Data Science Challenge : Predict Room Temperature for ENGIE Cofely

19 july 2016

TRACTEBEL

enGie

CONTENTS

1 PreProcessing

2 Data Exploration

3 Feature Engineering

4 Model

CONTENTS

5 Tools & frameworks

Pre-Processing

o   Data Used : Training.csv, Test.csv, measures.csv

o   No Alarm, no State info, no external data

o   Correct outliers

o   Interpolate missing values

o   Some strange data removed (ex : CN_RET for 2310320005, constant temperature for 2310320005 in october,…)

Data Exploration

o   Measures selected as predictors :

- o   BASE_VN_T.EXTER_T.REF_CORRIGE

- o   CIRN_AI_EAU_T.DEP_SONDE, N=1,…6

- o   CIRN_AI_EAU_T.RET_SONDE, N=1,…6

- o   PRIM_AI_EAU_T.DEP_SONDE

- o   PRIM_AI_EAU_T.RET_SONDE

- o   TIME

o   16 base variables
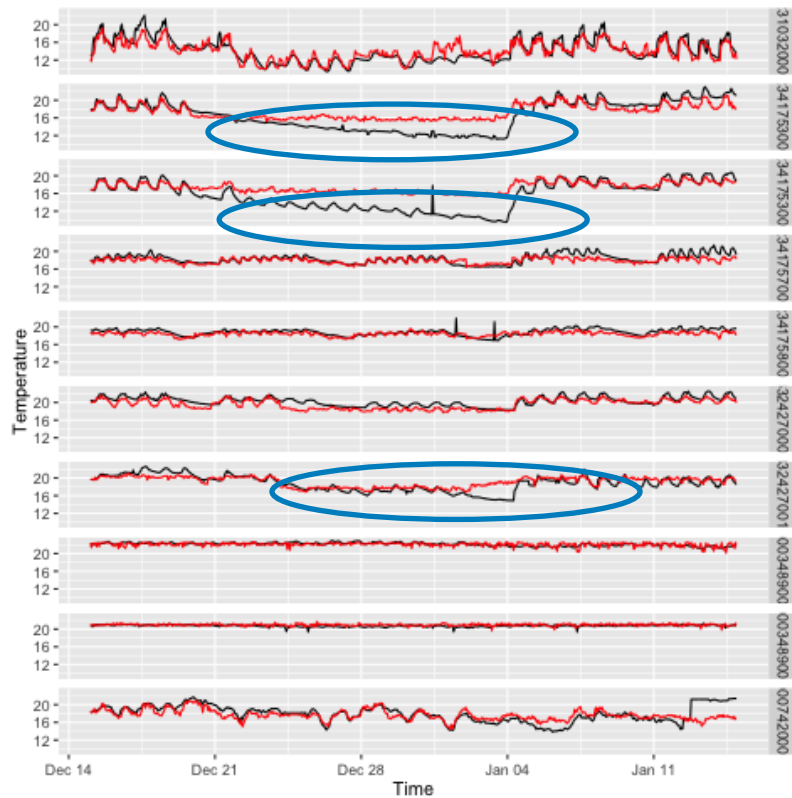
Data Exploration
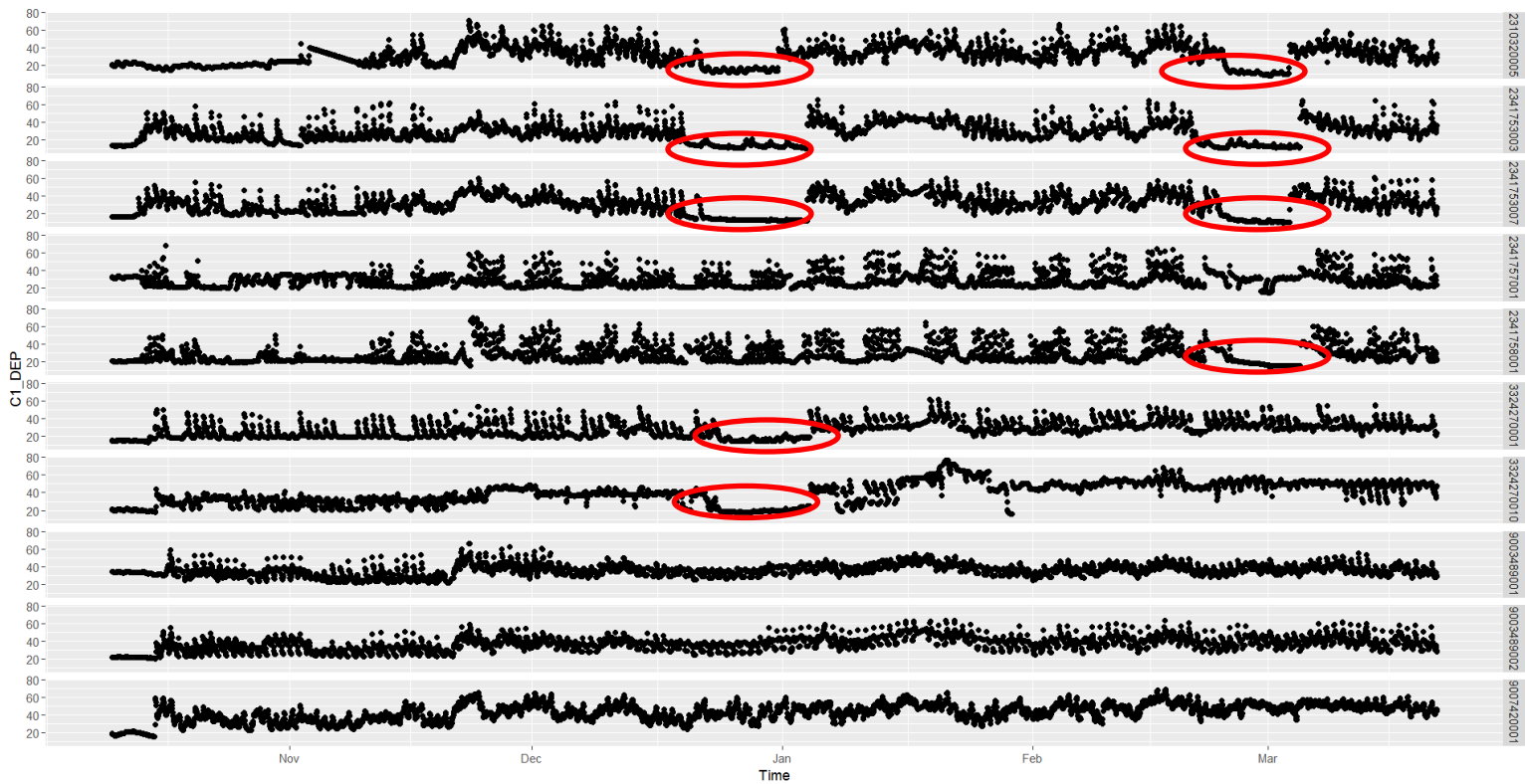
Holidays!!

Data Exploration

— Temperature
— Prediction

Feature Engineering

o   Additional features computed from base features:

    o   Values of base variables at 10 previous hours

    o   Binary feat for type of building

    o   Time index

    o   Hour in the week

    o   Holiday/weekend indicator variable

    o   Number of hours of holiday in the last 6,12,36,48,72 hours

o   177 predictors in total

Model

- o One single gradient boosted tree model (XGBoost)

- o Parameters : max_depth=16,eta=0.1. other parameters =default

- o No blending

Tools and frameworks

o Everything coded in R

o Libraries :

    o data.table (for data manipulation)

    o Lubridate (time manipulation)

    o xgboost (model)

    o Ggplot2 (plotting)

    o Forecast (for interpolation)