**Detection of Suspicious Behavior from Ship Transponder Data**

**Michael Bosch** (supervised by Frank Takes & Gerrit-Jan de Bruin)

Leiden Institute of Advanced Computer Science

s1875876@liacs.leidenuniv.nl

---

**Societal Relevance**

‘Urk has a problem’: Dutch fishing town caught up in cocaine-smuggling trial

Five people, including three fishermen from Urk, are suspected of attempting to smuggling 266 kilos of cocaine in their boat.

[Image: Urk has a problem: Dutch fishing town caught up in cocaine-smuggling trial]

It may be the Netherlands’ most ritually devoted community, where fishing and dancing are spawned by some as the devil’s work. But the wrath of God for indulging in those pursuits is unlikely to be the most pressing concern at the moment for some of the 20,000 residents of Urk, for centuries a major centre for Dutch fishing.

---

**Introduction**

The news item raised the question whether this was an one-time event or if this is happening more often?

Since there are many ships on The North Sea, the possibilities for smuggling are certainly there. But does this really happen?

And if so, is it possible to automatically detect smuggling and suspicious behaviour using data mining techniques and therefore stop smuggling in the future?

---

**Research Question**

In this research, data mining techniques on Automatic Identification System (AIS) data and examine if it is possible to detect suspicious behaviour. This research will contribute to a better understanding of AIS data and the the valuable information it contains. To do this, data mining techniques will be used to find patterns in the data and this patterns will be used to detect outliers. Therefore, the research question is as follows:

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

---

**Approach**

In order to give an answer to my research question, the following steps will be taken:

1. Do research in the relevant field to learn more about the automatic tracking system ships use and what kind of data it sends (and thus the data contains).
2. Perform simple analyses to calculate averages, plot distributions etc. in order to see what attributes are interesting for further analyses.
3. Analyze the interesting attributes in more detail. This can either be:
   (a) Calculate confidence intervals in order to detect outliers
   (b) Use machine learning to predict several characteristics of a ship using e.g. only the route as input

---

**The Data**

**General Information**

The data that will be used in this research came from the Ministry of Infrastructure and Water Management. The data consists of 448 million entries and was collected in several weeks. Moreover, there are 46 attributes for each record, which will be shown in more detail in the next paragraphs. The total size of the dataset is over 91 GB.

**Type of Data**

Each data entry includes information about the location, speed, call sign, name and heading of the ship as well as the ships’ length, breadth and the estimated arrival time at the destination.

---

**Sample of the Dataset**

**POSSIBLE OUTCOMES**

- A method to detect outliers for a specific numeric value. For example, the duration of a trip based on the route and the type of the ship or the average speed of a ship compared to its type.
- A machine learning model to predict several attributes of a ship based on one or a few attributes. For example, to predict the duration of a trip given only the route.

---

**Possible Outcomes**

- A method to detect outliers for a specific numeric value. For example, the duration of a tri based on the route and the type of the ship or the average speed of a ship compared to its type.
- A machine learning model to predict several attributes of a ship based on one or a few attributes. For example, to predict the duration of a trip given only the route.

---

**Figure 1:** Distribution of the average speed of each shiptype

**Figure 2:** Ship locations on The North Sea

---

**Article about Urk, The Guardian, 31 December 2017**

**Distribution of the average speed of each shiptype**

**Sample of the Dataset**

---

**Notes:**

- The data was collected over several weeks and contains 448 million entries.
- Each data entry includes information about the location, speed, call sign, name and heading of the ship as well as the ships’ length, breadth and the estimated arrival time at the destination.
- The total size of the dataset is over 91 GB.
- The data was obtained from the Ministry of Infrastructure and Water Management.
- The dataset contains 46 attributes for each record, which will be shown in more detail in the next paragraphs.

---

**References:**


---

**Acknowledgments:**

- Special thanks to Frank Takes and Gerrit-Jan de Bruin for their supervision and guidance on this research project.
- Thanks to the Ministry of Infrastructure and Water Management for providing the data used in this research.

---

**Contact:**

Michael Bosch
s1875876@liacs.leidenuniv.nl

---

**Permission:**

The author(s) have granted permission for the use of the images and data presented in this document.