Goal:

We would like to know about your expertise in the domain of timeseries statistical modeling and machine learning via this task.

Input:

The working package contains a historical dataset of 20 engines for past 100 days. We have provided two signals there, each signal has one data-point per day. Also note some engine's data may be erroneous, so clean them appropriately.

Explanation about CSV file: You can see following four columns in given CSV file.
time: Timestamps of data-points
operating_hours: Operating hours counter (cumulative) for engine
fuel_consumption(liter): Fuel level in tank for given date
engine_id: Unique identifier to identify equipment

Tasks:

a. Come up with a good "single variate timeseries forecasting" algorithm for "Operating Hours" signal, write steps for the same. Explain the characteristics / formulation of the algorithms chosen by you and also briefly tell us why they are the best for given dataset? Restrict your write up to two A4 pages.

b. Write a program in python for above algorithm and get 15 days forecasting for given dataset, for each engine (one data point per day). You can use whatever tooling / libraries you want.

c. What method will you use to evaluate your forecast against actual data-points? Write a python program which calculates this degree of precision.

Side Note: You have been provided with 100 days of data, you can split it into half or some fraction for training and testing purpose, and you can evaluate overall quality of your test results for applied algorithm on your own as well.

Bonus: Repeat all 3 tasks for "fuel_consumption(liter)" signal.

Time Constraint:

2 hours.

Result:

We expect following files from you:
1. Write up for task a. (max one A4 page long)
2. Python code for task b
3. Output CSV file containing 15 days forecast (format same as input file).