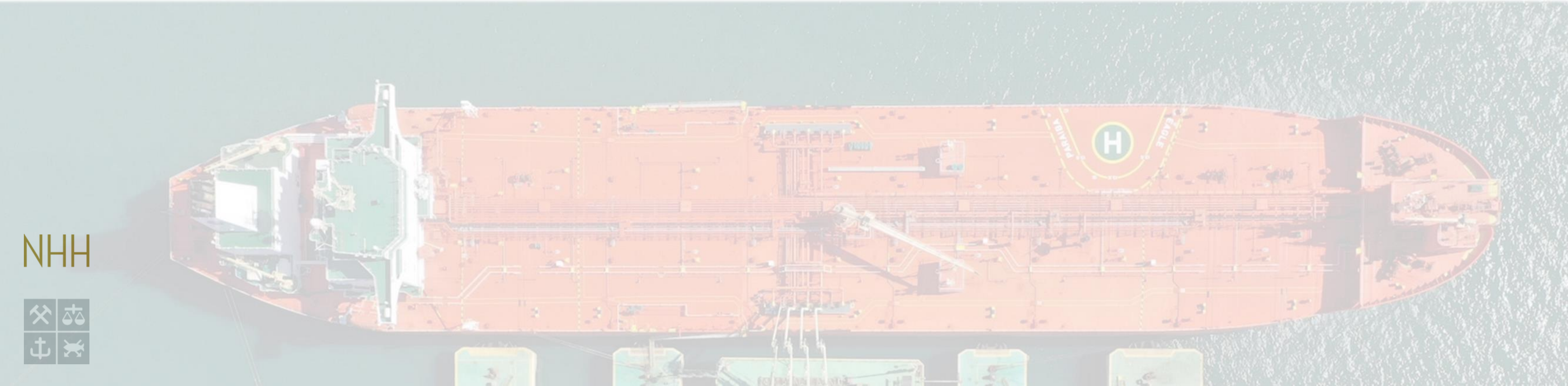
AIS Task team

July 2023

Master thesis presentation

Port waiting time for oil tankers –

Leveraging AIS data to predict port waiting time using machine learning



WHO AND WHY

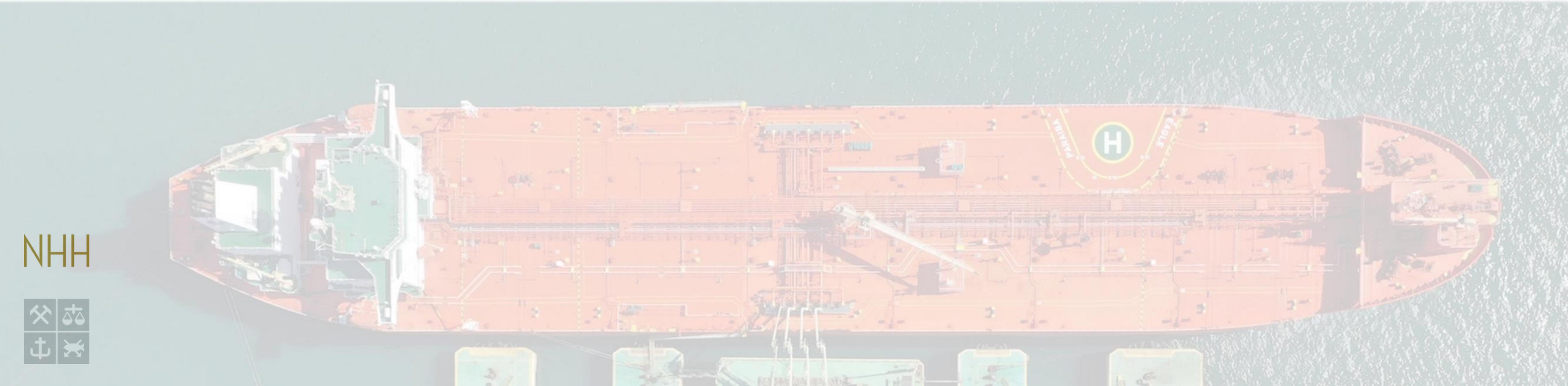
- The shipping industry is fascinating
- Perfect intersection between business analytics and domain knowledge
- Vast amounts of data available
- Possible to find underexplored niches

ERIC GLENJEN

KRISTOFFER SOLBERG

Problem statement:

"Can the waiting times in crude oil ports be predicted based on AIS data?"



NHH

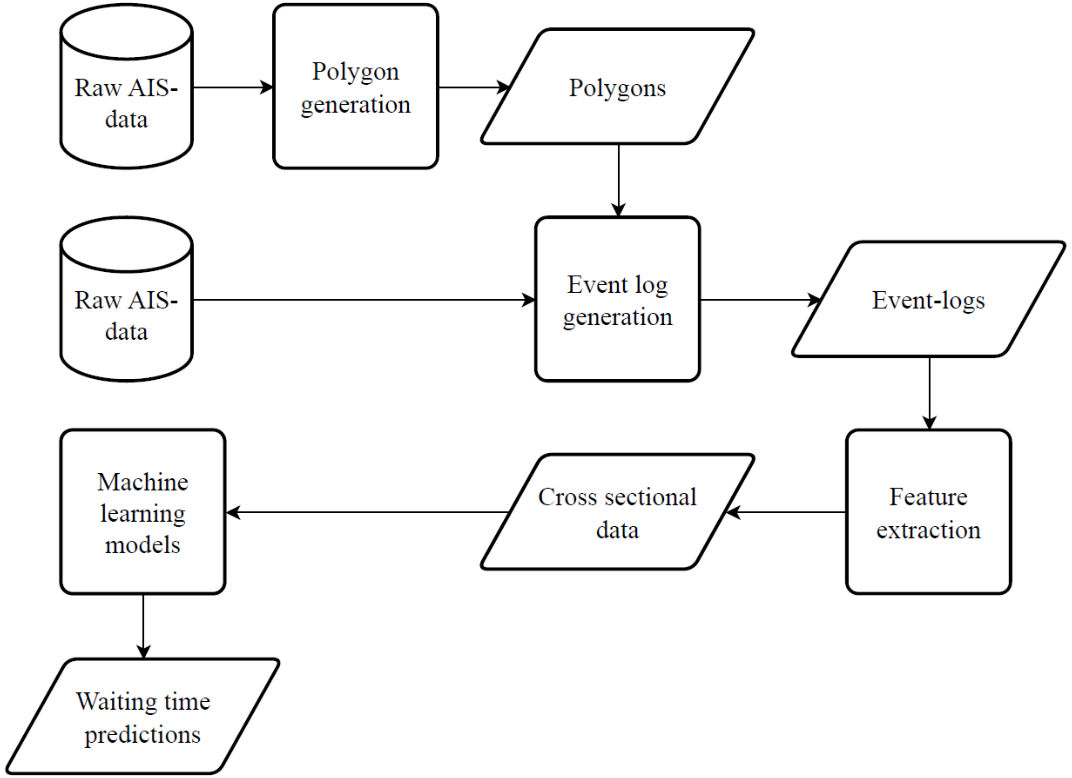
METHODOLOGICAL OVERVIEW

- Our goal was an automated scalable approach

- A three-step process
    1. Establish port areas

    2. Generate event logs

    3. Predict waiting times



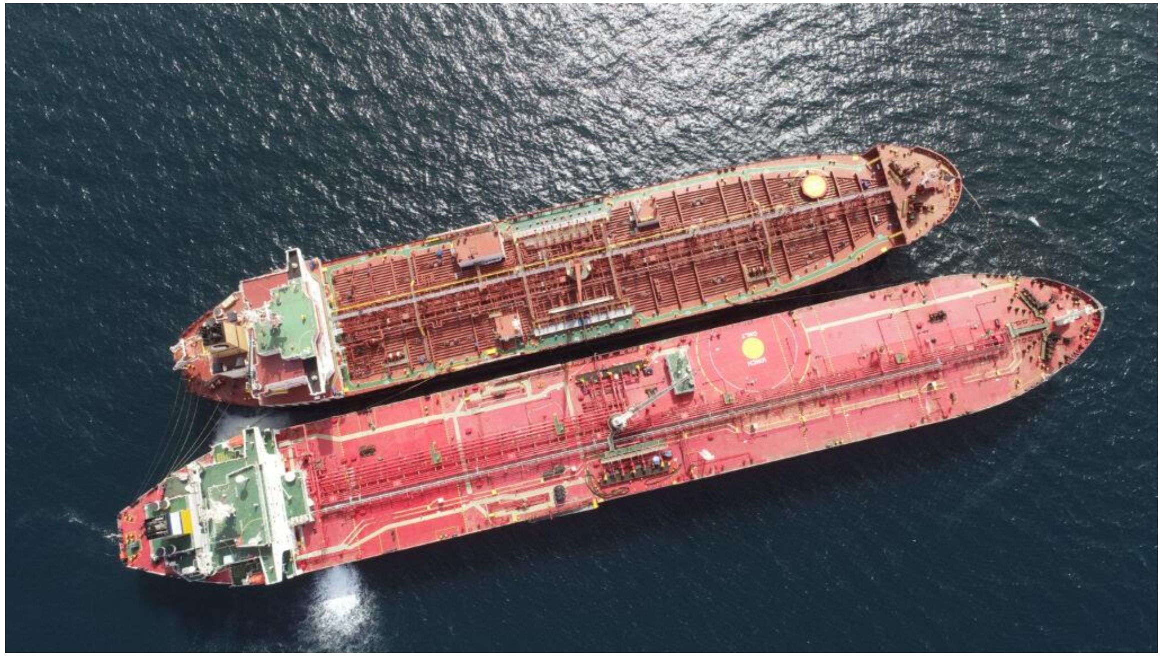
ESTABLISH PORT AREAS

GENERATE EVENT LOGS

PREDICT

ESTABLISH PORT AREAS

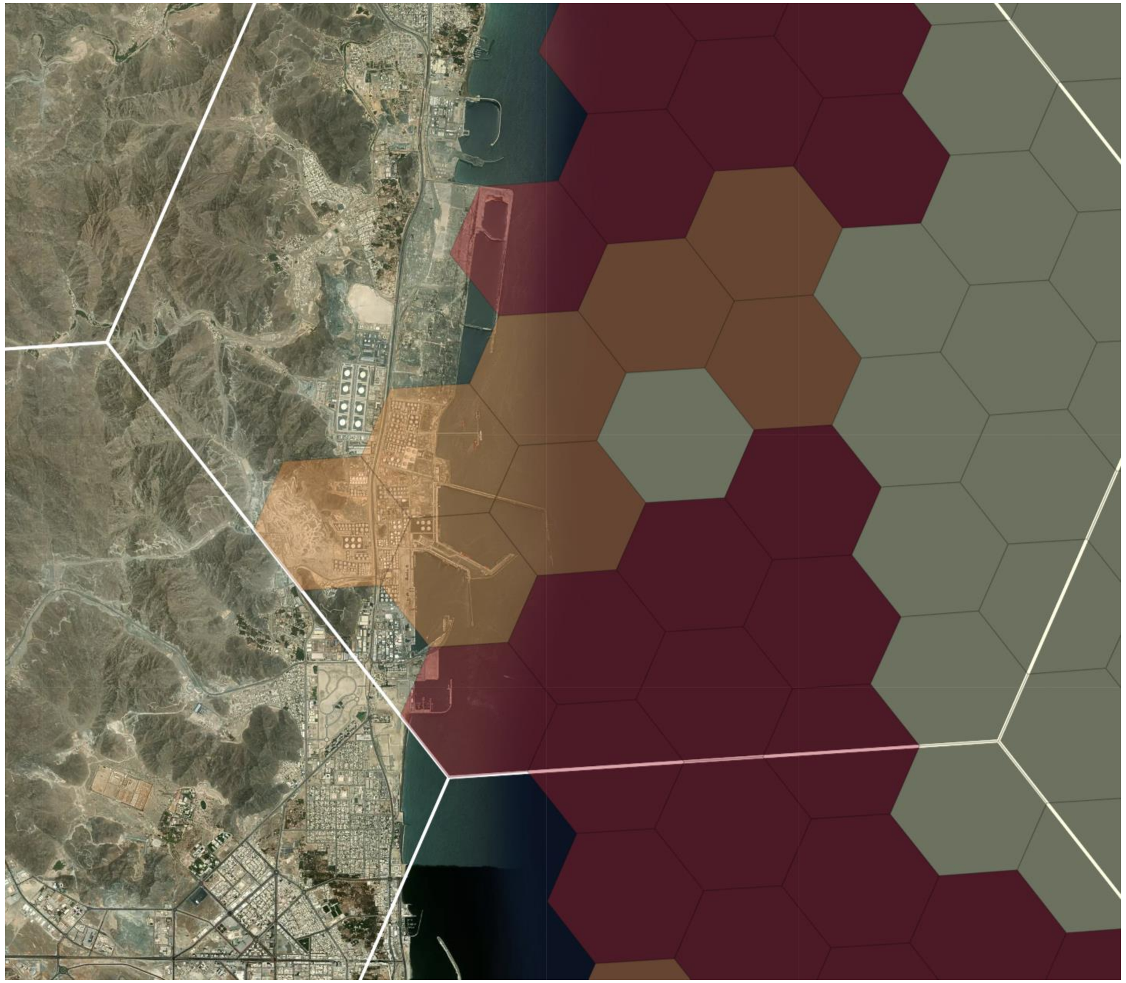
- Partition the port area
- Identify berth arrangements
- Generate clusters using Density-Based Spatial Clustering of Applications with Noise (DBSCAN)
- Identify vertices of the clusters (convex hull) using the Quickhull algorithm



PARTITION THE PORT AREA

- Based on the modal value for navigational status within each hexagon

- Orange – Moored
- Red – Under Way Using Engine
- Beige – At Anchor



NHH

ESTABLISH PORT AREAS

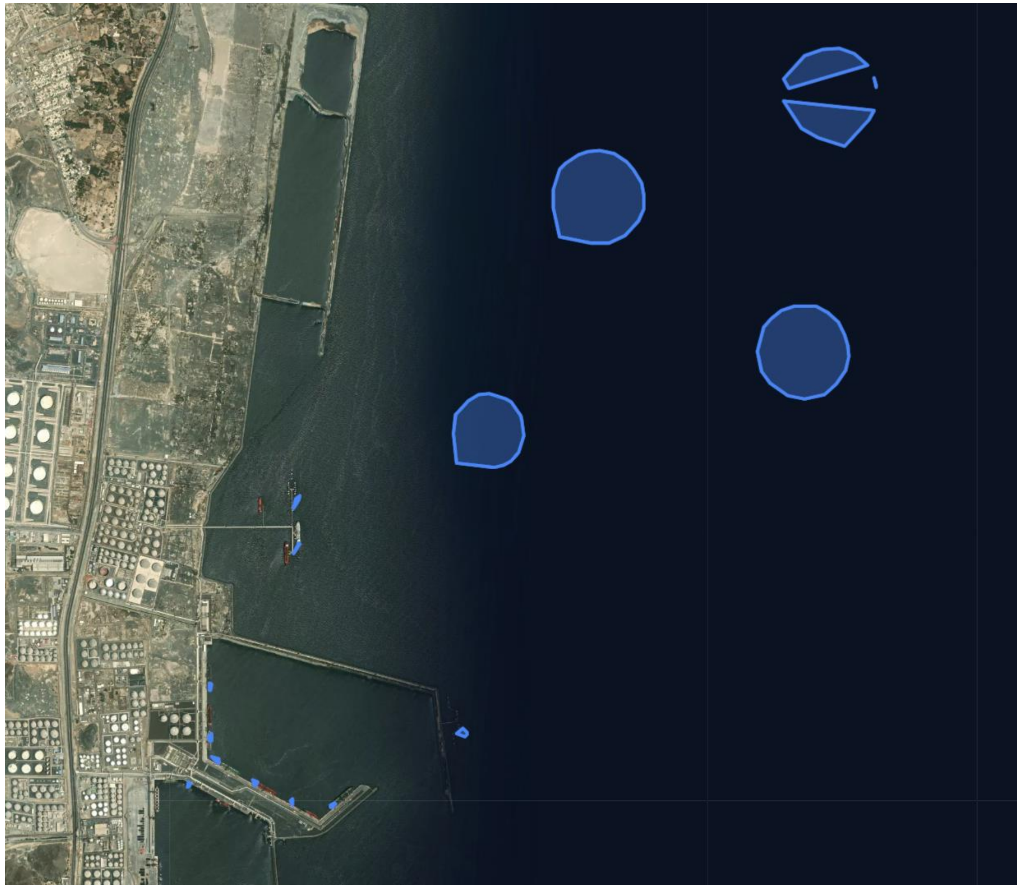
GENERATE EVENT LOGS

PREDICT

IDENTIFY BERTH ARRANGEMENTS

- Identify berth arrangements
  - Pier
  - Single point mooring (SPM)
  - Sea island
- Problem:
  - Difference in cluster density is the Achilles heel of DBSCAN

IDENTIFY BERTH ARRANGEMENTS



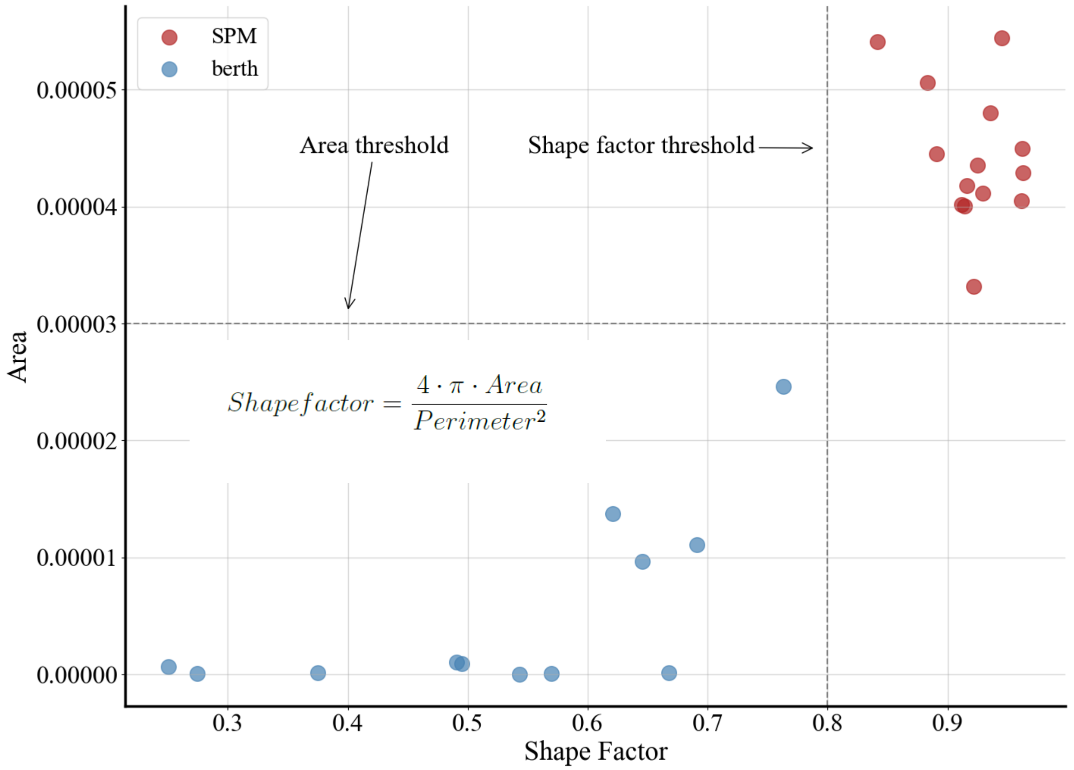
Epsilon = 100 meters

Epsilon = 250 meters

NHH

IDENTIFY BERTH ARRANGEMENTS

- Solution:
  - Use DBSCAN and convex hull to create temporary polygons that can be classified into SPMs or other mooring arrangements
  - The temporary polygons are classified using two parameters
    - Area
    - Shape factor
  - Re-run DBSCAN on SPMs and other mooring arrangements separately



GENERATE FINAL CLUSTERS

- Separating SPMs and other mooring arrangements makes it possible to tune the DBSCAN hyperparameters according to the mooring arrangement in question

- Silhouette score is used to tune the hyperparameters

- The convex hull of each cluster is found using the quickhull algorithm

NHH

ESTABLISH PORT AREAS

GENERATE EVENT LOGS

PREDICT

GENERATE EVENT LOGS

- Each observation constitutes one visit for one vessel within a polygon

- Event logs are a flexible starting point for feature engineering

1	imo	type	number	start	end	Deadweight	ShiptypeLevel5	typnum	vessel_class	time_in	trip_number
2	9779965.0	berth	3	2021-01-01 00:00:51	2021-01-01 15:27:06	113563.0	Crude Oil Tanker	berth3	Aframax	0 days 15:26:15	1
3	9853400.0	berth	1	2021-01-01 19:57:28	2021-01-03 03:32:22	112802.0	Crude/Oil Products Tanker	berth1	Aframax	1 days 07:34:54	1
4	9722900.0	berth	2	2021-01-03 13:31:37	2021-01-04 19:38:09	299629.0	Crude Oil Tanker	berth2	VLCC	1 days 06:06:32	1

DESCRIPTIVE STATISTICS

- The event log is also a great starting point for extracting descriptive statistics

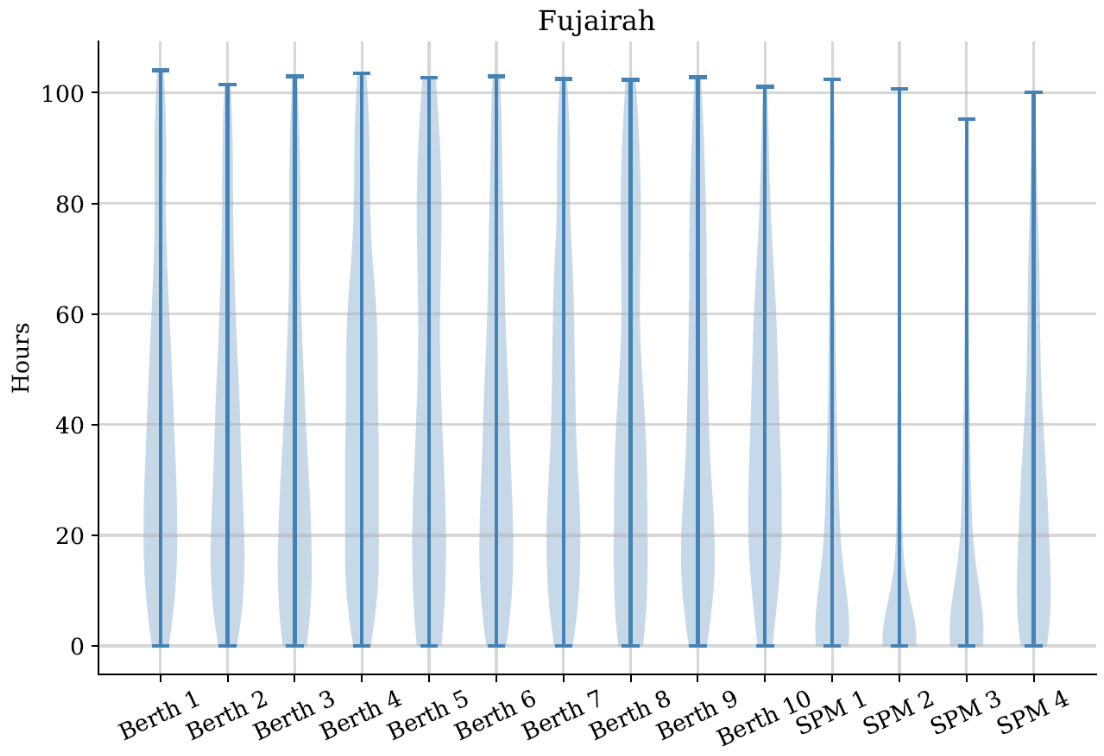
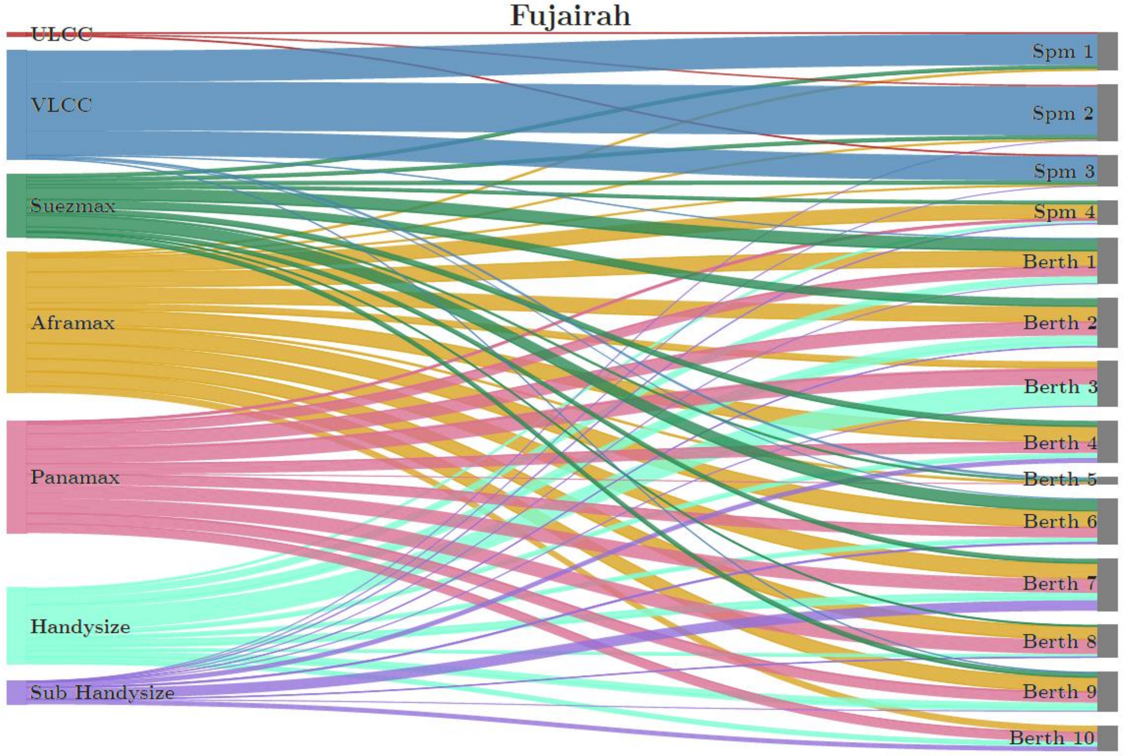
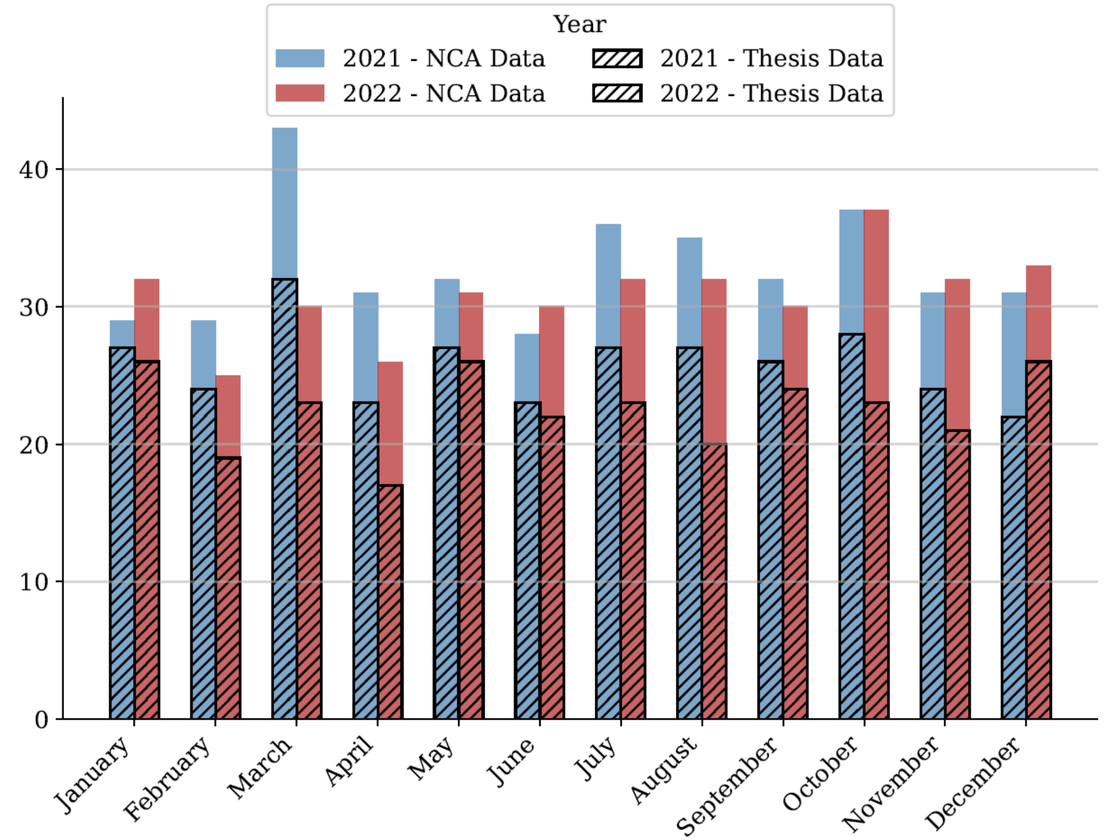


Figure 4.8: Waiting time Fujairah



VALIDATION

- The event log is validated using data from the Norwegian Costal Administration for the Norwegian oil terminals at Mongstad and Sture
- Trip duration are captured precisely
- Consistent underestimation of number of trips
  - Reduced transmission effect on AIS-transceiver when approaching the quay



PREDICT

- Between 85 and 111 features were extracted from the event log
- One observation is one visit
- Five different machine learning models were compared to a baseline model (simple mean)
  - Linear regression (OLS, LASSO, Ridge)
  - Support vector regression
  - Random forest
  - XGBoost
  - Feed forward neural network
- Mean average error (MAE) and mean squared error (MSE) used as evaluation metrics

Category	**Name**	**Type**
Vessel related	Vessel type	Categorical
	Vessel class	Categorical
	Vessel dwt	Continuous
	Predicted berth	Continuous
Temporal	Hour of day	Categorical
	Day of week	Categorical
	Month of year	Categorical
Berth based	Class of vessel i in berth j	Categorical
	Vessel i dwt in berth j	Continuous
	Time since vessel i moored at berth j	Continuous
Queue based	Total vessels in anchorage	Discrete
	Total dwt in anchorage	Continuous
	Number of VLCC in anchorage	Discrete
	Number of Suezmax in anchorage	Discrete
	Number of Aframax in anchorage	Discrete
	Number of sub-Aframax vessels in anchorage	Discrete
(6, 12, 24, 48 hours)	Lagged total vessels in anchorage	Discrete
(6, 12, 24, 48 hours)	Lagged total dwt in anchorage	Continuous
(6, 12, 24, 48 hours)	Lagged number of VLCC in anchorage	Discrete
(6, 12, 24, 48 hours)	Lagged number of Suezmax in anchorage	Discrete
(6, 12, 24, 48 hours)	Lagged number of Aframax in anchorage	Discrete
(6, 12, 24, 48 hours)	Lagged number of sub-Aframax vessels in anchorage	Discrete
(3, 6, 12, 24, 48, 72 hours)	Lagged mean waiting time	Continuous

NHH

ESTABLISH PORT AREAS

GENERATE EVENT LOGS

PREDICT

PREDICTION RESULTS

- Best model improvement (MAE) over the baseline model ranged from 15% to 46% depending on the port
- Random forest and XGBoost were overall the best performing models across both MAE and MSE
- Overestimating waiting time in the lower end of the spectrum, and underestimating in the higher end is the general trend across all models
- Ports with a relative homogenous vessel composition yielded the best prediction results

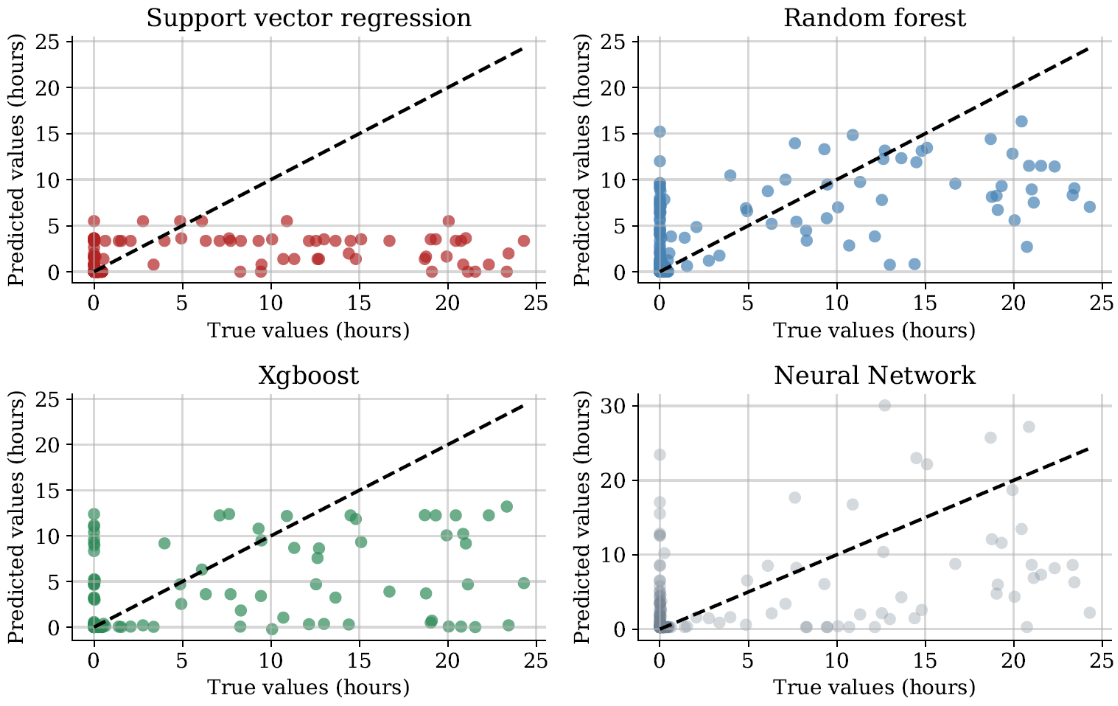


Figure 4.11: Predictions Al Jubail all vessels

NHH

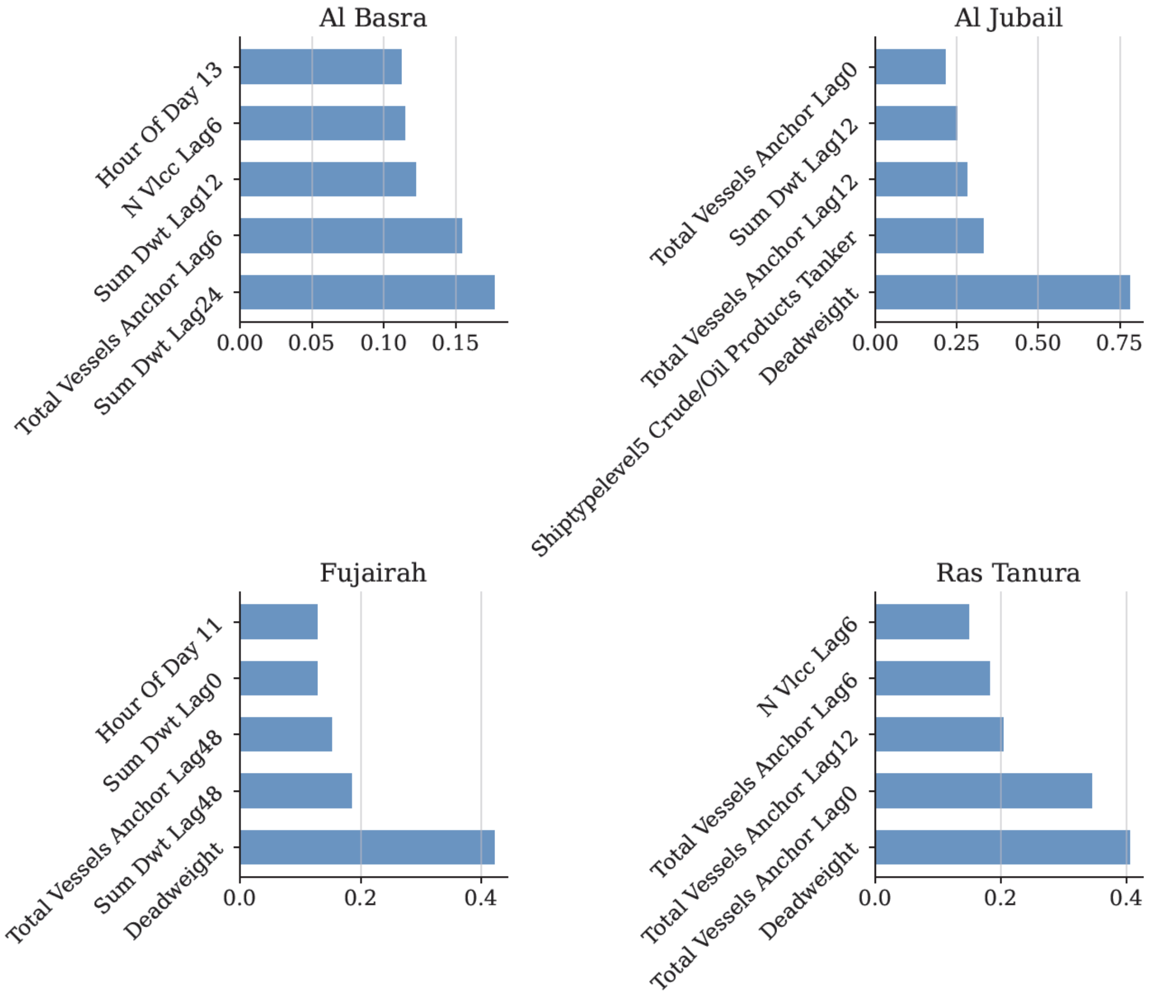
ESTABLISH PORT AREAS

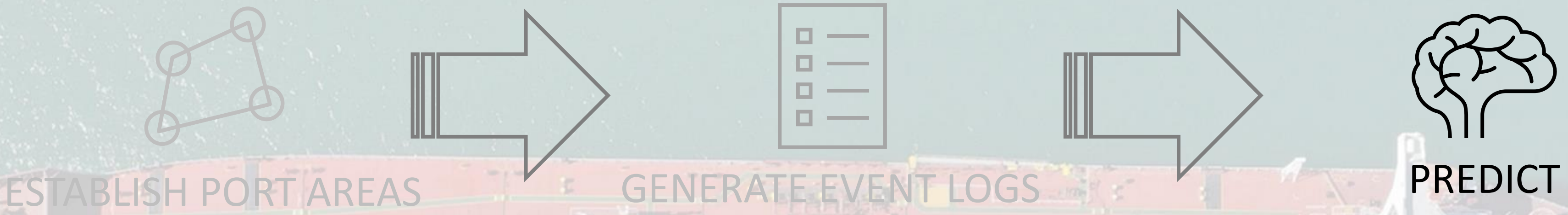
GENERATE EVENT LOGS

PREDICT

FEATURE IMPORTANCE

- Pairing features with descriptive statistics can be used to outline likely causal relationships
    - Deadweight implies assigned berth
    - Lagged tonnage in the anchorage has predictive power
    - Hour of arrival seems to affect waiting time





KEY TAKEAWAYS

- Automated polygon generation must be adapted to shipping segment

- DBSCAN-based methods appears robust, but is likely to underestimate number of trips

- Each port has its own dynamics, in terms of vessels composition and how vessels are served, this affects waiting time

- Waiting time can, to some extent, be predicted using AIS-data

ESTABLISH PORT AREAS → GENERATE EVENT LOGS → PREDICT

ERIC GLENJEN

THANK YOU!

KRISTOFFER SOLBERG

NHH