Predicting Global Ship Demolition Using Machine Learning Approach

S. M. Mizanur Rahman

a University of Bordeaux, CNRS, Arts et Metiers Institute of Technology, Bordeaux INP, INRAE, I2M Bordeaux, F-33400 Talence, France

Abstract:

Global ship demolition is mostly concentrated in south Asian countries, namely Bangladesh, India, Pakistan and China, since 1990’s, having competitive advantage for their high natural tide, and low environmental and social costs. Due to high social and environmental externalities, stakeholders increase monitoring of the externalities and continue to prescribe improvement towards sustainability, which put pressures on profitability and competitiveness. As a consequence, also seen in the past, a leakage effect may emerge, leading to shift of this activity to a region, with relatively less monitored and less strict on social and environmental impacts. Unfortunately, the leakage effect is never predicted in shipbreaking in order to understand the level of push compatible in the given socio-economic contexts. In this study, we have attempted to predict the future ship demolition landscape, applying machine learning technique to 34,531 in-service vessels worldwide, larger than 500 gross tonnage (GT), which is run against a learning model based on 3500 demolished vessels from 2014. This study shows that redistribution may occur among the top recycling nations: India may emerge out to be a dominant player in shipbreaking, surpassing Bangladesh by a margin of two-fold, while Pakistan and China are in decreasing trend. In addition, the leakage effect is observed, in that Vietnam is predicted to be the fourth largest ship demolition country, while China and Pakistan recede from the third and fourth place to 6th and 8th. Turkey is predicted to advance from fifth position to third position by vessel count but stays same in term of total GT dismantled. Although it is not clear if any leakage is to be observed in the near future, this study may be a
model for future predictive analytics and help stakeholders take evidence-based business decisions.

Keywords: Ship Recycling, Predictive Analytics, Big Data, Shipbreaking, Leakage Effect

Introduction:

Globally, in 2018, about 50,000 ocean-going vessels are commercially transporting more than 90% of the goods traded every year. After about thirty years of age, shipowners decide to dispose them at the recycling nations, an overwhelming majority (about 90% of the total ships) are dismantled in the south Asian nations. Bangladesh, India and Pakistan together constitute about 70% of the total ship recycling while China and Turkey constitute about 15-20% and rest of the world 10-15 percent (Demaria 2010). South Asian countries mostly dismantle large ships while Turkey and other EU countries dismantled smaller ships, demonstrating the capacity limitation that, among other factors, determines where a particular type of ships will be recycled (Rahman and Kim, 2019). Largely, three reasons explains why shipbreaking occurs in the south Asian region in the first place: 1) high natural tide that allows beaching easily and gives a competitive advantage for the yard managers, saving a high portion of investment cost, 2) low social and environmental compliance costs 3) lack of natural iron ore that makes these nations variably dependent on the steel scrap originated from the ship recycling (Rahman and Mayer 2015). From shipowner perspective, due to high scrap demand in the south Asian nations, selling ships in the south Asian nations is profitable compared to ships sold in standard facilities such as China and Turkey.

During 1950s, ships used to be dismantled in the developed countries, namely Japan, Korea, Taiwan and US (Figure 1). The high environmental compliance costs in those nations cause
this activity to move to a south Asian region in 2000s that have usually lower compliance costs. Overall, the resulting move worsened the environmental and social sustainability of the industry.

Figure 1: geographic shift of shipbreaking nations from 1975, adapted from alcaidea et al. 2016

International policy contexts and non-governmental organizations (NGOs) increasingly monitor the dangerous occupational and environmental aspects of the shipbreaking sites in the south Asian nations (NGO shipbreaking platform, 2019). For example, Hong Kong Convention (HKC) asks for a better coordination among stakeholders and attempt to introduce certification formalities in order to get the EOL ships free from hazardous content and to have it recycled in a risk-free condition (Rahman and Mayer 2016). EU ship recycling policy restrict owners not to sell their ships in substandard nations, known for their non-compliance. NGOs (Basel Ban, NGOs Shipbreaking platform, FIDH and others) are continuously reporting occupational accidents and casualties, forcing national and business level managers under pressure for improvement. Government and policy makers of the shipbreaking nations tend to formulate
policies to improve the occupational and environmental aspects, but the level of improvement is ever to reach at an accepted level, due to reportedly the uncertainty of profit making and poor law enforcement. Amid this heightened pressure for sustainability that will raise costs, is there any possibility that shipbreaking in south Asian region will lose competitive advantage and eventually shift to a new region of low environmental and social compliance?

To respond to this question, Cairns et al. (2014) evaluated the alternative scenario using qualitative scenario analysis method and conclude that any shift of this activities will worsen the socio-economic conditions of the host nations. Knapp et al. (2007) conducted econometric analysis in 2007 to identify the dynamics of market variables applying economic models. With this model, different relationship is tested against the dependent variable – decisions to scrapping. He conducted Logistic Regression (LR) against the recycling destinations, suggesting that the size matters in the selection of recycling destinations. For example, Bangladesh is more inclined to recycle larger ships and tanker. While Turkey and China mostly recycle smaller ships and India recycle relatively modest ships. In addition, they found that Bangladesh is more sensitive to market conditions than the other demolition nations. However, to the best of author knowledge, no study focused on the leakage potential of this activity. In this study, we therefore aim to predict future ship demolition based on sophisticated statistical algorithms and analyze how the future demolition landscape looks like.

Knapp et al (2008) conducted research to estimate the probability of scrapping of an in-service ship and that how the market features, ship features and system features determines the probability of scrapping. He did discuss however, the possibility of demolition locations based on those variables and also based on flagged countries variation. One of the variables they chose is the scrap price. This variable represents the market price of the scrap steel. The vessel selling process for scrapping does depend on the steel scrap price in the market. However, the buying price of vessel/ actual selling price of the vessel for scrapping differs heavily on the country, in
which vessel will be dismantled. This variation is somewhat captured in steel price variable when steel price is considered as a temporal pattern within a country case. For example, Turkey purchased a ship by 200 dollars in 2005 and a ship by 250 dollars in 2010. This is fine for the dependent variable of scrapping probably determination. But for the determination of which location the vessel has the possibility to breaking, steel scrap price is not enough. This require inter country price difference of the same vessel and thus, this study is contributing to more nuanced understanding of the factors underlying the location selection decisions.

The objective of this study is twofold: (1) to identify if leakage potentials exist and how the change look like and (2) to examine how realistic it is to apply the machine learning approach in this phenomenon.

**Related literature of big data and predictive analytics:**

Big data and predictive analytics (BPDA) are applied to enhance supply chain sustainability as well as positively impact on corporate financial performances (Hazen et al. 2016). Research also revealed that how BPDA can influence on attaining environmental and social sustainability is poorly understood and require an immediate attention. The use of advance prediction tools for converting data into information can explain uncertainties and help policy makers make informed decisions (Lee et al. 2014).

In the context of poor transparency and fuzzy international boundaries in which policy enforcement deemed to be challenging, routinized application of BPDA can help improve supply chain loopholes and organizational performances (Gunasekaran et al. 2017). For example, the results of BPDA can reduce supply chain management costs, achieve high efficiency, balance relationships among suppliers, respond quickly to specific events, provide planning capabilities and signal potential abuse.
Despite clear advantages of applying BPDA in the supply chain related problems, there is a lack of necessary skills that can merge data scientist skills and domain knowledge (Waller and Fawcett 2013). In addition, Hazen et al. (2014) presented data quality issues when dealing with advance prediction tools and proposed methods for monitoring and controlling data quality in the context of supply chain management. They mentioned that poor data quality costs company’s revenues by 8-12% and losses 40-60% of organization’s services. The dimensions of data quality are intrinsic (accuracy, timeliness, consistency and completeness) and contextual (relevancy, value-added, quantity, believability, accessibility, and reputation of the data). The quality-check of the dimensions of the data is addressed in the table (1).

Table 1: Challenges of data quality issues and approaches adopted to mitigate them in this study, adapted from Hazen et al. (2014)

<table>
<thead>
<tr>
<th>categories</th>
<th>dimensions</th>
<th>Descriptions</th>
<th>Level</th>
<th>Approaches adopted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic</td>
<td>Accuracy</td>
<td>Degree to which data are equivalent to the real values</td>
<td>High</td>
<td>Comparing two data sets from different sources. Individual ship data online</td>
</tr>
<tr>
<td></td>
<td>Timeliness</td>
<td>Degree to which data are updated</td>
<td>High</td>
<td>Marketplace data and more recent; frequently added every year</td>
</tr>
<tr>
<td></td>
<td>Consistency</td>
<td>Format and structure matching</td>
<td>Medium</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Completeness</td>
<td>Complete in content; no missing data</td>
<td>Medium</td>
<td>Important price variable missing</td>
</tr>
<tr>
<td>Contextual Relevance</td>
<td>Degree to which attributes are related to problem in hand</td>
<td>Medium</td>
<td>Price variable</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------------------------------------------</td>
<td>--------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>Value-added</td>
<td>How this data contributes to value generation</td>
<td>Medium</td>
<td>Shipowner decision making, policy level decision making, sustainability progress</td>
<td></td>
</tr>
<tr>
<td>Quantity</td>
<td>-</td>
<td>-</td>
<td>Train data limited; test data complete</td>
<td></td>
</tr>
<tr>
<td>Believability</td>
<td>How credible the sources are</td>
<td>High</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>accessibility</td>
<td>Are the data open or under paywall</td>
<td>-</td>
<td>One set of train data is public.</td>
<td></td>
</tr>
<tr>
<td>Reputation</td>
<td>How reputed the date generated company</td>
<td>High</td>
<td>Both data generators are well known</td>
<td></td>
</tr>
</tbody>
</table>

**Method:**

In this study, we have applied machine learning technique to predict ship demolition nations.

Detail description of how machine learning works is given below.

*Machine learning (ML):*
ML manages to discover complex structure from the data and yet, is flexible enough to choose functions that best fit the data. ML does not intend to produce regression coefficient (beta) between y and x. ML is technically easily accessible with conventional packages in R or Python built to deal with multiple techniques, for example, decision tree (DT), Random Forest Classifier (RFC), LASSO (Least Absolute Shrinkage and Selection Operator) regression coefficients. Easy accessibility may lead to naïve application and wrong output interpretation.

How does ML work? Supervised ML look for functions that predict well out of the given sample. For example, if we want to predict the value y of a house from input variables x on a sample (yi, xi). The algorithm finds a loss function \( L(y \text{ expected}, y) \) as an input and search for a function that has this low expected prediction loss. ML takes into account pairwise interactions among variables automatically through the splits of node and tend to be fit superbly. This high dimensionality is an important feature of ML that generates highly flexible functional forms in order to fit the varied structure of data. But picking up best functions is also a terrible choice that can be minimized by regularization, meaning that depth of the trees will be chosen in order not to be overfit.

Choosing the best regularization parameters, we apply cross validation approach that holds out one part of the sample for evaluation while fitting the prediction function for a range of regularization parameters for the remaining data. We conducted cross validation on LR, DT, RFC, SVM and XGBOOST. Each tree is fitted on an original set and transformed to a test sample with a chosen set of variables. The prediction of trees is estimated on an average. The regularization variables in this algorithm are: Complexity of each individual tree depth, number of variables used in each tree and size of the sample and the number of trees.

Data:
This study uses Clarkson.net data for both in-service and demolition data. Clarkson applies tracking technology and document the global vessel data in terms of owner, builder, flag country and demolition nation. In-service vessel and demolition dataset contain 48000 (from 2002 to 2018, April) and 3901 vessels (from 2014 to 2018, April). In service data is further refined and considered vessels that are larger than 500 GT. This reduces the dataset to 34531. Seven variables are considered from in-service data, which are Type, GT, Built year, Length, Builder nation, Owner nation, Flag nation. For the demolition dataset, 8 variables are considered with demolition nation as a target variable.

It appears that selling price of a vessel is a strong determinant on where the ships will be dismantled, however, the demolitions dataset contains only 1100 items out of total 3500, which is about one third of the total vessels. In order to use that variable, either one has to fill the null values by the average value of the ship type or demolition nation type. We did this by first applying group by function in python based on the demolition nation data that did only fill half of the null value. The remained fill-up is done by applying group by based on the average price value of the type of vessel. Alternatively, we look for similar dataset that have more price value but to no avail. For example, NGO shipbreaking platform contains ---dataset from 2012 to 2018, but those are highly inconsistent and lack price values. Given the crucial need of the use of the price variable, we have conducted RFC model on two different datasets: one without price variable and the other with price variable. The accuracy level for the model with price data yields more than 80% while it is only 60% for the model without price variable. Despite this importance, price data is still not useful in the model as a selling price cannot be attributed to the in-service vessels (Figure 2).
Descriptive analysis:

*GT vs Length of vessel:*

The in-service dataset and the demolition dataset are plotted GT mean against the length of the vessel. Both show similar relationship in that GT is increased as the length increases. But, it seems that GT per length is greater in the in-service dataset (Figure 3 (a) and (b)).
Figure 3: Scatterplot of GT against vessel length of the two datasets: a) in-service and b) demolition data set

*Number of vessels build each year:*

The figure (4) below shows that the number of vessels increased until 2010 from 2002 and then, begins to decline again at the same rate as it increased till 2017. In 2002, number of vessels added to the in-service vessels is 1000 and in 2010, the number is 3000, while in 2017, it is about 1300. The decline effect is probably due to the effect of financial crisis when the transport of goods and services reduced due to the drop of the demand of those goods.

![Graph showing the number of vessels built each year from 2002 to 2018.](image)

Figure 4: Number of vessels built every year from 2002 to 2018(April)

*Average GT per year:*

The figure (5) shows the average size of the vessel by year. It shows that till the year, 2010, the average GT (mean) was fairly constant which is then started to increase.
Collinearity among variables in demolition data and in-service data:

In in-service data, among the seven variables, apart from the GT and vessel length, other variables are very weakly correlated with values less than 0.1 in figure (6a). However, as expected, GT and vessel length are highly correlated with value more than 0.9. Flag nation and Owner nationality is observed to be quite related with a value of 0.52. Builder nationality is found to be related with flag nation and owner nation with value of 0.28 and 0.31. As expected, built year is found to be weakly related with other variables. However, and quite surprisingly, ship type is also weakly related with other variables with values less than 0.1, which slightly differs to the present knowledge on the correlation, for example, between ship type and destination nation. For example, oil tanker has relatively higher tendency to go to Bangladesh as this tends to be bigger and is having solid steel content than other types of vessels.
Looking at the demolition data, it is found that owner nations and flag nations have quite a high collinearity (0.4) as that for vessel length and GT. Other variables are not of great concern with low collinearity among them (Figure 6b).

![Figure 6: Multicollinearity of the variables of the in-service data (a) and demolition data (b)](image)

**Method application:**

This study applied machine learning approach using sci-kit library and spark. This study applied LR, DT, RFC and support vector machine (SVM) algorithm for fitting and predicting ship demolition nation (Table 2). The demolition data is fitted and then the in-service data is tested in the model for the prediction. The predicted code is then uncoded to return the demolition nation. The precision and accuracy is documented for each algorithm for the demolition dataset by a 30% random split. Using spark, these algorithms are applied ,with one additional model named K-means clustering. The classification report largely validate the results of the sci-kit learn model, assuring the minimal differences between the technology used for the modelling, detailed in the following section.

Table2: Classification result of different algorithm applied
<table>
<thead>
<tr>
<th>Model type</th>
<th>Data type</th>
<th>Test type</th>
<th>Classification report (precision)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>Demolition data</td>
<td>Train test</td>
<td>30%</td>
</tr>
<tr>
<td>DT</td>
<td>Demolition data</td>
<td>Train test</td>
<td>50%</td>
</tr>
<tr>
<td>RFC</td>
<td>Demolition data</td>
<td>Train test</td>
<td>59%</td>
</tr>
<tr>
<td>RFC</td>
<td>Demolition and In-service data</td>
<td>Fit- Demolition data</td>
<td>Predict- In-service data</td>
</tr>
<tr>
<td>SVM</td>
<td>Demolition data</td>
<td>Train test</td>
<td>58%</td>
</tr>
<tr>
<td>SVM</td>
<td>Demolition and In-service data</td>
<td>Fit- Demolition data</td>
<td>Predict- In-service data</td>
</tr>
<tr>
<td>Kmeans cluster</td>
<td>In-service data</td>
<td>Number of clusters</td>
<td>Root mean square error reduces with increasing number of K</td>
</tr>
</tbody>
</table>

**Results:**

*Train and test split and model precision:*
Table 3 shows that the four different models are run for the selection of the best model that can be run to the in-service dataset for the prediction. LR model provides a precision score 0.34 for a 1171 dataset. The recall score is slightly higher (0.40). DT (0.49) performs better than LR. RFC (0.59) performs even better than the DT model. SVM (0.59) performs somewhat similar to RFC. However, recall score is lower, which makes RFC a better model for this dataset than SVM, DT and LR.

Modelling same datasets in another powerful technology called spark demonstrated similar accuracy. Precisely, LR performs better from 34% to 57%, while DT and RFC perform slightly worse from 49% to 48% and from 59% to 49%. The results show consistency in the prediction pattern across algorithmic models and across technologies. This also points out that the limiting factor for not having higher accuracy may be data due to the lack of additional data of more relevant variables, for example selling price of the vessels.

Table 3: Classification report of the models applied in the demolition dataset.

<table>
<thead>
<tr>
<th>Model</th>
<th>Technology</th>
<th>Precision/Accuracy</th>
<th>Recall</th>
<th>F1-score</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>Sci-kit learn</td>
<td>0.34</td>
<td>0.40</td>
<td>0.33</td>
<td>1171</td>
</tr>
<tr>
<td>LR</td>
<td>Spark</td>
<td>57</td>
<td>NA</td>
<td>NA</td>
<td>1171</td>
</tr>
<tr>
<td>DT</td>
<td>Sci-kit learn</td>
<td>0.49</td>
<td>0.50</td>
<td>0.50</td>
<td>1171</td>
</tr>
<tr>
<td>DT</td>
<td>Spark</td>
<td>48</td>
<td>NA</td>
<td>NA</td>
<td>1171</td>
</tr>
<tr>
<td>RFC</td>
<td>Sci-kit learn</td>
<td>0.59</td>
<td>0.61</td>
<td>0.60</td>
<td>1171</td>
</tr>
<tr>
<td>RFC</td>
<td>Spark</td>
<td>49</td>
<td>NA</td>
<td>NA</td>
<td>1171</td>
</tr>
<tr>
<td>SVM</td>
<td>Sci-kit learn</td>
<td>0.59</td>
<td>0.41</td>
<td>0.32</td>
<td>1171</td>
</tr>
<tr>
<td>&quot;_&quot;</td>
<td>-</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1171</td>
</tr>
</tbody>
</table>
Feature importance:

Only about 1135 data points were existed with selling price of the vessels, while the rest is made up first by filling average value based on the destination nations. The rest were filled by the average selling price by type of vessels. Thus, the selling price information is highly uncertain. With this data deficit, we can still see the influence of this variable over the other variables in choosing the demolition nation. When considered variables without the selling price variable, most influential variable is owner nation with importance percentage, 43%, followed by the vessel length (15%) and flag nation (15%). However, with price column, the relative importance of the other variables abruptly reduced to attribute importance over 95% to the selling price variable (Table 4). The second position is occupied by the GT with importance percentage, 1.5%, while the rest are under 1% importance.

Table 4: Feature importance of the variables of the demolition dataset.

<table>
<thead>
<tr>
<th>Feature importance (%)</th>
<th>Type</th>
<th>GT</th>
<th>Vessel length</th>
<th>Built year</th>
<th>Builder nation</th>
<th>Owner nation</th>
<th>Flag nation</th>
<th>Year</th>
<th>Price fill</th>
<th>Demolition nation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without price fill</td>
<td>10.28</td>
<td>5.40</td>
<td>15.06</td>
<td>7.49</td>
<td>1.41</td>
<td>43.25</td>
<td>15.22</td>
<td>1.87</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>With price fill</td>
<td>1.11</td>
<td>1.58</td>
<td>0.16</td>
<td>0.19</td>
<td>0.14</td>
<td>0.85</td>
<td>0</td>
<td>0</td>
<td>95.97</td>
<td></td>
</tr>
</tbody>
</table>

Predicting demolition location:
Applying RFC in in-service data, we have found prediction for the in-service dataset. A total of 20 countries, instead of 40 countries listed in the demolition dataset, were predicted to be the future demolition nations. Countries that are excluded from the potential demolition nations are Curacao, Norway, Nigeria, Germany, Dominican Republic, Latvia, Philippines, France, South Korea, Sweden, Lithuania, Spain, Finland, Canary Islands, U. A.E, Ukraine, Portugal, Brazil and Romania. The excluded lists include some countries that are known for their shipbreaking practices, that put this prediction suspicious. For example, Norway, France, Brazil and others usually dismantled quite a significant number of smaller ships, but they may not dismantle ships that are bigger enough (larger than 500 GT) to be included in the prediction list.

The change in ship dismantling landscape is quite obvious. For example, India is predicted to dismantle 20000 ships, about half of the total vessel in the coming decades, surpassing Bangladesh by a margin of 15000 ships (Figure 7a). These two demolition nations have been in close competition in recent years. The reason behind this predicted acceleration may be due to their increasing trend of having their yards certified by EU agency. That will legally allow more ships to their yards. On the other hand, Bangladesh are found to be disinterested in any sort of certification process, thereby remaining behind in consumer confidence and reputation. The similar fate is predicted for Pakistan, dropping their ranking from third to six and can be attributable to the lack of certified yards. China is also anticipated to be declining their yard activities, the reason may be that China banned waste import in their territory and only stick to their own ship dismantling. The other reason may be that China is not being able to compete for the price negotiation against the cost originated from sustainability practice in their yards.

As a new demolition destination, Vietnam is predicted to be fourth among the nations, along with Japan and Indonesia, when considered by the number of vessels. Vietnam is predicted to dismantle about 2438 vessels, just after Turkey. While considering the total GT,
Vietnam is only after Turkey and China. The demolition dataset shows that Vietnam dismantled a total of 17 ships, with average 1835 GT in the last four years, all of them are owned by Vietnam. In in-service dataset, Vietnam owns 893 vessels, with an average 3984 GT. Recent report shows that the government of Vietnam approved the demolition of used ships in its territory and, in addition, issued an alert to deal with an appropriate licensed company, issued by the relevant Ministry (Safety4sea, 2019). This report showed that the government is serious about the development of shipbreaking industry and thus support our prediction.

![Figure 7: Predicted demolition nation against number of vessels (a) top eight countries and (b) Rest of the predicted countries](image)

Turkey seems to be consistently performing from ship count, GT mean and total GT to be dismantled (Figure 7a). Unknown ships share a significant portion and that requires to explore further why are they not known. Overall ship dismantling seems to be concentrated to a few nations while an internal redistribution is likely to occur. This change of demolition landscape looks like that the change are towards more sustainable practice with some caution due to decline of China.

**Discussion:**
This study has implications in three areas: First, the leakage effect previous occurred is not so drastic for the coming years. Relatively softer form of internal alteration among the big ship recycling nations may occur. This is more likely because south Asian countries such as Bangladesh offer high price for bigger ships due to arguably natural competitive advantage gained from beaching process and passing of externalities to the society. The fact that the introduction of stringent sustainability initiatives may hurt the profitability and cause to slip this industry to a new region where economic cost of handling would be less monitored is less probable as there are no other countries having such competitive advantages. Indonesia and Vietnam are two countries that may have some potential for attracting ships but not in a position to alter the ship recycling trajectory.

Second, findings of feature importance reveal that priced offered to yard owners at their EOL ships are a major driving factor and a by far influential factors. Shipowner responsibility is thus crucial to improve the cost sensitivity factor through a financial mechanism. Apart from this, decision to send the ships to south Asian nations are also found to be second most important features of this data set, strengthening the idea that shipowner will have to play a leading role to support the critical sustainable performances. NGOs discussion to attribute responsibility to the shipowner also supports this finding. It is also found that lot of missing values of the selling price of EOL ships is a limiting factor for getting a reliable prediction.

Finally, national and international stakeholders may think ahead how the possible substandard practice can be improved as the leakage effect does not seem to occur in near future. The Indian subcontinent continues to be the dominant shipbreaking players due to predominately market and capacity factors (Rahman and Kim 2020). Many factors such as economic downturn, COVID-19 effect and regulatory policies may shake the actual as well as predicted trend but an wider geographic change is not expected (Rahman et al. 2020, Rahman et al. 2021).
Conclusion:

This study is the first - to the best of the author knowledge- that applied sophisticated machine learning tools for the prediction of the most vibrant industry in the world and their recycling locations. In addition, this study applied the largest data sets of the currently in-service ships around the globe. The study finds that it is likely that an internal redistribution can occur within the major shipbreaking nations, with India capturing more vessels and while Pakistan, China and Bangladesh capturing less. Slight leakage effect is also predicted in that Vietnam may emerge as one of the five largest shipbreaking nations.

The limitation of this study is that the model fit is only 61% and more precision is expected. Generally, the model earns recognition when precision reaches beyond 80%. Further modelling with other ML packages, especially Deep Learning approaches, can be tried to validate the result. Relatively smaller demolition dataset compared to in-service dataset may be another reason that the prediction is not reaching the expected precision. In addition, adding more relevant variables, such as vessel price, are required for this study to predict with more accuracy. Data compilation can be improved eventually for more years and more variables can be added to enable more perfect predictions.

Acknowledgement:

The corresponding author states that there is no conflict of interest.

References:


