**Spoofing detection from AIS GPS data**

Detecting spam/spoofing ship data from AIS database. It’s a big data project process with parallel computing. AIS data file contains around 700000 data. Which I have to clean, preprocess and filter the data first. Step by step approach,

1. Load the data using panda
2. Clean the data and drop unnecessary/incomplete rows
3. Delete unnecessary column so data load is reduced
4. Change the timestamp to date format
5. Sort dataframe by shipid and timeframe
6. Then again groupby those by shipid and store them on tuples, so I can convert them to chunk by 10 ship data
7. Then chunk all those tuples data to chunk by 10 ship data combinedly
8. Now, chunk list contain: approx. 227 list components. Each component contains 10 ship all information timewise.
9. Identify Location anomalies so far
10. Multiprocessing done with 7 cores
11. There were almost 50%-time efficiency in multiprocessing
12. Multiprocessing was done by PyCharm as Notebook can’t handle it