

Aplicação de Técnicas de Mineração de Dados para Estimativa de Captura Incidental de Aves Marinhas Associadas à Pesca

Joana P. G. de Ataíde, Rafael Ballottin Martins, Rodrigo Sant'Ana

Laboratório de Inteligência Aplicada – Escola do Mar Ciência e Tecnologia -
Universidade do Vale do Itajaí (UNIVALI)
Rua Uruguai, 458 - 88302-202 – Itajaí- SC – Brasil

joana.godoy@edu.univali.br, ballottin@univali.br, rsantana@univali.br

Abstract. *Every year, thousands of albatrosses and petrels are accidentally killed in pelagic longline hooks when birds, drawn to fishing vessels for discards and baits, ingest hooked hooks and subsequently drown. Data recorded in on-board maps of the south and southeast regions of Brazil were collected from 1978 to 2012, containing information on incidental catch of seabirds only from 2004. The research identified, after the generation of some models of Data Mining, that the model with the weighting strategy and the algorithm Deep Learning obtained a good performance in addition to being able to generate a restoration of the missing data in accordance with the realized one, being able to consider it appropriate for the activity*

Resumo. Todos os anos, milhares de albatrozes e petréis são acidentalmente mortos na pesca de espinhel de superfície quando as aves, atraídas para embarcações de pesca por descartes e iscas, ingerem anzóis iscados e subsequentemente se afogam. Dados registrados em mapas de bordo das regiões sul e sudeste do Brasil foram coletados no período de 1978 a 2012, contendo informações de capturas incidentais de aves marinhas apenas a partir de 2004. A pesquisa identificou, após a geração de alguns modelos de Mineração de Dados, que o modelo com a estratégia de ponderação e o algoritmo Deep Learning obteve um bom desempenho além do mesmo conseguir gerar uma restauração dos dados faltantes, podendo o considerar apropriado para a atividade.

1. Introdução

O nível de mortalidade relacionada ao espinhel de superfície é tal que esse tipo de pesca foi identificado como uma grande ameaça que afeta albatrozes e petréis [Cooper et al. 1998].

As aves marinhas são atraídas pela operação de espinhel de superfície por descarte de iscas e miudezas, e as mortes ocorrem quando as linhas estão sendo iscadas e lançadas ao mar e as aves atacam os anzóis iscados, onde acabam se enroscando e/ou engolindo os anzóis, e, por fim, se afogando [Brothers 1991].

Embora tenha havido um maior foco na mortalidade de aves marinhas decorrentes de pescarias de espinhel de superfície nos últimos anos, são necessários mais dados sobre as taxas, causas e fatores que contribuem para a mortalidade de albatrozes e petréis gigantes como resultados de operações de arrasto.

Com base nesta necessidade, esta pesquisa teve como objetivo aplicar técnicas de Mineração de Dados sobre dados de diferentes instituições de pesquisa e ensino que atuam no monitoramento pesqueiro marinho no Brasil, referente a lances de pesca contendo informações de capturas incidentais de aves marinhas a partir de 2004 até 2012 a fim de estimar a captura incidental histórica de aves marinhas em número de

organismos e para que, de posse deste conhecimento, especialistas tenham subsídios no processo de decisão para aplicação de medidas que reduzam o índice de mortalidade desses animais.

Muitas técnicas estatísticas e matemáticas podem ser usadas para gerar previsões espacialmente contínuas para variáveis numéricas, mas muitas vezes são específicas para determinados tipos de dados e seu desempenho depende de muitos fatores [Li e Heap 2011].

Segundo Prasad, Iverson e Liaw (2006) novas técnicas de mineração de dados intensivas em computador baseadas em recursividade, reescalonamento, média e randomização podem descobrir estruturas ocultas nos dados e produzir melhores modelos preditivos. A mineração de dados é definida como o uso de técnicas automáticas de exploração de grandes quantidades de dados de forma a descobrir novos padrões e relações, que devido ao volume de dados, não seriam facilmente descobertas ao olho nu por um ser humano [Carvalho, 2005].

2. Materiais e Métodos

2.1 Características dos Dados

Os dados utilizados neste trabalho estão armazenados no Banco Nacional de Atuns (BNDA), mantido pelo Subcomitê Científico de Atuns e Afins, vinculado ao Comitê Permanente de Gestão (SCC Atuns/CPG) do Ministério da Agricultura, Pecuária e Abastecimento (MAPA) (Portaria n.º 222, de 27 de janeiro de 2017). Os dados contidos no BNDA são provenientes de (a) registros administrativos entregues pelos mestres de pesca de cada embarcação ao final de cada viagem (Mapas de Bordo) e (b) observadores de bordo embarcados principalmente na frota arrendada.

A base de dados contém uma série de informações coletadas, lance a lance, de pouco mais de 30 mil operações de pesca com espinhel de superfície da frota brasileira e arrendada, no período compreendido entre os anos de 1978 e 2012.

As informações destinadas a este trabalho compreendem dados de capturas, em quilogramas e número de indivíduos, de todos os recursos alvo, não alvo e incidentais, bem como, informações sobre esforço de pesca (número de anzóis), áreas (latitude, longitude) e épocas de operação (mês, ano), descrições gerais dos petrechos utilizados (distância entre anzóis) e condições oceanográficas no local e hora de cada lance (temperatura superficial da água do mar).

Os únicos atributos com valores faltantes são: (i) o atributo NR_AVES que possui 71,18% de valores faltantes, (ii) o atributo TEMPERATURA_MAR que possui 66,26% e (iii) o atributo BFT.KG com 10,77%.

Foi identificado que o atributo referente a quantidade de capturas de aves marinhas (NR_AVES) além de valores faltantes possui grande maioria dos seus valores sendo igual a zero. Enquanto que, dos 10.057 registros com captura incidental informada, apenas 391 registros possuem esta informação. Essa característica nos dados se dá devido a captura incidental em muitas pescarias ser um evento raro e difícil de detectar e o caracteriza como dados desequilibrados.

2.2 Pré-Processamento de Dados

O objetivo desta etapa foi a construção de um conjunto de dados final, sobre o qual as tarefas de Mineração de Dados foram executadas. A base de dados disponibilizada, por já ter sido utilizada no estudo de Marenzi (2017), possuía alguns tratamentos já realizados nos dados.

Antes de iniciar o pré-processamento, a base de dados foi dividida entre os registros com valor de captura informados e os sem, de modo que somente os dados com valor de captura informados serviram para treinamento dos modelos aplicados. Desta forma a base de dados com capturas incidentais informadas ficou com 10.057 registros e a base de dados com as capturas incidentais não informadas ficou com 25.598 registros a serem reconstruídos.

O principal foco deste estudo na limpeza de dados foi na identificação e remoção de *outliers*. Devido ao fato do evento a ser estimado ser raro, ou seja, possuir baixa porcentagem de registros com valor de captura incidental maior que zero, optou-se por realizar o processo manualmente para assim se evitar a remoção de registros com eventos raros presentes.

A identificação de *outliers* em atributos contínuos se deu com a utilização de diagramas de caixa, no qual foi analisado o diagrama de caixa de cada atributo, menos o atributo rótulo, e detectado os valores *outliers*. Não foi realizada a remoção de *outliers* nos atributos contínuos referentes as capturas de espécies alvos da pesca devido ao fato dos dados destes atributos possuírem característica de dados desequilibrados, do mesmo modo que o atributo referente a captura incidental. A identificação de *outliers* no único atributo nominal, BANDEIRA, foi realizada de acordo com a frequência de registros para cada valor. Ainda nesta etapa foram removidos registros duplicados da base de dados sendo, após a remoção dos registros *outliers*, resultando em uma base de dados com um total final de 9.512 registros.

As transformações de dados realizadas neste estudo se deram com a criação de novos atributos e a substituição de valores faltantes para melhorar os modelos. Foram criados neste estudo os atributos TEMPO_LUMINOSIDADE, DURACAO_PESCA e PERCENTUAL_LUMINOSIDADE. Além de criação de novos atributos também foi realizado a substituição de valores faltantes para o atributo BFT.kg.

A partir dos 21 atributos disponibilizados para este estudo e mais os 3 criados na etapa de transformação dos dados. Os atributos TRIMESTRE e TEMPO_LUMINOSIDADE foram removidos por estarem correlacionados com os atributos MÊS e PERCENTUAL_LUMINOSIDADE respectivamente, atributos correlacionados geralmente são removidos porque são semelhantes em comportamento e terão impacto semelhante nos cálculos de previsão, portanto, manter os atributos com impactos semelhantes é redundante. A remoção de atributos correlacionados economiza espaço e tempo de cálculo de algoritmos complexos. Além disso, também torna os processos mais fáceis de projetar, analisar e compreender.

O atributo TEMPERATURA_MAR foi removido pois possui grande quantidade de valores faltantes nas duas bases de treino e validação, do total de 35.660 registros, 23.625 não tem valor informado. Os atributos HORA_INICIO_LANCAMENTO e HORA_FIM_LANCAMENTO foram removidos devido a substituição dos mesmos pelo atributo DURACAO_PESCA criado na etapa de transformação de dados.

Ao final desta etapa, a base de dados obtinha 16 atributos sendo estes: BANDEIRA, DURACAO_LANCE, PERCENTUAL_LUMINOSIDADE, MES, ANO, ALB.kg, BET.kg, BFT.kg, SKJ.kg, SWO.kg, YFT.kg, ID_EMBARCACAO, ID_ESTRATEGIA DE PESCA, LATITUDE, LONGITUDE e NR_ANZOIS. Além do atributo rótulo NR_AVES.

2.3 Seleção de Algoritmos

A seleção de algoritmos foi realizada a partir de critérios pré-definidos, uma vez definidos foram realizadas pesquisas e ao final selecionados três algoritmos para serem aplicados nos dados.

Os critérios definidos foram baseados em pesquisas e no pré-processamento realizado no qual ficaram claras as características necessárias para a realização dos modelos. Sendo assim, os critérios foram os seguintes: 1) Método de aprendizado supervisionado; 2) Atributos numéricos; 3) Atributo rótulo numérico; 4) Ter sido utilizado e obtido sucesso em estudos utilizando predição / estimativa de dados;

O algoritmo *Random Forest* foi selecionado, pois foi amplamente utilizado, principalmente em estudos com dados ecológicos como no trabalho de Lennert-Cody e Berk (2007) e no trabalho de Pons et al. (2009), o algoritmo foi muito utilizado por obter baixo erro preditivo dos modelos usados.

O algoritmo que implementa Redes Neurais Perceptron Multicamada, chamado de *Deep Learning*, foi selecionado pois demonstrou desempenho promissor na mineração de dados de previsão com séries temporais, como no estudo de Brezak et al. (2012), onde foram obtidos resultados que indicaram características satisfatórias de previsão com redes neurais.

O algoritmo Árvore de Decisão foi selecionado por ser muito utilizado nas pesquisas sobre dados desequilibrados. E por final o algoritmo kNN também foi selecionado por ser muito utilizado na mineração de dados temporais, como no trabalho de Lucio et al, (2007)

2.4 Modelos

Identificou-se que domínios desequilibrados levantam desafios significativos ao construir modelos preditivos, mas que existem várias estratégias para resolver esse problema, principalmente em um cenário de classificação. Foram criados modelos com estratégias que resolvem o problema de dados desequilibrados, esses modelos utilizaram a tarefa de classificação no qual o atributo rótulo é binário, para isso foram criadas duas classes, sendo a classe “positivo” para quando houve captura e “negativo” caso o contrário. Os modelos foram testados com a técnica de validação cruzada.

As estratégias selecionadas para serem utilizadas neste trabalho foram estratégias realizadas no pré-processamento dos modelos. A estratégia denominada amostragem foi utilizada neste trabalho por ser a técnica mais explorada por estudos até o momento da pesquisa. Já estratégia de ponderação de dados foi escolhida por ser uma maneira de implementar aprendizado sensível ao custo que é outra estratégia que, segundo He e Garcia (2009), obtiveram resultados até melhores que o método de amostragem.

O método de sub-amostragem foi realizado removendo registros da classe majoritária aleatoriamente de modo que a quantidade de registros nas duas classes ficasse a mesma. Os modelos com diminuição de amostragem ao final estavam com 760 registros, pois a classe minoritária possui 360 registros.

O método de ponderação foi realizado aplicando peso para os registros de modo que registros que possuíam o atributo rótulo da classe minoritária (positivo) receberam o peso 0.013 e os registros da classe majoritária (falso), receberam o peso 0.001. Para distribuir o peso foi utilizado uma técnica no qual a partir de um valor de peso definido é dividido esse valor para cada registro da base de dados, de modo que as duas classes recebam a mesma quantidade de peso.

3. Resultados e Discussão

3.1 Resultados

Os modelos obtiveram grande maioria dos resultados de *f-measure* com desempenhos acima de 90%, considerados bons, sendo estes exibidos nas Figura 1 e Figura 2.

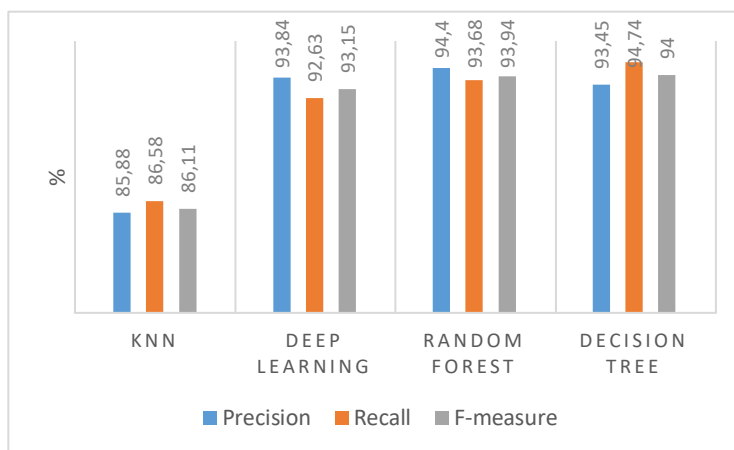


Figura 1. Resultados de modelos utilizando a estratégia de Sub-amostragem.

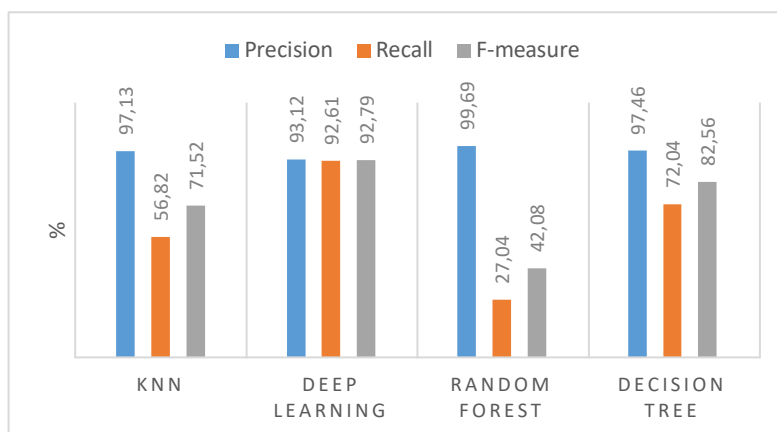


Figura 2. Resultados de modelos utilizando a estratégia de Ponderação.

Nota-se que, apesar da pequena diferença, o modelo com o melhor desempenho foi do algoritmo Árvore de Decisão com a estratégia de sub-amostragem. Na Figura 1 é possível observar as quantidades previstas por este modelo que contém apenas 760 registros no total e quase a mesma quantidade de acertos das duas classes.

Já o melhor modelo desta etapa com a estratégia de ponderação, resultado ilustrados na Figura 2, também com uma pequena diferença de performance entre os algoritmos, foi com o algoritmo *Deep Learning*. Este modelo obteve 8.678 acertos de um total de 9.368 registros, sendo 351 valores da classe verdadeira e 8.327 da classe falsa, no qual a classe verdadeira possui no total 380 valores e a classe falsa 8.988.

3.2 Discussão

A avaliação dos resultados dos modelos se deu pelos que obtiveram os melhores desempenhos, sendo que foram estes aplicados para a reconstrução da base de dados.

O modelo com o algoritmo Árvore de Decisão e a estratégia sub-amostragem teve como saída uma árvore, o Quadro 1 exibe as regras de maior destaque em relação a classe positiva geradas pela árvore deste modelo.

Quadro 1. Regras de destaque do modelo com o algoritmo Árvore de Decisão e a estratégia subamostragem.

Regra
Das 380 vezes em que houveram capturas, 166 foram quando o atributo NR_ANZOIS era maior que 2.467.
Das 380 vezes que houveram capturas, 44 foram quando o atributo NR_ANZOIS foi menor ou igual a 2.467, a latitude foi menor ou igual a -25, o ID_EMBARCACAO for menor que 176.5 e maior que 175, a bandeira foi BRA-BRA e a LONGITUDE foi menor ou igual a -42,5.
Das 380 vezes que houveram capturas 90 foram quando o atributo NR_ANZOIS foi menor ou igual a 2467, a latitude foi menor ou igual a -25, o ID_EMBARCACAO for menor que 176.5 e a BANDEIRA foi BRA-ESP ou BRA-JPN.
Das 380 vezes que houveram capturas 32 foram quando o atributo NR_ANZOIS foi menor ou igual a 2.467 e maior que 1.050, a latitude foi menor ou igual a -25, o ID_EMBARCACAO for menor que 176.5, a bandeira foi BRA-PRT, o MES foi menor ou igual 10.

O modelo com a estratégia de ponderação com o melhor desempenho foi o com o algoritmo *Deep Learning*. Apesar de ser um algoritmo com resultados *black box* foi possível identificar quais os atributos mais importantes do modelo, sendo estes por ordem de importância: 1) SWO.kg; 2) NR_ANZOIS; 3) BANDEIRA; 4) ALB.kg; 5) BET.kg; 6) DOL.kg; e 7) MES. Apesar do modelo com o algoritmo *Deep Learning* ter obtido o melhor desempenho com a estratégia de ponderação que os outros algoritmos, o algoritmo Árvore de Decisão também obteve bons resultados e foi considerado na avaliação de modo que este não é um algoritmo caixa preta e permite uma melhor interpretação dos resultados gerados.

O modelo com a estratégia de ponderação do algoritmo Árvore de Decisão obteve um *F-measure* de 82,56%, o Quadro 2 exibe as principais regras geradas em relação a classe positiva, nas quais exibe as quantidades de vezes em que a classificação foi positiva e as quantidades em que foi negativa. Observa-se que, apesar de ser gerado uma regra em relação a um comportamento que ocorreu na maioria dos casos, ainda houve uma grande quantidade de casos em que ocorreu o contrário.

Quadro 2. Regras de destaque do modelo com o algoritmo Árvore de Decisão e a estratégia de ponderação.

Regra
142 (positivo) contra 36 (negativo) vezes, quando o NR_ANZOIS foi maior que 2.562,5, a DURACAO_LANCE foi menor ou igual a 12 e a latitude menor ou igual a 19
51 (positivo) contra 45 (negativo) vezes, quando o NR_ANZOIS foi menor que 2.562,5, a LATITUDE foi menor ou igual -25, o ID_EMBARCACAO foi menor ou igual a 176,5 e maior que 175,5, a BANDEIRA foi BRA-BRA, a captura de SKJ.kg foi menor ou igual 1.216 e a captura de ALB.kg foi menor ou igual a 90.
19 (positivo) contra 8 (negativo) vezes, quando o NR_ANZOIS foi menor que 2.562,5, a LATITUDE foi menor ou igual -32, o ID_EMBARCACAO foi menor ou igual a 4,5, o MÊS foi menor ou igual a 10, a DURACAO_LANCE foi menor ou igual a 9, a captura de SKJ.kg foi menor ou igual 660 e a captura de BET.kg maior que 100.

Ao final foi possível constatar que os atributos NR_ANZOIS, BANDEIRA e MES obtiveram grande relevância nos três modelos avaliados nesta seção. Observou-se ainda que, com exceção do Random Forest, os algoritmos obtiveram pouca diferença de performance em relação as duas estratégias utilizadas.

4. Conclusões

O presente estudo relatou sobre o problema da captura incidental de aves marinhas durante atividades pesqueiras de espinhel de superfície e as dificuldades de se obter informações sobre estas capturas que poderiam gerar implantações de gerenciamento de pesca e conservação de espécies. Um caso desse problema é o proposto neste trabalho onde dados de bordo de pesca de espinhel de superfície foram coletados durante os anos de 1978 a 2012, porém sem a informação das capturas incidentais de aves marinhas levando a uma deficiência na percepção do impacto real dessas atividades no habitat das aves marinhas.

Buscando minimizar o impacto da falta de informação para especialistas da área de oceanografia foi realizada a estimativa dos dados faltantes com a aplicação do processo de descoberta de conhecimento em base de dados no qual possui a capacidade de, além de outras tarefas, estimar um atributo dependente a partir de atributos independentes na base de dados.

Pesquisas bibliográficas foram realizadas afim de encontrar trabalhos similares que realizaram estimativas de capturas incidentais em atividades pesqueiras. Estes trabalhos serviram de inspiração, principalmente, para a etapa de avaliação dos resultados, além de fortalecerem a ideia da importância da estimativa desses dados.

Todos os processos deste trabalho foram realizados na ferramenta RapidMiner [Hofmann e Klinkenberg 2013] na qual possui suporte para todas as etapas do KDD desde o pré-processamento com funções de análise geral e estatística dos dados até a geração de visualização dos mesmos em formato de gráficos, que foram utilizados nas etapas de domínio de aplicação e pré-processamento dos dados.

As etapas de domínio de aplicação e pré-processamento dos dados foram essenciais para a conclusão deste trabalho de modo que permitiram o entendimento tanto do contexto dos dados quanto os dados em si, nos quais foram analisados características e padrões dos mesmos, como a detecção de atributos com valores faltantes nos atributos TEMPERATURA_MAR e BFT.kg. Além de permitir a seleção dos melhores atributos para o desempenho dos modelos e a seleção dos algoritmos aplicados na etapa de mineração dos dados.

Ao longo da etapa de aplicação dos modelos de mineração foram encontrados desafios em relação a distribuição dos dados por serem desequilibrados, devido ao valor que caracteriza a presença de captura incidental possuir um comportamento esporádico.

A princípio foi realizado a tentativa de treinar os dados com a tarefa de regressão, o que acabou não gerando resultados satisfatórios em relação as saídas dos modelos que dificilmente conseguiam acertar algum valor previsto, e quando conseguiam era na maioria dos casos o valor zero.

Em seguida foram aplicados modelos de classificação afim de encontrar outro tipo de resposta, que seria a resposta positiva ou negativa em relação a captura incidental. Nesta segunda etapa notou-se que todos os algoritmos selecionados para as aplicações dos modelos não se comportavam bem com dados desequilibrados, conseguindo encontrar padrões e regras apenas para a classe majoritária.

Enfim foram encontradas estratégias para lidar com dados desequilibrados, sendo estas chamadas de amostragem e ponderação. Observou-se que na estratégia de amostragem foram geradas regras para a classe minoritária, mas ao aplica-las a

reconstrução da base de dados não foram validadas de modo que, ao remover grande parte dos padrões da classe majoritária dos dados o classificador perdeu informações importantes pertencentes a ela.

Já com a estratégia de ponderação, o classificador obteve sucesso em gerar uma reconstrução condizendo com a realidade, no caso do modelo com o algoritmo *Deep Learning*, podendo o considerar adequado para a aplicação e utilização nos dados de atividades pesqueiras desta pesquisa.

Em relação aos atributos de maior influência nos resultados dos modelos, no estudo de Marenzi (2017) foi constatado que os fatores que se mostraram explicativos para prever a captura incidental de aves e tartarugas marinhas foram 1) trimestre, 2) estratégia de pesca e 3) temperatura superficial da água do mar. Neste trabalho não foi possível constatar a influência do atributo TEMPERATURA_MAR devido sua grande quantidade de valores faltantes. No caso do fator trimestre, observado por Marenzi (2017), neste estudo também foi encontrada influência deste fator nas ocorrências de capturas incidentais, porém neste estudo foi utilizado o atributo MÊS para esta avaliação. Já sobre o fator estratégia de pesca, não foram encontrados evidências neste trabalho para com este atributo, constatando que o mesmo não se mostrou presente de forma significativa nos resultados dos modelos.

Ao final foi constatado neste estudo que os fatores que mais influenciaram nos resultados dos modelos, de modo que atribuíram maior peso e geração de regras para os mesmos, foram NR_ANZOIS, BANDEIRA e MÊS. Com estas informações e a reconstrução da base de dados, os especialistas terão subsídios para realizarem análises mais completas sobre o impacto da captura incidental no habitat das aves marinhas e com isso aplicar medidas que reduzam o índice de mortalidade desses animais.

5. Referências

- Prasad, A.M., Iverson, L.R., Liaw, A. (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*, 9, 181–199.
- BrezaK, Danko et al. (2012). A comparison of feed-forward and recurrent neural networks in time series forecasting. *IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFER)*.
- Brothers, N (1991). Albatross mortality and associated bait loss in the Japanese longline fishery in the Southern Ocean. *Biological Conservation, Southern Ocean*.
- Carvalho, L. A. V. (2005). *Datamining: a mineração de dados no marketing, medicina, economia, engenharia e administração*. Rio de Janeiro: Ciência Moderna.
- Cooper, J., Croxall, J.P., Fraser, W.R., Patterson, D.L., Poncet, S. e Woehler, E.J.; (1998). Breeding distribution and population status of the Northern Giant Petrel (*Macronectes halli*) and the Southern Giant Petrel (*M. giganteus*). *Marine Ornithology*, In Press.
- He, Haibo; Garcia, Eduardo A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, pages 1263-1284.
- Hofmann, M., Klinkenberg, R. (2013), *RapidMiner: Data Mining Use Cases and Business Analytics Applications*, CRC Press, 1st edition.
- Lennert-Cody, Cleridy E.; BERK, Richard A. (2007). Statistical learning procedures for monitoring regulatory compliance: an application to fisheries data. *Statistics in Society*, pages 671-689.

- Li, J.; Heap, A. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. *Ecological Informatics*, pages 228-241.
- Lucio, P. S. et al. (2007) Spatiotemporal monthly rainfall reconstruction via artificial neural network: case study: south of Brazil. *Advances in Geosciences*, pages 67-76.
- Marenzi, Maíra Carvalho. (2017). Análise espaço-temporal das capturas incidentais de mamíferos, tartarugas e aves marinhas por interação com a pesca de espinhel pelágico brasileira. Trabalho de Conclusão de Curso (Graduação de Oceanografia) - Universidade do Vale do Itajaí (UNIVALI), Itajaí.
- Pons, Maite et al. (2009). Machine learning procedures: an application to by-catch data of the marine turtles *Caretta Caretta*. *Collect. Vol. Sci. Pap. ICCAT*, pages 2443-2454.