Data Science Assignment: Ship Performance

Introduction

One of the fundamental problems in maritime transportation is predicting how much fuel will be used on an upcoming voyage. Imagine telling the vessel to go a certain speed, and you need to know how much fuel will be consumed across the various environmental conditions along the voyage. To begin to tackle this problem, we first need a performance model of the ship. Such a model will output a prediction of the rate of fuel usage given a set of features.

The speed of the ship (measured in knots) is the biggest predictor of the rate of main engine fuel consumption (measured in metric tonnes per day). In addition, other operating and environmental factors also influence main engine fuel usage: how loaded the ship is (measured via draft [https://en.wikipedia.org/wiki/Draft_(hull)]), wind speed and direction, acceleration/deceleration, rudder angle, time since last maintenance, sea state, current, and others. There are also other features like engine power, shaft RPM, and GPS distance, which may be useful. But think carefully about these features and how they relate to the problem of speed+env predicts fuel. For example, engine power is better thought of as an output of this model than an input.

Data

Included is a zip file of 16248 rows by 26 columns of CSV data from over 1 year on an LPG, VLGC ship (Liquefied Petroleum Gas, Very Large Gas Carrier [https://en.wikipedia.org/wiki/Gas_carrier]). The data contains just a subset of the data we've collected on the ship. It includes: fuel rate, fuel type, speed, wind, draft, etc. (More detail in the appendix below.) The original data from the ship was at 1-min samples, and has been downsampled to 1-hour samples in the included CSV. The samples are of both when the ship was at-sea and in-port.

Goal

Using this data, the goal of this assignment is to do the following:

1. Create a machine learning model that outputs a prediction of the rate of main engine fuel usage (mt/day) from a set of feature inputs. (You don't have to use all the columns of data as features.)
2. Use your model to score a range of speeds while holding other features constant (e.g. draft = 10m, wind = 0 knots, etc…) Plot multiple of these fuel-consumption versus speed curves in a single diagram to compare the effects of different drafts, different environmental factors, etc...
3. Submit a report detailing the work that was done.
4. Present this report and results during the onsite interview.

Just as important as the model and the graphs is a report documenting the process for how such a model was created. Such a report should be able to answer questions like: what was your process, what else was tried, what failed, how good are the results (describe the error), are the results better under some conditions compared to others, did you throw out some data points, what would you do next if you had more time, any other interesting insights. Communicating results is as important as getting these results. During the onsite interview, you will be present this report. Feel free to use slides or a "word" document for the report.

You do not need to send any code. We've already developed such models at Nautilus Labs, and we don't want to give any perception that we will steal your work. If your model is good, and described well, we want to hire you.

If you do send a Python Jupyter Notebook, we will run it on our end, so make sure it will run without any errors. If possible please include a requirements.txt file along with your submissions, to ensure we don't have any package dependency issues.

Please send your report 7 days from receiving this assignment. We will review it, and if it's good, we'll schedule an onsite interview for you to present this report.

Appendix

The columns in the attached CSV are described as follows:

- **Main Engine Fuel Consumption (MT/day) -** Fuel consumption rate in metric tonnes per day.
- **Main Engine Using HFO (bool) -** 1.0 if main engine is using Heavy Fuel Oil (a dirty fuel used in open ocean). 0.0 Otherwise. Value could be a fraction if only a portion of the hour is using HFO.
- **Main Engine Using MGO (bool) -** 1.0 if main engine is using Marine Gas Oil (a clean fuel required near coastal areas and in port). 0.0 Otherwise. Value could be a fraction if only a portion of the hour is using HFO.
- **Draft Forward (meters) -** Draft sensor in the front of the ship.
- **Draft Aft (meters) -** Draft sensor in the back of the ship.
- **Draft Mid-1 (meters) -** Draft sensor in the middle of the ship (unsure if left or right side).
- **Draft Mid-2 (meters) -** Draft sensor in the middle of the ship (the other side).
- ***Trim (meters)*** **-** (This value is not included as a column as it is trivial to calculate.) Difference between forward and aft draft. Trim has a significant impact on performance.
- **Shaft Speed (RPM) -** Propeller speed, propeller shaft speed, and main engine speed. There is no gear box. The engine turns at ~90 RPM max.
- **Shaft Torque (kNm) -** Torque on the above shaft.
- **Shaft Power (kW) -** Main engine output as measured in power.
- **Speed Over Ground (knots) -** Speed as measured from the frame of reference of GPS or an on-ground observer. A "stopped" ship could still have speed over ground as it is flowing with the current.

- **Speed Through Water (knots) -** Speed as measured from the frame of reference of a log floating in the water. Also called "log speed" for that exact reason. A "stopped" ship should have 0 speed through water. However, a moored ship may have speed through water as the water moves past the ship.
- **Heading (degrees) -** Ship heading.
- **Rudder Angle (degrees) -** Angle of the rudder.
- **Weather Service True Wind Speed (knots) -** Wind speed as measured and recorded by NOAA relative to a stationary object.
- **Weather Service True Wind Direction (degrees from north) -** Wind direction as measured and recorded by NOAA relative to a stationary object.
- **Weather Service Apparent Wind Speed (knots) -** Wind speed as measured and recorded by NOAA relative to the ship's vector.
- **Weather Service Apparent Wind Direction (degrees from bow) -** Wind direction as measured and recorded by NOAA relative to the ship's vector.
- **Weather Service Sea Current Direction (degrees from north) -** Ocean current as measured and recorded by NOAA.
- **Weather Service Sea Current Speed (knots) -** Ocean current as measured and recorded by NOAA.
- **Weather Service Temperature (celsius) -** Water surface temperature.
- **Water Depth (meters) -** Depth below the keel. Unfortunately, value is zero in deep water.
- **Weather Service Wind Wave Significant Height (meters)** - Wind-waves are produced (or were recently produced) by the local wind. This is a statistical measure that represents the average of the highest one-third (33%) of waves (measured from trough to crest) that occur in a given period, recorded by NOAA.
- **Weather Service Wind Wave Period (seconds)** - The peak period in seconds of the wind waves, recorded by NOAA.
- **Weather Service Wind Wave Direction (degrees from north)** - This is the direction that the wind-waves are coming from (recorded by NOAA).