

Anomaly and causality detection in Maritime Data

Daniel Blindu

Scope

Anomaly detection in maritime data can help with legal investigations or help commercial transport



MarineTraffic

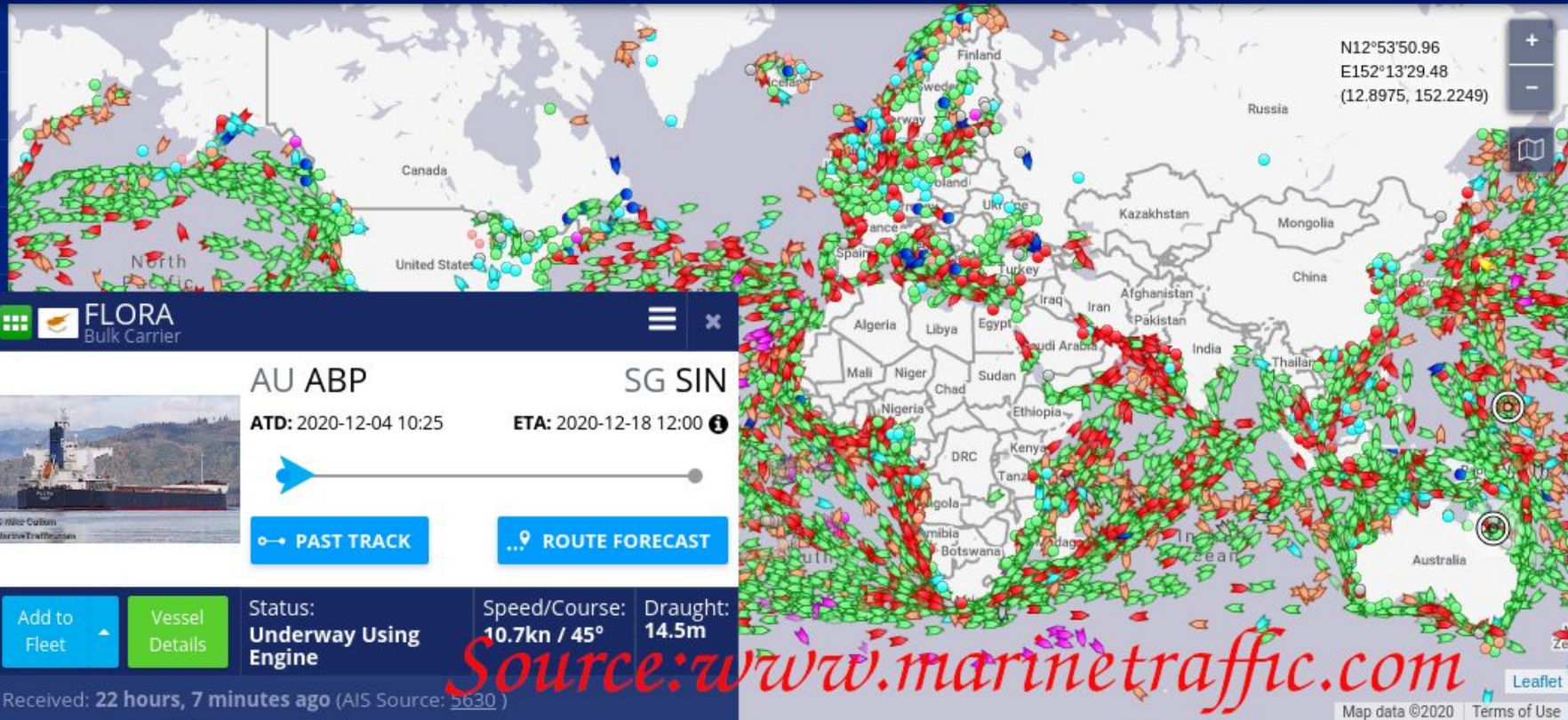
Live Map

Explore ▾

Community ▾

Pricing ▾

Search MarineTraffic



AIS and weather Data sample from Spire Maritime

	A	B	C	D	E	F	G	H	I	J	K	
1	created_at	timestamp	mmsi	msg_type	latitude	longitude	speed	course	heading	rot	imo	name
2	2020-08-20 23:20:52.957236 UTC	2020-08-20 23:05:17.4 UTC	352192000	1	16.711835	115.72252	13.5	32	31	-128		
3	2020-08-20 03:46:53.200501 UTC	2020-08-20 04:01:54.13 UTC	352192000	1	12.89456	113.18959	13.8	34	30	-128		
4	2020-08-20 18:58:54.146592 UTC	2020-08-20 18:44:20.27 UTC	352192000	1	15.86760166666667	115.192788333333	13.9	31	30	-128		
5	2020-08-20 11:48:53.639409 UTC	2020-08-20 11:40:05.2 UTC	352192000	1	14.4878566666667	114.269908333333	13.9	33	33	-128		
6	2020-08-20 14:49:54.11236 UTC	2020-08-20 14:43:06.96 UTC	352192000	1	15.0750733333333	114.668483333333	13.7	30	27	-128		
7	2020-08-20 01:13:53.884127 UTC	2020-08-20 01:05:37.26 UTC	352192000	1	12.4533616666667	112.89159	14	34	27	-128		
8	2020-08-20 00:10:56.31105 UTC	2020-08-20 00:07:47.63 UTC	352192000	1	12.2717	112.761751666667	13.7	35	30	-128		
9	2020-08-20 05:42:56.758928 UTC	2020-08-20 05:32:22.26 UTC	352192000	1	13.3128816666667	113.464995	13.8	33	34	-128		
10	2020-08-20 20:59:57.955051 UTC	2020-08-20 20:55:48.69 UTC	352192000	1	16.2948433333333	115.467185	14.1	31	28	-128		
11	2020-08-20 17:53:53.83088 UTC	2020-08-20 17:51:46.81 UTC	352192000	1	15.689925	115.08573	14	33	31	-128		
12	2020-08-20 14:34:54.146516 UTC	2020-08-20 14:21:16.35 UTC	352192000	1	15.005585	114.61994	13.7	34	34	-128		
13	2020-08-20 08:46:53.701766 UTC	2020-08-20 08:38:35.48 UTC	352192000	1	13.9049783333333	113.872846666667	13.6	32	36	-128		
14	2020-08-20 15:15:13.143997 UTC	2020-08-20 15:08:16.74 UTC	352192000	1	15.1574166666667	114.720548333333	13.7	36	35	-128		
15	2020-08-20 18:38:53.835201 UTC	2020-08-20 18:29:12.26 UTC	352192000	1	15.8165733333333	115.16147	14	30	29	-128		
16	2020-08-20 11:20:54.626487 UTC	2020-08-20 11:13:56.98 UTC	352192000	1	14.4041533333333	114.211083333333	13.9	34	35	-128		
17	2020-08-20 12:36:53.507044 UTC	2020-08-20 12:34:46.37 UTC	352192000	1	14.665675	114.38987	13.8	32	31	-128		
18	2020-08-20 11:27:57.562808 UTC	2020-08-20 11:19:51.44 UTC	352192000	1	14.4234883333333	114.224431666667	13.9	34	36	-128		
19	2020-08-20 08:14:54.779876 UTC	2020-08-20 08:10:06.66 UTC	352192000	1	13.8162466666667	113.809603333333	13.8	35	40	-128		
20	2020-08-20 04:57:54.358743 UTC	2020-08-20 04:53:37.69 UTC	352192000	1	13.1887216666667	113.383323333333	13.3	34	33	-128		
21	2020-08-20 11:01:01.940505 UTC	2020-08-20 10:58:55.65 UTC	352192000	1	14.3559433333333	114.177885	13.8	34	36	-128		
22	2020-08-20 20:55:54.411512 UTC	2020-08-20 20:43:55.28 UTC	352192000	1	16.2545916666667	115.44332	14	30	28	-128		
23	2020-08-20 19:58:53.511434 UTC	2020-08-20 19:47:33.69 UTC	352192000	1	16.0639716666667	115.33987	13.5	46	44	-128		
24	2020-08-20 13:11:23.490459 UTC	2020-08-20 13:05:11.92 UTC	352192000	1	14.764465	114.454695	13.6	35	34	-128		

DAIS_SouthChinaSea_20200820

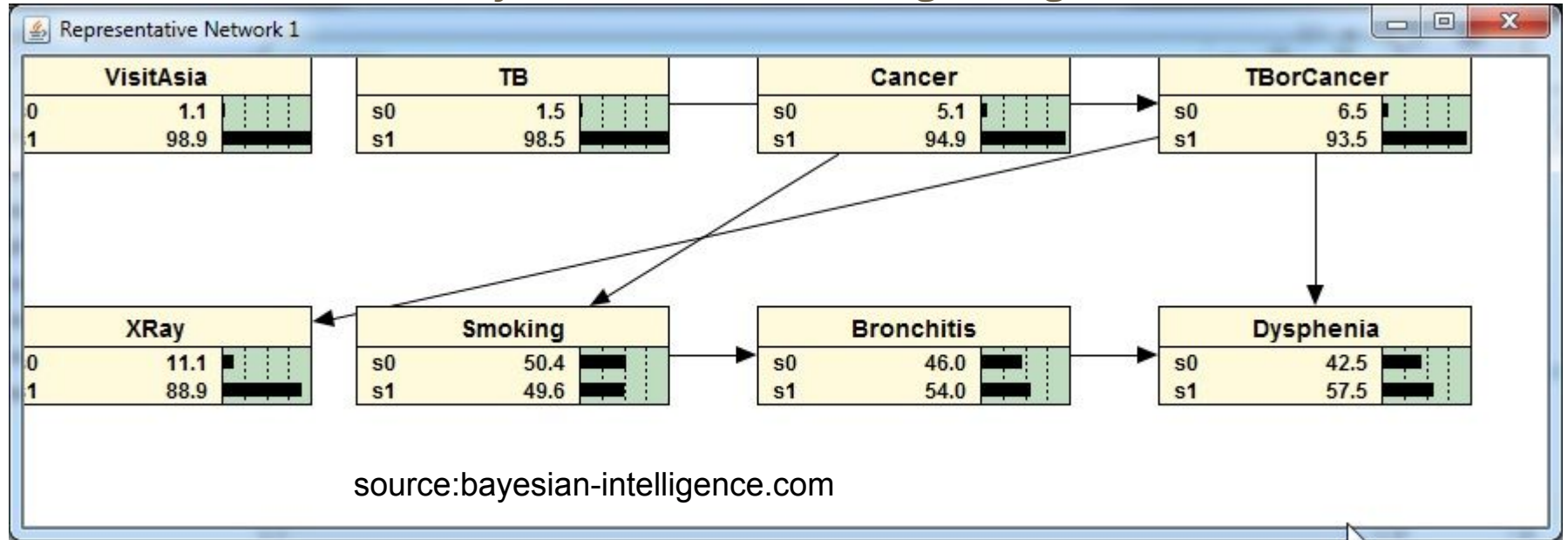
source:spire.com

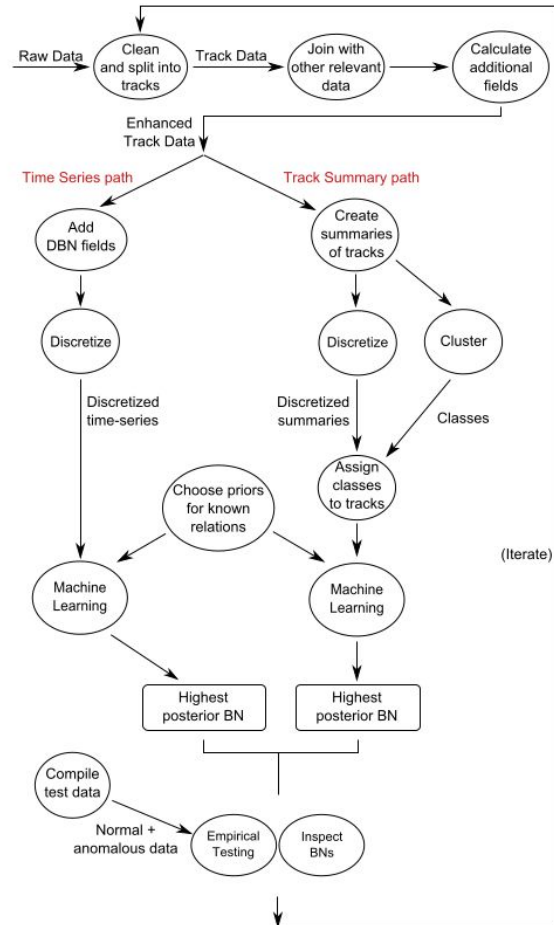
Bayesian Networks using two models:

1. **Time series** model
2. **Track summary** model

Tools

CaMML (Causal discovery via Minimum Message Length)





Data is combined, cleaned, fed to the tool and also refined and re-introduced in the learner via the two methods:

- Time Series
- Track Summary

Usage

You can use **discrete_time_series.py** or **discrete_track_summaries.py**

One is taking AIS data and checks anomalies based on time (one ship in the sea can be missing at 10 AM at specific location, which is bad because usually it's there at that time)

Another one is track records and its' used to show for example that ships in the middle of the ocean usually are bigger

Usage

When you run one of the scripts it will download some AIS sample and

- Clean data
- Make continuous variable into discrete (instead of ship 500m -> ship has value 'large')
- Output a new CSV file

Usage

Run `camml_gui.sh` with the new CSV