



Universiteit
Leiden
The Netherlands

Computer Science & Economics

Detection of suspicious behaviour
from ship transponder data

Michael Bosch

Supervisors:
Frank Takes & Gerrit Jan de Bruin

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

24/06/2019

Abstract

Since the 1990s, data of vessels has been collected and sent by a so-called Automatic Identification System (AIS). The data contains information about the vessels such as the speed, the location and the vessel type (for example cargo, tanker or fishing). In this research, we aim to detect suspicious ship behaviour from this AIS data. To do so, data mining techniques are applied to get more insight in this relatively unknown domain. Thereafter, machine learning is applied to predict the vessel type based on 11 features. This prediction is compared with the real vessel type to identify vessels with untypical behaviour: suspicious vessels.

The derived prediction model has a F1 score of 0.70 using all features and a F1 score of 0.58 using only dynamic (radar) features. The model can be applied in the real world to perform vessel controls based on facts instead of performing them randomly, which can lead to a more efficient use of resources by the water police.

Contents

1	Introduction	1
1.1	Context	1
1.2	Related work	3
1.3	Research question	4
1.4	Thesis outline	4
2	Data set	6
2.1	Overview of the data set	6
2.2	Data preprocessing	7
3	Methods	9
3.1	Machine learning	9
3.2	Our approach	10
3.3	Features	11
3.4	Tools	12
4	Experiments	13
4.1	Experimental setup	13
4.2	Descriptive results	15
4.2.1	Features	15
4.2.2	Measurements per vessel	17
4.2.3	Different vessel types	17
4.2.4	Travelled distance	18
4.2.5	Speed	20
4.2.6	Heading stability	22
4.3	Predictive results	23
5	Conclusions	27
5.1	Conclusions	27
5.2	Applied in the real world	27
5.3	Discussion	28
5.4	Further research	29

1 Introduction

In the introduction, we will describe the context of the problem this research will address. We will also explain the techniques that will be used and what difficulties need to be handled. Moreover, we will discuss some related work, the goal of the research and the research question, and how this research differs from previous research. The introduction will end with the thesis outline.

1.1 Context

Vessels have been sailing on the sea for thousands of years. Since technology has grown enormously, it also became possible to collect data from these ships. This data is collected and sent by a so-called Automatic Identification System (AIS). The AIS is developed in the 1990s [25] and is compulsory for all vessels above 15 meters in length since 2014 [6]. It is intended to assist the vessel's crew in navigating and allows maritime authorities to track and monitor a vessel's movements. The AIS automatically broadcasts information via a transmitter built into the system. AIS data has a high variability. It contains among other things the speed, the location and the vessel type (for example cargo, tanker or fishing) [17].

The AIS data may contain valuable information about the vessels themselves, but also about their behaviour on sea. In this research, a data-driven approach will be used to analyse this behaviour. This means there won't be a hypothesis like in traditional research, but the data will be analysed with a wide perspective. These results will reveal patterns and create insight in the entire domain. We have chosen for this approach over the traditional hypothesis-driven approach to get more insight in the domain instead of just validating a single hypothesis.

By applying this data-driven approach, we will use data mining techniques on the AIS data to examine the vessel's behaviour. Specifically, we will examine if it is possible to detect suspicious behaviour. Suspicious behaviour refers in general to behaviour that indicates someone might be doing something illegal, for example bringing a large empty bag into a department store dressing room as if you might steal the clothes [15]. 'Suspicious' is a wide concept, since many different activities can be called suspicious. Examples of suspicious behaviour in the context of this research are vessels sailing on forbidden places, vessels with a deviant speed and vessels that do not follow the typical routes.

In this research, we assume vessels of the same vessel type have similar behaviour (speed, length etc.). Therefore, we call a vessel suspicious if it has deviant behaviour compared to other vessels of the same vessel type, for example a cargo vessel with other behaviour than a 'normal' cargo vessel.

There are a number of technical techniques and terms that are used within this research. One of the relevant techniques is *data mining*. Data mining is the field of research that focuses on getting a better understanding of a (large) data set in an automated way, for example by searching for patterns in the data [29]. By using these automated techniques, information will be retrieved that is impossible, or at least very difficult, to discover without machines. Data mining is mostly used for descriptive analysis, which focuses on gaining insight from historical data.

A related technique is *machine learning*. Machine learning is the science of getting computers to learn and act like humans do, and improve their learning over time in autonomous fashion, by feeding them data and information in the form of observations and real-world interactions [7].

Basically, machine learning tries to find a relation between the input characteristics, so-called features, and the output, so-called class or target. The goal of machine learning is to build models, which can make predictions about never seen instances. In contradiction to data mining, machine learning is used for predictive analysis.

Machine learning has been implemented successfully in many domains. Some examples are:

- Health care for recognising the size and type of tumours. Machine learning speeds up this process from at least 30 minutes to less than a minute [8].
- Construction industry for risk assessment. Machine learning has made this process faster and more accurate [10].
- Car industry for self-driving cars [1].
- Police for predictive policing. Machine learning shows predictions of places where and times when crimes are most likely to occur [31].

Data mining and machine learning both use data for their analyses. Unfortunately, real world data is often incomplete and contains errors, which makes it unsuitable for analysis [28]. Examples of these issues are missing values, impossible values (human length of 5 meter) and impossible combinations (sex is male and pregnant is yes). Another issue that often occurs is *class imbalance*. This means the class, which is the target of the prediction, is very *skewed distributed*. This happens when the proportions of the possible values are far from equal, for example when 90% of the data set is male and only 10% female. Machine learning algorithms tend to classify every person as male since it will already achieve an accuracy of 90% in that way. If we want information about both sexes, the results will not be useful. To solve these issues, *data preprocessing* can be used. There are numerous data preprocessing techniques, for example deleting invalid and incomplete instances, correcting these instances and using oversampling (increasing the size of the minority class) or undersampling (decreasing the size of the majority class).

Another relevant term is a *ROC curve*. This is a graph that shows the possible ratios between the True Positive Rate (TPR) and the False Positive Rate (FPR) [2]. The TPR and FPR indicate to what extent a positive prediction is indeed true (TPR) and to what extent a positive prediction is actually false (FPR). This is shown in Table 1. The TPR and FPR are calculated via: $TPR = \frac{TP}{TP+FN}$ and $FPR = \frac{FP}{FP+TN}$. The area under a ROC curve (AUC) is a performance measurement, often used in classification problems. In general, a high AUC implies a model that makes good predictions. You can set your model to achieve any point on the ROC curve, depending on your preference.

ROC curve and F1 score mogen naar Methods. Misschien nieuwe kopje: "Assess performance models?"

	Real value		
		+	-
Predicted value	+	TP	FP
	-	FN	TN

Table 1: TPR and FPR.

Other relevant performance metrics are [23]:

- *Precision*: $\frac{TP}{TP+FP}$
From those predicted positive, how many of them are actual positive. 0 is the worst possible score and 1 the best possible score.
- *Recall*: $\frac{TP}{TP+FN}$
From those actual positive, how many were predicted positive. 0 is the worst possible score and 1 the best possible score.
- *F1 score*: $2 \cdot \frac{Precision * Recall}{Precision + Recall}$
Balance between Precision and Recall. 0 is the worst possible score and 1 the best possible score.

1.2 Related work

Although there has not been much research in this domain, there are some researches that might be relevant for this research.

Recent work used machine learning on AIS data to automatically detect SAR (Search And Rescue) activity [3]. From AIS data, they tried to predict whether a vessel was a SAR vessel or not and if it was carrying out a SAR operation. To do this, the authors split the data in a training and a test set and subsampled the data to avoid having an imbalanced evaluation metric of the classifier. On the training data, they applied the Random Forest Classifier. This model labelled 77.5% of the SAR activities correctly.

Another research that combined data mining and machine learning with AIS data, tried to improve fishing pattern detection [5]. They used three dominant fishing gear types and investigated whether it was possible to detect and identify potential fishing behaviour for these types. For each fishing gear type, they used another model. The accuracy for the three types were 83%, 84% and 97%. There has been more research that focused on fishing, for example automatically discovering fishing areas [19] and visualising the fishing effort [22]. Although these researches focused only on fishing vessels, it shows there has been more research that combined data mining and AIS data.

There has also been research that tried to detect suspicious behaviour. The first research used historical AIS data and process mining to extract a reference model of the normal behaviour of vessels per vessel type [16]. This reference model can be compared with current data to detect whether a vessel has similar behaviour to other vessels of the same vessel type. The other example used clustering to group vessels that have comparable features [27]. When a vessel starts exhibiting different features compared to the cluster, it is called suspicious. In this research, only the features speed, course and shipping density (to detect common routes) were used. The authors also used, same as the previous research, historical AIS data to create a reference model.

Other researches that used AIS data analysed the location of vessels to avoid collision [11, 20] and extracted motion patterns to use it for anomaly detection [26].

1.3 Research question

This research will contribute to a better understanding of AIS data and the valuable information it contains. Since there has not been done much research in this domain, this research will be partly exploratory. This part of research will focus on finding relations between characteristics of vessels and determine to what extent these relations can be used to detect patterns in the data. The intended outcome of the research is an overview of characteristics that are the best indicators for suspicious behaviour. Moreover, there will be a predictive part of this research in which a prediction model will be developed. This model will use the findings of the first part to detect suspicious behaviour.

The research question we will answer is:

To what extent can we detect suspicious behaviour of vessels, using data mining techniques on ship transponder data?

Compared to the researches described in Section 1.2, this research will have the most similarities with the researches that also tried to detect suspicious behaviour [16, 27]. These researches used AIS data combined with respectively process mining and clustering. The disadvantage of using clustering to detect suspicious vessels is that some outliers might be normal. If a specific vessel type, for example a SAR vessel, does not often occur, it will be marked as suspicious. It is indeed an outlier due to its frequency but it is not a suspicious vessel. The lack of using the vessel types to avoid having this problem is a disadvantage of this research.

The research that used process mining and a reference model focused exclusively on the activities of a vessel e.g. ‘under_way_sailing’, ‘under_way_using_engine’ and ‘mooring’ (because it was a process mining approach). The disadvantage of this approach is that it does only check for deviant activities. For example, if a vessel slows down to drop off a package (but keeps under way using their engine), the model would not notify this. Therefore, the suspicious vessel would not be marked as suspicious. The lack of using vessel characteristics is a disadvantage of this research.

With our approach, as described in Section 3.2, we won’t have these disadvantages. In that way, this research will contribute to a better understanding of how suspicious behaviour can be detected from AIS data.

1.4 Thesis outline

This chapter contains the introduction and a general description of relevant terms such as data mining, machine learning, data preprocessing, class imbalance and some performance metrics. Moreover, it describes relevant work, how it differs from this research and the research question. Since it is a data-driven research, we will first describe the data set and the executed preprocessing steps in Section 2. Thereafter, in Section 3, the approach, the features and the used tools will be explained. Section 4 will describe the setup of the experiments and the results of the descriptive and predictive experiments. Section 5 contains the conclusions of the research, what these conclusions mean in the real world, the discussion and finally some ideas for further research.

This bachelor thesis was written as part of research done at LIACS. It has been supervised by dr. Frank Takes and Gerrit Jan de Bruin MSc. To obtain domain knowledge, we worked together with a small team of the Human Environment and Transport Inspectorate of the Dutch Ministry of Infrastructure and Water Management. This team consists of Jasper van Vliet, Paul Merkx and Candy Reebroek.

2 Data set

In this section, we will give a description of the data set, in which we separate the attributes into 3 types. Moreover, we will describe the data preprocessing steps we will take to make the data suitable for analyses.

2.1 Overview of the data set

The data set used in this research was provided by the Dutch Ministry of Infrastructure and Water Management. It contains raw data which is automatically transmitted by the AIS. In total, the data set consists of 448,015,339 rows and 46 attributes. Each row is a measurement that contains numerous characteristics of one vessel on a specific moment in time. The data has a timespan of 2 months, from 01-04-2107 up to 31-05-2017 and contains only data from vessels in The North Sea. There are three types of attributes in this data set: static attributes, voyage related attributes, and dynamic attributes.

- **Static attributes**
Static data is data that changes rarely. Static data contains the characteristics of the vessel.
- **Voyage related attributes**
Voyage data is static during the voyage but differs per voyage.
- **Dynamic attributes**
The rest of the data is dynamic data, which means it changes during the voyage.

An overview of the relevant attributes is shown in Tabel 2. We will only use static and dynamic attributes because voyage related attributes can be manipulated easily. Examples of voyage related attributes are the destination and the ETA (Estimated Time of Arrival). These attributes are easier to manipulate than attributes such as the speed or the location.

Attribute	Unit	Description	Type
MMSI	-	Unique identification number of the vessel	Static
Length	Meters	Length of the vessel in meters	Static
Breadth	Meters	Breadth of the vessel in meters	Static
Vessel type	-	The type of vessel (possible types described in Section 2.2)	Static
Latitude	Degrees	Latitude of the location of the vessel	Dynamic
Longitude	Degrees	Longitude of the location of the vessel	Dynamic
Speed	Knots*	Speed of the vessel over ground in knots	Dynamic

* 1 knot is 1.852 km/h

Table 2: Attributes.

2.2 Data preprocessing

Before the data can be analysed, it has to be clean. The cleaning will be done by using preprocessing techniques. The following preprocessing techniques will be applied on the data set:

- Remove rows with missing values

The data set contains many rows with empty values. Since rows with missing values are not useful in analyses, they can be removed. In this step, we checked only for empty cells in relevant columns to keep as much data as we could. In total, 287,647,934 rows were deleted. Only the columns 'MMSI' and 'Vessel type' contained empty values (respectively 255,566,466 and 287,568,660) and caused the deletion of rows.

- Remove outliers

Outliers were also deleted from the data set. In this step, 279,887 rows were deleted. The outliers were found in the following columns:

- Speed

The data set contains rows where the speed of the vessel is 365 knots, which is 676 km/h. The fastest vessel in the world has a speed of 216 knots (400 km/h) [13], which means the row contains an invalid number. Therefore, these rows were removed from the data set.

- Length

There were similar outliers in the column that contains the length of the vessel. The longest vessel ever built was 458.45 meters long [12] but the data set contains vessels with a length of 500 meters. Therefore, these rows were removed.

- Location

At last, the data set contains locations which are not located in The North Sea. The most extreme cases were vessels with a location in Russia. These outliers were removed from the data set as well. The heat map of all the locations is shown in Figure 1. The locations in The North sea are shown in Figure 2.

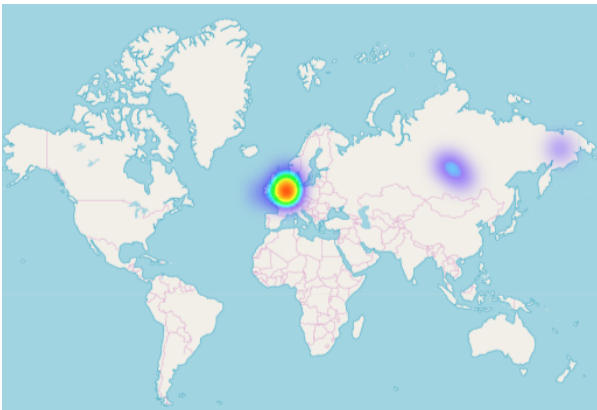


Figure 1: Global heat map of the locations of the vessels in the data set.

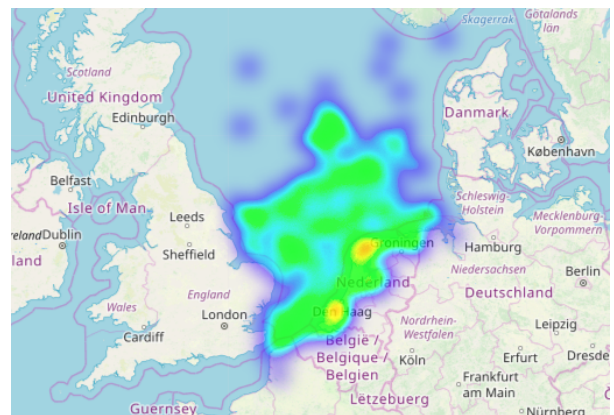


Figure 2: Heat map of the locations of the vessels in the data set zoomed in on The North Sea.

- Handle imbalance

As will be described in Section 3.1, the vessel type will be the target of our prediction model. Therefore, we only checked for imbalance on this attribute. The vessels consist of 22 different vessel types. The distribution of the vessel types is shown in Figure 3.

The distribution of the vessel types is skewed, which will have negative effects when we build a prediction model for these vessel types [24]. To handle this imbalance, some vessel types are merged into one type. According domain experts, these vessel types are very similar and could therefore be merged into one vessel type. We merged the vessel types ‘Tug (31.0)’, ‘Tug (52.0)’ (multiple variants of ‘Tug’, depending on the usage of the vessel), ‘Pilot Vessel’ and ‘Port Tender’ into the type ‘Service Vessel’. After this, only vessel types with a frequency of at least 3% have been selected. The new distribution is shown in Figure 4.

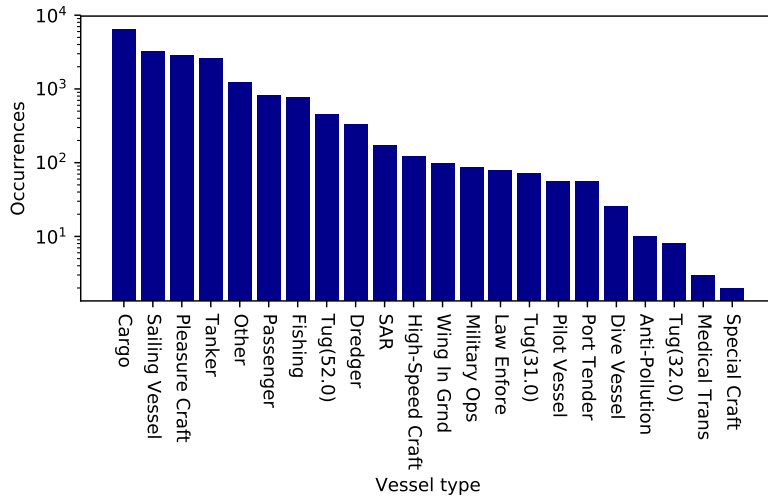


Figure 3: Distribution of the vessel types (before handling imbalance).

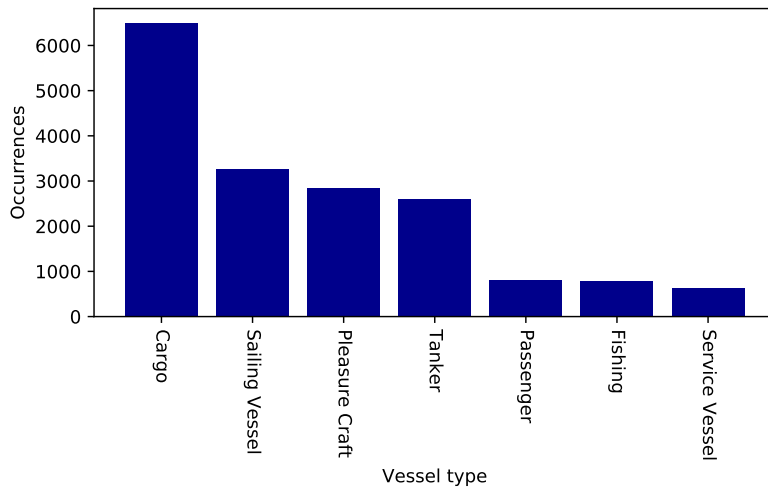


Figure 4: Distribution of the vessel types (after handling imbalance).

3 Methods

In this section, we will discuss how machine learning will be used to detect suspicious behaviour. We will also describe our approach to answer the research question as described in Section 1.3. This section will end with the features and the tools we will use.

3.1 Machine learning

As described in Section 1.1, machine learning will be used to detect suspicious behaviour. We call a vessel suspicious if it has deviant behaviour compared to other vessels of the same vessel type. To detect this, we will use the vessel type as target. The prediction model will predict the vessel type based on the features described in Section 3.3. When a vessel does not match the prediction, i.e. it does not sail like the ‘normal’ vessel of its type, it will be marked as an anomaly and thus a suspicious vessel. Without machine learning, it would be impossible or at least more difficult to detect such suspicious vessels. This shows the value of combining machine learning and AIS data.

In this research, we will use supervised machine learning, which means we will use instances with already known labels (the vessel type) [14]. We will use supervised machine learning because the model can learn how a vessel of a specific vessel type behaves. This model can be applied on never seen instances to predict the vessel type. Since these never seen instances are also labelled, the prediction can be compared to the actual vessel type, where a mismatch means the vessel is suspicious. As discussed in Section 1.3, we think this approach does not have the disadvantages of earlier researches.

When the same data set...

To apply machine learning, we first constructed the described features for each vessel. We split the constructed feature set in a training set (80%) and a test set (20%). Then cross-validation is applied on the training set to reduce the bias and overfitting of the model [21]. When a single data set is used to train and verify the data mining results, there is a high possibility that the model is perfectly trained on the data set, but will not make good predictions on other data sets. This will be prevented by splitting the data set in a training set and a set, and by using cross-validation. This partitioning is described in more detail in Section 4.1. Figure 5 shows the partitioning schematically.

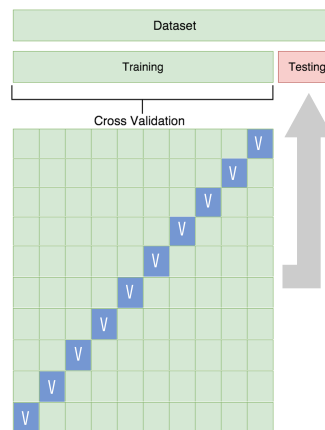


Figure 5: Schematic partitioning of the feature set.

What is v and what is a green cell? Explain in the caption.

3.2 Our approach

To answer the research question mentioned in Section 1.3, we will take the following steps:

1. Define ‘suspicious behaviour’
This is described in Section 1.1. **Let op je `en`. Je vergeet ook een punt aan het einde van de zin.**
2. Data preprocessing to prepare the data for analysis
This is described in Section 2.2.
3. Feature selection
In this step, we will determine which features we will use. These features will be used in step 4 to get more insight and in step 5 to build a prediction model. The features will be described in Section 3.3.
4. Descriptive experiments
To get more insight in the data, we will perform some descriptive experiments. This means we will investigate how the features are distributed and what the relation between the features is. The results of these experiments will be shown in Section 4.2
5. Predictive experiments
The features of step 3 will also be used to create prediction models. We will create multiple prediction models, separated into 2 types. The first type of model predicts the vessel type from all possible vessel types. Based on the features, it will predict whether it is, for example, a cargo vessel, a tanker vessel or a fishing vessel. These models will be called ‘Model type 1’. As classifier, we will use the Decision Tree Classifier because it is relatively easy to understand and it also shows the impact of each feature on the target. Since the main goal of this research is to get more insight in the data, the Decision Tree Classifier is the most suitable classifier. Other often used machine learning classifiers will be applied too to see how accurate they can make predictions, but they will not be discussed in detail. Moreover, the default parameters will be used.

Furthermore, we will create a prediction model for each vessel type. These models will make binary predictions. This means there will be a model that predicts whether it is a cargo vessel or not, a model that predicts whether it is a tanker vessel or not, a model that predicts whether it is a fishing vessel or not etc. These models will be called ‘Model type 2’. As classifier, we will use the Decision Tree Classifier and the best performing classifier of ‘Model type 1’. The results will be shown in a ROC curve.

With both models, we will make predictions using all features and using only dynamic features. Predictions with only dynamic features are relevant because these features can be derived from radar data. In the results, we will distinguish between the results using all features and the results using radar features. Additional details about the setup of the experiments will be described in Section 4.1 and the results will be shown in Section 4.3.

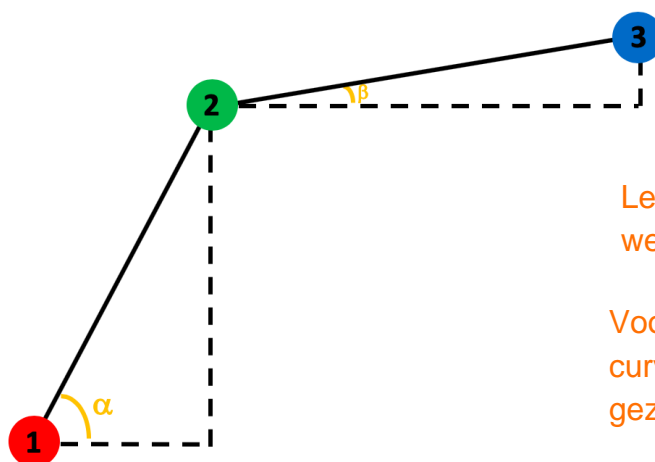
6. Interpretation of results

The last step is to interpret the results and answer the research question of Section 1.3. This will be described in Section 5.1. We will also describe when which model type can be useful. This will be described in Section 5.2.

3.3 Features

To get more insight in the data and predict the vessel type, we will use the following features:

- Length of the vessel
- Breadth of the vessel
- Median, standard deviation and the maximal speed of the vessel
- Median, standard deviation and the maximal heading stability of the vessel. The heading stability means to what extent a vessel sails in the same direction. Assume we have a situation as shown in Figure 6. The vessel starts in location 1, sails to location 2 and ends in location 3. To calculate the heading stability, we first calculate angle α with arctan. Then we calculate angle β and subtract angle α from it. This is the heading difference. In this way, a vessel that sails in a straight line will have a heading difference of 0 (because angle $\alpha =$ angle β). We can do this for all locations of a vessel (in a chronological order) and calculate the median. The standard deviation and the maximal heading difference. The AIS data already contains the attribute ‘heading’, but this is very sensitive for the place of the transponder on the vessel. Therefore, we used another approach to determine the heading and the heading stability.



Leg duidelijker uit in de caption wat we zien.

Voor de volledigheid: je neglect curvature from earth. (En dat is prima, gezien korte afstanden.)

Figure 6: Example of heading stability.

- Days. This feature describes the number of days the vessel occurred in the data set.
- Total and average travelled distance of the vessel. The total travelled distance is the distance the vessel travelled in the entire timespan of the data set. This is calculated using the location of the vessels (latitude and longitude). To do this, we used the Pythagorean theorem. In this calculation, we also added the radius of the Earth to correct for the curvature of the Earth [30].
The average travelled distance is the total travelled distance divided by the number of days the vessel occurred in the data set.

3.4 Tools

In this research, we used Python or more specifically the Python module Pandas. The Python console used is Jupyter Notebook. For our tools, we used the following versions:

- Jupyter Notebook version 4.4.0 to run Python.
- Python version 3.7.1 as programming language.
- Pandas version 0.23.4 is a Python library that provides easy-to-use data structures and data analysis tools.
- Scikit-learn version 0.20.3 for data mining and machine learning tools.

4 Experiments

In this section, we will describe ~~some additional implementations details~~ for the experiments. Thereafter, we will give the results of the descriptive experiments and the results of the predictive experiments.

4.1 Experimental setup

As described in Section 3.1, the constructed feature set is split into a training set and a test. Thereafter, cross-validation is applied on the training set. This approach is described in more detail below.

- The constructed features
Once the features are constructed, we have an overview that contains the MMSI, the features as described in Section 3.3 and the vessel type of each vessel. An (anonymous) sample is shown in Figure 7.
- Cross-validation
During the training of the model, cross-validation with 10 folds is applied to make the training as unbiased as possible. With cross-validation, the data set is split in 10 folds. In each iteration, 9 folds are used for training and 1 fold for validation. This process is repeated 10 times, with each time another fold as validation fold. Because of this implementation, every fold and thus every row has been used for training (9 times) and validation (once). This make the models more general applicable.
- Testing
After applying cross-validation, the 20% test set is used to provide an unbiased evaluation of our final model. The performance of this testing will be used as the performance of the model.

	t_mmsi	length	breadth	Median speed	Std speed	Max speed	Distance	Days	Median heading	Std heading	Max heading	Avg distance	shipType
0	#	12.0	4.0	0.0	1.345339	4.8	5692.004331	61	15.671279	53.831674	180.000000	93.311546	Sailing Vessel
1	#	120.0	20.0	6.7	1.671458	8.4	763.468400	3	1.273614	4.744751	102.892454	254.489467	Cargo
2	#	22.0	5.0	0.0	0.981455	6.1	2282.065663	59	5.052219	52.392952	180.000000	38.679079	Fishing
3	#	80.0	8.0	3.4	1.839299	6.2	1406.634441	24	1.695825	38.248420	180.000000	58.609768	Cargo
4	#	12.0	4.0	2.3	1.155731	3.8	744.851362	8	6.384227	32.920871	180.000000	93.106420	Sailing Vessel
5	#	80.0	9.5	3.6	0.688194	3.9	106.584646	1	0.746361	7.655326	138.590943	106.584646	Cargo
6	#	199.0	30.0	0.2	2.420283	8.4	794.642523	7	6.395054	44.294322	180.000000	113.520360	Cargo
7	#	175.8	31.3	5.9	2.304991	8.6	624.557456	2	0.955461	9.463443	143.682799	312.278728	Cargo
8	#	84.0	9.0	0.2	2.321696	5.8	353.884184	8	11.611659	40.029324	180.000000	44.235523	Tanker
9	#	17.0	5.0	1.5	0.463739	6.3	302.316471	12	1.388013	11.579260	161.118682	25.193039	Fishing

Figure 7: Sample of the features set.

- Performance metrics

For ‘Model type 1’, we use the F1 score, which is the weighted average of the ‘Precision’ and the ‘Recall’. This performance measure also takes the TPR and FPR in account and is therefore better than for example the accuracy.

For ‘Model type 2’, we used the AUC as performance measure because it is one of the most important evaluation metrics for checking any classification models performance [2].

- Prediction

After training, validating and testing the model, it can be used to predict the vessel type. This can be applied on new data. This research focuses on creating the model and not on actually applying it, so this step will not be executed.

To find the best parameters for the Decision Tree Classifier in ‘Model type 1’, the algorithm ‘Grid Search’ is used. Using Grid Search, the Decision Tree Classifier will be executed with different parameters to find the ones that result in the highest F1 score.

‘Model type 2’ will use a ROC curve to show the results and calculate the performance of the model. Because a ROC curve handles only binary classes (classes with only 2 possible values such as ‘yes’ or ‘no’), we first applied a ‘Label binarizer’. This means the column ‘Vessel type’ will be replaced by a column of each possible vessel type that indicates whether it is true or not. An example is shown in Figure 8.

Fishing	Tanker	Cargo	Passenger	Pleasure Craft	Service Vessel	Sailing Vessel
True	False	False	False	False	False	False
False	True	False	False	False	False	False
False	False	True	False	False	False	False
False	False	True	False	False	False	False
True	False	False	False	False	False	False
False	False	True	False	False	False	False
False	False	True	False	False	False	False
False	False	True	False	False	False	False
False	False	False	True	False	False	False
False	True	False	False	False	False	False
False	False	True	False	False	False	False

Plaatje is niet nodig.

Figure 8: Vessel types after applying a label binarizer.

4.2 Descriptive results

In this section we will describe the results of the descriptive experiments. This includes among other things a heat map, a pair plot and the distributions of the dynamic features ‘Travelled distance’, ‘Speed’ and ‘Heading stability’.

4.2.1 Features

In the heat map of Figure 9, the correlation between the features and the possible vessel types is shown. The highest correlation for a vessel type is the correlation between ‘Length’ and a cargo vessel, which is 0.51. Overall, we see cargo and tanker vessels have the highest correlations with the features, which means these vessel types will probably be the easiest to predict.

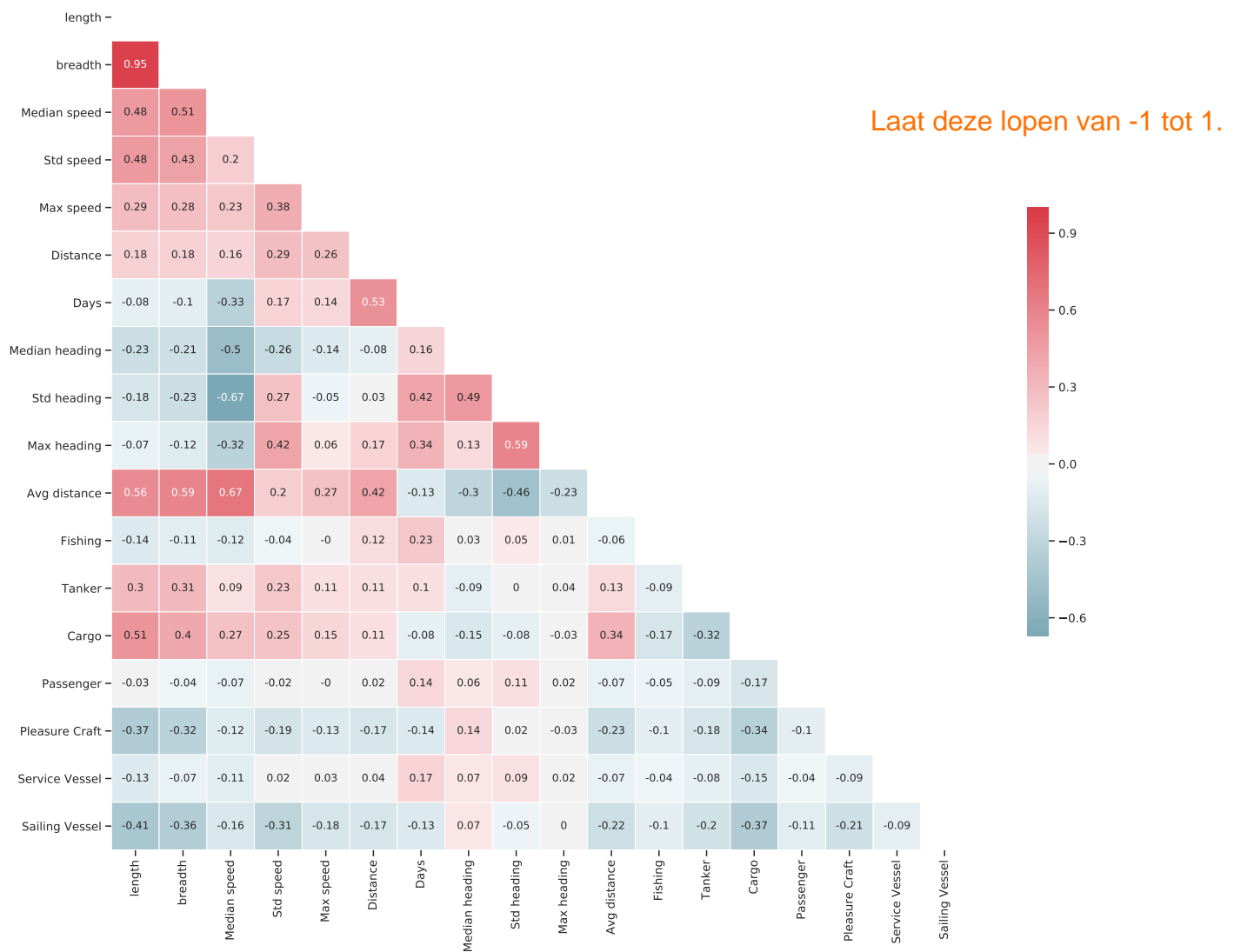


Figure 9: Heat map with correlations between attributes for each vessel type.

Furthermore, we see the static features ‘Length’ and ‘Breadth’ have the highest correlations with the vessel types. We also see that not all features are very relevant for predicting the vessel type. Especially ‘Std heading’ and ‘Days’ have a low correlation with the vessel types, compared with the other features. A side note is that these features have individually a low correlation with the class attribute (the vessel type), but might have a high impact when they are used in combination with other features.

In Figure 10, a pair plot with the distribution of all features is shown. We see for example a strong positive correlation between the ‘Length’ and ‘Breadth’ of a vessel. This confirms the 0.95 correlation the heat map shows. Another observation is the distribution of ‘Days’ e.g. how many days the vessel occurred in the data set. This distribution is right skewed, which means there are relatively a lot of vessels which are just a few days in the data set.

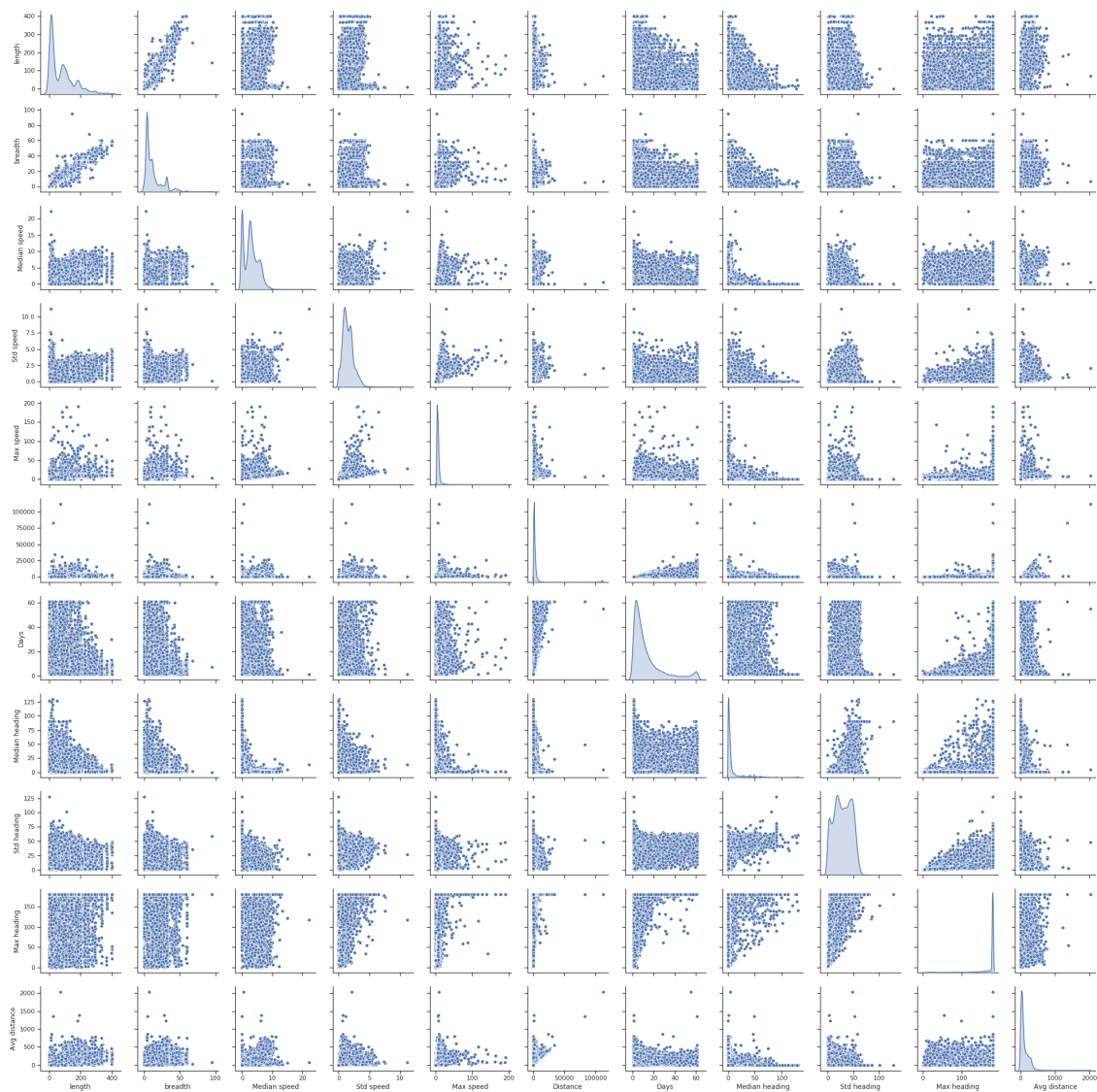


Figure 10: Pair plot of correlations between attributes.

4.2.2 Measurements per vessel

After the preprocessing steps as described in Section 2.2, we find that there are 21,857 vessels. The distribution of the number of measurements per vessel is shown in Figure 11. The figure shows that most of the vessels have many measurements (because the distribution is left skewed). This is convenient for analysing the vessels' behaviour, because the more measurements we have, the better we can find patterns in this measurements.

4.2.3 Different vessel types

After merging and deleting the vessel types as described in Section 2.2, there are 7 different vessel types left. The frequency of these types differ from 642 to 6502. The distribution is shown in Figure 12.

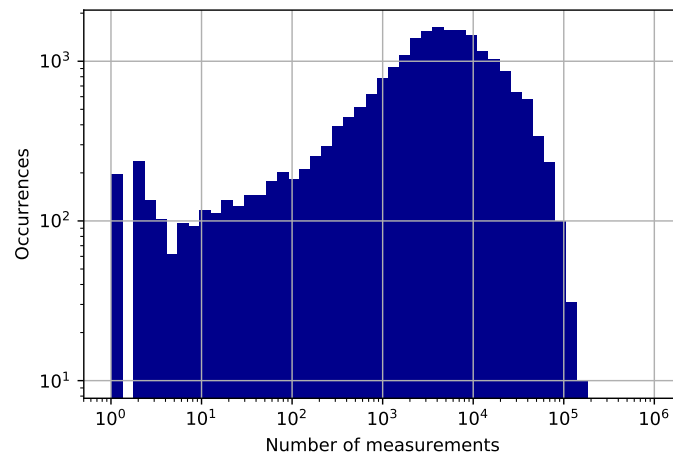


Figure 11: Number of measurements per vessel.

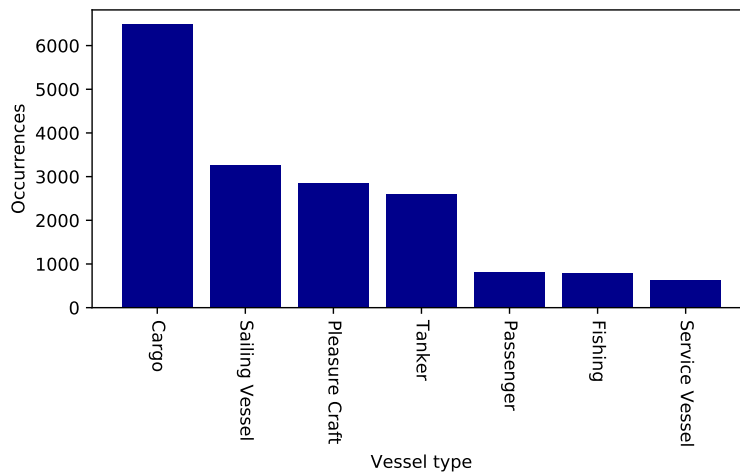


Figure 12: Distribution of the vessel types (after handling imbalance).

4.2.4 Travelled distance

As described in Section 3.3, we calculated the total travelled distance based on the location of the vessel. The distances in Figures 13 and 14 are in kilometers.

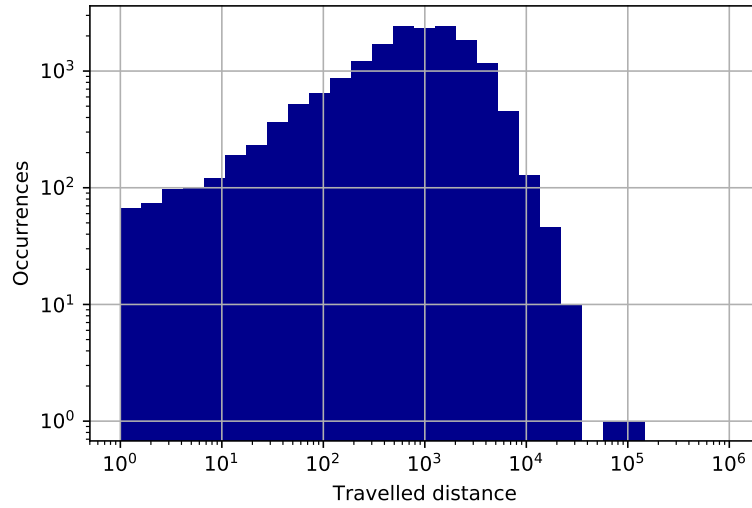


Figure 13: Distribution of the total travelled distance.

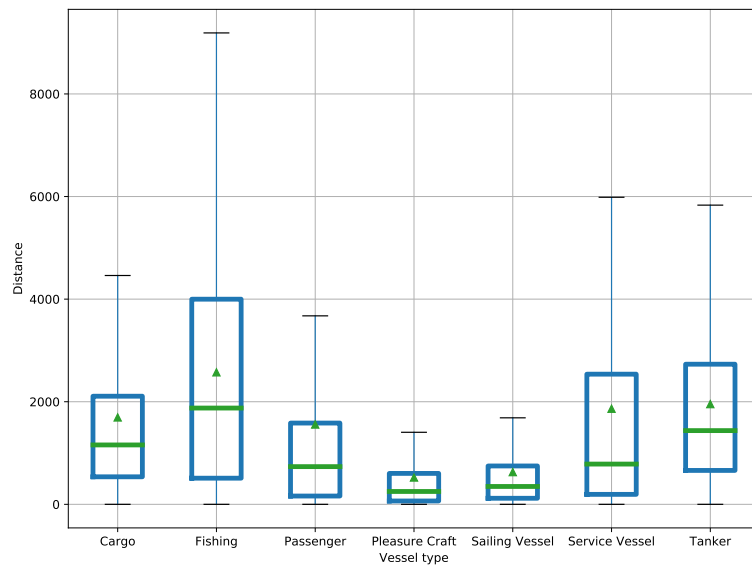


Figure 14: Total travelled distance per vessel type. The green line is the median of the total travelled distance and the green triangle is the average of the total travelled distance.

As shown in Figure 13, the distribution of the distance is skewed to the left, which means most vessels travelled a relative long distance. To get some more insight, we also plotted the travelled distance per vessel type. As shown in this Figure (14), the pleasure crafts and sailing vessels travel a relative constant distance (since the box plot is narrow). The variance is largest within the fishing vessels. This might be caused by the different fishing techniques: either sailing with 10 knots with long nets deployed hanging vertically from floats from or drifting slowly while having a setting of fishing lines equipped with several hundred to several thousands of hook [5]. Another insight from Figure 14, we did not expect, is that cargo vessels have a lower median distance than fishing vessels. This can be explained by Figure 15 and 16.

As shown in Figure 15, the median of the average distance of cargo vessels is higher than for the fishing vessels (in contrast to the total distance). The explanation for this is shown in Figure 16. This figure shows the average number of days a vessel type occurred in the data set. Since the average distance is calculated by dividing the total distance by the number of days, a higher number of days (for fishing vessels) results in a lower average distance.

Besides the distance and average distance, the number of days a vessel occurred in the data set can be interesting as well. It might be that some vessel types, for example cargo vessels, typically sail to The Netherlands, moor and go back to The United Kingdom. In that case they will only occur in the data set for a few days. Other vessel types, for example fishing vessels, will have a relatively constant area where they sail, so if they occur in the data set, they will probably have data on many more days.

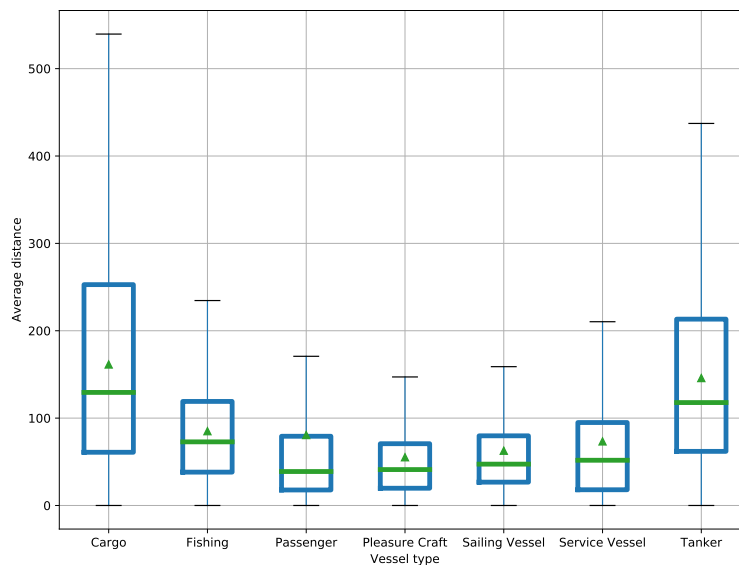


Figure 15: Average distance per vessel type.

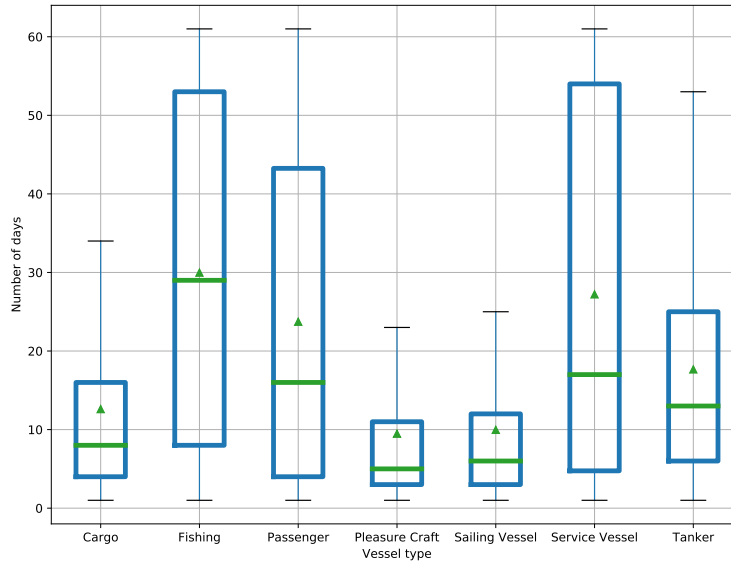


Figure 16: Number of days per vessel type.

4.2.5 Speed

Another attribute of the data set is the speed of the vessel in knots. The distribution of the ‘Median speed’ per vessel is shown in Figure 17. The biggest part of the vessels has a relative low speed, which results in a right skewed distribution. This distribution is also divided into the different vessel types, which is shown in the box plot of Figure 18. Here, we see the cargo and tanker vessels have the highest median speed of approximately 4 knots (7.41 km/h). These vessel types also have the largest standard deviation. This might be caused by the high speed of the vessels when they are at sea and the low speed of the vessels when they are mooring (which takes a long time).

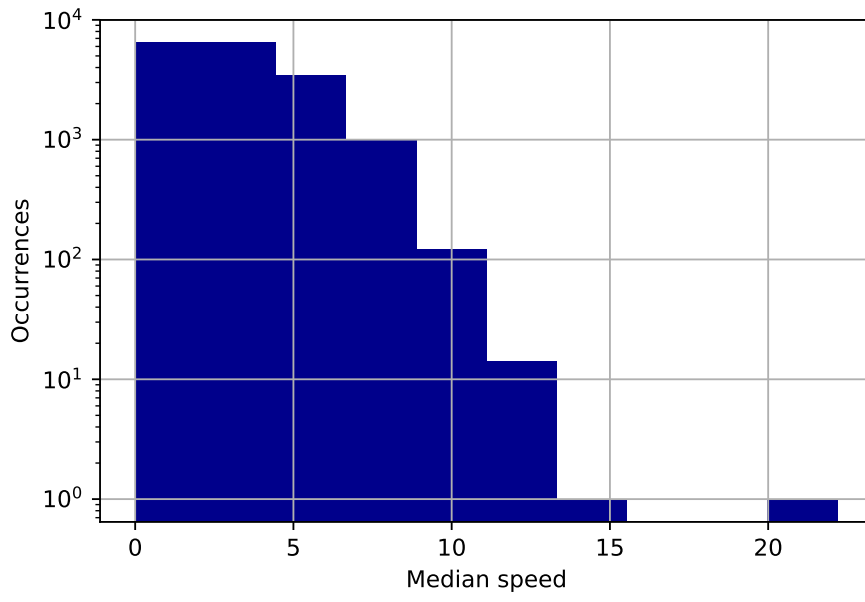


Figure 17: Distribution of the 'Median speed'.

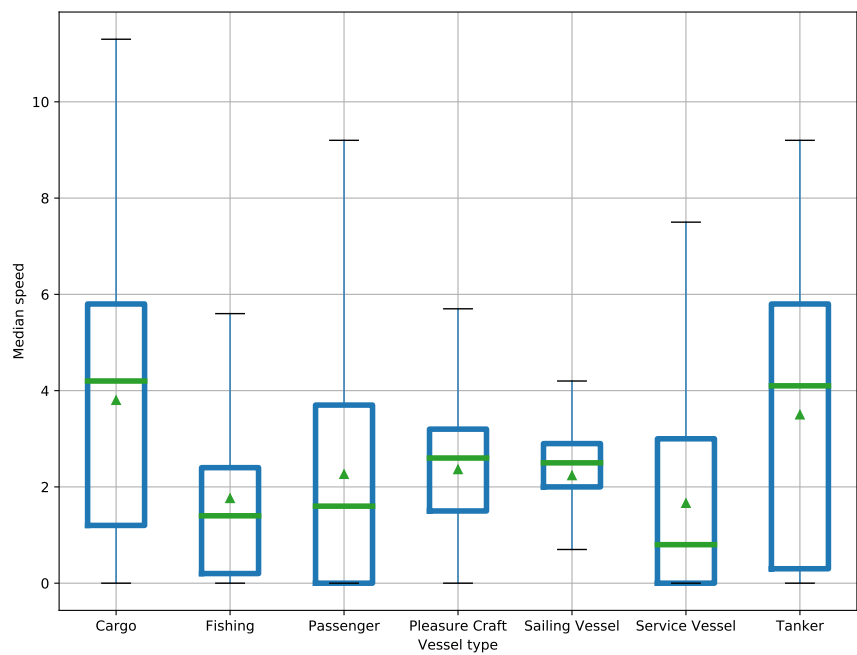


Figure 18: 'Median speed' per vessel type.

4.2.6 Heading stability

We also calculated the heading stability of the vessels i.e. to what extent a vessel sails in the same direction. The distribution of the ‘Median heading’ and the box plot in which this feature is distributed over the vessel types are shown in Figures 19 and 20.

Using domain knowledge, we initially assumed the heading stability would differ per vessel type. A large cargo vessel will not make as many and as sharp turns as smaller vessels. This is confirmed by Figure 20. We see the ‘Median heading’ of cargo and tanker vessels (which are the longest vessels) is lower than the ‘Median heading’ of the other vessels.

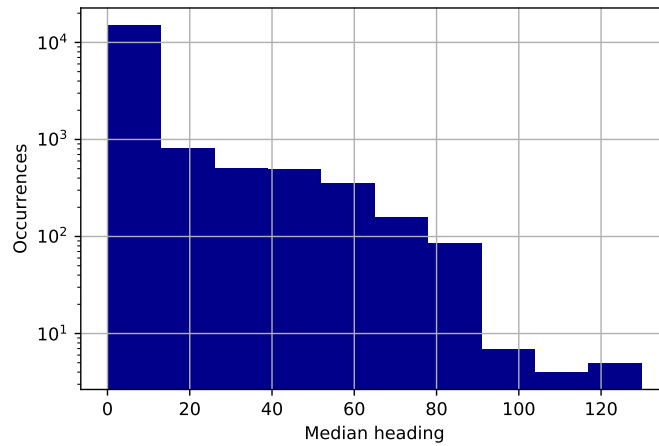


Figure 19: Distribution of the ‘Median heading’.

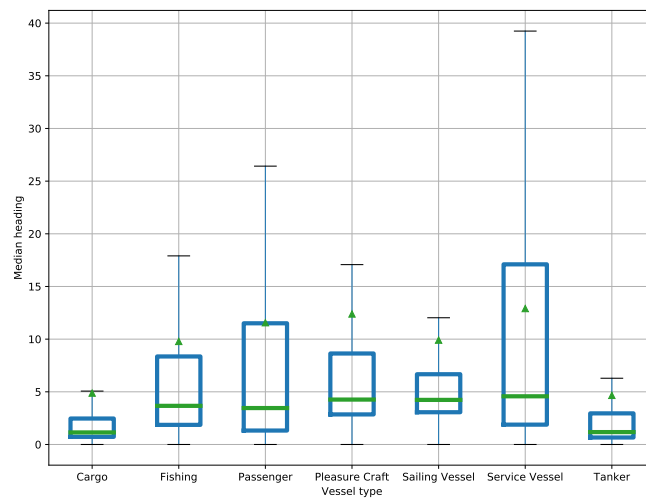


Figure 20: ‘Median heading’ per vessel type.

4.3 Predictive results

- ‘Model type 1’

To predict the vessel type, we applied 8 machine learning algorithms. The scores are shown in table 3. The highest F1 score is 0.704, which is achieved by the Random Forest Classifier. We also performed the experiment with only radar features. These F1 scores are shown in table 4.

To get more insight in how the prediction is done, we looked deeper into the Decision Tree Classifier. The optimal parameters found by ‘Grid Search’ lead to a huge tree of depth 10 and 1159 nodes. To give an idea of the decision tree, a smaller tree of depth 2 is shown in Figure 21.

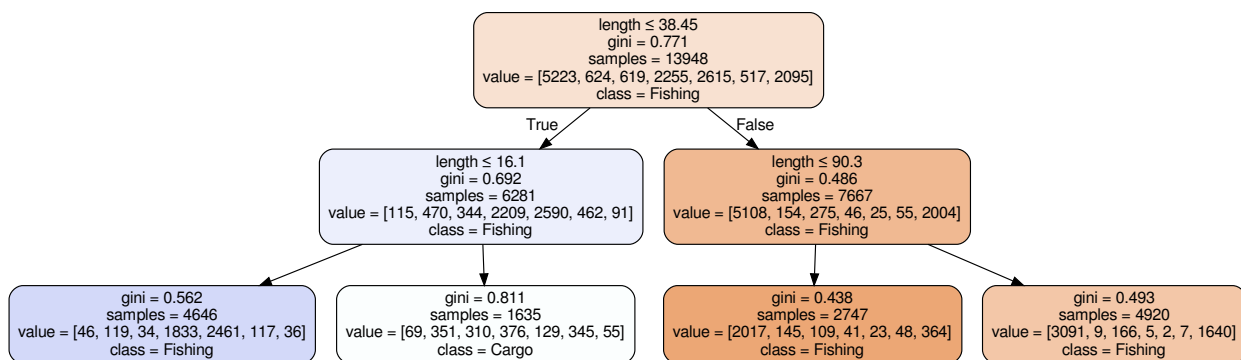


Figure 21: Decision tree of depth 2.

With the tree of depth 10, we achieved an F1 score of 0.659. The importance of each feature to make the prediction is shown in Tabel 5. The static features are the most important features to predict the vessel type. When the static features are removed, the F1 score drops to 0.519. The importance of each feature of that prediction is shown in Tabel 6.

	Model	F1 score
1	Random Forest	0.704
2	Gradient Boosting Classifier	0.688
3	Decision Tree	0.659
4	Neural network (MLP)	0.618
5	Logistic Regression	0.542
6	KNN	0.527
7	Gaussian	0.504
8	Support Vector Machine	0.379

Table 3: F1 scores of different machine learning algorithms using all features.

	Model	F1 score
1	Random Forest	0.582
2	Gradient Boosting Classifier	0.555
3	Neural network (MLP)	0.525
4	Decision Tree	0.519
5	Gaussian	0.383
6	Logistic Regression	0.372
7	KNN	0.368
8	Support Vector Machine	0.283

Table 4: F1 scores of different machine learning algorithms using radar features.

	Feature	Importance
1	length	0.382
2	breadth	0.090
3	Std speed	0.085
4	Std heading	0.069
5	Median heading	0.068
6	Distance	0.063
7	Avg distance	0.062
8	Max speed	0.056
9	Median speed	0.056
10	Days	0.053
11	Max heading	0.016

	Feature	Importance
1	Median heading	0.219
2	Std speed	0.191
3	Distance	0.129
4	Avg distance	0.111
5	Std heading	0.107
6	Max speed	0.107
7	Median speed	0.106
8	Max head	0.030

Table 5: Feature importance using all features. Table 6: Feature importance using radar features.

- ‘Model type 2’

After the prediction models of ‘Model type 1’, we created the prediction models of ‘Model type 2’. The created ROC curves are shown in Figure 22 and 23. The average AUC using the Decision Tree Classifier is 0.74 and the average AUC using the Random Forest Classifier is 0.93. This shows the Random Forest Classifier is also better than the Decision Tree Classifier in making binary predictions.

The ROC curves where only radar features are used, are shown in Figure 24 and 25. These ROC curves have an average AUC of 0.66 and 0.86. The score is, just like the results of ‘Model type 1’ lower than when all features are used.

Niet zo'n informatief plaatje. Ik mis dan ook de uitleg van wat we hier zien.

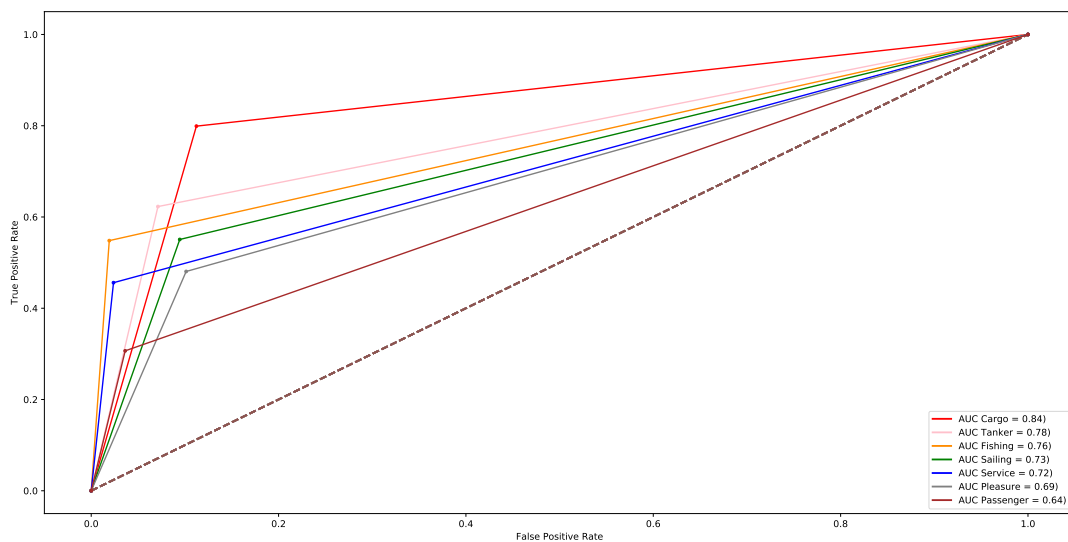


Figure 22: ROC curve using Decision Tree Classifier.

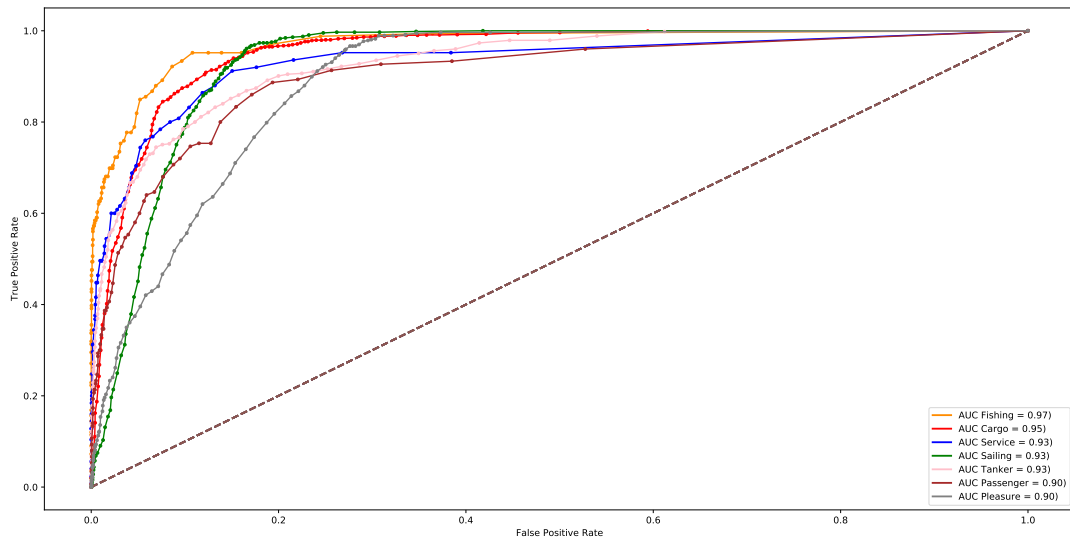


Figure 23: ROC curve using Random Forest Classifier.

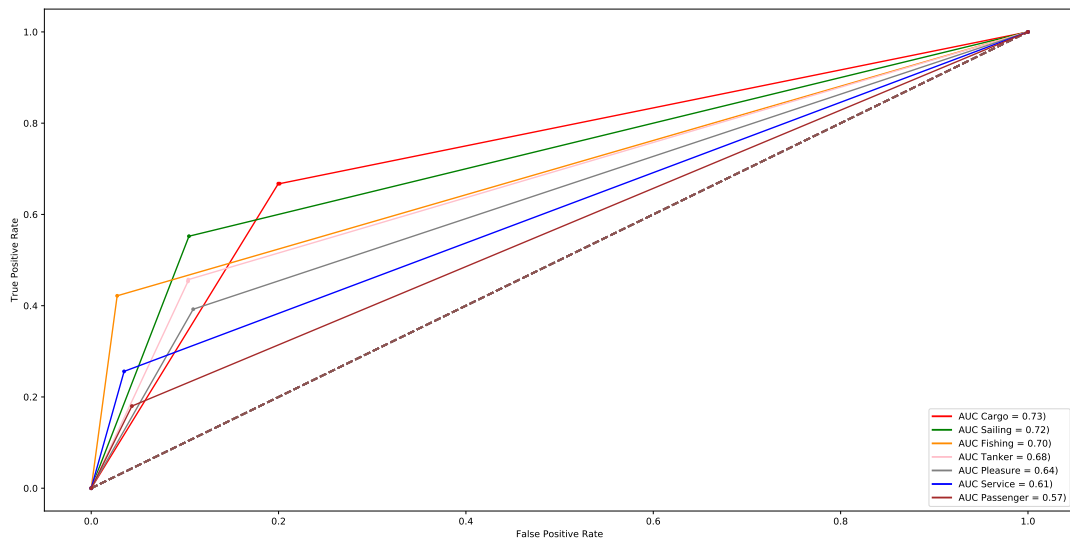


Figure 24: ROC curve using Decision Tree Classifier and only radar features.

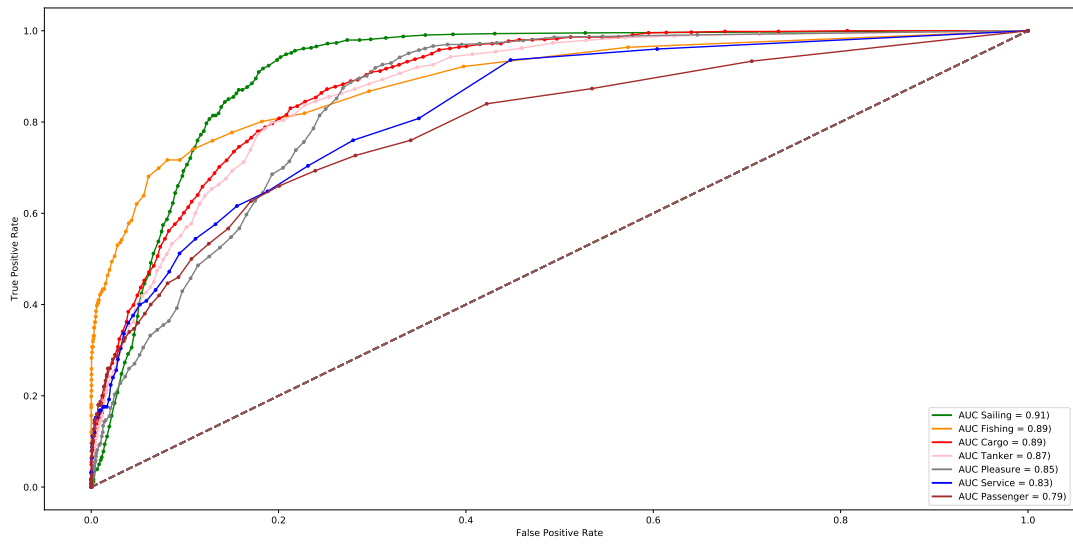


Figure 25: ROC curve using Random Forest Classifier and only radar features.

5 Conclusions

In this last section, we will give the conclusions of the research. We will also describe what these conclusions mean in the real world and when the different models can be useful. Thereafter, there will be a discussion and some ideas for further research.

5.1 Conclusions

As shown in Section 4.3, the highest obtained F1 score is approximately 0.70. If the vessel type would be guessed without using any model, for example by always ‘predicting’ the majority class (cargo), the F1 score would be 0.37. If all described features are used, the ‘Length’, the ‘Breadth’ and the ‘Standard deviation of the speed’ are the most important features for prediction. When only radar features are used, the ‘Median heading’, the ‘Standard deviation of the speed’ and the ‘Median speed’ are the most important features. The highest obtained F1 score is in that case approximately 0.58. This means that using the static feature leads to an increase of 20%.

The average AUC obtained by using the Random Forest Classifier and all features is 0.93, which is high. Both ROC curves, using all features and using only radar features, show that cargo and fishing vessels are the best predictable vessel types. This means that the behaviour of these vessel types differs the most from the behaviour of other vessel types. This is in accordance with the heat map shown of Figure 9, in which saw cargo has a high correlation with most features. On the other hand, pleasure and passenger vessels are the hardest vessel types to predict.

We also see that the Random Forest Classifier makes significantly better predictions than the Decision Tree Classifier and both classifiers score better when all features are used.

The research question of this research is: To what extent can we detect suspicious behaviour of vessels, using data mining techniques on ship transponder data? Using the definition for suspicious behaviour as described in Section 1.1 and the approach as described in Section 3.2, we can conclude that suspicious behaviour can be detected with an accuracy of approximately 0.70, which is 95% more accurate than a random guess.

5.2 Applied in the real world

So how and why should these models be applied in the real world? The models can be applied to predict with a higher certainty whether a vessel has suspicious behaviour or not than when this would be ‘predicted’ randomly. If the water police would do random controls, chances are suspicious vessels are being undetected. On the other hand, innocent vessels are being controlled for nothing, which is a waste of resources. Using a data-driven approach, such as the derived prediction models, can decrease this waste.

To apply these models in the real world, it is important to think about what the goal is and what the model should return. When to use which model depends on the context. When it is relevant to get an overview of the vessel types, it is best to use ‘Model type 1’. This model can distinguish between the possible vessel types. It can be a useful model when someone is interested in getting an impression of the type of vessels on sea.

‘Model type 2’ can be used in more specific situations. If you are interested in the vessel types that are sailing in a fishing-only area, it only matters if the vessel is or is not a fishing vessel. In that case it does not really matter if it is a cargo vessel or a tanker vessel.

In general, ‘Model type 1’ should be used for general, high-level questions and ‘Model type 2’ is more useful in specific situations such as fishing-only areas.

Especially ‘Model type 2’ requires more attention before it can be applied in the real world. Although a high AUC indicates that the model can make good predictions, it is also useful to check the shape of the ROC curve. It depends on the situation whether it is acceptable to make a wrong prediction. If we take a look at Figure 25, we see the AUC of sailing vessels is 0.91, which indicates a very good model. However, to achieve a high TPR (for example 0.9+), the FPR will also be relatively high (0.2+). A FPR of 0.2+ is not always acceptable or even doable. For example, if the water police has limited resources or the damage of wrongly accusing someone is high, a FPR of 0.2+ won’t be acceptable to them.

There will be other situations where the resources are limited or the damage, in terms of time or money, of a false prediction is high. Therefore, it is important to be aware of the ratio of TPR and FPR and to determine which ratio is acceptable in the current situation.

5.3 Discussion

In this research, there are a few factors that might have influenced the F1 score, either in a positive or a negative way. The choice with the biggest impact on the score, is the definition of suspicious behaviour. We assumed vessels of the same type will have similar behaviour in terms of speed, heading stability and travelled distance. Vessels that do not show the typical behaviour of their type will be marked as suspicious. Other definitions of suspicious behaviour might lead to a completely different score.

Another important factor is the used vessel types. We took the vessel types with a frequency of at least 3%, which were seven types. It might be possible that there is not much difference between the behaviour of these vessel types (for example tanker and cargo vessels have very similar behaviour), but there is more difference between the ignored vessel types. This could also be vice versa. This means the F1 score depends heavily on the used vessel types and could be higher or lower when other vessel types are used.

A downside of AIS data in general is that it is possible to turn the AIS transponder off [4]. This means the vessel can not be tracked and there will be gaps in the data set. In this way, people who plan to carry out illegal activities can turn off their AIS transponder and do these activities unnoticed. Vessels with suspicious behaviour will therefore not be in the data set at the moments they carry out illegal or suspicious activities. There are other reasons, like weak signals, that can create gaps in the data. Vessels with gaps in their data are therefore not automatically suspicious vessels. This disadvantage can not be solved but it is still good to be aware of it.

5.4 Further research

The amount of available AIS data will keep growing and we expect the number of researches in this domain to grow as well. Besides completely different researches on the same AIS data set, further research can also focus on finding more relevant features related to the vessel type. This should lead to a significant higher F1 score than mentioned in Section 4.3. A possible extension of this research is to use more or different vessel types. This is related to the second item mentioned in the discussion of Section 5.3. By using more of different vessel types, it can be determined if the features mentioned in this research also relate to the other vessel types. If this is the case, the model would be general applicable.

Another interesting extension of this research is to use maps with sailing, fishing and anchorage areas. Using a map with sailing and anchorage areas, it can be detected if vessels only sail and anchor in the allowed areas. Using a map with fishing areas, it can be detected if only fishing vessels are in this area (using ‘Model type 2’). Moreover, it can be detected if fishing vessels stay in the fishing areas in order to detect illegal fishing activities. This could all be an indicator for suspicious behaviour.

It could also be interesting to calculate how many times and for how long a vessel does not sail. It could be that some vessel types, for example cargo vessels, will keep sailing from the begin to the end of their journey, while smaller vessels, for example fishing vessels, have more locations where they stop. This could be a feature to increase the F1 score.

The feature heading stability as described in 3.3 can also be extended. In this research, the heading stability is calculated over the entire data set. It could also be interesting to calculate this per time period, for example per minute, per hour or per day, or per voyage. This could give more insight in how a vessel sails over time and during a voyage.

A last recommendation is to look for suspicious location patterns. Suspicious location patterns can be strange patterns like sailing circles but also vessels that do not follow the typical tracks.

According to the team of the Dutch Ministry of Infrastructure and Water Management, only radar data will be used in the near future. This research shows that the vessel type can already be predicted with an accuracy of 0.58 when only dynamic (which can be obtained via radar) data is used. To respond to the development of using radar data, future research should focus on creating models that use only radar data. These models can be applied directly when it receives radar data in order to make real time predictions.

References

- [1] Bosch. (n.d.). Self-driving car technology - Between man and machine. Retrieved June 20, 2019, from <https://www.bosch.com/stories/autonomous-driving-interview-with-moritz-dechant/>.
- [2] Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, *30*(7), 1145-1159. [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2).
- [3] Chatzikokolakis, K., Zisis, D., Spiliopoulos, G., & Tserpes, K. Mining Vessel Trajectory Data for Patterns of Search and Rescue. In *Proceedings of the EDBT/ICDT Workshops*, pages 117-124, 2018.
- [4] Cutlip, K. (2016, July 30). Going Dark: When Vessels Turn Off AIS Broadcasts. Retrieved May 15, 2019, from <https://globalfishingwatch.org/data/going-dark-when-vessels-turn-off-AIS-broadcasts/>.
- [5] de Souza, E. N., Boerder, K., Matwin, S., & Worm, B. (2016). Improving Fishing Pattern Detection from Satellite AIS Using Data Mining and Machine Learning. *PLOS ONE*, *11*(9). <https://doi.org/10.1371/journal.pone.0163760>.
- [6] European Commission. (2016, October 7). Control technologies - Fisheries - European Commission. Retrieved April 28, 2019, from https://ec.europa.eu/fisheries/cfp/control/technologies_en.
- [7] Faggella, D. (2016, September 20). What is machine learning? Retrieved April 27, 2019, from <https://emerj.com/ai-glossary-terms/what-is-machine-learning/>.
- [8] Faggella, D. (2019, May 19). Machine Learning Healthcare Applications 2018 and Beyond. Retrieved June 20, 2019, from <https://emerj.com/ai-sector-overviews/machine-learning-healthcare-applications/>.
- [9] Ford, J. H., Peel, D., Kroodsma, D., Hardesty, B. D., Rosebrock, U., & Wilcox, C. (2018). Detecting suspicious activities at sea based on anomalies in Automatic Identification Systems transmissions. *PLOS ONE*, *13*(8). <https://doi.org/10.1371/journal.pone.0201640>
- [10] GenieBelt. (2019, April 12). AI and machine learning: Whats in it for the construction industry? Retrieved June 20, 2019, from <https://geniebelt.com/blog/ai-and-machine-learning-whats-in-it-for-the-construction-industry>.
- [11] Hexeberg, S., Flten, A. L., & Brekke, E. F. AIS-based vessel trajectory prediction. In *Proceedings of the 20th International Conference on Information Fusion*, pages 1-8, 2017.
- [12] Interesting Engineering. (2019, February 27). 9 of the World's Largest Ships. Retrieved April 27, 2019, from <https://interestingengineering.com/9-of-the-worlds-largest-ships>.
- [13] Koman, S. (2018, October 30). Fastest Navy Ships in the World. Retrieved April 27, 2019, from <https://owlcation.com/misc/Fastest-Navy-Ships-in-the-World>.

- [14] Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, 26(3), 159190. <https://doi.org/10.1007/s10462-007-9052-3>.
- [15] LoveToKnow. (n.d.). Suspicious dictionary definition - suspicious defined. Retrieved June 13, 2019, from <https://www.yourdictionary.com/suspicious>.
- [16] Maggi, F. M., Mooij, A. J., & van der Aalst, W. M. (2013). Analyzing vessel behaviour using process mining. In *Situation Awareness with Systems of Systems* (pp. 133-148). Springer, New York, NY.
- [17] Mao, S., Tu, E., Zhang, G., Rachmawati, L., Rajabally, E., & Huang, G. B. An automatic identification system (AIS) database for maritime trajectory prediction and data mining. In *Proceedings of ELM-2016*, pages 241-257, 2018.
- [18] MarineTraffic. What kind of information is AIS-transmitted? Retrieved April 10, 2019, from <https://help.marinetraffic.com/hc/en-us/articles/205426887-What-kind-of-information-is-AIS-transmitted->.
- [19] Mazzarella, F., Vespe, M., Damalas, D., & Osio, G. Discovering vessel activities at sea using AIS data: Mapping of fishing footprints. In *Proceedings of the 17th International Conference on Information Fusion*, pages 1-7, 2014.
- [20] Mou, J. M., Van Der Tak, C., & Ligteringen, H. Study on collision avoidance in busy waterways by using AIS data. In *Proceedings of Ocean Engineering*, pages 483-490, 2010.
- [21] Overfitting in machine learning: What It Is and How to Prevent It. (2017, September 7). Retrieved April 12, 2019, from <https://elitedatascience.com/overfitting-in-machine-learning>.
- [22] Natale, F., Gibin, M., Alessandrini, A., Vespe, M., & Paulrud, A. (2015). Mapping Fishing Effort through AIS Data. *PLOS ONE*, 10(6). <https://doi.org/10.1371/journal.pone.0130746>.
- [23] Ping Shung, K. (2018, May 15). Accuracy, Precision, Recall or F1? Retrieved June 19, 2019, from <https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9?gi=3503eb12c789>.
- [24] Provost, F. machine learning from imbalanced data sets 101. In *Proceedings of the AAAI2000 workshop on imbalanced data sets*, volume 68, pages 1-3, 2000.
- [25] Ratinaud, N. (2017, August 14). AIS. Retrieved April 25, 2019, from <https://mods.marin.nl/display/MIOD/AIS>.
- [26] Ristic, B., La Scala, B., Morelande, M., & Gordon, N. Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction. In *Proceedings of the 11th International Conference on Information Fusion*, pages 1-7, 2008.
- [27] Scholte, K. A. (2013). Detecting suspicious behaviour in Marine traffic using the Automatic Identification System.

- [28] Sharma, M. (2018, July 25). What Steps should one take while doing Data Preprocessing? Retrieved June 15, 2019, from <https://hackernoon.com/what-steps-should-one-take-while-doing-data-preprocessing-502c993e1caa?gi=5f05138b1f6f>.
- [29] Takes, F. W. (2014). Algorithms for analyzing and mining real-world graphs. Leiden Institute of Advanced Computer Science (LIACS), Faculty of Science, Leiden University.
- [30] Williams, D. (n.d.). Earth Fact Sheet. Retrieved May 5, 2019, from <https://nssdc.gsfc.nasa.gov/planetary/factsheet/earthfact.html>.
- [31] Williams , M. (n.d.). Machine Learning and the Age of Predictive Policing. Retrieved June 20, 2019, from <https://www.herox.com/crowdsourcing-news/455-machine-learning-and-the-age-of-predictive-policin>.