



# Predicting the Time of Arrival at Port for Maritime Surface Ships in the Baltic Sea Using Recurrent Neural Networks

Jonatan Lahtivuori  
40601

## **Abstract**

Here is where I would say what is in this document.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Maritime vessels and traffic</b>	<b>2</b>
2.1	Automatic identification system . . . . .	2
2.1.1	Estimated Time of Arrival . . . . .	2
2.2	HELCOM dataset . . . . .	2
2.2.1	Description of HELCOM dataset . . . . .	2
2.2.2	Statistical information . . . . .	2
2.3	Port of Naantali . . . . .	2
<b>3</b>	<b>Recurrent Neural Networks</b>	<b>3</b>
3.1	Long Short Term Memory . . . . .	3
<b>4</b>	<b>Implementation</b>	<b>4</b>
4.1	Libraries used . . . . .	4
4.2	Data pre processing . . . . .	4
4.2.1	Algorithm for extracting routes from dataset . . . . .	4
4.2.2	Coordinate accuracy . . . . .	5
4.2.3	Feature selection . . . . .	5
4.2.4	Time series data preparation and cleanliness . . . . .	5
4.3	ML model . . . . .	5
4.4	Comparison model . . . . .	5
<b>5</b>	<b>Results</b>	<b>6</b>
5.1	ETA prediction . . . . .	6
5.2	Navigation in the archipelago . . . . .	6
<b>6</b>	<b>Discussion</b>	<b>7</b>
<b>7</b>	<b>Conclusion</b>	<b>8</b>

# **1 Introduction**

Describe the need for accurate ETA predictions for maritime traffic and what data is available what has been done already and what the thesis will contribute.

Structure of the thesis.

## 2 Maritime vessels and traffic

Vessels navigating the Baltic Sea

Importance of maritime traffic globally and timely operation

### 2.1 Automatic identification system

What is AIS, what are its features, problems. Example of raw AIS data.

#### 2.1.1 Estimated Time of Arrival

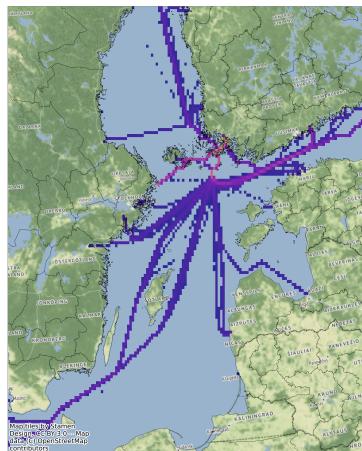
Importance of ETA and what current systems there are.

### 2.2 HELCOM dataset

Introduction to HELCOM

Describe the features and quirks with the HELCOM dataset as it is a processed dataset that contains AIS data already processed from the raw format.

Showing examples of the data as plots of the Baltic, also shows the area covered



**Figure 1:** Heatmap of the Baltic Sea with a limit of more than 10 unique messages per grid cell.

#### 2.2.1 Description of HELCOM dataset

#### 2.2.2 Statistical information

Vessels physical features and the grouping of different vessels by size.

### 2.3 Port of Naantali

The amount of data collected travelling to FINLI from the HELCOM dataset, rather busy port

### **3 Recurrent Neural Networks**

Introduction to what neural networks and more precisely recurrent neural networks and why they are ideal for solving this problem (remembering the order of inputs and time is linear which means all previous timesteps have an impact on the next)

#### **3.1 Long Short Term Memory**

Description of what exactly LSTM networks are

## 4 Implementation

Python

### 4.1 Libraries used

Keras

Tensorflow (GPU)

Pandas

Numpy

Geopandas

Packages and programming language used.

### 4.2 Data pre processing

Processing the HELCOM data to extract routes that can be used to train a model with the goal in mind.

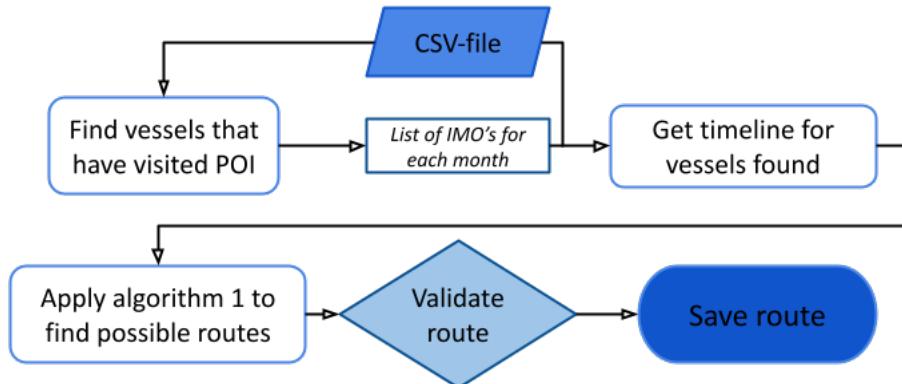
#### 4.2.1 Algorithm for extracting routes from dataset

Explain the algorithm behind getting the routes from the raw data set and that the idea can be applied to any port or area of interest really by defining the area of interest.

The largest gaps allowed

The port chosen for the evaluation has its caveats and should perhaps be identified.

Validation of the routes extracted i.e gaps, no shorter than n number of messages where the vessel is going, not returning to port, longer than n hours, no faulty data etc.



**Figure 2:** Getting routes.

POI (port of interest could also be thought of as Area of Interest). Timeline is the vessels historical data which is fed into algorithm 1. Route validation according to rules and then save the complete route.

#### **4.2.2 Coordinate accuracy**

Scaling the accuracy from raw data to a suitable accuracy. GIS decimal degrees

Vessel travelling at 14 knots (25 km/h) and a message interval of 10 minutes approx. will move 2.2 nmi (4.1 km) per message

#### **4.2.3 Feature selection**

The features chosen for the final training

#### **4.2.4 Time series data preparation and cleanliness**

Further details on how the processed routes are handled to generate the training data to utilize multiple timesteps per prediction which improves performance. Also why the number of timesteps per prediction was chosen, with the time normalized data and how many steps then per time window.

The largest allowed difference between messages in the time normalized data

### **4.3 ML model**

Description of the neural network model used and tested to find the optimal performer

### **4.4 Comparison model**

!!! If the travel distance left to destination is used in nmi for example, test the accuracy against simply calculating time left by the current speed and distance left

## **5 Results**

### **5.1 ETA prediction**

### **5.2 Navigation in the archipelago**

Models difficulties to predict ETA within the archipelago

## **6 Discussion**

Discussion about the results and validity future work

## 7 Conclusion

The possibility to use historical AIS data, example HELCOM dataset, to train a model on predicting the ETA for ports that have good coverage of vessels inbound or some amount of routes coming to the port