Data Scientist Recruitment Exercise 2020

The Client is a marine electronics company. They produce hardware and software for sailing yachts. They have installed some sensors on one of the boats and provide the dataset they've collected. They don't have any experience with machine learning, and they don't have a solid understanding of what exactly they want to do with this data.

The .csv file with the data and data dictionary can be found in an attached ZIP

Your **first task** is to analyse the data, give the client some feedback on data collection, cleaning and handling process.

Your supervisor told you that on top of whatever you come up with, what you should definitely do is 'tack prediction'. "A tack is a specific maneuver in sailing and alerting the sailor of the necessity to tack in the near future would bring some advantage to them compared to other sailors, who would have to keep an eye out on the conditions all the time to decide when to tack," he writes in his email. The supervisor, who has some experience in sailing labels the tacks in the data from the client (added as 'Tacking' column in the data).

Your **second task** is to build a model that will alert sailors of tacking event happening in the future.

The **third task** is present an architecture that would be used to ingest the training data, train a model and deploy it out to a low power device that would run onboard a boat.

Your **deliverables:**

- one coherent presentation of your work on tasks **all three tasks**. The presentation should be no longer than 30 minutes, excluding the questions that will arise during and after the presentation
- All the code (scripts/notebooks) you have produced published on your public GitHub (or equivalent) repository. You should provide the link to this repository no less than 2 days prior to your interview

During the exercise you can ask clarifying questions like you would ask the client. The subject of your email should be "*<first name>: exercise*".