Data Science Internship

ExactSpace

You are provided a dataset of a cyclone preheater which is part of an industrial process. In the duration of operation there are instances of abnormal operations.

Objective:

Using python and any algorithm of your choice, highlight time periods where this abnormality can be observed.

About the data

There are 6 variables and 370k records. Data is rcorded once every 5 minutes over a duration of 3 years.

1. Cyclone_Inlet_Gas_Temp – Temperature of Hot gas entering the cyclone.
2. Cyclone_Gas_Outlet_Temp – Temperature of Hot gas leaving the cyclone.
3. Cyclone_Outlet_Gas_draft – Draft (pressure) of gas at outlet of cyclone.
4. Cyclone_cone_draft – Draft (pressure) of gas at cone section of cyclone.
5. Cyclone_Inlet_Draft – Draft (pressure) of gas at inlet of cyclo
   ne.
6. Cyclone_Material_Temp – Temperature of the material at the outlet of the cyclone.

Expected output:

Prepare a **zip/rar folder** with the following files and share with the ExactSpace team, as indicated. Name of the folder must be candidate's "FirstName_LastName_DataScience"

1. Your resume
2. Provide the source code file/s of your work
3. A ppt with 3-5 slides detailing the following:
   - Data preparation – What kind of treatment or processing did you apply on the raw data possible. What was the reasoning behind your specific decisions?
   - Analysis strategy – Detail the methodology you followed to analyze the data.
   - Insights – What did you find out from the data provided. Where are the abnormal periods and how did you identify them?