

ML and Data Science Selection process

Hello everyone,

There are two separate tasks crucial to our organization being tested, please follow the guidelines mentioned and complete the tasks.

Note: Please use Python for this task. You are free to use any library necessary as long as it is open sourced.

You are encouraged to share your solutions:

- On Github(version controlling is a must), keep track of you models and params. A end file model.py is required which can be ran to recreate the model.
- Use Jupyter notebooks for Studying the data and drawing important conclusion.
- Use of Docker containers will be appreciated (bonus points for this).

Both the assignments are quite simple, with the first assignment dealing with Data science, while the second assignment being a regression model. Please feel free to reach out to us in case of any further queries. You can download the data from:

<https://drive.google.com/drive/folders/1YOj9vJ262vpnVtzRHnxLa-8PjZ2Sn3ok?usp=sharing>

Assignment 1 - Used Cars Price Prediction and Evaluation(Car_price.csv)

- Draw important conclusion from the data (eg. priceUSD vs Mileage or year and segment vs price):
 - Use of proper function like groupby, to draw decisive factors from the data
 - Show plots to show trends in data
- Statistical analysis of data using p values to draw important conclusions is appreciated (eg. T-test or Chi Square Test)
- Create a model to predict the priceUSD columns.
 - Show how you are doing data cleaning and Encoding the data.
 - Use of function over redundant codes for everything is important.

Assignment 2 - Propulsion Plants Decay Evaluation(propulsion.csv)

- Develop a model, to predict the column
 - 'GT Compressor decay state coefficient'
 - GT Turbine decay state coefficient minimizing 'rmse' loss function
- You are encouraged to try out different techniques from ML/DL and share comparative results
- Please refrain from using very complex architectures in case of deploying deep learning algorithms
- Use of proper feature selection process and hyperparameter tuning (if required) is necessary.

- Properly commenting the code is required
- We will use a test set to check the model performance