

A Random Forest Algorithm-based Accident Prediction to Prevent Marine Pilot Occupational Accidents

Gokhan Camliyurt* · Won Sik Kang** · Daewon Kim*** · Sangwon Park**** · † Youngsoo Park

*Department of Navigation, Graduate School, Korea Maritime, and Ocean University
**Transportation Safety Assessment Office, Korea Maritime Transportation Safety Authority
***Shipping and Logistics Research Division, Korea Maritime Institute
****, † Division of Navigation Convergence Studies, Korea Maritime, and Ocean University

Abstract : Marine pilot occupational accidents during transfer to/from the ship are at the top of the agenda after several safety campaigns by IMPA and individual attempts. There is multiple transfer method for the marine pilot, but a most common way is to use the pilot cutter. This paper aims to predict marine pilot occupational accidents before it occurs by using historical data. Since the problem depends on several variables, this paper develops a model by using the random forest method to predict marine pilot accidents before happening with the random forest method by using RStudio software

Key words : marine pilot, pilot ladder, transfer safety, occupational accident, random forest

1. Introduction

This study aims to develop a marine pilot occupational accident prediction model using a methodology of random forest. Hence aim the purpose of a prediction method to prevent accidents to be occurred by determining the circumstances and safety factors of the accident as a basis for making recommendations, and to prevent further marine pilot casualties and incidents from occurring in the future. (····(omit)····

2. Literature Review

2.1 Merchant shipping is known to be an occupation with a high rate of fatal accidents (Carter, Williams, & Roberts, 2019) caused by maritime disasters and occupational accidents. (····(omit)····

In the maritime industry, random forest usage is also very common. Marine environmental and ecological research (····(omit)····

3. Methodology

The study aims to predict marine pilots' occupational

accidents during their transfer to/from the ship using tree-based machine learning algorithms, and describe the ranking of contributing factors with random forest methods. First, a multi-source dataset with metadata of 99 accidents is created from the reports. Then, decision tree and random forest models are trained on the data to predict the target outcome; both are also tuned using parameter search to achieve the best possible performance. Finally, interpretation of the tree's decisions and explanation of the random forest predictions are visualized, discussed, and contrasted with maritime domain knowledge. (····(omit)····

4. Results

As a result of the analysis trials, it has been seen that the best algorithm is obtained with 1200 decision trees and when the number of variables that can be used for partition is 6 (mtry) (Scornet, 2018) in each tree node as shown in table 3. mTry is the number of variables for splitting at each of the nodes and depends on which R package is used. The Out-of-bag (OOB) error rate of the random forest obtained as a result of the final study was found to be 16.67%. (····(omit)····

Table 3 Parameters of the Random Forest Model

N of Tree	mtry	OOB Error
1200	6	16,67

Figure 13. N of Tree selection graph. After reaching 1100-1200 classification error is tending to be stable. (K. Li et al., 2017)As stated before R package default setting gives the decision of 1200.

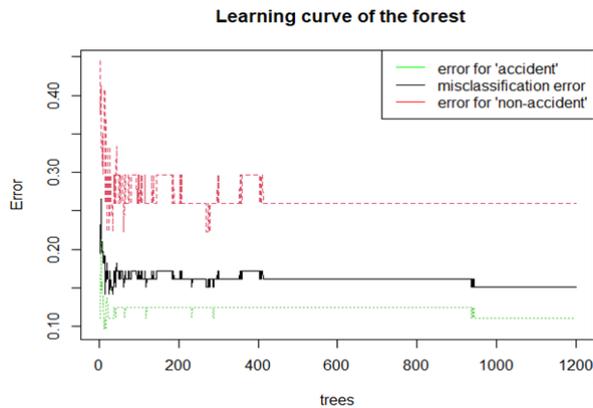
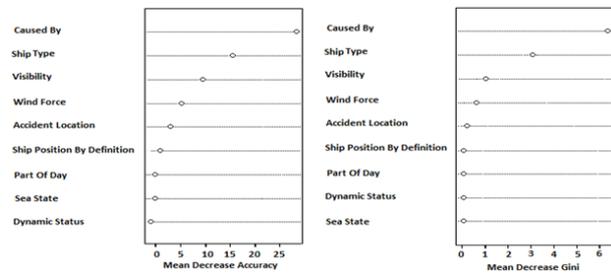


Fig. 14 Contribution of Variables importance of Accuracy and The Purity Values of The



5. Conclusion

In this study, the effects of nine different factors on whether an unwanted situation on a ship results in an accident or not were analyzed. In the data analysis, the classification model was created by using the Random Forest algorithm. It is observed that the final model obtained predicted the training data at a very good level with 84% and test data as high as 88%.

....(omit)....

References

- [1] Carter, T., Williams, J. G., & Roberts, S. E. (2019). Crew and passenger deaths from vessel accidents in United Kingdom passenger ships since 1900. *International Maritime Health*, Vol. 70, No. 1, pp. 1 - 10. <https://doi.org/10.5603/IMH.2019.0001>
- [2] Li, K., Yu, N., Li, P., Song, S., Wu, Y., Li, Y., & Liu, M. (2017). Multi-label spacecraft electrical signal classification method based on DBN and random forest. *PLoS ONE*, Vol. 12 No.5, pp. 1 - 19. <https://doi.org/10.1371/journal.pone.0176614>
- [3] Scornet, E. (2018). TUNING PARAMETERS IN RANDOM FORESTS. *ESAIM: PROCEEDINGS AND SURVEYS*, Vol. 60, pp. 144 - 162. <https://doi.org/10.1051/proc/201760144>