INTEGRATION OF SATELLITE-BASED ENVIRONMENTAL DATA FOR SKIPJACK TUNA FISHING GROUND DETERMINATION

Adillah Alfatinah (1), Hone-Jay Chu (2)

[1] Spatial Analysis and Intelligent Management Laboratory, Geomatics Department, National Cheng Kung University, 1 University Road, 70101, Tainan City, Taiwan, R.O.C.

[2] Associate Professor, Spatial Analysis and Intelligent Management Laboratory, Geomatics Department, National Cheng Kung University, 1 University Road, 70101, Tainan City, Taiwan, R.O.C.

Email: adillah.alfatinah@gmail.com; honejaychu@geomatics.ncku.edu.tw

KEY WORDS: Skipjack Tuna, Remote Sensing, Environmental Parameters, Decision Tree (DT), Generalized Linear Model (GLM)

ABSTRACT: Skipjack Tuna fishing ground area was studied using satellite remotely sensed environment and catch data. Skipjack Tuna tend to aggregate in ocean areas that exhibit specific environmental conditions with water quality parameters such as Sea Surface Temperature (SST) and dissolved oxygen. Weekly resolved remotely sensed SST, surface chlorophyll (Chl-a), and Sea Surface Height (SSH) in 2017 being used as the environmental parameters to determine Skipjack tuna fishing ground. Machine Learning method which is Decision Tree (DT) were constructed with the environmental parameters as model covariates to examine Skipjack Tuna fishing ground determination. As a comparison, Generalized Linear Model (GLM) also was applied. This study exhibits the ability of both model performances for the fishing ground determination. Both models appear to be appropriate for fishing ground determination, while the DT acquires a better performance than GLM. The mean accuracy and Area Under Curve (AUC) of DT was about 0.876 and 0.877, respectively. In other hand, GLM only able to acquire 0.6907 and 0.8397 for its mean accuracy and AUC, respectively.

1. INTRODUCTION

Skipjack tuna (*Katsuwonus pelamis*) is a highly migratory pelagic species inhabiting all tropical and subtropical waters of the world’s oceans (Matsumoto, 1975). It is commercially valuable and exploited by many coastal countries along the Indian and the Pacific Oceans (Fonteneau et al, 2013). The species is commercially important, ranked among the first 10 species that have contributed highly to global catches in previous years (FAO, 2009). Skipjack tuna fish inhabits the upper layers of the ocean, spread into the water column above the thermocline, particularly below 100 m of the water column, rarely penetrating the thermocline layer. Even though they have a wide horizontal migration movement, it vertical movement were restricted caused by its limitation with several environmental parameter qualities. The recent studies concluded that Sea Surface Temperature (SST) and surface chlorophyll-a (Chl-a) are the most two important habitat predictors for skipjack tuna migration (Mugo et al, 2010). SST plays an important role in tuna physiology, and temperature variations are often linked with the biological richness of an oceanic area (Bertrand et al, 2002). In the oceanic environment, Chl-a is often considered as an index of biological productivity and it could be related to fish production (Gower, 1972). A Combination of the SST and Chl-a can complement one another for determining potential fishing ground for tuna (Zainuddin et al, 2004).

Sustainable and effective method required for an effective monitoring and management of marine resources. Conventional approaches of sampling the ocean using research vessels are limited in both time and space scales of coverage, making it difficult to study entire ecosystems. Since the advent of satellite remote sensing, especially remote sensing of ocean color and temperature, it has become possible to sample the global ocean over large areas and with acceptable temporal resolutions (Klemas, 2012). Satellite remote sensing has been an important technique in fishery research, management and harvesting, because it provides synoptic ocean measurements for evaluating environmental influences on the abundance and distribution of fish populations and allows ecological analyses at community and ecosystem scales (Chassot et al., 2011; Stuart et al., 2011). There were many practical fisheries-related application of environmental remotely sensed data for efficient management and controlled exploitation of marine resources. In the previous study about the remote sensing data application in marine resources monitoring discussed by (Mugo et al., 2010), he provided two case. One case was about the application of remote sensing environmental data and vessel monitoring technology in a skipjack tuna fishery in the western North Pacific by including a simple algorithm, for determining fishing activity from vessel speed. For another case, by using the remote sensing data they focus on the impact of climate change on scallop (*Mizuhopecten yessoensis*) aquaculture in Funka Bay, Hokkaido, Japan. Valavanis et al., (2008) modelled fish habitat using fishery and environmental remote sensing dataset to demonstrate that the identification of oceanographic features such as Potential Fishing Zone (PFZs) is feasible in different oceans basins. Another study by Hartog et al (2011) also considered how management decisions based on SRS-derived habitat preferences of southern bluefin and yellowfin tuna (*Thunnus albacares*) could be influenced by ocean warming. These examples proved if environmental remote sensing data has been taken from the recent past for the marine resources monitoring.

As mentioned previously, Skipjack tuna aggregation was affected by environmental marine parameters and by taking it into consideration, this research aims to determinate the fishing ground by integrating remote sensing environmental data. Considering their custom to inhabit the upper layers of the ocean, Sea Surface Height (SSH) also being taken as one of the parameter beside SST and Chl-a. Furthermore, two different method wills be executed to obtain this aim. First method is using machine learning, with the environmental parameters as model covariates, to examined and finding out which parameter affected skipjack tuna aggregation the most. And for the comparison, the Generalized Linear Model (GLM) method will also applied as the second method. This study will be executed using all data from the first week of January 2013 and expected to exhibits the ability of both model performances for the fishing ground determination.

2. MATERIAL AND METHODS

Work on case study was conducted in Southwest part of North Pacific Ocean (30^{o}S -10^{o}N 140-180^{o}E).

2.1 Skipjack Tuna Fishery Data

Catch points of skipjack tuna from 2013 until 2017 was obtained for pursuing this research, however as previously stated, the models performances will be tested solely using data from the first week of January 2013. There were 70 catch points of skipjack tuna being recorded during this period and being plotted as displayed in Figure 1.

The catch points data consist of latitude and longitude position and fish total catch of the fishing vessel. A spatial filter excluded all points positioned out of the boundary (within study area). The points then will be processed using kernel density function in ArcMap 10.5 with the intention of getting the possible fishing area. The area with high density will be assumed as Fish Area, while the area with less density will be No Fish Area.

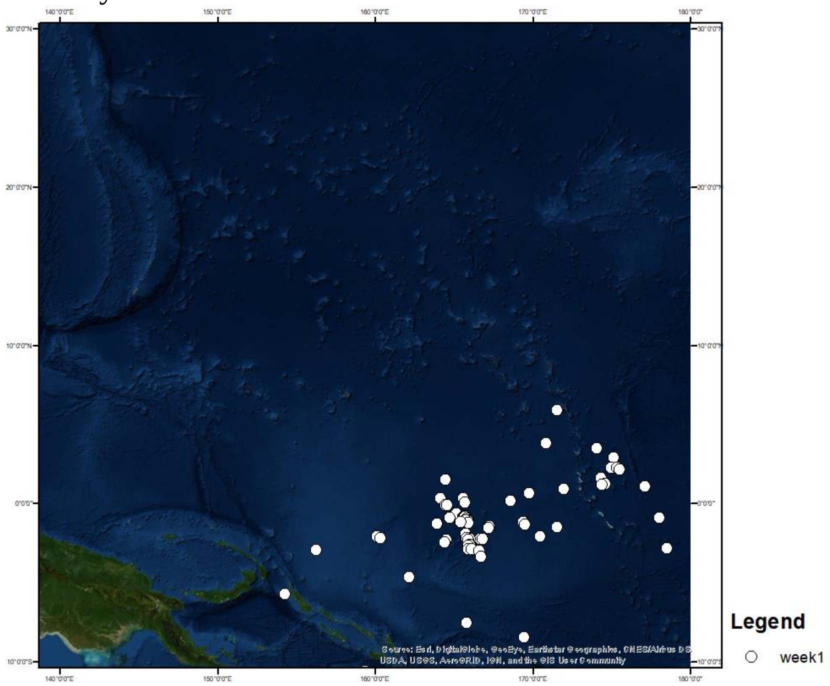

Figure 1. Recorded Catch Points of Skipjack Tuna
(c ESRI Image Copyright)

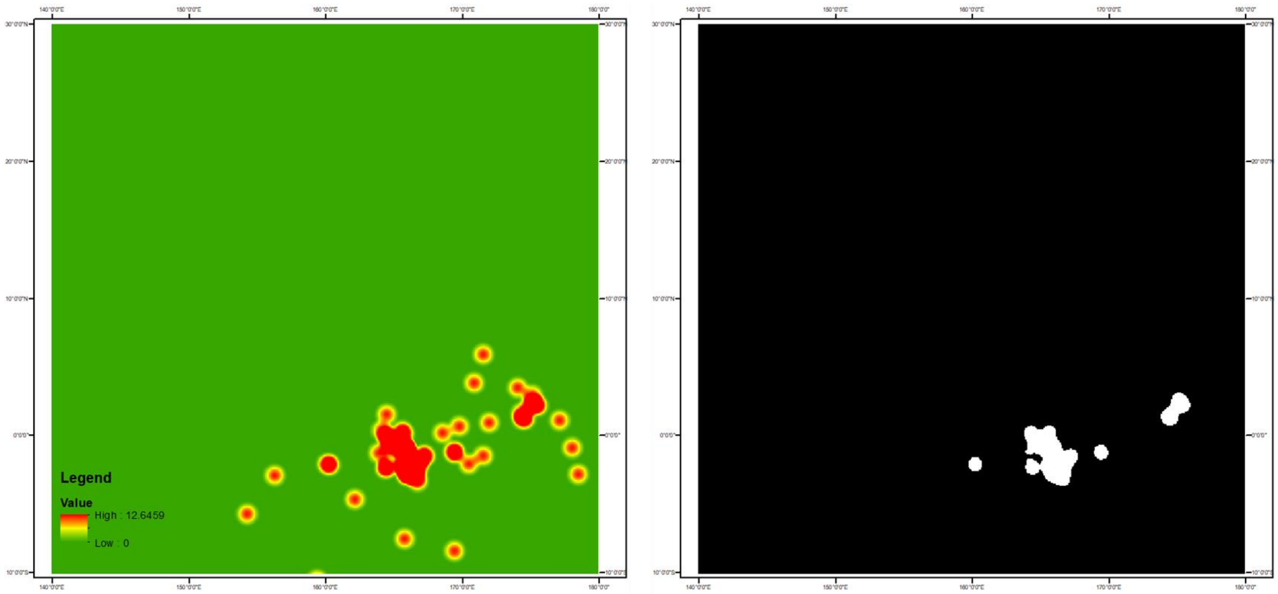

Figure 2. (a) Density Map based on Fish Catch Point; (b) Fish and No Fish Area Probability Map

Figure 2 showed the density and fish area probability map based on the catch point. In the fish area map, black color area indicating No Fish Area, while the white color indicating the Fish Area.

2.2 Remotely Sensed Environmental Data

Remotely sensed environmental data used to describe the oceanographic conditions around the skipjack tuna fishing grounds were SST, Chl-a and SSH. All SST and Chl-a data were estimated from Aqua/MODIS Standard Mapped Image (SMI) level 3 binary data with daily average temporal resolution and 9 km of spatial resolution. The delayed time and merged SSH product obtained from Global Ocean Physics Reanalysis provided by Copernicus Marine Environment Monitoring Service (http://marine.copernicus.eu/) with daily-mean temporal resolution and 0.083^o x 0.083^o spatial resolution. All the data recorded in the same period with the fishery data.

The environmental dataset will be acted as model covariates variables in both prediction methods to acquire the most possible fishing ground area. Those environmental data later will be resampled to seize them into the same cell size. Figure 3 displayed the environmental remotely sensing data at the first week of January 2013. SST value range around 20.5^{o}C until 31.6^{o}C, with the temperature near-shore appear to be warmer than off-shore. Similar case also happened to Chl-a, with the highest concentration located in the on-shore.

In the Chl-a satellite data (Figure 3c), we can detect several areas with null value appear as white color. This instance caused by the inability of satellite sensor to capture the reflected Chl-a value because of several circumstances. In sake of model construction, it better to avoid the null value to optimizing the model performance.

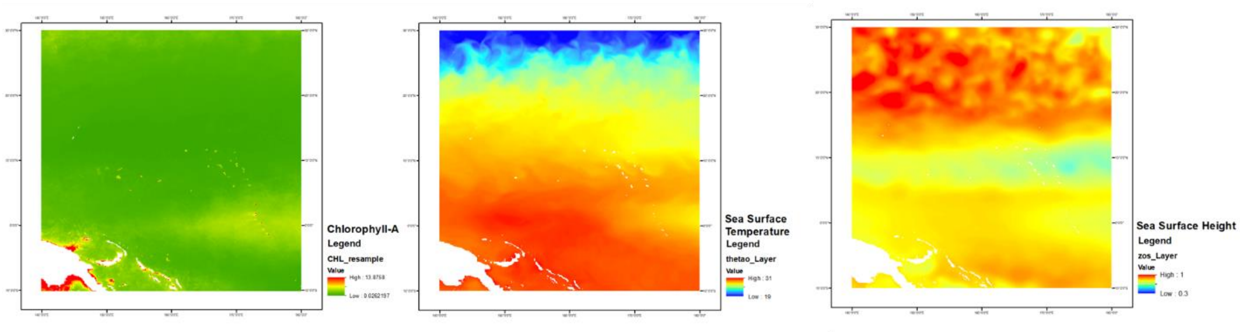

Figure 3. Environmental Remotely Sensed Data (a) SST; (b) SSH; and (c) Chl-A
(c MODIS Image Copyright 2013 NASA; CMEMS Image Copyright 2013)

2.3 Construction of Decision Tree and GLM

As described previously, the catch fish points then used to generate fish map probability which consist of Fish Area and No Fish Area. Several points will be located on both Fish and No Fish Area. These points will be referred as Fish Class points from now on. The points then will be separated in two, one served as training points for model construction sake, while another being used as the testing points for checking the model performance. There were 298 and 118 total points for training and testing points, respectively. Furthermore, by using the training and testing points, the environmental data value also being extracted and stacked. Data filtering was needed in this process for convincing the null value already excluded from the both dataset.

In this study, the training model construction using machine learning method will be executed using Decision Tree. Decision tree is a data mining technique for solving classification and prediction problems. Data mining consists of different methods and algorithms used for discovering the knowledge from large data sets (Šebalj et al, 2017). Decision trees are used for solving classification, as well as regression problems. When a decision tree is used for classification tasks, it is most commonly referred to as a classification tree, and when it is used for regression tasks, it is called a regression tree. Speaking of the classification problems, the learning scheme is presented with a set of classified examples (training set) from which it is expected to learn a way of classifying unseen examples (testing set). Decision tree has a simple hierarchical structure easy to understand, consisted of nodes and leaves. Each node in the tree involves testing a particular attribute and each leaf of the tree denotes a class. Decision tree classifies instances by sorting them down the tree from the root to some leaf node, which gives a classification that applies to all instances that reach the leaf. The tree complexity is measured by one of the following metrics: the total number of nodes, total number of leaves, tree depth and number of attributes used (Hssina et al, 2014; Rokach & Maimon, 2014; Witten et al, 2011). The goal of the decision tree is to create a training model that can predict the classes or values of target variables by learning decision rules inferred from training data. Following its goal, the training data consist of environmental data will served as independent variables or target variables and Fish Class points will be the dependent variables. Decision tree procedure was run under default parameters within the '*r-part*' package in Program R.

As mentioned, GLM also being conducted for model performance comparison. Generalized linear models (GLMs) are a means of modeling the relationship between a variable whose outcome we wish to predict and one or more explanatory variables. The predicted variable is called the target variable and is denoted y (Goldburd et al, 2008). GLMs model the relationship between μ_i (the model prediction) and the predictors as follows:

$$g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} \quad (1)$$

Equation 1 states that some specified transformation of μ_i (denoted $g(\mu_i)$) is equal to the intercept (denoted β_0) plus a linear combination of the predictors and the coefficients, which are denoted $\beta_1 \ldots \beta_p$. By applying this concept, environmental data acted as predictor variables which will calculate the value of $g(\mu_i)$ as the model prediction. Similar with decision tree, the GLM also being proceed in Program-R using '*glm*' package.

3. RESULTS AND DISCUSSION

The models performance was applied in the same time, which is the first week of January 2017. Based on the constructed decision tree showed in Figure 2, we obtained SST as the most impacting variables in the model construction followed by SSH and Chl-a, while for GLM, Chl-a was the most impacting variable followed by SST and SSH.

Results from both model showed their capability to determinate skipjack tuna fishing ground and the maps generated by the models displayed in Figure 4. With the intention to knowing the fittest model with this study purpose, several accuracy assessment being carried out by using the testing points. We then compared the confusion matrix, Area Under Curve (AUC) and mean accuracy from them to find out which one having a better performance. Comparison between the mean accuracy and AUC value showed in Table 1 and the confusion matrix of their performance showed in Table 2. Based on the accuracy assessment, we stated if decision tree able to exceed GLM performance with AUC value and mean accuracy about 0.876 and 0.931, respectively. GLM only acquired 0.690 and 0.860 for its accuracy and AUC value.

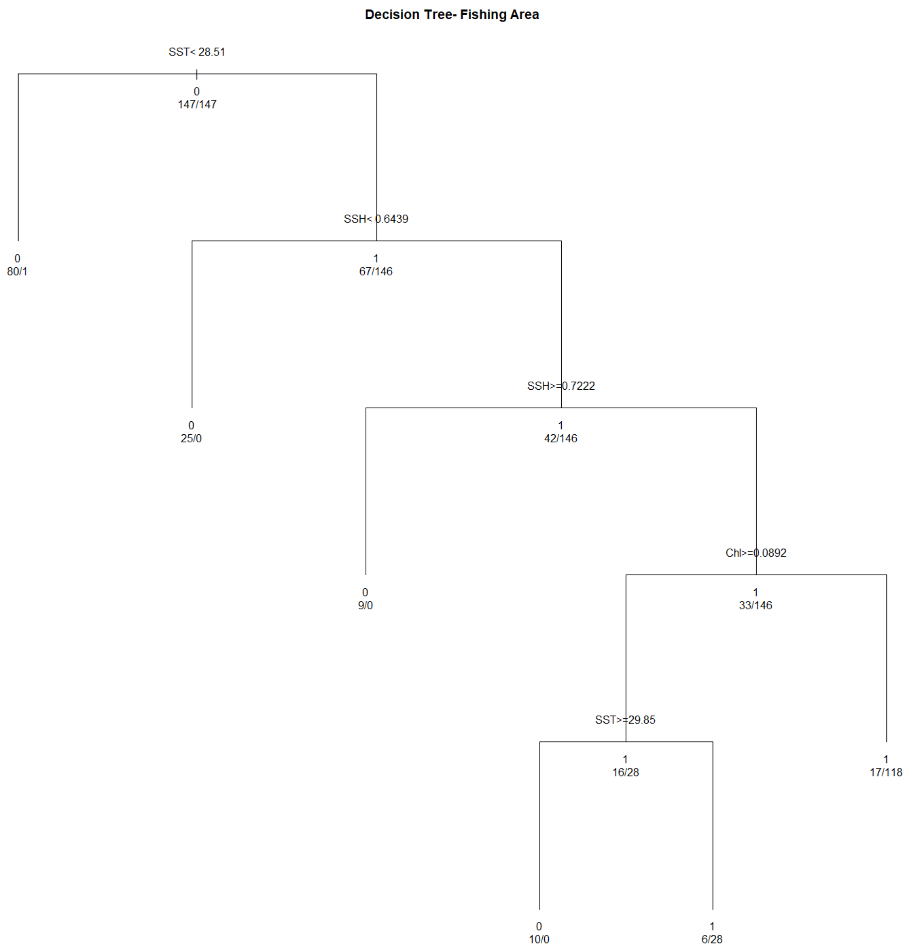

Figure 4. Decision Tree Construction

Table 1. Accuracy Assessment Comparison

Prediction Model	Mean Accuracy	AUC
Decision Tree	0.876	0.931
GLM	0.690	0.860

Table 2. Confusion Matrix Model Performances

<table>
<tr><td rowspan="2">Models</td><td rowspan="2" colspan="2"></td><td colspan="2">Observed</td></tr>
<tr><td>Fish Area</td><td>No Fish Area</td></tr>
<tr><td rowspan="2">Decision Tree</td><td rowspan="4">Predicted</td><td>Fish Area</td><td>73</td><td>0</td></tr>
<tr><td>No Fish Area</td><td>6</td><td>38</td></tr>
<tr><td rowspan="2">GLM</td><td>Fish Area</td><td>30</td><td>4</td></tr>
<tr><td>No Fish Area</td><td>43</td><td>40</td></tr>
</table>

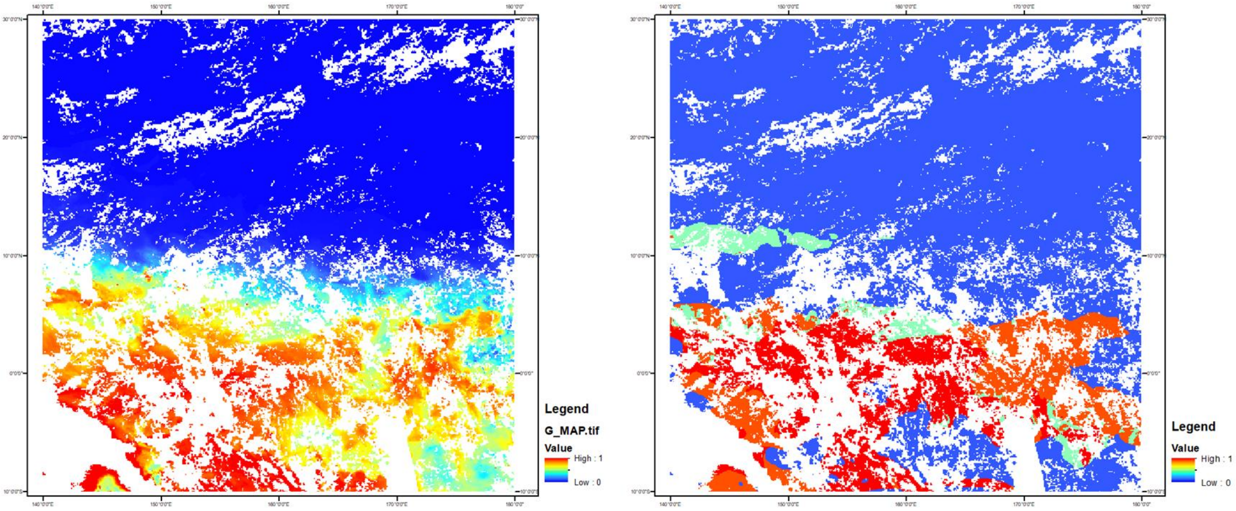

Figure 5. (a) Fishing Ground Prediction using Decision Tree; (b) using GLM

4. CONCLUSION

Based on study in the recent past, skipjack tuna aggregation appears to be severely affected by marine environmental parameter. Besides, the catch points of skipjack tuna displayed in Figure 1 indicate if the catching activity by the vessel happen in area with SSH lower than 0.5 m. Each variable appears to have different importance level in each model construction. This was likely reasonable happened caused by the models different algorithm in predicting. From the accuracy assessment of each model, we can state if decision tree performance surpassed GLM, proving it as the fittest model to be used for this study purpose. But other than that, fishing ground probability maps showed in Figure 5 still indicated area with null data value. This issue occurred due to null value in the Chl-a data. in order to get a better result, it best to obtain the data with less null value in the future

REFERENCES

Bertrand, A., Josse, E., Bach, P., Gros, P., &Dagorn, L. (2002). Hydrological and trophic characteristics of tuna habitat: consequences on tuna distribution and longline catchability. *Canadian Journal of Fisheries and Aquatic Sciences*, *59*(6), 1002–1013. https://doi.org/10.1139/f02-073

Chassot, E., Bonhommeau, S., Reygondeau, G., Nieto, K., Polovina, J. J., Huret, M., …Demarcq, H. (2011). Satellite remote sensing for an ecosystem approach to fisheries management. *ICES Journal of Marine Science*, *68*(4), 651–666. https://doi.org/10.1093/icesjms/fsq195

FAO. (2009). *The State of World Fisheries and Aquaculture 2008*. Rome: FAO.

Fonteneau, Alain, Chassot, Emmanuel, &Bodin, Nathalie. (2013). Global spatio-temporal patterns in tropical tuna purse seine fisheries on drifting fish aggregating devices (DFADs): Taking a historical perspective to inform current challenges. *Aquat. Living Resour.*, *26*(1), 37–48. https://doi.org/10.1051/alr/2013046

Goldburd, M., Khare, A., &Tevet, D. (2008). *Generalized Linear Models for Insurance*

Rating. *CAS Monograph*. Casualty Actuarial Society. https://doi.org/10.1017/CBO9780511755408

Gower, J. F. R. (1972). *A Survey of the Uses of Remote Sensing from Aircraft and Satellites in Oceanography and Hydrography*. Sidney.

Hartog, J. R., Hobday, A. J., Matear, R., &Feng, M. (2011). Habitat overlap between southern bluefin tuna and yellowfin tuna in the east coast longline fishery – implications for present and future spatial management. *Deep Sea Research Part II: Topical Studies in Oceanography*, *58*(5), 746–752. https://doi.org/https://doi.org/10.1016/j.dsr2.2010.06.005

Hssina, B., Merbouha, A., Ezzikouri, H., &Erritali, M. (2014). A comparative study of decision tree ID3 and C4.5. *International Journal of Advanced Computer Science and Applications*, *4*(2), 13–19. https://doi.org/10.14569/specialissue.2014.040203

Klemas, V. (2012). Remote sensing of environmental indicators of potential fish aggregation: An overview. *Baltica*, *25*(2), 99–112. https://doi.org/10.5200/baltica.2012.25.10

Matsumoto, W. M. (1975). Distribution, relative abundance, and movement of skipjack tuna, Katsuwonus pelamis, in the Pacific Ocean based on Japanese tuna longline catches, 1964-67. National Oceanic and Atmospheric Administration.

Mugo, R., Saitoh, S. I., Nihira, A., &Kuroyama, T. (2010). Habitat characteristics of skipjack tuna (Katsuwonus pelamis) in the western North Pacific: a remote sensing perspective. *Fisheries Oceanography*, *19*(5), 382–396. https://doi.org/10.1111/j.1365-2419.2010.00552.x

Rokach, L., &Maimon, O. (2014). *Data Mining With Decision Trees: Theory and Applications* (2nd ed.). River Edge, NJ, USA: World Scientific Publishing Co., Inc.

Šebalj, D., Franjkovic, J., &Hodak, K. (2017). Shopping intention prediction using decision trees. *Millenium - Journal of Education, Technologies, and Health*, *2*, 13–22. https://doi.org/10.29352/mill0204.01.00155

Stuart, V., Platt, T., &Sathyendranath, S. (2011). The future of fisheries science in management: a remote-sensing perspective. *ICES Journal of Marine Science*, *68*(4), 644–650. https://doi.org/10.1093/icesjms/fsq200

Valavanis, V. D., Pierce, G. J., Zuur, A. F., Palialexis, A., Saveliev, A., Katara, I., &Wang, J. (2008). Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS. *Hydrobiologia*, *612*(1), 5–20. https://doi.org/10.1007/s10750-008-9493-y

Witten, I. H., Frank, E., &Hall, M. A. (2011). Chapter 1 - What's It All About? In I. H.Witten, E.Frank, &M. A. B. T.-D. M. P. M. L. T. and T. (Third E.Hall (Eds.), *The Morgan Kaufmann Series in Data Management Systems* (pp. 3–38). Boston: Morgan Kaufmann. https://doi.org/https://doi.org/10.1016/B978-0-12-374856-0.00001-8

Zainuddin, M, &Jamal, M. (2009). Satellite remote sensing and geographic information system of potential fishing zones and migration pattern of Skipjack Tuna in Bone Bay, South Sulawesi. In *International Proceeding of World Ocean Conference*.

Zainuddin, Mukti, Saitoh, S.-I., &Saitoh, K. (2004). Detection of potential fishing ground for albacore tuna using synoptic measurements of ocean color and thermal remote sensing in the northwestern North Pacific. *Geophysical Research Letters*, *31*.