



Port Detection in AIS Vessel Data (2025-06-04)



Dominykas Venclovas, Ugnė Kniukškaitė, Paulina Leščinskaitė

Vilnius University
Faculty of mathematics and informatics

2025

General Filtering Logic:

- Only take useful columns, drop NA values, cast to appropriate data types.
- Remove ships with non-standard coordinate types and helicopters by filtering the "Ship type" column.

Filtering For Port Detection Logic:

- Filter out moving vessels that send GPS messages for speed of 0.5 by "SOG" or higher for more than 5 times. The "SOG" < 0.5 decision was made to include nearby "waiting to port" ships. For message count > 5 decision was made to exclude ships "traveling in channel" and slowing down for a small period of time.
- Filter out vessels that have traveled more than 10 kilometers in the given day. Traveled $< 10\text{km}$ decision was made to include "fishing near port" ships.

Port Detection Methods

Spatial Binning:

- Spatial bins were created based on latitude and longitude from the dataset. Grid cell size was approximately 0.64×1.11 km rectangles (roughly 0.7 km squared). The vessels were assigned to each bin by their GPS coordinates. Vessel count was summed in each bin to decide the port size.

K-means Clustering:

- Initially, the optimal cluster size was determined using the elbow method. Subsequently, clustering was also performed with 15 and 20 clusters. Only clusters with more than 20 vessels were left, to filter out fishing vessel spots. The best result was with 19 ports.
- The vessels were assigned to each cluster and summed unique vessels for each port to find port sizes.

DBSCAN:

- DBSCAN was used to cluster Spatial Binning clusters to larger ports. This allowed to give each port just one port/terminal as opposed to multiple that was seen in just Spatial Binning method.
- The maximum distance between two samples for one to be considered as in the neighborhood was selected as 0.1, equal to approximately 11 kilometers, while minimum samples in a cluster was set to 1, as some ports were

Ports On A Map

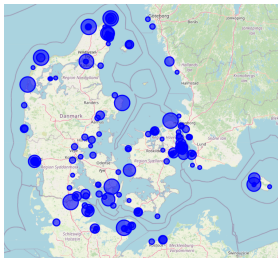


Figure: Spatial Binning

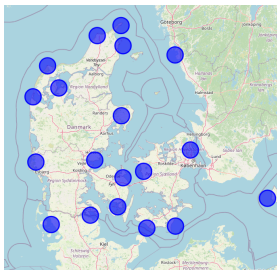


Figure: 19 ports with a K-means clustering

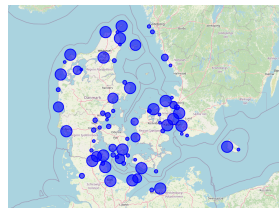


Figure: Spatial Binning aggregated with DBSCAN

Table: Largest ports based on DBSCAN clustering

Port	Vessel Count
København	131
Hirtshals	76
Frederikshavn	68
Skagen	63
Heiligenhafen	62