 2001,

NÃO CLASSIFICADO

p. 211) Assim, para reduzir a dificuldade de processamento, admite-se que os atributos são independentes entre si, logo sendo b^k o valor do atributo X^k por instância B, tem-se:

$$P(B|C^i) = \prod_{k=1}^{n} P(b^k|C^i) = P(b^1|C^i) \times P(b^2|C^i) \times \ldots \times P(b^n|C^i) \quad (35)$$

Para se calcular a probabilidade de $P(b^k|C^i)$ é necessário ter em conta se os atributos X^k são categóricos ou contínuos.

Se X^ké uma variável categórica, então $P(b^k|C^i)$ corresponde ao número de instâncias da classe C^i no conjunto de treino *T*, que tenham o valor b^k para o atributo X^k, dividido pelo número total de instâncias da classe C^i em *T*.

Se X^ké uma variável contínua, recorre-se a uma distribuição normal, com média μ e desvio padrão σ, definida por:

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (36)$$

Logo,

$$P(b_k|C_i) = g(b_k, \mu_{C_i}, \sigma_{C_i}). \quad (37)$$

Para se estimar a classe de *B*, calcula-se então $P(B|C^i)\,P(C^i)$ para cada classe C^i e escolhe-se a classe com maior probabilidade, ou seja onde $P(B|C^i)\,P(C^i)$ é máximo,

$$P(B|C^i)\,P(C^i) > P(B|C^j)\,P(C^j), \forall\, 1 \leq j \leq m, j \neq i. \quad (38)$$

Considerando novamente o exemplo, para se saber se uma dada embarcação que foi fiscalizada há 6 anos e é um arrastão está legal $B_6(6, ARRASTÃO, Legal)$ ou em infração $B_6(6, ARRASTÃO, Infrator)$. Considere-se X^1 o atributo tempo, variável contínua e X^2 o atributo tipo de embarcação, variável categórica. Assim, para $B_6(6, ARRASTÃO, Infrator)$ no atributo X^1, temos:

NÃO CLASSIFICADO

$$\mu^{C^1} = \frac{(2+5)}{2} = \frac{7}{2}, \sigma^{C^1} = \sqrt{\frac{(2-3.5)^2 + (5-3.5)^2}{2-1}} = \sqrt{2(\frac{3}{2})^2} = \sqrt{\frac{9}{2}} \cong 2.12 \quad (39)$$

$$P(6|C^1) = g\left(6, \mu_{C_1}, \sigma_{C_1}\right) = \frac{1}{\sqrt{2\pi}\sqrt{\frac{9}{2}}} e^{-\frac{(6-3.5)^2}{2\times\sqrt{\frac{9}{2}}^2}} \cong 0.094$$

De forma semelhante para $B_6(6, ARRASTÃO, Legal)$,

$$\mu^{C^2} = \frac{(1+3+4)}{3} = \frac{8}{3}, \sigma^{C^2} = \sqrt{\frac{\left(1-\frac{8}{3}\right)^2 + \left(3-\frac{8}{3}\right)^2 + \left(4-\frac{8}{3}\right)^2}{3-1}} \quad (40)$$

$$= \sqrt{\frac{\left(-\frac{5}{3}\right)^2 + \left(\frac{1}{3}\right)^2 + \left(\frac{4}{3}\right)^2}{3-1}} = \sqrt{\frac{\frac{14}{3}}{2}} = \sqrt{\frac{7}{3}} \cong 1.528$$

$$P(6|C^2) = g\left(6, \mu_{C_2}, \sigma_{C_2}\right) = \frac{1}{\sqrt{2\pi}\sqrt{\frac{7}{3}}} e^{-\frac{\left(6-\frac{8}{3}\right)^2}{2\times\sqrt{\frac{7}{3}}^2}} \cong 0.0241,$$

Para X^2, temos:

$$P(ARRASTÃO|C^1) = 1, P(ARRASTÃO|C^2) = 0. \quad (41)$$

Desta forma,

$$P(B|C^1) = P(6|C^1) \times P(ARRASTÃO|C^1) \cong 0.094 \quad (42)$$

$$P(B|C^2) = P(6|C^2) \times P(ARRASTÃO|C^2) \cong 0$$

Logo,

$$P(B|C^1) \times P(C^1) = 0.094 \times \frac{2}{5} = 0.0376 \quad (43)$$

$$P(B|C^2) \times P(C^2) = 0 .$$

NÃO CLASSIFICADO

Assim sendo, pelo *naive de Bayes*, a embarcação *Neptuno* seria classificada como "Infrator", pois $P(B|C^1)\,P(C^1) > P(B|C^2)\,P(C^2)$.

Para conferir os resultados, introduziram-se os dados do problema no *software* WEKA. A Figura 18 mostra que os resultados obtidos coincidem.

```
Naive Bayes Classifier

                    Class
Attribute           Legal Infrator
                   (0.57)   (0.43)
===================================
1
  mean              2.6667      3.5
  std. dev.         1.2472      1.5
  weight sum             3        2
  precision              1        1

2
  CERCADORA            2.0      1.0
  ARRASTÃO             1.0      3.0
  PALANGRE             2.0      1.0
  POLIVALENTE          2.0      1.0
  [total]              7.0      6.0


Time taken to build model: 0 seconds

=== Predictions on test set ===

inst#,actual,predicted,error,prediction
1,1:?,2:Infrator,,0.945
```

Figura 18 – Resultados obtidos no WEKA para o exemplo considerado.

Em teoria estes classificadores têm a menor taxa de erro quando comparados com todos os outros classificadores (Lewis, 1998, p. 1). Contudo, na prática nem sempre é assim, devido às suposições que são feitas para possibilitar o seu uso, como por exemplo a independência de atributos e assumir que os atributos têm uma distribuição gaussiana. Estes classificadores são ainda bastante uteis para ajudar a entender outros classificadores, por exemplo, sob algumas suposições, pode mostrar que muitas redes neuronais originam resultados baseados nas máximas probabilidades condicionadas, à semelhança do *naive de* Bayes (Amor, Benferhat, & Elouedi, 2004, pp. 420-421).

4.3.3. Árvores de Decisão

Uma árvore de decisão, é um dos métodos lógicos mais utilizados, consiste num fluxograma com estrutura em árvore, onde cada nó representa uma opção para um dado atributo, cada ramo a opção escolhida e onde no final existe um nó que atribui uma classe

NÃO CLASSIFICADO

às opções que foram sendo tomadas (Han et al., 2011, p. 330). A Figura 19 é um exemplo de árvore de decisão, utilizada para identificar um navio.

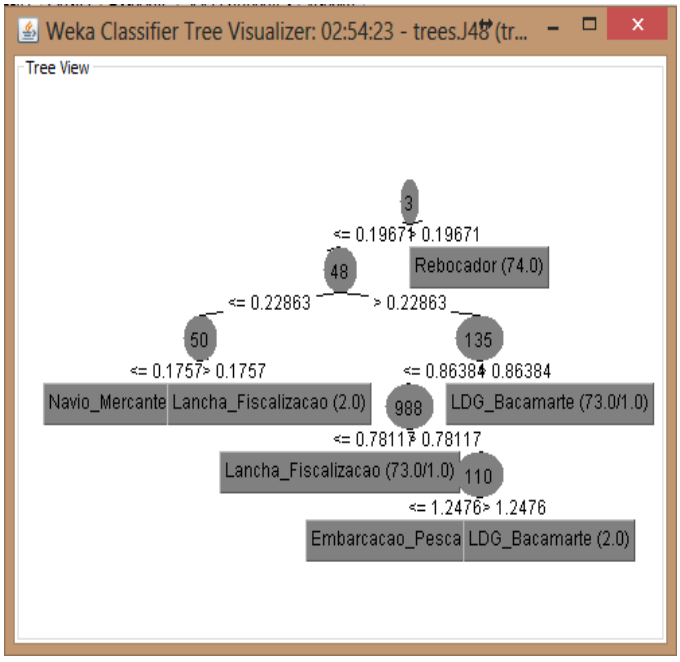

Figura 19 - Exemplo de árvore de decisão, obtida no software WEKA.

Por norma, na criação de árvores de decisão visa-se reduzir ao máximo o número de nós necessários. No caso da Figura 19, a título de exemplo, para saber se determinado navio é um navio mercante, bastava ter como regra "ser igual ou inferior a 0.1757 para o atributo 50".

Para se obter uma dada classificação para uma instância *B*, basta observar os valores dos atributos e compará-los com os nós de decisão, até se chegar a uma dada classificação. As árvores de decisão podem facilmente ser convertidas em regras de classificação (Freitas, 2013, pp. 45-46).

Em relação às principais vantagens, são bastante fáceis de se construir e apropriadas para análise exploratória de conhecimento. Estas suportam dados multidimensionais e a sua representação é intuitiva e fácil de assimilar. Os processos de aprendizagem e classificação são simples e rápidos. Geralmente têm um bom nível de eficácia, contudo dependendo dos dados e da sua sensibilidade a perturbações no conjunto de treino podem ter maiores ou menores taxas de erros (J. Ross Quinlan, 1986, pp. 82-83).

NÃO CLASSIFICADO

De entre os algoritmos que podem ser utilizados destacam-se o *Iterative Dichotomiser* (ID3) e o C4.5. O algoritmo ID3 é um algoritmo utilizado na criação de árvores de decisão, onde inicialmente se seleciona um atributo do conjunto de treino que permite dividi-lo em duas partes. De seguida cria-se um ramo para cada valor do atributo escolhido e distribuem-se as instâncias correspondentes. O algoritmo é aplicado sucessivamente até todas as instâncias num nó terem apenas uma classificação. (Kantardzic, 2011, p. 171).

Na construção de árvores de decisão, como foi referido, o objetivo consiste sempre em construir uma árvore de decisão com o menor número de nós possível, para tal é necessário selecionar sempre o atributo que tenha maior ganho de informação, ou seja, o que minimiza a informação necessária na subárvore, para classificar a instância (Freitas, 2013, pp. 47-49).

Existe ainda uma extensão do ID3, chamada algoritmo C4.5 (também conhecida por J48), que para além da classificação de atributos categóricos, permite também a classificação de atributos numéricos. Este método favorece atributos que dividam os dados em subconjuntos com uma baixa classe de entropia, isto é, com a maioria das instâncias pertencentes apenas a uma classe. Desta forma, o algoritmo escolhe os atributos que possibilitam maior grau de descriminação entre as classes (J Ross Quinlan, 2014, p. 25).

4.3.4. Avaliação do Desempenho do Classificador

Para se poder determinar qual é efetivamente o melhor método a utilizar para cada situação, é necessário comparar os classificadores. A melhor forma de o conseguir, é testando-os. Existem basicamente três formas simples de o fazer: através do método *holdout*, o método de validação cruzada e o método de *bootstrap* (Leite, 2007, pp. 53-56).

- **Método *Holdhout***

Neste método dividem-se os dados em dois conjuntos, o conjunto de treino e o conjunto de teste. Ao fazer-se esta divisão é necessário ter em conta que a escolha da percentagem destinada a treino afetará a de teste e vice-versa, isto é, se selecionar uma grande percentagem para treino, o classificador à partida será mais fiável, contudo não

NÃO CLASSIFICADO

podemos garantir que será infalível, uma vez que foram utilizadas poucas instâncias para teste. Se por outro lado optarmos por escolher um conjunto de teste demasiado grande, a taxa de acerto será fiável, contudo muito menor, pois o classificador criado utilizou poucas instâncias para treino. Assim sendo, este método deve-se utilizar com uma percentagem de conjunto de treino a variar entre os 70% e os 80% (Gupta, 2011, p. 136).

- **Método da Validação Cruzada**

Neste método dividem-se os dados em n amostras. De seguida treinam-se os dados com as amostras a_2, a_3 ... a_n, testam-se com a amostra a_1, é registada a taxa de erro e junta-se a amostra a_1 ao conjunto de treino. Após a amostra a_1 ser colocada novamente no conjunto de treino, retira-se a amostra a_2, testam-se os dados com esta, regista-se novamente a taxa de erro e volta-se a juntar a amostra a_2 no conjunto de treino. Este processo é repetido até todas as amostras serem testadas e no final calcula-se a média aritmética de cada taxa de erro (nenhum dos classificadores é melhor que o outro). Assim a taxa de erro deste método pode ser definida por:

$$\overline{Erro} = \frac{e_1 + e_2 + \ldots + e_n}{n} = \frac{1}{n}\sum_{i=1}^{n} e_i \,. \tag{44}$$

O método de validação cruzada utiliza-se particularmente para conjuntos de dados muito pequenos, em que não se pode utilizar o método de treino-teste, podendo no limite o número de n amostras ser igual ao número de i instâncias (Freitas, 2013, pp. 86-87). Este método é ainda eficaz para comparar as taxas de erro dos classificadores, pois tem-se a certeza que os dados utilizados são exatamente os mesmos e que a taxa de erro é fiável (dado o número de dados utilizados para teste).

- **Método de *Bootstrap* ou Resubstituição**

Existe ainda um método semelhante a este, conhecido por *bootstrap*, utilizado geralmente para conjuntos de dados pequenos, em que em vez de se irem retirando amostras de dados para teste, copia-se o conjunto de treino completo e testa-se o conjunto de treino com ele próprio (Gupta, 2011, p. 137). Trivialmente se conclui que ao se testarem os dados com eles próprios se obtêm estimativas otimistas, contudo existem

NÃO CLASSIFICADO

várias técnicas estatísticas para se obterem estimativas de erro realistas (Efron & Tibshirani, 1994, p. 6).

NÃO CLASSIFICADO

Capítulo 5 - Descrição e pré-processamento dos dados

5.1. Descrição

Dias 09 a 12, 16 e 17 de julho de 2013, no âmbito do exercício REP13, com o objetivo de testar novos sensores e plataformas, aplicados em missões de reconhecimento GEOINTEL e em levantamento das condições ambientais (METOC) num teatro de operações costeiras e estuarinas, foram efetuadas gravações do ruído submarino com um hidrofone SR-1 (*Marsensing*). O hidrofone foi colocado à entrada da barra de Setúbal, na posição $\varphi = 38°27'40.5''N\ , L = 008°57'28.1''W$, a uma profundidade de cerca de 11m. Cada ficheiro de som foi registado com uma duração de cinco minutos. Para o mesmo período, uma máquina fotográfica registou todos os contactos que transitavam sobre o local de fundeamento, com uma cadência de disparos de dois em dois minutos.

Numa primeira fase, procedeu-se à seleção dos dados. Foram excluídos todos os ficheiros de imagem que não tinham embarcações ou navios que pudessem ser corretamente identificados, por motivos de distância, tamanho ou condições visibilidade e de luminosidade insuficientes. Numa segunda fase, foram associados os ficheiros de imagem aos do som através da correspondência dos seus grupos data-hora. Foram excluídos todos os ficheiros de som em que o hidrofone não teve estabilizado na posição. Com os ficheiros selecionados (som e fotografia) foi construída uma base de dados. Nela os ficheiros tomam o nome do número do contacto na sequência analisada, associando os *links* dos ficheiros de imagem e de som correspondentes, os grupos data-hora de início e fim, o tipo de navio e outras observações a ter conta na análise (ex. existência de mais do que um navio na área). Como o número de ficheiros para cada navio não é representativo, optou-se por dividir cada série em diferentes blocos, sobre os quais se calculou o respetivo espectro, cuja média aumentou a significância estatística da análise.

A Tabela 1 representa os tipos de navios ou embarcações considerados na seleção dos dados, bem como o número de ficheiros de som obtidos para cada tipo.

NÃO CLASSIFICADO

Embarcações/Navios	Nº de elementos distintos	Nº de ficheiros de som
Embarcações à vela	10	33
Embarcações de pesca	16	71
Embarcações de piloto	1	27
Embarcações de recreio	9	26
Lanchas	5	62
Navios mercantes	20	57
Reabastecedor	1	3
Rebocadores	4	7
Total	66	286

Tabela 1 - Ficheiros de som selecionados para cada tipo de navio definido.

5.2. Pré-processamento

Para a criação da rotina de pré-processamento dos dados recorreu-se ao *MATrix LABoratory* (MATLAB), um dos sistemas para cálculo científico mais utilizados na análise de dados acústicos (Vieira, 2003, p. 4).

Foram selecionados ficheiros de som de navios que não estivessem sujeitos à influência de outro navio nas proximidades. A cada ficheiro de som corresponde uma série temporal $x(t)$. O sinal $x(t)$ tem as seguintes características:

$\boldsymbol{Tamanho\ da\ série}$	$\mathbf{30468096}\ \boldsymbol{pontos}$
$\boldsymbol{Frequência\ de\ amostragem\ (fa)}$	$101562\ Hz$
$\boldsymbol{Duração}$	$4{,}9999\ minutos$

Tabela 2 - Características do sinal x(t).

NÃO CLASSIFICADO

Foram identificados dois problemas fundamentais durante a fase inicial do pré-processamento dos sinais acústicos:

• O primeiro prende-se com a dimensão da amostra. A frequência de amostragem elevada, aliada ao tamanho atual do sinal, faz com que seja demasiado exigente computacionalmente processar e armazenar a informação. Assim sendo, decompôs-se a série original em blocos de 1s, uma fez que as frequências mais significativas contidas em sinais acústicos submarinos de origem antropogénica situam-se entre 1Hz e 1 kHz (Jesus, Silva, & Zabel, 2005, p. 9). Por esta mesma razão foram consideradas na análise espectral as primeiras 1024 frequências que cobrem esta gama mais significativa.

• O segundo diz respeito ao ruído indesejado associado ao sinal recebido pelo hidrofone. Para tal, calcularam-se espectros médios de oito blocos sucessivos de 1 s, sobrepostos em 50% da sua duração, utilizando uma análise FFT. Estes foram depois normalizados pela energia média dos oito. Ao todo cada ficheiro de som deu origem a 74 espectros correspondentes a 74 instâncias diferentes ao longo dos 5 minutos de observação. A Figura 20 mostra o exemplo dos espectros médios das 74 instâncias de 3 registos distintos. Foram utilizados dois ficheiros de som seguidos no tempo, da lancha de fiscalização Auriga (um para introduzir no conjunto de treino e outro para testar) e um ficheiro de som da lancha de desembarque grande Bacamarte.

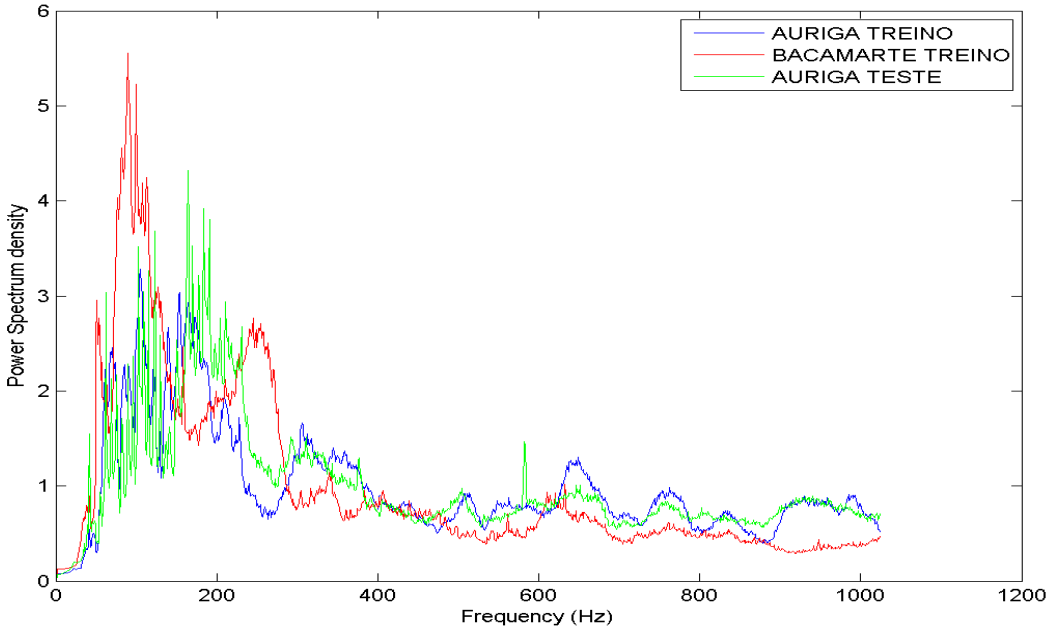

Figura 20 - Espectros Normalizados de dois ficheiros de som da lancha hidrográfica NRP Auriga e um da lancha de desembarque grande NRP Bacamarte.

NÃO CLASSIFICADO

Numa última fase do pré-processamento, exportam-se os ficheiros obtidos para o formato *Comma-separated values* (CSV), para que possam ser lidos pelo WEKA.

O *software Waikato Environment for Knowledge Analysis* (WEKA), é um sistema de referência no que diz respeito a *data mining* e *machine learning*. Disponibiliza algoritmos *de machine learning* e ferramentas de pré-processamento dos dados, permitindo comparar com facilidade diferentes métodos em novos conjuntos de dados (Hall et al., 2009, p. 1).

Dos dados obtidos, retira-se que o número de atributos será sempre constante e igual a 1024 (uma vez que foi o limite em frequência considerado para o espectro), enquanto o número de instâncias depende do número de navios que se introduzirem nos ficheiros de teste e de treino, tendo cada navio associadas 74 instâncias. A cada instância faz-se corresponder uma classe, neste caso, o nome do navio a que pertence.

Cada atributo está portanto associado a uma frequência do espectro e cada instância corresponde à variação da amplitude em função da frequência. Cada instância não é mais do que uma linha horizontal do espectro.

É com base nesta informação que os classificadores utilizados no WEKA vão comparar e classificar os dados.

A Figura 21 esquematiza de forma geral o pré-processamento dos sinais antes de serem exportados para o *software* de classificação.

NÃO CLASSIFICADO

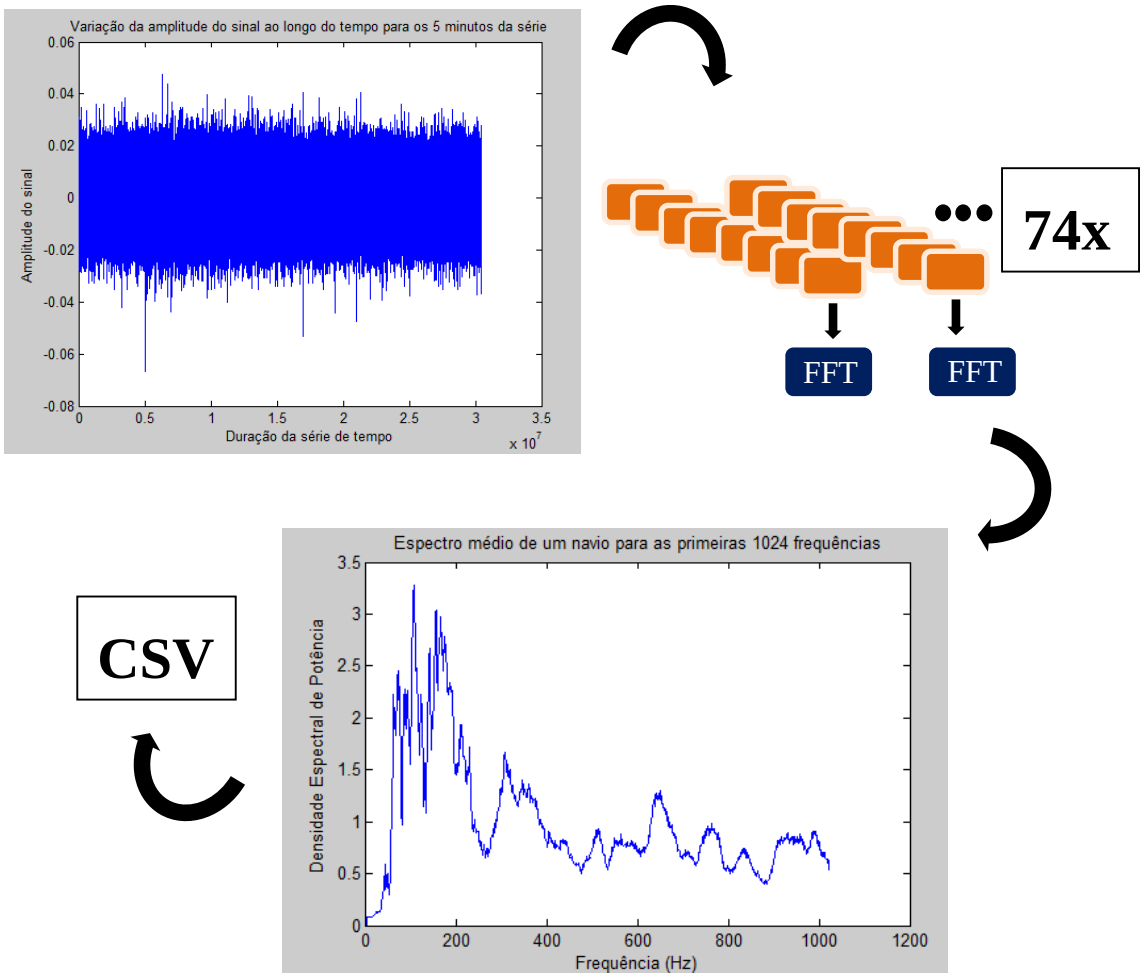

Figura 21 - Esquema do pré-processamento dos sinais acústicos.

NÃO CLASSIFICADO

Capítulo 6 - Testes e Resultados

Estando já definido um método para pré-processar os sinais, é então necessário classificá-los, para que se possam distinguir uns dos outros.

Neste sentido, utilizou-se o software WEKA para testar se, recorrendo a diferentes classificadores, é possível reconhecer o sinal de um novo ficheiro de um navio, que tenha sido previamente identificado e classificado.

6.1. Testes com dois navios

Numa primeira análise introduziu-se um ficheiro CSV com a informação de dois navios como conjunto de treino no WEKA, para treinar os classificadores e adicionou-se um ficheiro de um dos navios num período diferente, como conjunto de teste. O ficheiro de treino contém um total de 148 instâncias, 74 das quais pertencentes ao espectro da lancha hidrográfica NRP Auriga e os restantes 74 respeitantes à lancha de desembarque grande Bacamarte. Para testar o conjunto de treino, utilizou-se 20% do conjunto de treino para teste e o método de validação cruzada, para os classificadores k-vizinhos, *naive de Bayes* e árvore de decisão J48.

Por fim, utilizaram-se os vários classificadores para verificar se um terceiro navio, neste caso outro espectro da lancha Auriga, na sequência do primeiro, era ou não corretamente classificado.

6.1.1. Método *Holdhout*

Começou-se por testar os diferentes classificadores recorrendo ao método *holdhout*, utilizando 20% do conjunto para teste.

- **K-vizinhos**

Com o classificador k-vizinhos, obtiveram-se os seguintes resultados:

NÃO CLASSIFICADO

k=1 a k=11	Número total de intâncias classificadas - 30
	Percentagem de instâncias corretamente classificadas - 100%
	Matriz de confusão: 20 0 - 20 Verdadeiros Positivos, Auriaga 0 10 - 10 Verdadeiros Negativos, Bacamarte

Tabela 3 - Classificador k-vizinhos com 20% de conjunto de teste.

Nos casos considerados o classificador teve sempre 100% de sucesso na classificação dos dados de teste.

- **Naive de Bayes**

Recorrendo-se ao classificador *naive de Bayes* para o conjunto de treino, reservando 20% do conjunto para teste, obteve-se:

20% Teste	Número total de intâncias classificadas - 30
	Percentagem de instâncias corretamente classificadas - 100%
	Matriz de confusão: 20 0 - 20 Verdadeiros Positivos, Auriga 0 10 - 10 Verdadeiros Negativos, Bacamarte

Tabela 4 - Classificador naive de Bayes com 20% de conjunto de teste.

Como se pode observar, todas as instâncias foram corretamente classificadas.

NÃO CLASSIFICADO

- **Árvore de decisão J48´**

Reservando 20% do conjunto de treino para teste com este classificador, obteve-se:

20% Teste	Número total de intâncias classificadas - 30
	Percentagem de instâncias corretamente classificadas - 100%
	Matriz de confusão: 20 0 - 20 Verdadeiros Positivos, Auriga 0 10 - 10 Verdadeiros Negativos, Bacamarte

Tabela 5 - Classificador árvore de decisão J48 com 20% de conjunto de teste.

Obteve-se 100% de sucesso na classificação das instâncias.

6.1.2. Método Validação Cruzada

Recorrendo à validação cruzada com 10 partições, em que 9 são utilizadas para treino e a 10ª para teste, percorrendo todas as partições, testaram-se os diferentes classificadores.

- **K-vizinhos**

Utilizando10 partições, em que 9 são utilizadas para treino e a 10ª para teste, obteve-se:

NÃO CLASSIFICADO

k=1 a k=11	Número total de intâncias classificadas - 148
	Percentagem de instâncias corretamente classificadas - 100%
	Matriz de confusão: 74 0 - 74 Verdadeiros Positivos, Auriga 0 74 - 74 Verdadeiros Negativos, Bacamarte

Tabela 6 - Classificador k-vizinhos com validação cruzada.

- **Naive de Bayes**

Utilizando 10 partições, em que 9 são utilizadas para treino e a 10ª para teste, para este classificador obteve-se:

Validação Cruzada	Número total de intâncias classificadas - 148
	Percentagem de instâncias corretamente classificadas - 100%
	Matriz de confusão: 74 0 - 74 Verdadeiros Positivos, Auriga 0 74 - 74 Verdadeiros Negativos, Bacamarte

Tabela 7 - Classificador naive de Bayes com validação cruzada.

Neste caso, obteve-se igualmente 100% de instâncias bem classificadas.

- **Árvore de decisão J48**

Recorrendo à validação cruzada com os mesmos critérios, obteve-se:

NÃO CLASSIFICADO

Validação Cruzada	Número total de intâncias classificadas - 148	
	Percentagem de instâncias corretamente classificadas - 99.3243 % Percentagem de instâncias incorretamente classificadas - 0.6757 %	
	Matriz de confusão:	
	73 1	- 73 Verdadeiros Positivos, Auriga
	0 74	- 74 Verdadeiros Negativos, Bacamarte
		- 1 Falso Negativo, Bacamarte

Tabela 8 - Classificador árvore de decisão J48 com validação cruzada.

Obteve-se uma elevada taxa de sucesso na classificação das instâncias.

6.1.3. Introdução de um novo ficheiro para teste

Nesta fase, adicionou-se um ficheiro da lancha Auriga, com 74 instâncias, na sequência temporal da primeira gravação, para testar os classificadores e analisar se o 'novo' navio é ou não identificado positivamente como sendo a Auriga.

- **K-vizinhos**

Introduzindo o ficheiro de teste para classificação, que corresponde a outra gravação da Auriga, obteve-se:

NÃO CLASSIFICADO

k=1	Número total de intâncias classificadas - 74
	Percentagem de instâncias corretamente classificadas - 79.7297% Percentagem de instâncias incorretamente classificadas - 20.2703%
	Matriz de confusão: 59 15 - 59 Verdadeiros Positivos, Auriga 0 0 - 15 Falsos Negativos, Bacamarte

Tabela 9 - Classificador k-vizinhos com a introdução de um novo ficheiro para teste, com k=1.

Na Figura 22 pode-se observar os erros de classificação obtidos. Os quadrados a azul representam os dados que foram erradamente classificados como pertencentes à LDG Bacamarte, enquanto as cruzes a azul representam os dados que foram corretamente classificados como pertencentes à LH Auriga.

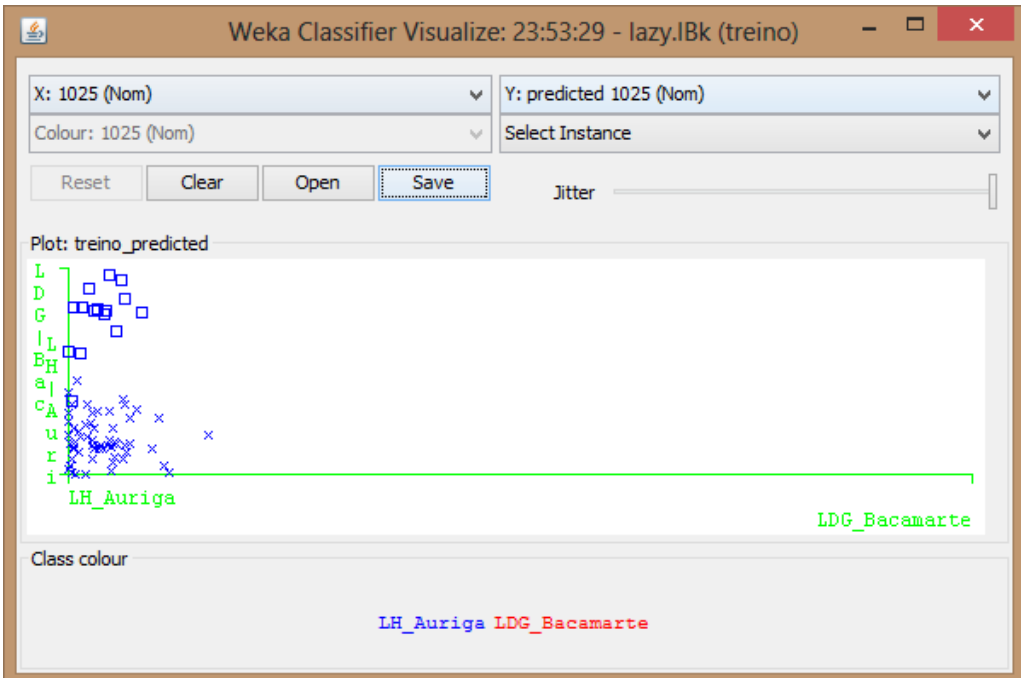

Figura 22 - Erros de classificação do método k-vizinhos com introdução de um ficheiro de teste, com k=1.

Como não se obteve 100% de sucesso na classificação do ficheiro de teste, testaram-se vários valores de k, com as seguintes percentagens de instâncias corretamente classificadas:

NÃO CLASSIFICADO

Valor de K	% Corr. Class.
K=3	79.7297 %
K=4	87.8378 %
K=5	79.7297 %
K=6	89.1892 %
K=7	81.0811 %
K=8	85.1351 %
K=9	83.7838 %
K=10	83.7838 %
K=11	79.7297 %
K=20	82.4324 %
K=21	82.4324 %
K=30	86.4865 %
K=31	86.4865 %
K=50	94.5946 %
K=51	94.5946 %
K=100	95.9459 %
K=110	98.6486 %
K=120	98.6486 %
K=130	98.6486 %
K=135	98.6486 %
K=136	100%

Tabela 10 - Classificador k-vizinhos com a introdução de um novo ficheiro para teste, para vários valores de k.

Constata-se que o k é ótimo para k=136. A tendência para a percentagem de instâncias corretamente classificadas aumentar com o aumento de k revela que pode haver algum ruído para valores de k mais próximos. De qualquer forma, para os vários k, as percentagens de acerto obtidas são bastante aceitáveis.

- ***Naive de Bayes***

Introduzindo o ficheiro de teste para classificação, obtém-se:

NÃO CLASSIFICADO

Navio Teste	Número total de intâncias classificadas - 74
	Percentagem de instâncias corretamente classificadas - 48.6486 % Percentagem de instâncias incorretamente classificadas - 51.3514 %
	Matriz de confusão: 36 38 - 36 Verdadeiros Positivos, Auriga 0 0 - 38 Falsos Negativos, Bacamarte

Tabela 11 – Classificador naive de Bayes com a introdução de um novo ficheiro para teste.

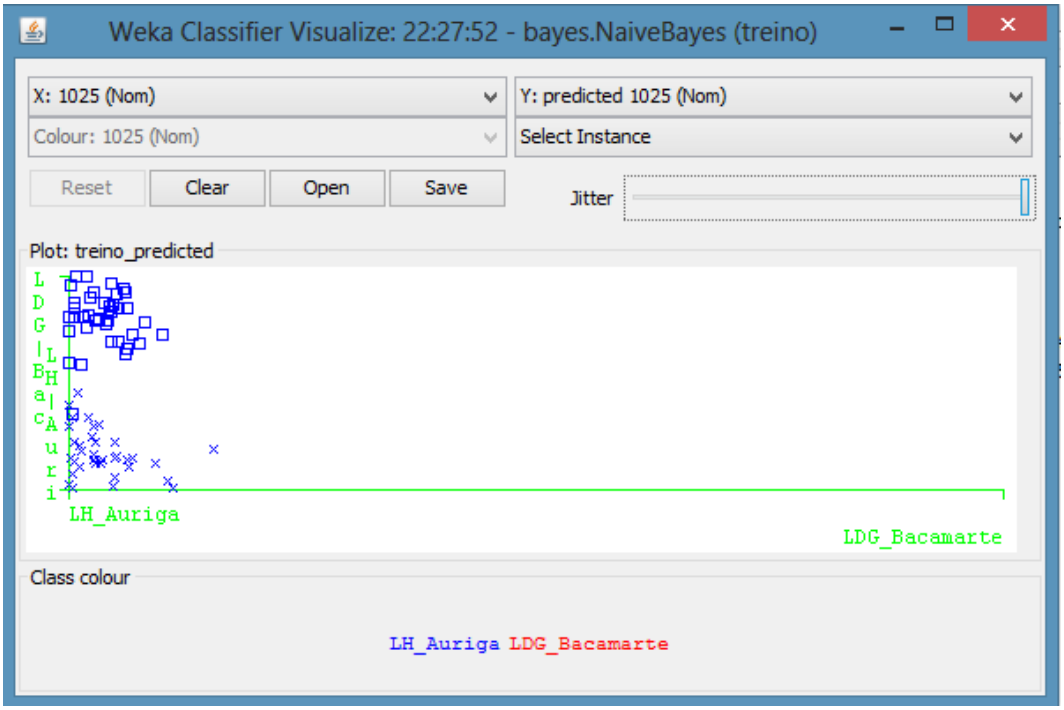

Figura 23 - Erros de Classificação naive de Bayes

Dos resultados obtidos, pode-se constatar que o *naive de Bayes* é um mau classificador para o caso em questão. Pela observação do gráfico de erros de classificação, Figura 23, verifica-se que dos 74 dados, 38 estão mal classificados e apenas 36 se encontram bem classificados.

- **Árvore de decisão J48**

Introduzindo o ficheiro de teste para classificação, obtém-se:

NÃO CLASSIFICADO

Navio Teste	Número total de intâncias classificadas - 74
	Percentagem de instâncias corretamente classificadas - 86.4865 % Percentagem de instâncias incorretamente classificadas - 13.5135 %
	Matriz de confusão: 64 10 - 64 Verdadeiros Positivos, Auriga 0 0 - 10 Falsos Negativos, Bacamarte

Tabela 12 - Classificador árvore de decisão J48 com a introdução de um novo ficheiro para teste.

Apesar de não se obter um classificador com 100% de sucesso, tem uma elevada taxa de instâncias corretamente classificadas, recorrendo ao critério ilustrado na Figura 24. Como se pode observar, o classificador detetou um valor barreira no atributo 52.

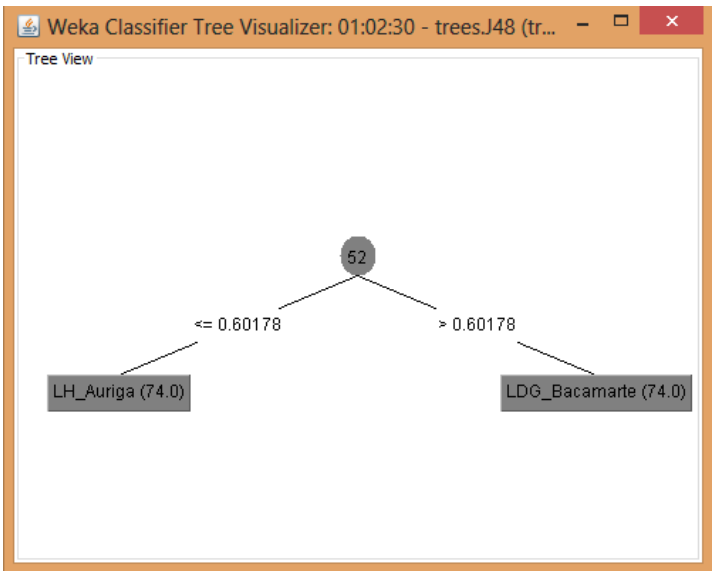

Figura 24 - Critério de decisão da árvore de decisão J48

NÃO CLASSIFICADO

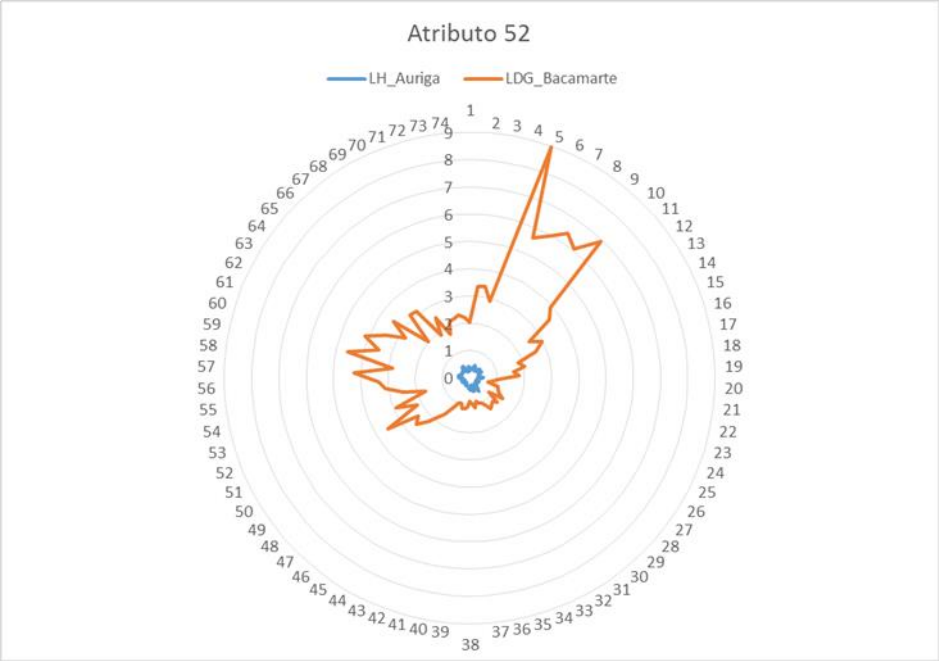

Figura 25 - Radar plot da amplitude do atributo 52 em cada uma das gravações da LH Auriga e da LDG Bacamarte.

O gráfico radar representado na Figura 25 demonstra como facilmente se distinguem os dois navios em relação ao atributo representado.

6.2. Testes com cinco navios

Numa segunda fase de teste, criaram-se mais uma vez dois ficheiros CSV. Um correspondente ao conjunto de treino, com informação relativa a cinco navios, uma lancha de fiscalização da classe Centauro, um navio mercante, uma embarcação de pesca, um rebocador e a lancha de desembarque grande Bacamarte e outro simplesmente com a lancha de desembarque, num período distinto, para ser utilizado como conjunto de teste. O ficheiro com instâncias de treino contém um total de 370 instâncias, 74 para cada uma das classes referidas anteriormente, enquanto o ficheiro do navio a testar contém 74 instâncias obtidas do espectro da Bacamarte. Utilizaram-se os mesmos classificadores e os mesmos métodos de teste para avaliar a sua capacidade de distinguir e classificar corretamente os diferentes navios do conjunto de treino e em último lugar identificar de forma assertiva o navio ‘desconhecido’.

6.2.1. Método *Holdhout*

Utilizou-se novamente este método para testar o conjunto de treino, com 20% do conjunto para teste.

NÃO CLASSIFICADO

- **K-vizinhos**

Reservando 20% do conjunto de treino para teste, obtiveram-se os seguintes resultados:

k=1 e k=2	Número total de intâncias classificadas - 74
	Percentagem de instâncias corretamente classificadas - 100%
	Matriz de confusão: a b c d e <-- classified as 7 0 0 0 0 | a = Lancha_Fiscalizacao 0 16 0 0 0 | b = Navio_Mercante 0 0 22 0 0 | c = Embarcacao_Pesca 0 0 0 17 0 | d = Rebocador 0 0 0 0 12 | e = LDG_Bacamarte

Tabela 13 - Teste 2: Classificador k-vizinhos com 20% de conjunto de teste, para k=1 e k=2.

Para k=1 e k=2, da matriz de confusão tem-se que todos os dados foram classificados positivamente dentro de cada classe.

Para maiores valores de k, obteve-se a seguinte percentagem de instâncias corretamente classificadas:

Valor de K	**% Corr. Class.**
K=3	98.6486 %
K=4	100 %
K=5	98.6486 %
K=6	98.6486 %
K=11	97.2973 %
K=30	89.1892 %
K=31	83.7838%
K=51	67.5676 %

NÃO CLASSIFICADO

K=100	50%
K=101	48.6486 %
K=131	48.6486 %
K=201	25.6757 %

Tabela 14 - Teste 2: Classificador k-vizinhos com 20% teste, para maiores valores de k.

Obtêm-se melhores resultados para menores valores de k, isto porque integram um domínio cada vez mais abrangente.

- ***Naive de Bayes***

Recorrendo-se ao classificador *naive de Bayes* para o conjunto de treino, reservando 20% do conjunto para teste, obteve-se:

20% Teste	Número total de intâncias classificadas - 74
	Percentagem de instâncias corretamente classificadas - 100%
	Matriz de confusão: a b c d e <-- classified as 7 0 0 0 0 | a = Lancha_Fiscalizacao 0 16 0 0 0 | b = Navio_Mercante 0 0 22 0 0 | c = Embarcacao_Pesca 0 0 0 17 0 | d = Rebocador 0 0 0 0 12 | e = LDG_Bacamarte

Tabela 15 - Teste 2: Classificador naive de Bayes com 20% teste.

Pela Tabela 15, obteve-se 100% das instâncias corretamente classificadas.

- **Árvore de decisão J48**

Reservando 20% do conjunto de treino para teste com este classificador, obteve-se:

NÃO CLASSIFICADO

20% Teste	Número total de intâncias classificadas - 74
	Percentagem de instâncias corretamente classificadas - 94.5946 % Percentagem de instâncias incorretamente classificadas - 5.4054 %
	Matriz de confusão: a b c d e <-- classified as 7 0 0 0 0 | a = Lancha_Fiscalizacao 0 14 0 0 2 | b = Navio_Mercante 0 0 20 0 2 | c = Embarcacao_Pesca 0 0 0 17 0 | d = Rebocador 0 0 0 0 12 | e = LDG_Bacamarte

Tabela 16 - Teste 2: Classificador árvore de decisão J48 com 20% teste.

Pela Tabela 16, verifica-se que 70 das 74 instâncias foram corretamente identificadas. A Figura 26 ilustra a classificação das instâncias para este caso, em que os quadrados são as instâncias mal classificadas.

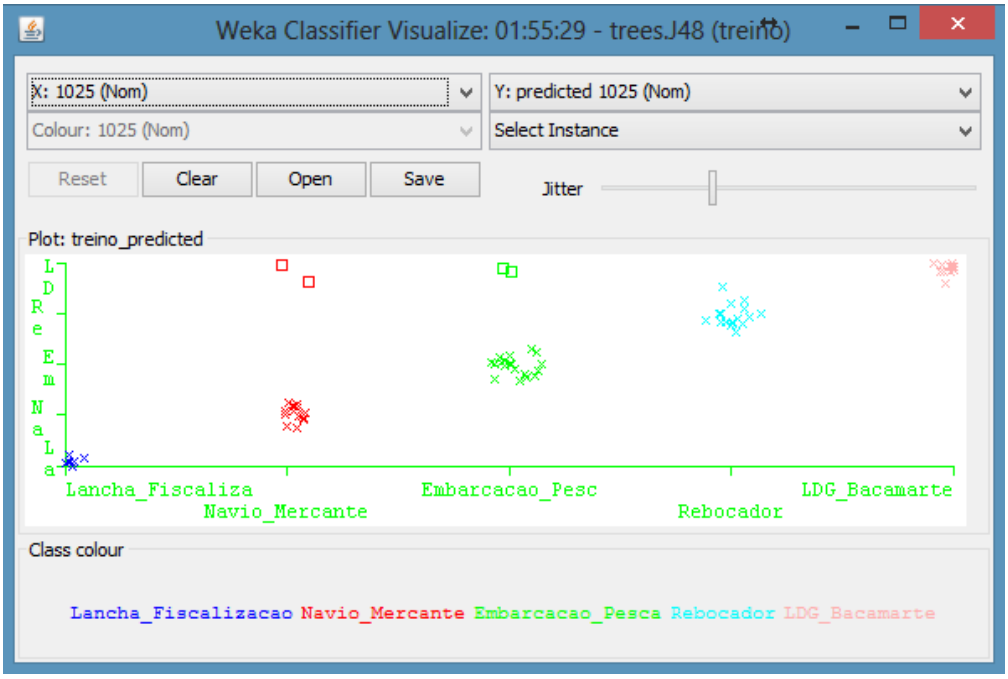

Figura 26 - Teste 2: Erros de classificação do método árvore de decisão J48

NÃO CLASSIFICADO

6.2.2. Método Validação Cruzada

Recorrendo à validação cruzada, com os mesmos critérios utilizados para o primeiro teste, testaram-se novamente os classificadores.

- **K-vizinhos**

Aplicando o método de validação cruzada, obteve-se:

k=1 e k=2	Número total de intâncias classificadas - 370
	Percentagem de instâncias corretamente classificadas - 100%
	Matriz de confusão: a b c d e <-- classified as 74 0 0 0 0 | a = Lancha_Fiscalizacao 0 74 0 0 0 | b = Navio_Mercante 0 0 74 0 0 | c = Embarcacao_Pesca 0 0 0 74 0 | d = Rebocador 0 0 0 0 74 | e = LDG_Bacamarte

Tabela 17 - Teste 2:- Classificador k-vizinhos com validação cruzada.

Como se pode observar pela Tabela 17, recorrendo à validação cruzada para k=1 e k=2, obtém-se 100% de sucesso na classificação dos dados.

Para maiores valores de k, obteve-se a seguinte percentagem de instâncias corretamente classificadas:

Valor de K	**% Corr. Class.**
K=3	99.7297 %
K=11	99.1892 %
K=31	93.7838 %
K=50	87.5676 %
K=51	87.2973%
K=100	72.1622 %

NÃO CLASSIFICADO

K=101	72.1622%
K=131	66.7568 %
K=201	76.7568 %
K=301	23.2432 %
K=333	18.9189 %

Tabela 18 - Teste 2: Classificador k-vizinhos com validação cruzada.

Como se pode observar pela Tabela 18 a maior percentagem de instâncias corretamente classificadas verifica-se, mais uma vez, para menores valores de k.

- ***Naive de Bayes***

Recorrendo à validação cruzada com 10 partições, em que 9 são utilizadas para treino e a 10ª para teste, percorrendo todas as partições, obteve-se 100% de sucesso na classificação como demonstra a Tabela 19:

Validação Cruzada	Número total de intâncias classificadas - 370
	Percentagem de instâncias corretamente classificadas - 100%
	Matriz de confusão: a b c d e <-- classified as 74 0 0 0 0 | a = Lancha_Fiscalizacao 0 74 0 0 0 | b = Navio_Mercante 0 0 74 0 0 | c = Embarcacao_Pesca 0 0 0 74 0 | d = Rebocador 0 0 0 0 74 | e = LDG_Bacamarte

Tabela 19 - Teste 2: Classificador naive de Bayes com validação cruzada.

NÃO CLASSIFICADO

- **Árvore de decisão J48**

Recorrendo à validação cruzada, segundo os mesmos critérios, obteve-se:

Validação Cruzada	Número total de intâncias classificadas - 370
	Percentagem de instâncias corretamente classificadas - 93.7838 % Percentagem de instâncias incorretamente classificadas - 6.2162 %
	Matriz de confusão: a b c d e <-- classified as 69 1 2 0 2 | a = Lancha_Fiscalizacao 1 71 1 0 1 | b = Navio_Mercante 2 4 64 0 4 | c = Embarcacao_Pesca 0 0 0 74 0 | d = Rebocador 2 0 2 1 69 | e = LDG_Bacamarte

Tabela 20 - Teste 2: Classificador árvore de decisão J48 com validação cruzada.

Por observação da Tabela 20 conclui-se que as instâncias na sua maioria são bem identificadas.

6.2.3. Introdução de um novo ficheiro para teste

À semelhança do teste anterior, introduziu-se um novo conjunto de 74 instâncias para testar os classificadores, desta vez da LDG Bacamarte.

- **K-vizinhos**

Introduzindo o ficheiro de teste para classificação, que corresponde a outra gravação da lancha de desembarque Bacamarte, obteve-se:

Valor de K	**% Corr. Class.**
K=1	100%
K=2	100%
K=3	100%
K=11	100%
K=51	100%

NÃO CLASSIFICADO

K=101	98.6486 %
K=131	98.6486%
K=201	98.6486%
K=301	9.4595 %
K=369	0%
K=370	0%

Tabela 21 - Teste 2: Classificador k-vizinhos com introdução de um novo ficheiro para teste.

Como se pode observar da Tabela 21, o classificador utilizado é bastante eficaz para menores valores de k. Para os últimos valores de k considerados, nenhuma instância foi classificada com sucesso. Todas foram associadas à lancha de fiscalização, conforme se pode confirmar pela Tabela 22.

Navio Teste	Número total de intâncias classificadas - 74
	Percentagem de instâncias corretamente classificadas - 0% Percentagem de instâncias incorretamente classificadas - 100 %
	Matriz de confusão: a b c d e <-- classified as 0 0 0 0 0 | a = Lancha_Fiscalizacao 0 0 0 0 0 | b = Navio_Mercante 0 0 0 0 0 | c = Embarcacao_Pesca 0 0 0 0 0 | d = Rebocador 74 0 0 0 0 | e = LDG_Bacamarte

Tabela 22 - Teste 2: Classificador k-vizinhos com introdução de um novo ficheiro para teste, com k=369 e k=370.

- ***Naive de Bayes***

  Introduzindo o ficheiro de teste para classificação, obteve-se:

NÃO CLASSIFICADO

Navio Teste	Número total de intâncias classificadas - 74
	Percentagem de instâncias corretamente classificadas - 67.5676 % Percentagem de instâncias incorretamente classificadas - 32.4324 %
	Matriz de confusão: a b c d e <-- classified as 0 0 0 0 0 | a = Lancha_Fiscalizacao 0 0 0 0 0 | b = Navio_Mercante 0 0 0 0 0 | c = Embarcacao_Pesca 0 0 0 0 0 | d = Rebocador 9 0 3 12 50 | e = LDG_Bacamarte

Tabela 23 - Teste 2: Classificador naive de Bayes com introdução do ficheiro de teste.

Neste caso (Tabela 23), 50 das 74 instâncias foram classificadas corretamente, tendo diminuído a taxa de sucesso do classificador em relação aos resultados anteriores.

- **Árvore de decisão J48**

Introduzindo o ficheiro de teste para classificação, tem-se:

Navio Teste	Número total de intâncias classificadas - 74
	Percentagem de instâncias corretamente classificadas - 93.2432 % Percentagem de instâncias incorretamente classificadas - 6.7568 %
	Matriz de confusão: a b c d e <-- classified as 0 0 0 0 0 | a = Lancha_Fiscalizacao 0 0 0 0 0 | b = Navio_Mercante 0 0 0 0 0 | c = Embarcacao_Pesca 0 0 0 0 0 | d = Rebocador 1 0 4 0 69 | e = LDG_Bacamarte

Tabela 24 - Teste 2: Classificador árvore de decisão J48 com introdução de navio teste.

NÃO CLASSIFICADO

Com a introdução do ficheiro de teste, obtiveram-se igualmente bons resultados, segundo a Tabela 24, 69 das 74 instâncias foram bem classificadas. A Figura 27 mostra os critérios que definem a árvore de decisão nesta situação.

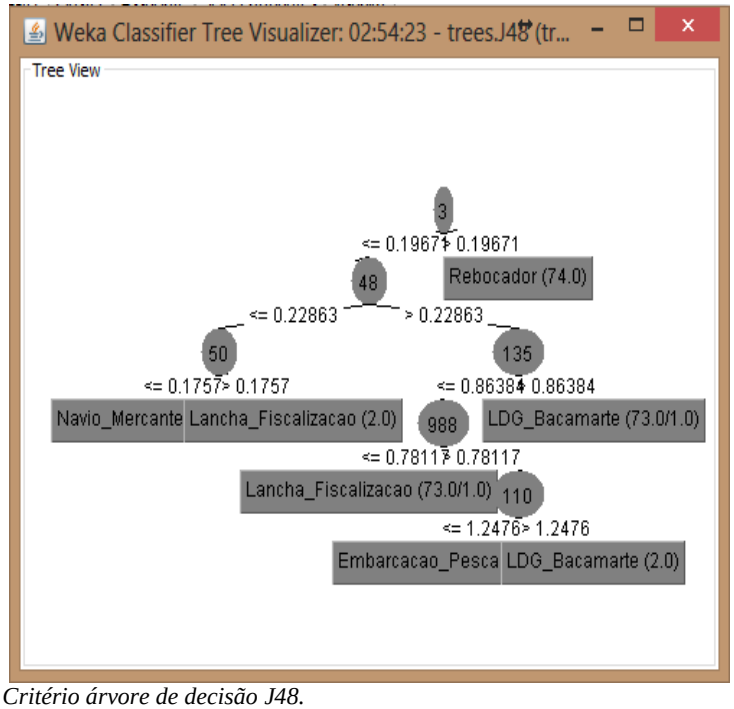

Figura 27 - Teste 2: Critério árvore de decisão J48.

```
3 <= 0.19671
        48 <= 0.22863
                50<=0.1757: navio_mercante
                50>0.1757:lancha_fiscalizacao
         48 > 0.22863
                135 <= 0.86384
                        988<=0.78117:lancha_fiscalizacao
                        988 > 0.78117
                                110<=1.2476: embarcacao_pesca
                                110 > 1.2476: ldg_bacamarte
                135 > 0.86384: ldg_bacamarte
3 > 0.19671: rebocador
```


NÃO CLASSIFICADO

Capítulo 7 - Conclusões

7.1. Primeiro Teste

Recorda-se que no primeiro teste se utilizaram três ficheiros diferentes, dois obtidos do espectro da lancha hidrográfica Auriga e um da lancha de desembarque grande Bacamarte. Um dos ficheiros da Auriga foi considerado conjunto de teste e os restantes dois constituíram o conjunto de treino.

Para o classificador k-vizinhos, testando o conjunto de treino pelos métodos *holdhout* e validação cruzada para os primeiros 11 vizinhos mais próximos utilizados, obteve-se 100% de sucesso na classificação. Ao introduzir o 3º ficheiro com um navio para teste, a percentagem de instâncias bem classificadas passou para cerca de 80% para valores de k mais baixos e atingiu os 100% para k=136.

O classificador *naive de Bayes*, na fase de teste do conjunto de treino pelos mesmos métodos obteve igualmente 100% de instâncias corretamente classificadas. Ao introduzir o conjunto de teste para classificar, esta percentagem baixou drasticamente para cerca de 49%, ou seja é praticamente igual a uma decisão aleatória.

Com recurso ao classificador árvore de decisão J48, aplicando o método *holdhout* ao conjunto de treino, todas as instâncias foram bem classificadas e pelo método de validação cruzada, cerca de 99%. Introduzindo o conjunto de teste, cerca de 87% das instâncias deste conjunto foram corretamente classificadas.

A Tabela 25 e o gráfico da Figura 28, fazem o resumo das percentagens de instâncias corretamente classificadas para os três classificadores, pelos diferentes métodos.

1º Teste	**K-vizinhos**	***Naive de Bayes***	**Árvore de decisão J48**
Método Holdout	100%	100%	100%
Método Validação Cruzada	100%	100%	99.3243%
Navio para teste	100%	48.6485%	86.4865%

Tabela 25 - Balanço dos resultados obtidos no primeiro teste.

NÃO CLASSIFICADO

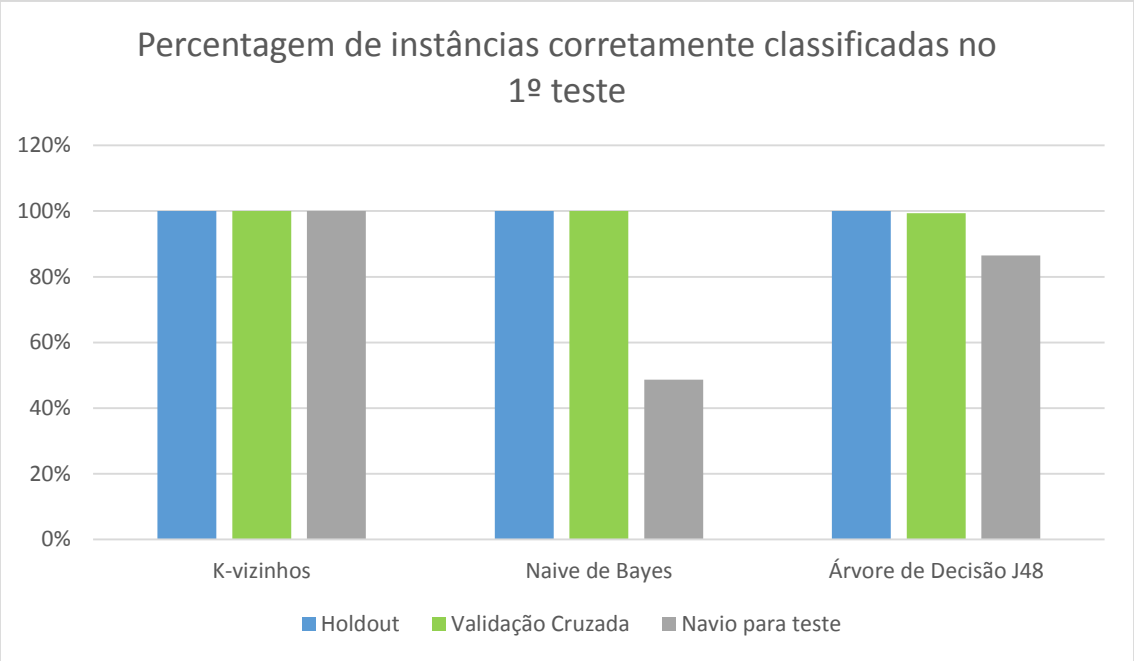

Figura 28 - Gráfico representativo da percentagem de instâncias corretamente classificadas no 1º teste, para os vários classificadores e métodos de teste.

7.2. Segundo Teste

Desta vez, o conjunto de treino é composto por cinco navios, uma lancha de fiscalização da classe Centauro, um navio mercante, uma embarcação de pesca e a lancha de desembarque grande Bacamarte, enquanto o conjunto de teste contém outro ficheiro da lancha de desembarque Bacamarte.

Pelo classificador k-vizinhos aplicado ao conjunto de treino com os dois métodos de teste já mencionados, obteve-se em ambos, uma percentagem de instâncias corretamente classificadas equivalente a 100% para os dois primeiros valores de k (k=1 e k=2), com tendências a diminuir com o aumento de k. Introduzindo o ficheiro de teste, obteve-se 100% de instâncias bem classificadas para valores de k até 50.

O classificador *naive de Bayes* obteve 100% de instâncias corretamente classificadas pelos dois métodos de teste do conjunto de treino. Adicionando o conjunto de teste para classificação a percentagem baixou para cerca de 68%.

Utilizando a árvore de decisão J48, o método *holdhout* classificou bem cerca de 95% das instâncias do conjunto de treino e o método de validação cruzada cerca de 94%. Com a introdução do conjunto de teste, aproximadamente 93% das instâncias foram corretamente classificadas.

NÃO CLASSIFICADO

A Tabela 26 e a Figura 29 ilustra as melhores percentagens de acerto obtidas na classificação das instâncias para os diferentes classificadores, recorrendo aos diferentes métodos de teste.

2º Teste	K-vizinhos	*Naive de Bayes*	Árvore de decisão J48
Método *Holdout*	100%	100%	94.5946%
Método Validação Cruzada	100%	100%	93.7838 %
Navio para teste	100%	67.5676 %	93.2432 %

Tabela 26 - Balanço dos resultados obtidos no segundo teste.

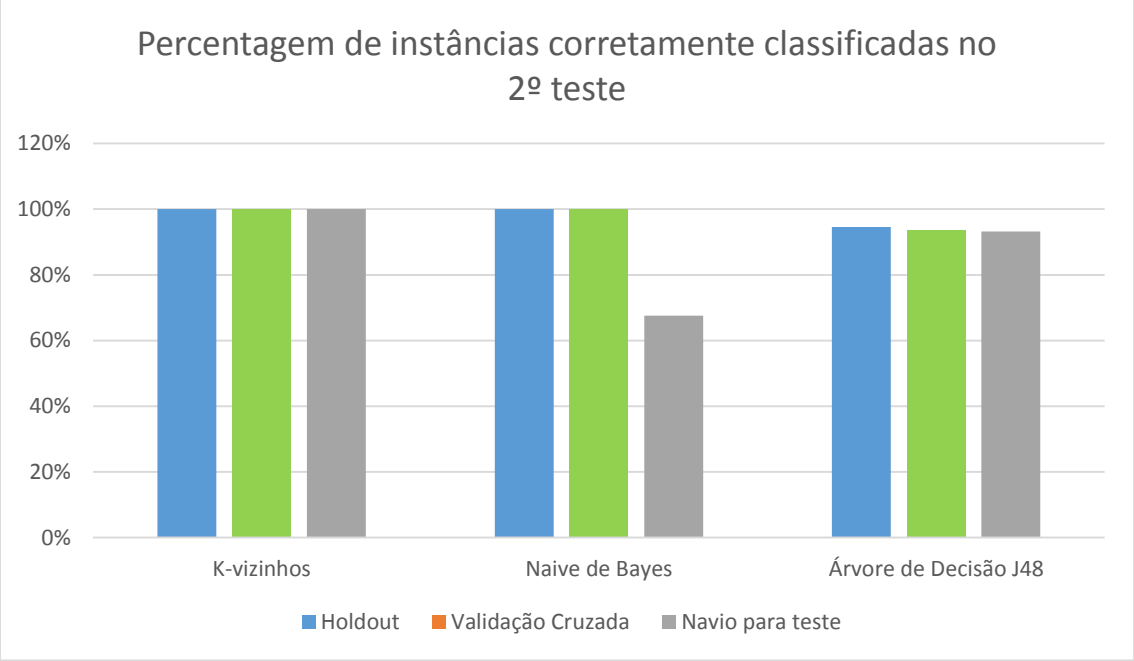

Figura 29 - Gráfico representativo da percentagem de instâncias corretamente classificadas no 2º teste, para os vários classificadores e métodos de teste.

7.3. Balanço

Por análise dos testes efetuados pede-se retirar que o classificador que obtém melhores resultados na identificação de um navio a partir do seu espectro é o classificador vizinho mais próximo, contudo está muito dependente do valor de k escolhido. Apesar de se detetar uma tendência para obter um maior número de instâncias bem classificadas para valores de k mais pequenos (principalmente na ordem das unidades e das dezenas),

NÃO CLASSIFICADO

é difícil escolher o k ótimo. Contudo a decisão não é crítica (as taxas de erro são muito semelhantes) e um valor de k=3 geralmente dá bons resultados.

A árvore de decisão J48 chegou igualmente a percentagens de acerto na classificação do navio 'desconhecido' bastante satisfatórias. Este classificador tem a vantagem de ser simples e não ter necessidade de ser parametrizado (escolha do valor de k ou número de neurónios) para alcançar bons resultados.

Pode-se também verificar que os resultados obtidos nestes dois classificadores, com a introdução de um navio para teste, são muito semelhantes aos obtidos ao testar o conjunto de treino, ao contrário do que acontece para o classificador *naive de Bayes*.

O classificador *naive de Bayes* revelou-se particularmente sensível a pequenas variações do sinal, sendo completamente inútil nos casos reais. É de esperar que os resultados obtidos para este classificador não tenham sido os mais favoráveis uma vez que faz uma aproximação gaussiana para cada classe baseada em valores de média e desvio padrão estimados.

7.3.1. Limitações

Deve-se ter em consideração que as gravações dos sinais acústicos foram obtidas sempre no mesmo local e ao longo do mesmo período temporal. Assim sendo, é normal que dois sinais emitidos por um mesmo navio tendam a ser muito semelhantes, já que as condições de propagação não tiveram muita variabilidade. Os principais fatores que podem influenciar o sinal recebido pelo hidrofone são a alteração das condições meteorológicas e a variação do regime de máquinas ou operação de determinados equipamentos a bordo do navio de que se obtém o registo acústico.

7.3.2. Trabalho Futuro

Para potenciar a classificação, fará sentido construir uma base de dados maior, com maior número de navios e maior diversidade de registos de cada navio em condições distintas de propagação do som. Desta forma, aplicar-se-ia um dos classificadores com melhores resultados a uma maior quantidade de dados, para que pudesse correr uma base de dados com os espectros de diferentes navios e assim permitir a classificação de um navio a partir do seu registo acústico.

NÃO CLASSIFICADO

Na construção da base de dados, seria de ponderar complementar a identificação dos navios recorrendo ao *Automatic Identification System*[12] (AIS), uma vez que os registos fotográficos não são os mais esclarecedores, havendo vários fatores que contribuem para a perda de definição da imagem.

Na componente de identificação propriamente dita, devem-se desenvolver algoritmos e implementá-los em sistemas de classificação em tempo real. Para a passagem dos registos acústicos no momento, poderá recorrer-se por exemplo ao hidrofone TP-1 da empresa *Marsensing*.

Estes desenvolvimentos poderão dar origem a sistemas passivos de monitorização de navios com emissão de alertas em tempo real.

[12] Sistema Automático de Identificação de navios. Sendo obrigatório para navios de 300 toneladas brutas ou mais envolvidos em viagens internacionais, navios de carga de 500 toneladas não envolvidos em viagens internacionais e navios de passageiros (mais de 12 passageiros) independentemente do tamanho (Council, 2003).

NÃO CLASSIFICADO

Referências

Althage. (2004). Acoustic Intelligence: Charting the Undersea Frontier. *Underseawarfare, The Official Magazine of the U.S. Submarine Force*. http://www.navy.mil/navydata/cno/n87/usw/issue_22/ai.htm

Brandt, A. (2011). *Noise and vibration analysis: signal analysis and experimental procedures*: John Wiley & Sons.

CIRA. (2014). *Centre d'Interprétacion et de Reconnaissance Acoustique*. http://www.defense.gouv.fr/marine/ressources-humaines/ecoles-et-formations/ecoles-de-specialistes/ecole-du-domaine-sous-marin-et-nucleaire/cira2/cira

Collier, R. D. (1998). *Ship and platform noise, propeller noise. Handbook of acoustics*: M. J. Crocker, John Wiley & Sons.

Costa, N. (2013). Lei do Inverso do Quadrado da Distância, Pressão Eficaz. Retrieved 24/01/2015, 2015, from http://www.soundzonemagazine.com/2013/10/lei-do-inverso-do-quadrado-da-distancia.html

Council, N. R. (2003). *Shipboard Automatic Identification System Displays: Meeting the Needs of Mariners*: Transportation Research Board.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*: CRC press.

Filho, W. S. (2011, Dezembro). Sistema de Detecção, Acompanhamento e Classificação de Contatos (SDAC). *Pesquisa Naval*, 06.

Freitas, A. A. (2013). *Data mining and knowledge discovery with evolutionary algorithms*: Springer Science & Business Media.

Gama, J., & Brazdil, P. (1995). Characterization of classification algorithms *Progress in Artificial Intelligence* (pp. 189-200): Springer.

Gupta, G. (2011). *Introduction to data mining with case studies*: PHI Learning Pvt. Ltd.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter, 11*(1), 10-18.

NÃO CLASSIFICADO

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques: concepts and techniques*: Elsevier.

Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining*.

Hartigan, J. (1983). Bayes theory.

Herman Medwin, C. S. C. (1997). *Fundamentals of Acoustical Oceanography*. Califórnia: Academic Press.

Hodges, R. P. (2010). *Underwater Acoustics. Analysis, Design and Performance of Sonar*. Reino Unido: WILEY.

Imperadeiro, P. (2010). Sistemas de sensores para o tanque de acústica submarina da Escola Naval. Lisboa.

Jesus, S., Silva, A., & Zabel, F. (2005). Acoustic oceanographic buoy data report Makai Ex 2005.

Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*: John Wiley & Sons.

Larose, D. T. (2005). *Discovering knowledge in data: an introduction to data mining*: John Wiley & Sons.

Leite, R. M. S. R. (2007). Selecçao de Algoritmos de Classificaçao.

Li, Q. (2012). *Digital Sonar Design in Underwater Acoustics: Principles and Applications*: Springer Science & Business Media.

Lobo, V. J. d. A. e. S. (2002). *Ship Noise Classification. A contribution to prototype based classifier design*. Lisboa.

Lurton, X. (2002). *An introduction to Underwater Acoustics. Principles and Applications.* Reino Unido: Praxis.

Lyons, R. G. (2010). *Understanding digital signal processing*: Pearson Education.

MetEd, C. (2015a). Introduction to Ocean Acoustics. Retrieved 28/06/2015, 2015, from https://www.meted.ucar.edu/sign_in.php?go_back_to=http%253A%252F%252Fwww.meted.ucar.edu%252Foceans%252Facoustics%252Fprint.htm#

NÃO CLASSIFICADO

MetEd, C. (2015b). Teaching and Training Resources for the Geoscience Community. Retrieved 03/06/2015, 2015, from https://www.meted.ucar.edu/sign_in.php?go_back_to=http%253A%252F%252Fwww.meted.ucar.edu%252Foceans%252Facoustics%252Fprint.htm#

OGP. (2008). Fundamentals of underwater soud: International Association of Oil & Gas Producers.

PEEAS 43 (A).

Platform Signature Monitoring System (2014). *Platform Signature Monitoring System*. http://www.drumgrange.com/media/30209/platform-signature-monitoring-system.pdf

Press, W. H. (1992). *Power Spetrum Estimation Using the FFT*: Cambridge University Press.

Quinlan, J. R. (1986). Induction of decision trees. *Machine learning, 1*(1), 81-106.

Quinlan, J. R. (2014). *C4. 5: programs for machine learning*: Elsevier.

Richardson, W. J., Greene Jr, C. R., Malme, C. I., & Thomson, D. H. (2013). *Marine mammals and noise*: Academic press.

Rumsey, F., & McCormick, T. (2012). *Sound and recording: an introduction*: CRC Press.

Sciences, N. A. o. (2003). What are common underwater sounds? Retrieved 01/07/2015, 2015, from http://www.dosits.org/science/soundsinthesea/commonsounds/?CFID=3944073&CFTOKEN=99391732

SICLA - Sistema de Clasificación Acústica - SAES. (2014). *SICLA - Sistema de Clasificación Acústica - SAES*. http://www.electronica-submarina.com/defensa/sonar-y-sistemas-embarcados/sicla-sistema-de-clasificacion-acustica-submarina/#

Silva, T. (2014). Hidroaústica - propagação do som na água. http://www.ebah.pt/content/ABAAAAp3kAI/hidroacustica

Urick, R. J. (1984). Ambient noise in the sea: DTIC Document.

NÃO CLASSIFICADO

Vieira, J. M. N. (2003). *Matlab num Instante.* Universidade de Aveiro.

Waite, A. D. (2002). *Sonar for practising engineers* (Vol. 3).

Wenz, G. M. (1962). Acoustic ambient noise in the ocean: spectra and sources. *The Journal of the Acoustical Society of America, 34*(12), 1936-1956.

William J. Emery, R. E. T. (2001). *Data Analysis Methods in Physical Oceanography*. Holanda: Elsevier Science.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann.

Zezula, P., Amato, G., Dohnal, V., & Batko, M. (2006). *Similarity search: the metric space approach* (Vol. 32): Springer Science & Business Media.

NÃO CLASSIFICADO

Índice remissivo

NÃO CLASSIFICADO

R

S

T

V

NÃO CLASSIFICADO

Apêndices

AP 1. Obtenção do espectro dos sinais

function A=AAA_Convert_Recording(filename, code)

OBJETIVO: Ler um ficheiro de som e calcular o espectro médio, produzindo um espectrograma. Cada espectro tem a resolução de 1 Hz e vai de 0 a 1024 Hz. Cada espectro é obtido da média de 8, com 50% de sobreposição.

INPUT:

Filename

Code

OUTPUT:

A – matriz 1025x 'N', 1024 mais o código da gravação

%%%

```
[x fs]=wavread(filename);
N=max(size(x));
NPONTOS = 1024; % o numero de pontos a usar no espetro deve ser 1024
NMEDIAS = 8; % cada espectro é a media de 8 espectros

namostras_instantaneas=floor(N/fs)*2-1; % temos 50% de overlap
namostras=floor(namostras_instantaneas/NMEDIAS);
for i=1:namostras
   media=zeros(NPONTOS,1);
   for j=1:NMEDIAS
      inicio = (i-1)*(NMEDIAS/2)*fs + (j-1)*fs/2 +1;
      fim = inicio+fs-1;
      xt=x(inicio:fim);
```


NÃO CLASSIFICADO

```
        y=fft( xt-mean(xt));  % calcular a transformada
        y=y(1:NPONTOS); % primeiros NPONTOS
        y=abs(y);  % transformar o sinal complexo no modulo da amplitude
        media=media+y;
    end;
    media=media/NMEDIAS;
    volume=mean(media);
    media=media/volume;
    if i==1
        A=media;
    else
        A=[A media];
    end;
end;
A = [ A ; code*ones(1,namostras)];
return;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% exportar para weka
G=AAA_Convert_Recording('filename',code);
todos = [G]';
cabecalhos=[1:1025];
todos= [cabecalhos; todos];
csvwrite( 'C:\Users\Catarina Santos\Desktop\Tese2\Weka\Entrada_teste.csv',todos);
```

AP 2. Localização do hidrofone SR-1

A imagem do Google Earth representa a posição do hidrofone SR-1, colocado à entrada da barra de Setúbal, bem como os limites das fotografias capturadas na área.

Figura 30 - Imagem obtida do Google Earth que ilustra a localização do hidrofone SR-1.

NÃO CLASSIFICADO

Anexos

AX 1. Imagem ilustrativa das frequências típicas para fontes de ruído ambientais e antropogénicas.

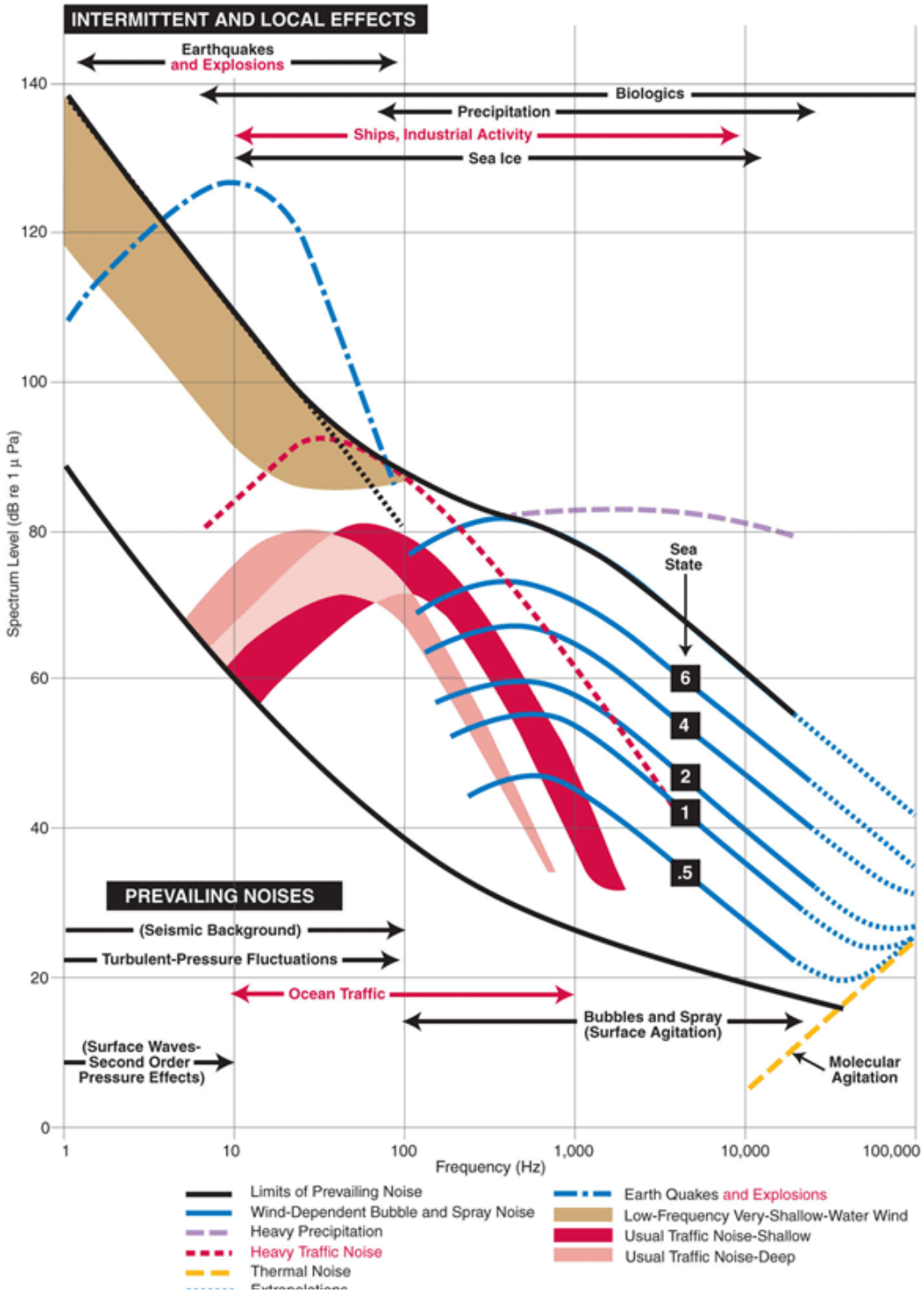

Figura 31 - Níveis sonoros típicos do ruído de fundo do oceano a diferentes frequências, segundo as medições de (Wenz, 1962). Gráfico adaptado por (Sciences, 2003).

NÃO CLASSIFICADO