

Course Project 2020/2021

Rules and Dataset Explanation



Course Project

Dataset Description

This year project is based on a **predictive maintenance** dataset for the prediction of faults on air conditioning equipment installed mobile network transmission sites in a 14-days forecast window. Available information are related to weather conditions (past and forecast), alarms and faults occurred on site, static features of the site.

- Daily data (April 2019 – Gen 2020 ~ 10 months) for each site
- Data are available for 2605 sites, distributed between training set (2071) and test set (534)

Variable/Variable pattern	Description	Type
SITE_ID	Unique identifier for the site belonging to the network	<u>int</u>
DATE	Reference date of the sample	<u>date</u>
N_TRANSPORTED_SITES	Number of neary sites for which the radio signal is transported through the site	<u>int</u>
CELL_TYPE_X	Indicates if the transmission cell of type X is mounted on site	<u>binary</u>
GEOGRAPHICAL_CLUSTER_K_x	Membership in the geographic cluster x (network clustered in 10 regions from 0 to 9)	<u>binary</u>
<i>mean/max/min_w_prevXd</i>	<i>Mean, max or min</i> of the weather condition <i>w</i> in the <i>previous X</i> days	<u>float</u>
<i>mean/max/min_w_f_nextXd</i>	<i>Mean, max or min</i> of the forecasted weather condition <i>w</i> in the following <i>X</i> days	<u>float</u>
<i>cat_sum_alarms_prevXd</i>	Number of alarms associated to the category <i>cat</i> observed in previous <i>X</i> days. Alarms are classified in 9 categories. Details are available on attached excel	<u>int</u>
<i>cat_mean/max/min_persistance_prevXd</i>	<i>Mean, max or min</i> alarm duration (in minutes) of <i>cat</i> alarms in the previous <i>X</i> days	<u>float</u>
<i>skew_cat_alarms_prev14d</i>	Skewness indicator of <i>cat</i> alarms distribution in time in the previous <i>X</i> days	<u>float</u>
<i>kurt_cat_alarms_prev14d</i>	Kurtosis indicator of <i>cat</i> alarms distribution in time in the previous <i>X</i> days	<u>float</u>
aircon_sum_wo_target_next_14d	Binary target variable indicating the presence of a fault in the following 14 days	<u>binary</u>



Alarms categories

Course Project

Testing and Evaluation

Performances will be evaluated by means of the average daily **Weighted Recall** computed considering the **10 sites with highest fault probability** as predicted with fault weighted by `N_TRANSPORTED_SITES` column. The thresholding level is therefore defined day by day by the top sites and not by a fixed threshold level.

Data Available in csv format: <https://we.tl/t-n6DOILHf49> (archive password: *DMTMChallenge2020*)

Deadline: **Monday December 21 23:59**

We ask you to prepare and upload on Beep platform an archive containing:

- **Prediction.csv**: you can find an **example in the archive**. The output of the prediction must be a **fault probability value between 0 and 1**.
- **Report.pdf**: **4 pages** to describe in detail your approach, data processing techniques, prediction model, performance computation methods and analytical results
- **Presentation.pptx**: **5 slides** for the final project presentation describing your approach, data processing, prediction model (a summary of the report you wrote, imagine to present to a potential customer your approach)
- **Scripts.zip**: any notebook or script you wrote will be evaluated, we expect **a ordered list of script from data processing to prediction.csv file** output.

The evaluation considers: Recall Score obtained on the Test Predictions, Report Quality, Presentation Speech, Quality of the code.

If you have any doubt you can ask directly on Beep forum.

THANK YOU

HERE TO DARE

