# Pipeline for open AIS data with filtering based on vessel class

Mirjam Bayer, Tabea Fry, Sören Dethlefsen, Peer Kröger
University of Kiel, Germany

Image from
NASA
STS-52

# AIS Data?

Image by Pixabay _marion


Image from
blue novation


Pipeline
Pipeline

# AIS Data

## AIS: **A**utomatic **I**dentification **S**ystem

- Satellite Based Communication
- Features include:
  - Unique Vessel ID (MMSI)
  - GPS Position
  - Speed
  - Navigational Status
- History of messages build trajectory

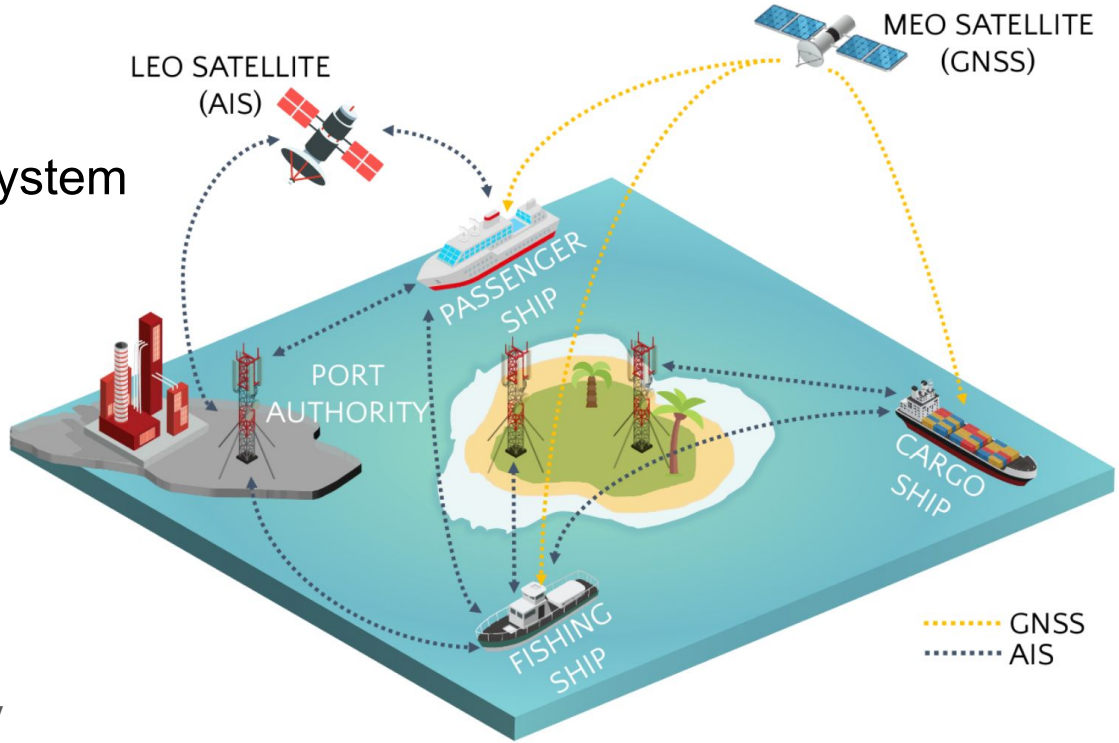



Image by ??

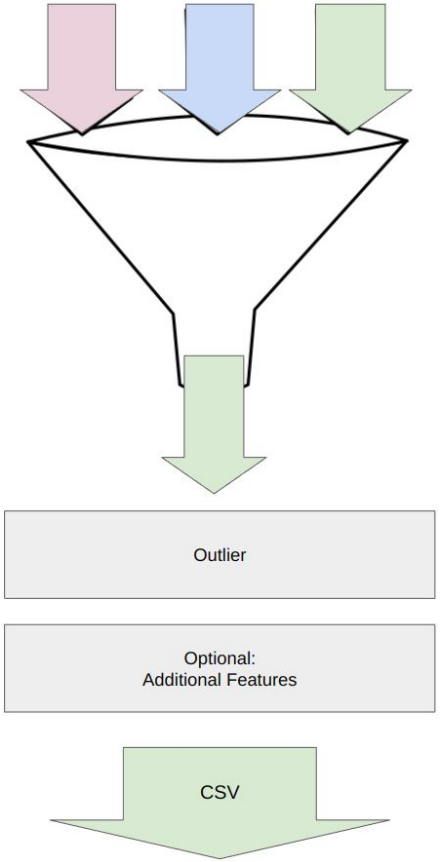
# Pipeline

Goal 1: Volume Reduction

- Filtering to contain only messages of desired ship type

Goal 2: Curate Dataset

- Cleaning
- Feature Engineering



## Fishing Vessel

- Navigational status
  - LIKE “*% fishing%*”
- Ship Type
  - LIKE “*%fishing%*”

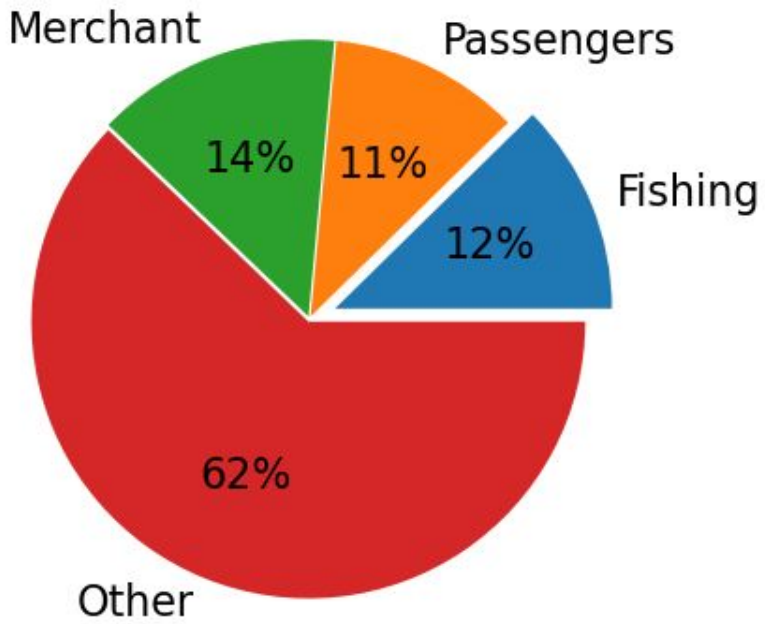
## Merchant Ship

- Cargo Type not *Nan*
- Possible distinction between defined classes

## Passenger Ship

- Ship Type LIKE “%passenger%”

# Volume reduction ✔



# “Dark Vessels”

- Hiding activity and traces
- Database containing MMSI of ships previously detected as one of the aforementioned type
- Growing database with usage

# Other Features of Pipeline

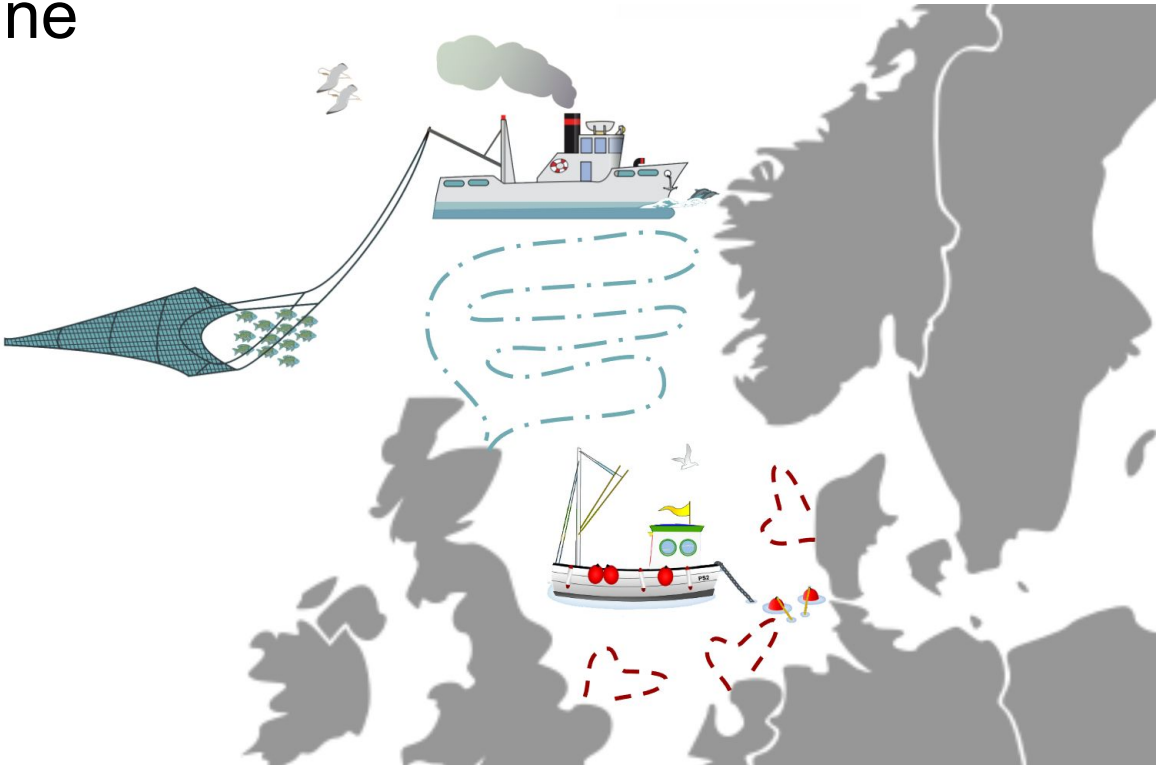
- Cleaning
  - unrealistic values
  - exceeding thresholds
- Enhancement with other features
  - Water depth
  - computed speed
- Output:

Curated dataset ✔

# How I use this pipeline

## Fixed Gear Distinction

- Gill net
- Crab pots
- Stationary longlines



# How you can use this pipeline

- Open source code (of course)
- Instruction for running on Git
- Data stream sources
  - DMA
  - Marine Cadastre
  - Links in paper
- Data sets
  - Brest
  - Piraeus
  - GFW “Anonymous AIS training data”

# Open AIS Data Pipeline


Mirjam Bayer, Tabea Fry, Sören Dethlefsen, Daniyal Kazempour

# Drawbacks

- Regional restrictions based on open sources
- not real time
- Long processing time for adding water depth
- Written in Python, future in R?

# Classes of Cargo in AIS messages

**Based on ship type number:**

**1** = Major Hazard (Haz A)

**2** = Hazard (Haz B)

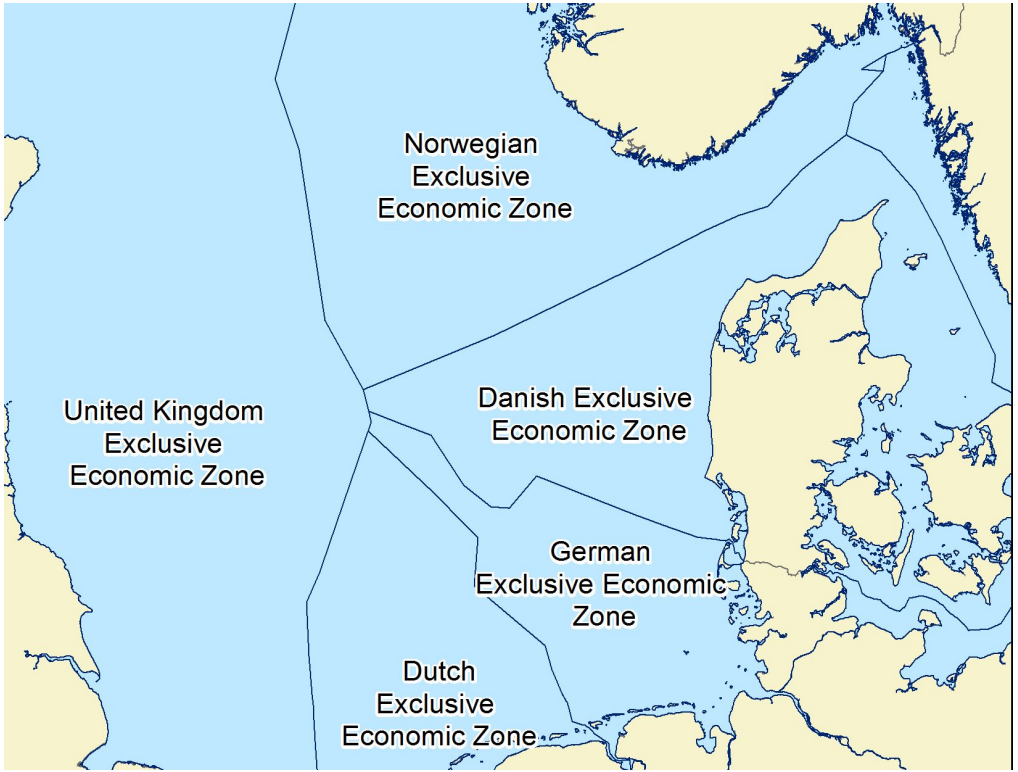
**3** = Minor Hazard (Haz C)

**4** = Recognisable Hazard (Haz D)

# Contribution to database

- Contact for newest SQLite3 database
- Currently 5 GB
- If outgrows shareability split into three based on class

# Region of DMA




## Marine Cadastre

vessels in U.S. and international waters in near real time