# Predicting detention and deficiencies using random forests

**Sabine Knapp[1] and Michel Van de Velden[2]**

**Econometric Institute Report 2022-08**

**Erasmus University**

**Abstract**

The aim of this exploration study is to predict detention and twelve deficiency types which can be used to enhance port state control targeting as well as domain awareness for coastal administrations. A total of 234 combinations of random forest variants are explored evaluating over 400 covariates. The study uses a comprehensive and unique, global inspection dataset of over 200k inspections and 400k deficiencies (2014 to 2019) and out of sample data from 2020 to 2021 for evaluation. The results show that based on the used data, normal random forests outperform other variants and overall detention has the highest decile lift with 3 or higher compared to random selection. This is followed by the deficiency groups safety of navigation, certificates and qualification and the Maritime Labor Convention. Deficiencies related to newer areas such as MARPOL Annex VI, ballast water treatment and anti-fouling are more difficult to predict and are also more difficult to detect compared to other areas where detection often depend on the training and background of inspectors. Future work will evaluate further model variants and evaluate inspection policies by filtering out high risk vessels that were missed.

**Keywords:** detention, deficiencies, random forests, top-decile lift, predicting probabilities

[1] Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, Netherlands, phone +61-466827029, email knapp@ese.eur.nl

[2] Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, Netherlands, email vandevelden@ese.eur.nl

## 1. Introduction

This study develops and tests thirteen prediction models based on random forest algorithms and extends previous work on improved targeting for Port State Control (PSC) and domain awareness which can be used in conjunction with incident type models (Knapp and Van de Velden, 2021, Knapp and Heij, 2020). Currently the status quo of the industry is to use detention only while deficiency types and incident types are not considered (Knapp 2006, Knapp and Franses 2007, Knapp and Heij 2020). Knapp and Heij (2020) demonstrate that targeting based on combined risk dimensions (using detention, incident and incident types) can improve hit rates and reduce false negative events (missing a risky vessel). This approach adds to this philosophy by adding twelve deficiency types in addition to detention. Eighteen random forests variants are explored based on a unique global inspection dataset of over 200k inspections and 400k deficiencies (2014 to 2019) and over 400 risk factors are considered. Out of sample data for evaluation of a period of 2 years (2020 to 2021) is used.

## 2. Data and variables used in this study

Datasets used are spit into matrix data (train data) and the out of sample data (test data). For the train data, 212,228 global inspections (6,070 detentions) from January 2014 to June 2019 form the basis of this study to develop the formulas. Please refer to Table 1 for basic statics of the inspection data used as train data. For the test data to evaluate the models and inspection policies, the out of sample data comprises of 1,029,726 observation (January 2020 to December 2021) of the world fleet with 133,252 inspections and 2,704 detentions.

**Table 1: Detention and mean deficiencies by year (train data)**

| | Not Detained | Detained | Inspections | Det rate | Deficiencies | | |
|---|---|---|---|---|---|---|---|
| **Year** | **Count** | **Count** | **Sum** | **%** | **Indicator*)** | **Sum** | **Mean** |
| **2014** | 38,322 | 1,320 | 39,642 | 3.3% | 54,459 | 88,464 | 2.23 |
| **2015** | 39,279 | 1,337 | 40,616 | 3.3% | 51,228 | 79,909 | 1.97 |
| **2016** | 38,777 | 1,145 | 39,922 | 2.9% | 48,737 | 75,990 | 1.90 |
| **2017** | 39,412 | 1,052 | 40,464 | 2.6% | 47,638 | 73,543 | 1.82 |
| **2018** | 38,899 | 963 | 39,862 | 2.4% | 47,612 | 71,161 | 1.79 |
| **2019** | 11,469 | 253 | 11,722 | 2.2% | 12,004 | 18,193 | 1.55 |
| **Total** | **206,158** | **6,070** | **212,228** | **2.9%** | **261,678** | **407,260** | **1.92** |

**) Note: at least 1 deficiency per deficiency group*

The inspection data covers data from the main PSC MoU's (Paris MoU, Tokyo MoU, Vina del Mar, USCG) and missing data such as ship particulars at the time of the inspection is complemented by using data from IHS Maritime. Only initial inspections are considered and follow up inspections were excluded to reduce a possible source of bias. Inspection data is biased since the inspection selection is guided by the various target factors of the various Port State Control Memoranda of Understanding (MoU)s. For this reason, it is better to combine data from various MoU's rather than just one country or one region (Knapp, 2006).

Table 2 provides an overview of the number of variables in each main group and the data type for each main group. The initial selection of variables is based on the literature such as Knapp (2006), Knapp and Heij (2020) and Knapp and Van de Velden (2021) and are as follows:

1. Ship particulars such size, the age, the ship type, flags, main engine builder and designed and classification societies.
2. The country where the ship was built which is grouped into four groups and interaction effects with 2 age groups (0-2 and above 14 years of age representing high age risk and 3-14 years of ship age represent low age risk).

3. The country of location of the Safety Management Companies (DoC company) and group beneficial owner which are classified according to income based on the World Bank classification such as: 1) high income, 2) upper middle income, 3) lower middle income, 4) low income and 5) unknown.
4. The year of existence of safety management and beneficial owner which serves as proxy to their experience and quality. This is further complemented by an indicator that expressed the concentration of maritime industries such as ownership companies, safety management companies, engine designers and builders. The concentration acts as proxy to knowledge spill over and safety quality.
5. Lagged inspection, deficiency and incident history of the vessel (within 1 year prior to event date) and changes of ship particulars overtime such as flag changes, ownership changes, DoC company changes and class changes within 3 years prior to event date of interest

**Table 2: High level overview of number of variables**

| Variable groups | Type | Nr of factors |
|---|---|---|
| Size, age | continuous | 2 |
| Ship Types | categorical | 9 |
| Flag | categorical | 151 |
| Class | categorical | 81 |
| Main engine designer | categorical | 115 |
| Main engine builder location | categorical | 30 |
| Safety management company (county) | categorical | 5 |
| Owner company location (country) | categorical | 5 |
| **Maritime expertise** | | |
| Company presence and years of existence | categorical | 6 |
| **Previous histories:** | | |
| Previous inspections, detentions, incidents (VS, S, and LS) | continuous | 6 |
| Previous changes in ship particulars | | 4 |
| **Interaction variables** | | |
| Shipyard country groups with age groups | categorical | 8 |
| **Total variables evaluated** | | **422** |

There are over 600 individual deficiency codes and 29 main deficiency groups. The deficiency codes groups were regrouped into 12 groups reflecting inspection areas that are found to be useful for inspections and for domain awareness and which could also be combined with the incident types used by Knapp and Van de Velden (2021) or Knapp and Heij (2020). Since vessels can have more than one deficiency for each deficiency group during inspections, the variables are reclassified into 0 and 1 indicating at least one deficiency (or no deficiency).

For a high-level overview of the dependent variables, please refer to Table 3. The only deficiency group excluded from the analysis is ISPS (security) since the dataset does not have enough observations for this type of deficiency group. The endpoints of interest (dependent variables) are as follows:

1. *detained*
2. *Group 2: Certificates and Qualifications (Code groups 01100, 01200, 01300)*
3. *Group 3: Maritime Labor Convention (Code groups 18100, 18200, 18300, 18400)*
4. *Group 4: Structural Conditions and Watertight Integrity (Code groups 02100, 03100)*
5. *Group 5: Propulsion and Auxiliary Machinery (Code group 13100)*
6. *Group 6: Life Saving Appliances and Fire Safety (Code groups 1100, 07100)*
7. *Group 7: Emergency Systems and Alarms (Code groups 04100, 08100)*
8. *Group 8: Safety of Navigation and Radio Communications (Code groups 10100, 05100)*
9. *Group 9: Safety Management (ISM-15100, Cargo Operations-06100 and Dangerous Goods-12100, Other – 99101, 99102)*

*10. Group 10: MARPOL Annex 1 to 3 (Oil-14100, Chemicals-14200, 14300)*
*11. Group 11: MARPOL Annex 4 and 5 (Sewage-14400, Garbage-14500)*
*12. Group 12: MARPOL Annex 6 (Air Pollution-14600)*
*13. Group 13: Ballast Water and Anti Fouling (Code groups 14700, 14800)*

**Table 3: High level overview of dependent variables used**

| | **Matrix data (2014-2019)** | | **Out of sample (2020-2021)** | |
|---|---|---|---|---|
| | **indicator** | **sum** | **indicator** | **sum** |
| Inspected | 212,228 | - | 99,944 | 133,252 |
| Detained | 6,070 | - | 2,602 | 2,704 |
| Certificates and Qualifications | 32,394 | 50,947 | 9,783 | 16,729 |
| Maritime Labor Convention | 28,583 | 45,965 | 12,063 | 22,413 |
| Structural and Watertight Integrity | 29,001 | 43,048 | 8,934 | 14,433 |
| Propulsion and Machinery | 14,466 | 19,223 | 5,117 | 7,938 |
| Life Saving and Fire Appliances | 56,622 | 108,966 | 16,521 | 33,349 |
| Emergency systems and alarms | 19,190 | 24,104 | 7,265 | 10,140 |
| Safety of Nav. and Radio Com. | 36,102 | 59,261 | 10,929 | 19,527 |
| Safety Management (ISM) | 24,415 | 31,555 | 6,272 | 9,065 |
| Marpol A1 to A3 | 6,351 | 7,119 | 1,992 | 2,397 |
| Marpol A4 and 5 | 8,578 | 9,402 | 2,897 | 3,548 |
| Marpol A6 | 4,429 | 4,863 | 997 | 1,122 |
| Ballast Water and Antifouling | 1,547 | 1,724 | 1,677 | 2,002 |
| **Total deficiencies** | **261,678** | **406,177** | **84,447** | **142,663** |

*Note: indicator means for matrix data at least 1 deficiency per inspection and for out of sample data at least one inspection, detention or deficiency per period*

## 3. Combination of model variants and model evaluations

The present study considers 13 end points of interest in total (detention plus 12 deficiency types). Table 4 provides a list of the model combinations that were used – a total of 18 variants for the 13 endpoints of interest, hence a total of 234 combinations. For a general overview of random forests, class-imbalance and tuning please refer to Knapp and Van de Velden (2021) and Breiman (1996, 2001) and Breiman et al (1984).

Table 4 shows three model groups. 1) Regular random forest (RF), 2) Balanced random forests (BRF) by Chen et al. (2004) and 3) Random Forest on balanced samples (RF_BS) using under sampling of the training data. For each group, variants are considered. Based on initial experiments on tuning where different values for *m* were assessed. It was decided to consider three options for *m* given in Table 4.

Moreover, for all random forests, aggregation of results is considered using both - majority voting as well as averaging of probabilities. For majority voting, the class predictions of each tree are considered and the proportions of predicted classes over all trees is calculated. For probability aggregation, the average predicted leaf proportions over all trees in the forest is calculated. To estimate and evaluate the models, R is used.

**Table 4: Summary of model variants used**

| Group | Variant | Explanation |
|---|---|---|
| 1 | RF_m_16 | Regular RF, m =16, majority votes aggregation |
| 1 | RF_p_16 | Regular RF, m =16, probability votes aggregation |
| 1 | RF_m_32 | Regular RF, m =32, majority votes aggregation |
| 1 | RF_p_32 | Regular RF, m =32, probability votes aggregation |
| 1 | RF_m_8 | Regular RF, m =8, majority votes aggregation |
| 1 | RF_p_8 | Regular RF, m =8, probability votes aggregation |
| 2 | BRF_m_16 | Balanced RF, m =16, majority votes aggregation |
| 2 | BRF_p_16 | Balanced RF, m =16, probability votes aggregation |
| 2 | BRF_m_32 | Balanced RF, m =32, majority votes aggregation |
| 2 | BRF_p_32 | Balanced RF, m =32, probability votes aggregation |
| 2 | BRF_m_8 | Balanced RF, m =8, majority votes aggregation |
| 2 | BRF_p_8 | Balanced RF, m =8, probability votes aggregation |
| 3 | RF_BS_m_16 | RF balanced training data, m =16, majority votes aggregation |
| 3 | RF_BS_p_16 | RF balanced training data, m =16, probability votes aggregation |
| 3 | RF_BS_m_32 | RF balanced training data, m =32, majority votes aggregation |
| 3 | RF_BS_p_32 | RF balanced training data, m =32, probability votes aggregation |
| 3 | RF_BS_m_8 | RF balanced training data, m =8, majority votes aggregation |
| 3 | RF_BS_p_8 | RF balanced training data, m =8, probability votes aggregation |

*Notes: m= majority voting aggregation, p=probability votes aggregation, the numbers correspond to m, the number of variables considered for splitting. The default value for the data sets is 16. The number of trees for all models is 500*

The out of sample data for evaluation comprises of 1,029,726 observation (January 2020 to December 2021) of the world fleet with 99,944 inspections and 2,602 detentions. First, the probabilities are estimated at a certain time with the assumption that they are valid for 3 months (see Knapp and Heij, 2020). Second, observed data is matched with the estimated probabilities and evaluation metrics are calculated using the following setup and eight periods:

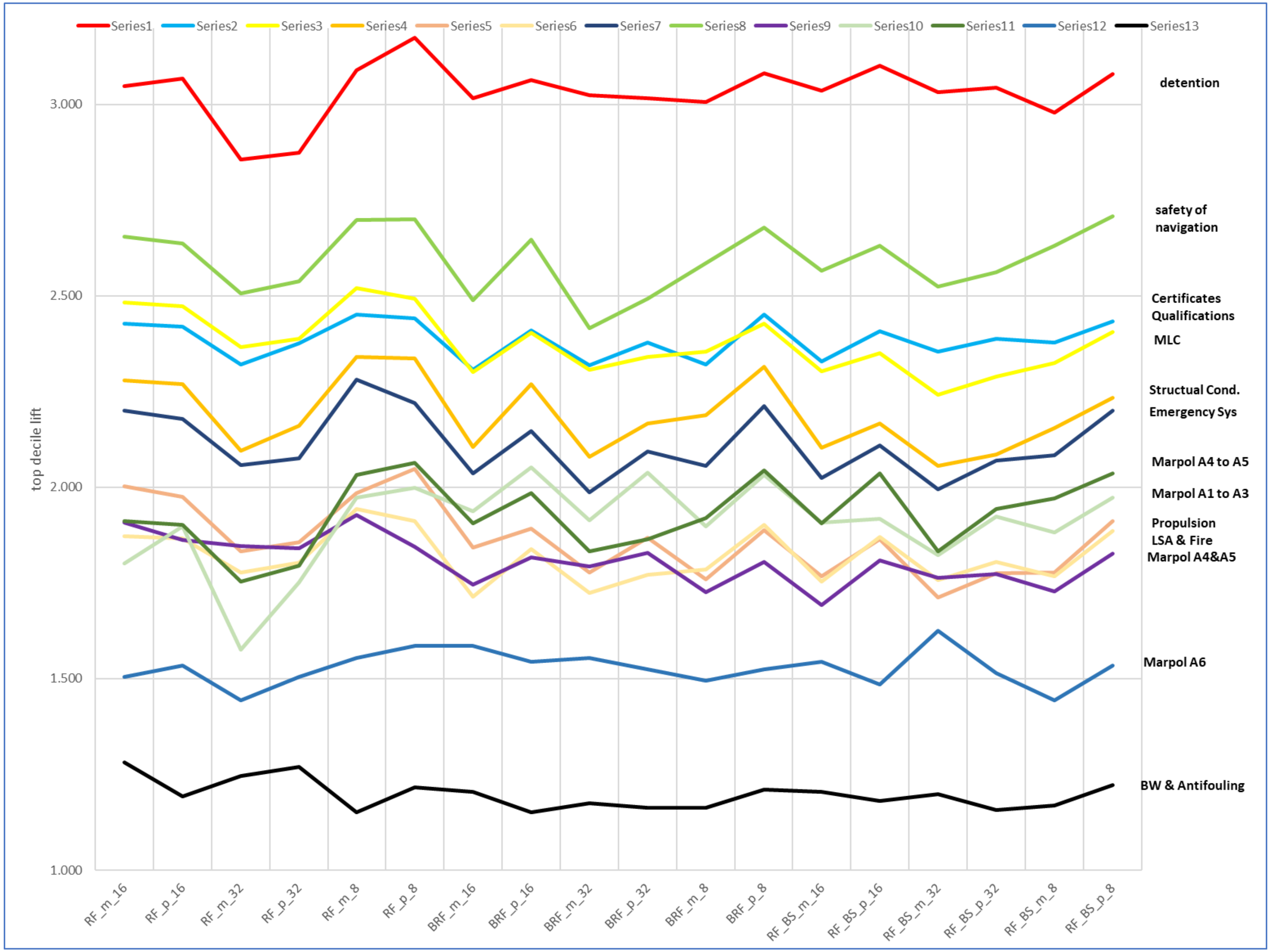
- *P1: Probabilities as of January 2020 – empirical data from January 2020 to March 2020*
- *P2: Probabilities as of April 2020 – empirical data from April 2020 to June 2020*
- *P3: Probabilities as of July 2020 – empirical data from July 2020 to September 2020*
- *P4: Probabilities as of October 2020 – empirical data from October to December 2020*
- *P5: Probabilities as of January 2021 – empirical data from January to March 2021*
- *P6: Probabilities as of April 2021 – empirical data from April to June 2021*
- *P7: Probabilities as of July 2021 – empirical data from July to September 2021*
- *P8: Probabilities as of October 2021 – empirical data from November to December 2021*

The main interest for targeting of vessels is to classify vessels and to reduce false negative events. Based on Knapp and Van de Velden (2021) who explain the various evaluation metric limitations, the top-decile lift is considered. It compares of the 10% highest estimated probabilities and compares the hit rate to random selection. If the predicted probabilities are good, the top decile lift is large.

## 4. Results and Discussions

Figure 1 and Table 5 summarize the top decile lift for each of the model variants. The higher the top decile lift, the better the model variant performs on out of sample data for the test data (2020 to 2021) compared to random selection. A value of below 1 performs worse than random selection. Overall, the detention model outperforms the various deficiency type models with a decile lift of 3 and higher compared to 1 or 2.

**Figure 1: Top decile lift for each model variant (out of sample data 2021 and 2020)**




*Abbreviations: 1: detained, 2: Certificates and Qualifications, 3: Maritime Labor Convention, 4: Structural Conditions and Watertight Integrity, 5: Propulsion and Auxiliary Machinery, 6: Life Saving Appliances and Fire Safety, 7: Emergency Systems and Alarms, 8: Safety of Navigation and Radio Communications, 9: Safety Management (ISM), 10: MARPOL Annex 1 to 3 (Oil and chemicals), 11: MARPOL Annex 4 and 5 (Sewage and Garbage-14500), 12: MARPOL Annex 6 (Air Pollution), 13: Ballast Water and Anti Fouling*

**Table 5: Summary of results for detention and deficiency models – top decile lift**

| | Dependent Variable | | Deficiency Groups | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Model variants** | **Detention** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** |
| 1 | RF_m_18 | 3.048 | 2.428 | 2.482 | 2.279 | 2.003 | 1.873 | 2.200 | 2.655 | **1.908** | 1.802 | 1.912 | 1.505 | **1.282** |
| 2 | RF_p_18 | 3.067 | 2.420 | 2.472 | 2.269 | 1.976 | 1.867 | 2.179 | 2.637 | 1.863 | 1.898 | 1.902 | 1.535 | 1.193 |
| 3 | RF_m_32 | 2.856 | 2.320 | 2.366 | 2.095 | 1.833 | 1.778 | 2.058 | 2.506 | 1.847 | 1.576 | 1.754 | 1.444 | 1.246 |
| 4 | RF_p_32 | 2.875 | 2.376 | 2.388 | 2.161 | 1.857 | 1.804 | 2.076 | 2.538 | 1.841 | 1.752 | 1.795 | 1.505 | 1.270 |
| 5 | RF_m_8 | 3.090 | **2.451** | **2.520** | **2.341** | 1.984 | **1.944** | **2.282** | 2.698 | 1.927 | 1.973 | 2.033 | 1.555 | 1.151 |
| 6 | RF_p_8 | **3.175** | 2.441 | 2.492 | 2.337 | **2.048** | 1.912 | 2.220 | 2.700 | 1.844 | 1.998 | **2.064** | 1.585 | 1.217 |
| 7 | BRF_m_16 | 3.017 | 2.307 | 2.301 | 2.106 | 1.843 | 1.714 | 2.037 | 2.489 | 1.745 | 1.938 | 1.905 | 1.585 | 1.205 |
| 8 | BRF_p_16 | 3.063 | 2.410 | 2.404 | 2.270 | 1.892 | 1.838 | 2.146 | 2.647 | 1.817 | **2.053** | 1.985 | 1.545 | 1.151 |
| 9 | BRF_m_32 | 3.025 | 2.318 | 2.306 | 2.079 | 1.778 | 1.725 | 1.986 | 2.415 | 1.794 | 1.913 | 1.833 | 1.555 | 1.175 |
| 1 | BRF_p_32 | 3.017 | 2.378 | 2.341 | 2.167 | 1.868 | 1.772 | 2.094 | 2.493 | 1.828 | 2.038 | 1.864 | 1.525 | 1.163 |
| 1 | BRF_m_8 | 3.006 | 2.321 | 2.354 | 2.189 | 1.759 | 1.785 | 2.057 | 2.586 | 1.726 | 1.898 | 1.919 | 1.495 | 1.163 |
| 1 | BRF_p_8 | 3.082 | **2.451** | 2.427 | 2.315 | 1.888 | 1.902 | 2.213 | 2.679 | 1.806 | 2.033 | 2.044 | 1.525 | 1.211 |
| 1 | RF_BS_m_16 | 3.036 | 2.329 | 2.304 | 2.104 | 1.767 | 1.753 | 2.025 | 2.566 | 1.692 | 1.908 | 1.905 | 1.545 | 1.205 |
| 1 | RF_BS_p_16 | 3.102 | 2.408 | 2.351 | 2.166 | 1.864 | 1.870 | 2.110 | 2.631 | 1.809 | 1.918 | 2.037 | 1.485 | 1.181 |
| 1 | RF_BS_m_32 | 3.032 | 2.355 | 2.242 | 2.056 | 1.712 | 1.758 | 1.995 | 2.525 | 1.763 | 1.822 | 1.833 | **1.625** | 1.199 |
| 1 | RF_BS_p_32 | 3.044 | 2.388 | 2.290 | 2.086 | 1.775 | 1.805 | 2.069 | 2.562 | 1.774 | 1.923 | 1.943 | 1.515 | 1.157 |
| 1 | RF_BS_m_8 | 2.979 | 2.379 | 2.324 | 2.155 | 1.777 | 1.768 | 2.084 | 2.632 | 1.729 | 1.883 | 1.971 | 1.444 | 1.169 |
| 1 | RF_BS_p_8 | 3.079 | 2.434 | 2.406 | 2.234 | 1.911 | 1.887 | 2.200 | **2.708** | 1.826 | 1.973 | 2.037 | 1.535 | 1.222 |

*Abbreviations:1=detained, 2: Certificates and Qualifications, 3: Maritime Labor Convention, 4: Structural Conditions and Watertight Integrity, 5: Propulsion and Auxiliary Machinery, 6: Life Saving Appliances and Fire Safety, 7: Emergency Systems and Alarms, 8: Safety of Navigation and Radio Communications, 9: Safety Management (ISM), 10: MARPOL Annex 1 to 3 (Oil and chemicals), 11: MARPOL Annex 4 and 5 (Sewage and Garbage-14500), 12: MARPOL Annex 6 (Air Pollution), 13: Ballast Water and Anti Fouling*

Detention is easier to predict than individual deficiency groups of which detection depends upon the training and background of the inspector (Knapp, 2006). It is therefore no surprise to see this difference. Deficiency groups related to safety of navigation and certificates and qualifications as well as the Maritime Labor Convention follow in second and third place while MARPOL Annex 6 (air emissions) and Ballast Water and Antifouling have the worst top decile lift as they are harder to predict and inspectors are less experiences in these areas as they are relative new areas.

In Table 5, the best performing model based on the two-year test data is highlighted in bold. Not surprisingly, the result varies across the dependent variable but overall, the normal random forests variants RF (m8 and p8) outperform the other variants for most dependent variables. Variants BRF (m8 and p8) are possible alternatives. For Safety of Navigation, variant RF_BS (p8) is the best and for MARPOL Annex VI, variant RF_BS (m32) performs best. It is recommended to choose the best five models and re-evaluate their performance every year with new out of sample data. Especially for the areas that are relatively new for inspectors and where inspections are not as straight forward, detection and prediction is more difficult compared to classic deficiencies such as certificates, qualifications or areas related to the safety of navigation.

Future work will entail the evaluation of additional model variants and dependent variables as well as the evaluation of inspection regimes by filtering out risky vessels that could have been inspected but were missed.

## References


Breiman, L. Bagging predictors. Mach Learn **24,** 123–140 (1996). https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random Forests. Machine Learning **45,** 5–32 (2001). https://doi.org/10.1023

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (1984), Classification And Regression Trees (1st ed.). Routledge.

Chen, C., Liaw, A., and Breiman L. (2004), Using random forest to learn imbalanced data, 110, Tech-report of University of California, Berkeley, pp. 1–12.

Knapp S, (2006), The Econometrics of Maritime Safety – Recommendations to improve safety at sea, Doctoral Thesis, Erasmus University Rotterdam.

Knapp, S, Franses, P.H (2007), A global view on port state control - econometric analysis of the differences across port state control regimes. Maritime Policy and Management 34, 453-483.

Knapp S, Heij C (2020), Improved strategies for the maritime industry to target vessels for inspections and to select inspection priority areas, Safety 2020, 6, 18.

Knapp S, Michel Van de Velden (2021), Exploration of machine learning algorithms for maritime risk applications, EI 2021-03, http://hdl.handle.net/1765/137081

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/