

Teletraan Takehome Data Challenge

Part I

Goal

Exploration and forecasting using the *car_sensor_data* dataset.

Data description

car_sensor_data

This dataset is captured by the numerous sensors installed in an electronic car.

Tasks / Questions

1. For what period is this set of data captured?
2. How many of the features appear to be constant machine settings?
3. Plot the daily mean for 温度检测点1号温度 and 温度检测点2号温度.
 1. Are the two trends significantly different?
 2. Which days are missing?
 3. Fill the missing values with a reasonable method.
 4. Is the data stationary? If not, what does it imply and what would you do?
 5. What period appears to have an abnormal value compared to a baseline? Describe and implement a statistical method that detects temporal anomaly.
4. Build a model to forecast 温度检测点1号温度 for the next day. Briefly explain how prediction confidence level is obtained for the model of your choice.

Part II

Goal

Cell towers transmit and receive signals from telecommunication devices. Certain Sometimes, a cell tower may “go to sleep”, ie enter into periods of inactivity, causing disruptions in communication. Given the dataset, build a model to predict whether the tower will “fall asleep” 10 hours in advance.

Data description

cell_tower_data

ecell_int_id : cell tower ID

period_start : the hour at which certain health metrics of the cell tower is captured

CellAvailability - IncomingHO : health metrics of the cell tower measured by default.

Tasks / Questions

1. There may exist missing time steps in the data. How would you deal with the missing hours? Explain your logic.
2. The “target” column is currently empty (all zeroes) and you have to create it (bear in mind you’re predicting 10 h in advance). Given the rule that a tower is considered “sleeping” if
 1. CellAvailability’ = 100 and
 2. RrcConnEstabSuccRate = 0 and
 3. S1SigConnEstabSuccRate = 0 and
 4. ErabEstabSuccessRate = 0
3. Your classification model may not have perfect performance. Given that the telecommunications company cares more about preventing any sleeping events from happening, and doesn’t care so much about wasting time and labor to check out false alarms, tune your model to perform at the desired threshold. (Again don’t worry about your final metrics not looking perfect. Just show your reasoning.)