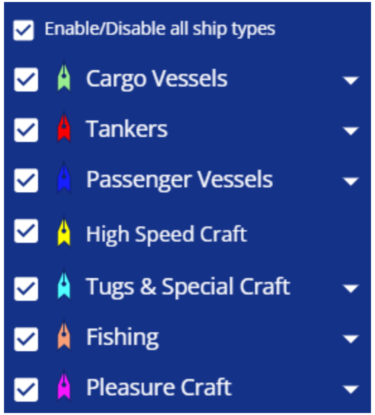
Membership Inference Attacks in a Maritime Context

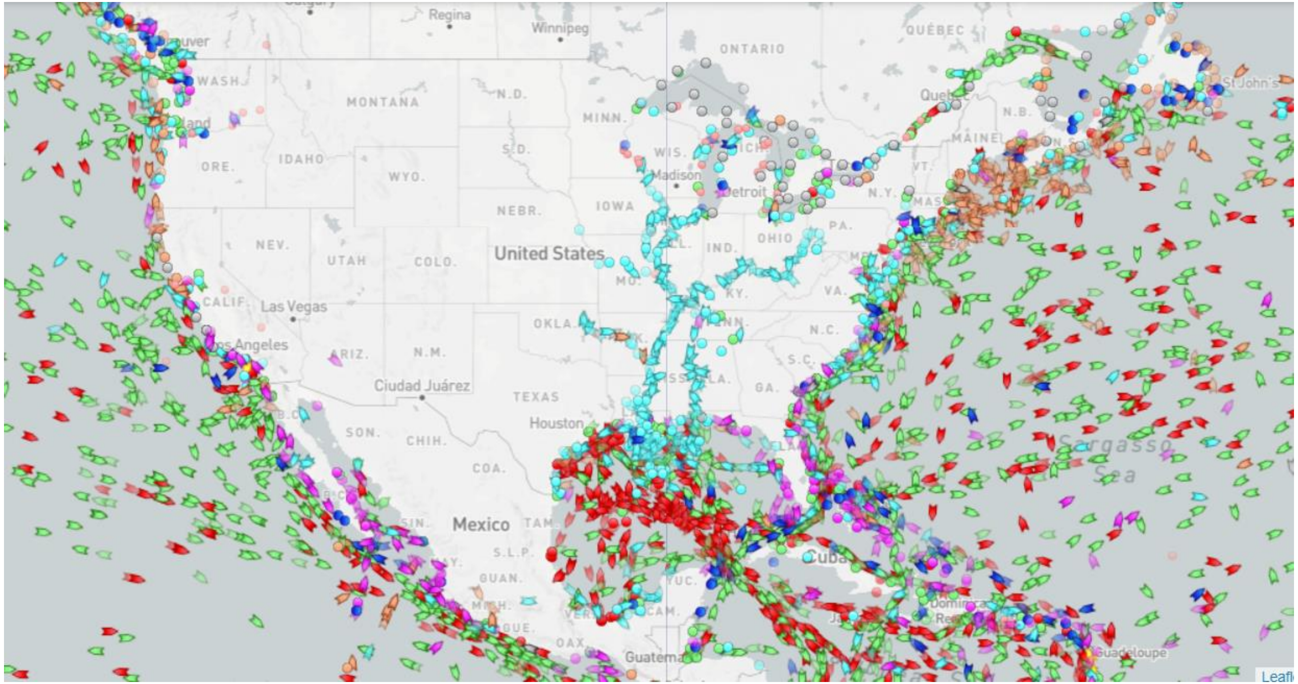
Anonymization, Privacy (2021-2022) - Final Project

Pierre Onghena

Background

Maritime Transport





https://www.marinetraffic.com

Problem Statement

Example: A spoofed ship that carries out fishing in a restricted area

I'm no tanker but a fishing vessel

https://mpatlas.org/zones/

→ To identify the vessel type, AIS data could be used to develop a classification model

Privacy AIS data

- Automatic Identification System (AIS) is a tracking system to supplement information about each moving vessel

- This data could be exploited for anomaly detection or classification of vessel types

- However, it represents business operations of ports and shipping companies

Data for classification

AIS data

- Past data is available for U.S. coastal waters for calendar years 2009 through September 2021 *(https://marinecadastre.gov/ais/)*

- It represents a time-series as every ship, with unique identifier MMSI, transmits AIS data every 1 minute

Geolocation of datapoints for day 01/01/2019

Data Dictionary

Following features are selected for classification:

1. MMSI (Identifier)
2. LAT
3. LON
4. SOG
5. COG
6. Heading
7. Length
8. Width

VesselType

	Name	Description	Example	Units	Resolution	Type	Size
1	MMSI	Maritime Mobile Service Identity value	477220100			Text	8
2	BaseDateTime	Full UTC date and time	2017-02-01T20:05:07		YYYY-MM-DD:HH-MM-SS	DateTime	
3	LAT	Latitude	42.35137	decimal degrees	XX.XXXXX	Double	8
4	LON	Longitude	-71.04182	decimal degrees	XXX.XXXXX	Double	8
5	SOG	Speed Over Ground	5.9	knots	XXX.X	Float	4
6	COG	Course Over Ground	47.5	degrees	XXX.X	Float	4
7	Heading	True heading angle	45.1	degrees	XXX.X	Float	4
8	VesselName	Name as shown on the station radio license	OOCL Malaysia			Text	32
9	IMO	International Maritime Organization Vessel number	IMO9627980			Text	16
10	CallSign	Call sign as assigned by FCC	VRME7			Text	8
11	VesselType	Vessel type as defined in NAIS specifications	70			Integer	short
12	Status	Navigation status as defined by the COLREGS	3			Integer	short
13	Length	Length of vessel (see NAIS specifications)	71.0	meters	XXX.X	Float	4
14	Width	Width of vessel (see NAIS specifications)	12.0	meters	XXX.X	Float	4
15	Draft	Draft depth of vessel (see NAIS specifications)	3.5	meters	XXX.X	Float	4
16	Cargo	Cargo type (see NAIS specification and codes)	70			Text	4
17	TransceiverClass	Class of AIS transceiver	A			Text	2

Classification: Vessel Type

- A total of 10 classes be retained to construct an informative probability vector
- The 10 most common vessel types are the following:

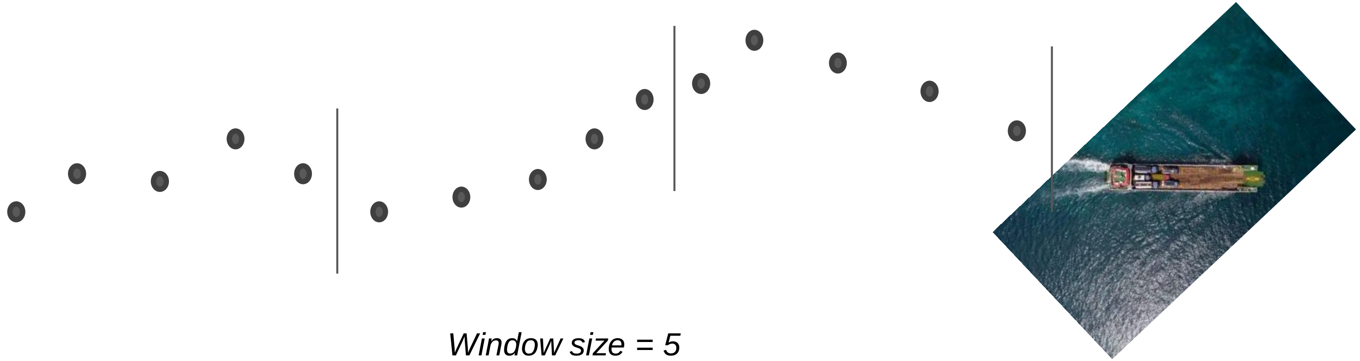
    1. Fishing
    2. Towing
    3. Sailing
    4. Pleasure craft
    5. Pilot vessel
    6. Tug
    7. Passenger
    8. Cargo
    9. Tanker
    10. Other

Research Method

Architecture: LSTM

- An LSTM is designed to capture relationships on sequence data. Therefore, it could monitor the changes in a ship's trajectory

- As neural networks require inputs to have the same shape, a ship trajectory is divided into multiple windows of the same size



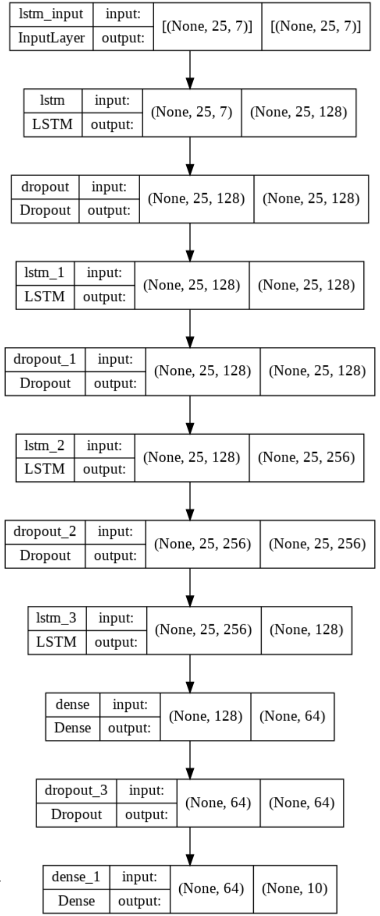
Window size = 5

Target Model

Training data: 01/01/2019
Test data: 01/01/2021

Approx. half of the ships reoccur in both datasets, but still with different trajectory

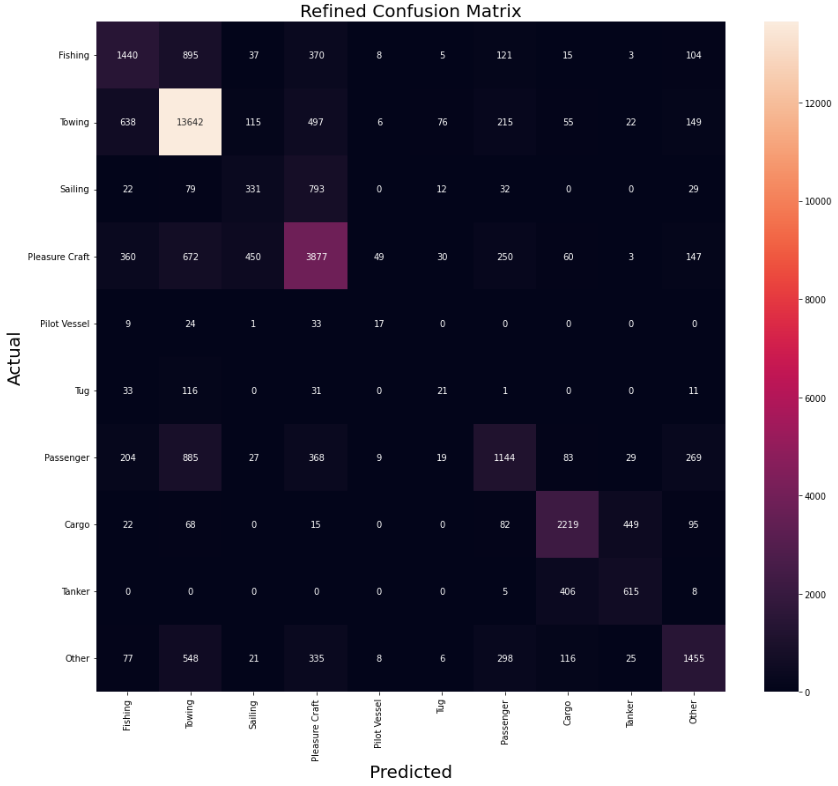
- Input size: (window size = 25, features = 7)
- 4 LSTM layers with dropout to support regularization
- Probability distribution for the 10 classes:



Training and Test Accuracy

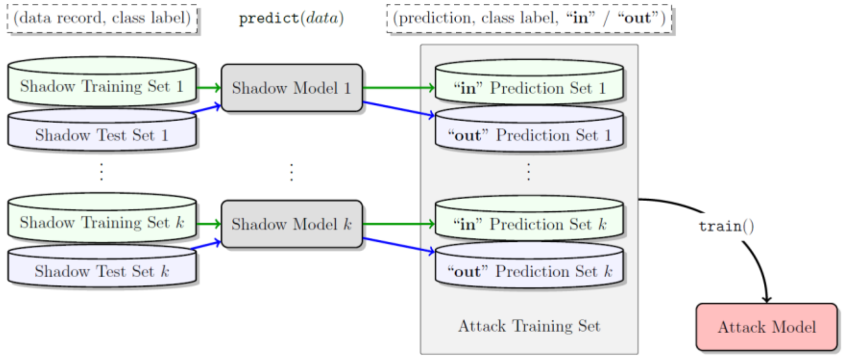
Training accuracy: 93%
Test accuracy: 70%

→ Overfitting causes the target model to be vulnerable to membership inference



Membership Inference

- To conduct membership inference, an attack model must recognize training examples
- Therefore, to distinguish training from test examples, several shadow models that mimic the target output are created by the attacker
- From the shadow model, it’s known which the training and test data are, so we could label them for binary classification of the attack model

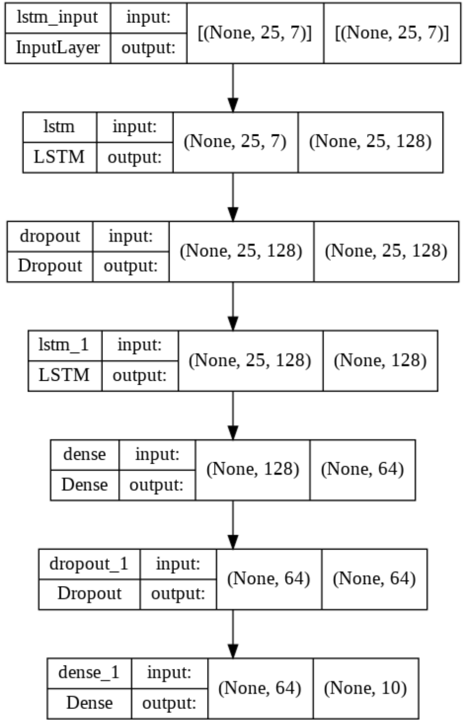


Shadow Models

- Similar structure to target model but primarily with less layers and different window size
- Due to time restrictions, only able to train two LSTM shadow models
- Training data: 17/04/2018 and 11/05/2017
  Test data: 18/12/2018 and 19/02/2017

Training accuracy:	86%	93%
Test accuracy:	82%	62%

→Much less overfitting for shadow model 1, which exploits only 2 LSTM layers



Shadow model 1

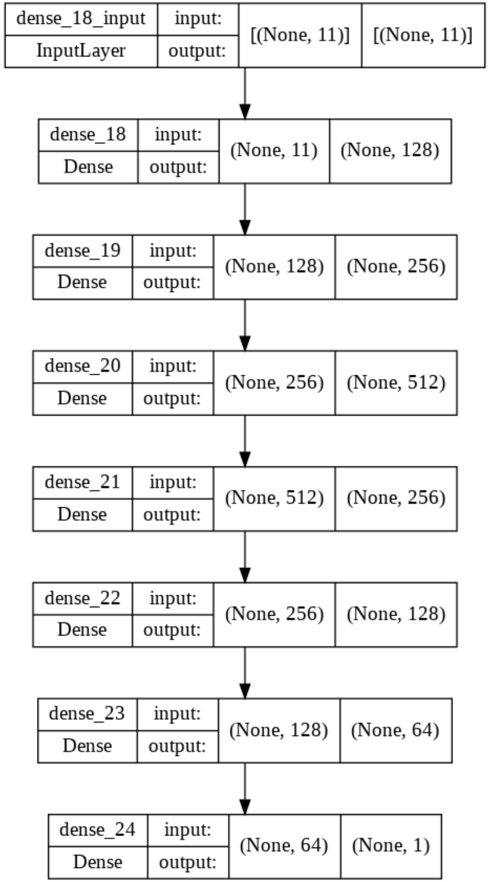
Attack Model

- Binary classification based on the probability vector of the (10 classes + correct class label)

```
,0,1,2,3,4,5,6,7,8,9,Class,Label
0,0.9998586,3.152706e-05,6.668778e-10,1.11297195e-05,1.0989926e-05,1.7170145e-06,6.705457e-07,2.0995101e-05,2.398303e-11,6.436373e-05,0.0,1
1,0.9994924,3.9057642e-05,3.6029206e-09,3.359317e-05,9.34889e-06,1.4630317e-06,3.0384215e-06,0.00031571605,5.458346e-10,0.00010536124,0.0,1
2,0.99948174,1.7052345e-05,1.887342e-09,3.506276e-05,1.0168921e-05,1.7199261e-06,3.4476573e-06,0.00034366574,3.3681197e-10,0.000107281805,0.0,1
```

- 7 dense layers

- Sigmoid activation to determine whether the example was used during training or not



Experimental Results

Evaluation

- Training data: probability vectors, members / non-members, of both shadow models get concatenated and shuffled

→ Total of 155.000 examples for training attack model

- The attack model will be evaluated on an equal number of members and non-members:

  $1/2$ Training data: 01/01/<u>2019</u> – members

  $1/2$ Test data: 01/01/<u>2021</u> – non-members

→ Total of (35.000 + 35.000=) 70.000 examples for classification

- The limitation above implies that the accuracy would normally be greater than 50% of random guessing

- Confusion matrix to access the performance

Confusion Matrix

Training accuracy: 63%
Test accuracy: 60%

→ Slightly above random guessing of 50%

Precision: 57%
Recall: 80%

→ Low precision, also actual non-members are often interpreted as members
→ High recall, which means that actual members are rarely missed by the model