

AI PROJECT

Energy consumption

Croissant

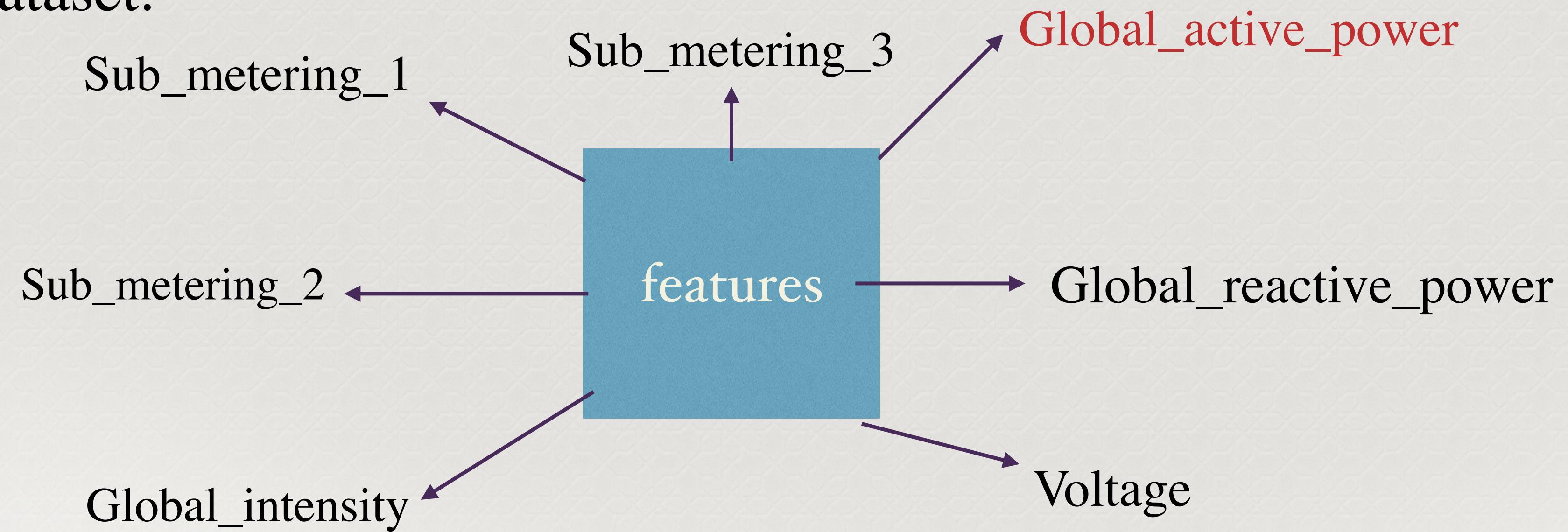
Sparsh, sindhoor,dikshita

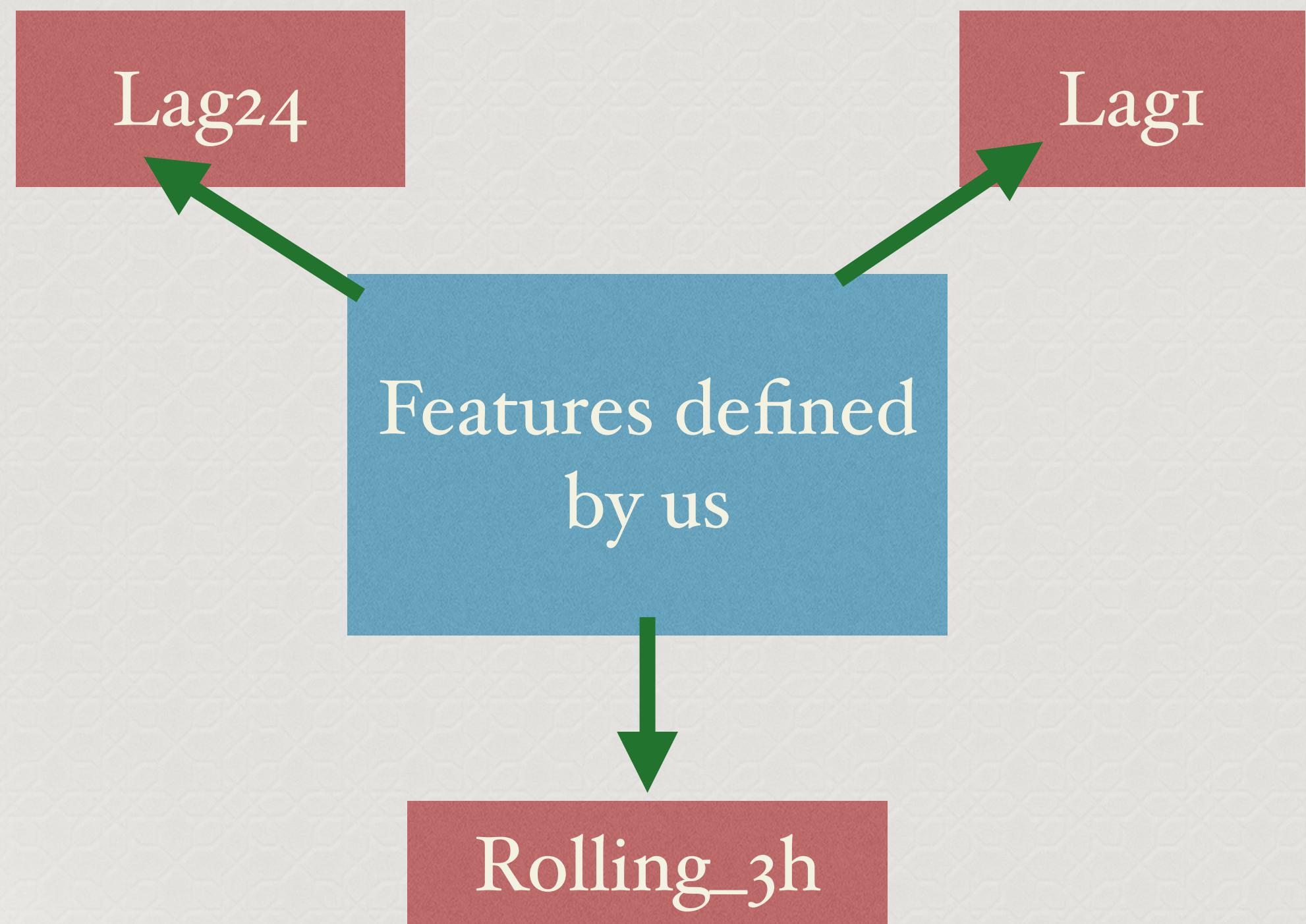


Energy consumption model

This code is a time-series regression pipeline built to predict **Global Active Power** consumption using past household energy data.

About the dataset:





⋮

Time series split

The data is split such that we train the model on data from before a certain date and the testing is done on data taken after that date, so that essentially our model isn't “cheating”

How does the model work?

- 
- 1) *load the data set*
 - 2) *Exploratory data analysis*
 - 3) *Feature engineering*
 - 4) *Feature scaling and splitting*
 - 5) *XGboost algorithm / random forest*
 - 6) *Model predicts..*

Getting to
know your data
before building
the model

```
==> Summary Statistics ==>
   Global_active_power  Global_reactive_power      Voltage \
count          103640.000000          103640.000000  103640.000000
mean           1.204582                  0.125108    239.056833
std            1.184427                  0.112474     3.658287
min            0.096000                  0.000000    224.620000
25%            0.308000                  0.000000    236.350000
50%            0.672000                  0.106000    239.530000
75%            1.666000                  0.194000    241.810000
max           10.670000                  0.986000    250.020000

   Global_intensity  Sub_metering_1  Sub_metering_2  Sub_metering_3
count          103640.000000          103640.000000  103640.000000
mean           5.138283                  1.281455     1.692956    6.066316
std            5.007758                  6.551113     6.825293    8.260284
min            0.400000                  0.000000     0.000000    0.000000
25%            1.400000                  0.000000     0.000000    0.000000
50%            3.000000                  0.000000     0.000000    0.000000
75%            7.000000                  0.000000     1.000000   17.000000
max           46.400000                 75.000000    75.000000   20.000000
```

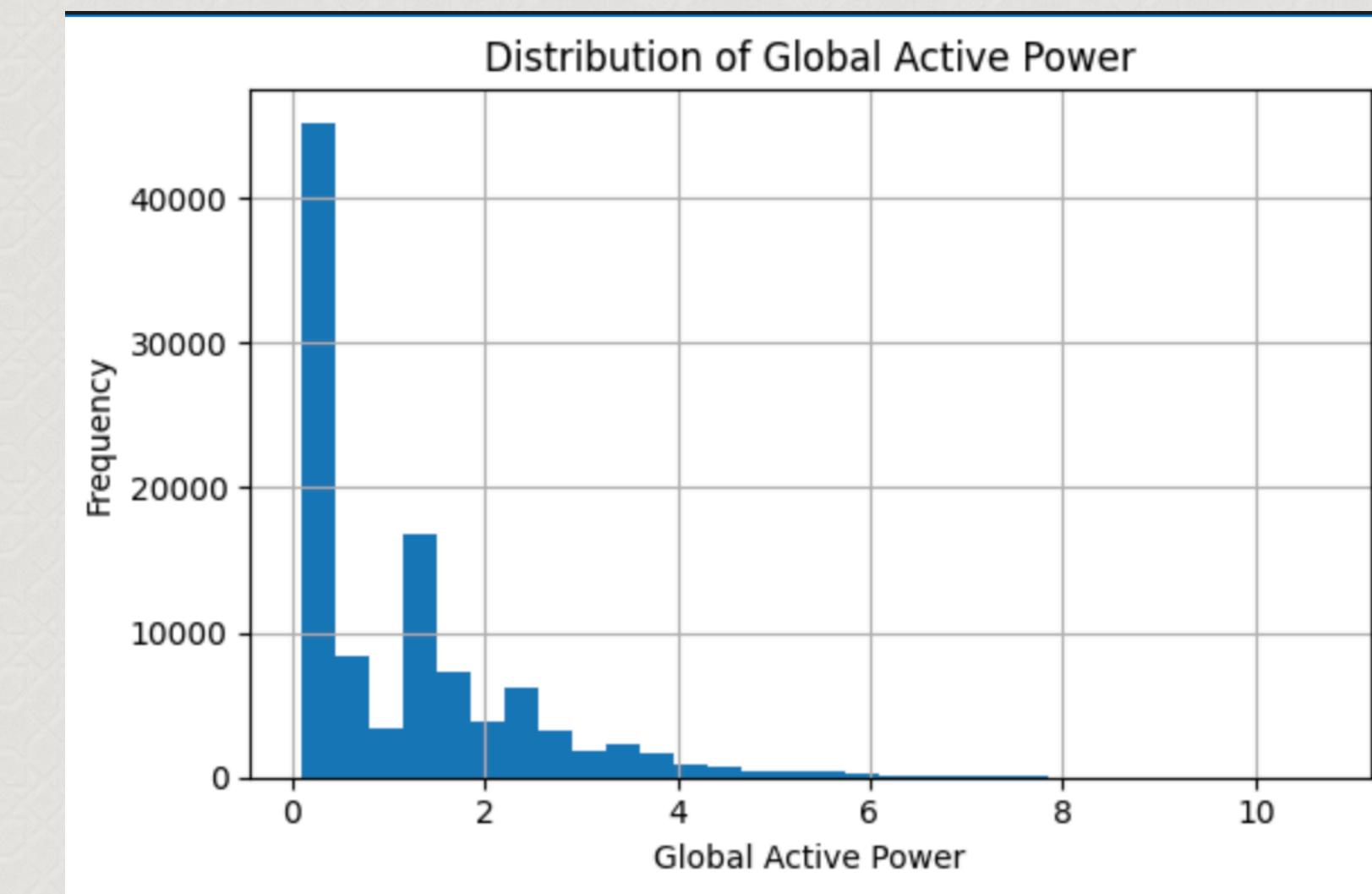
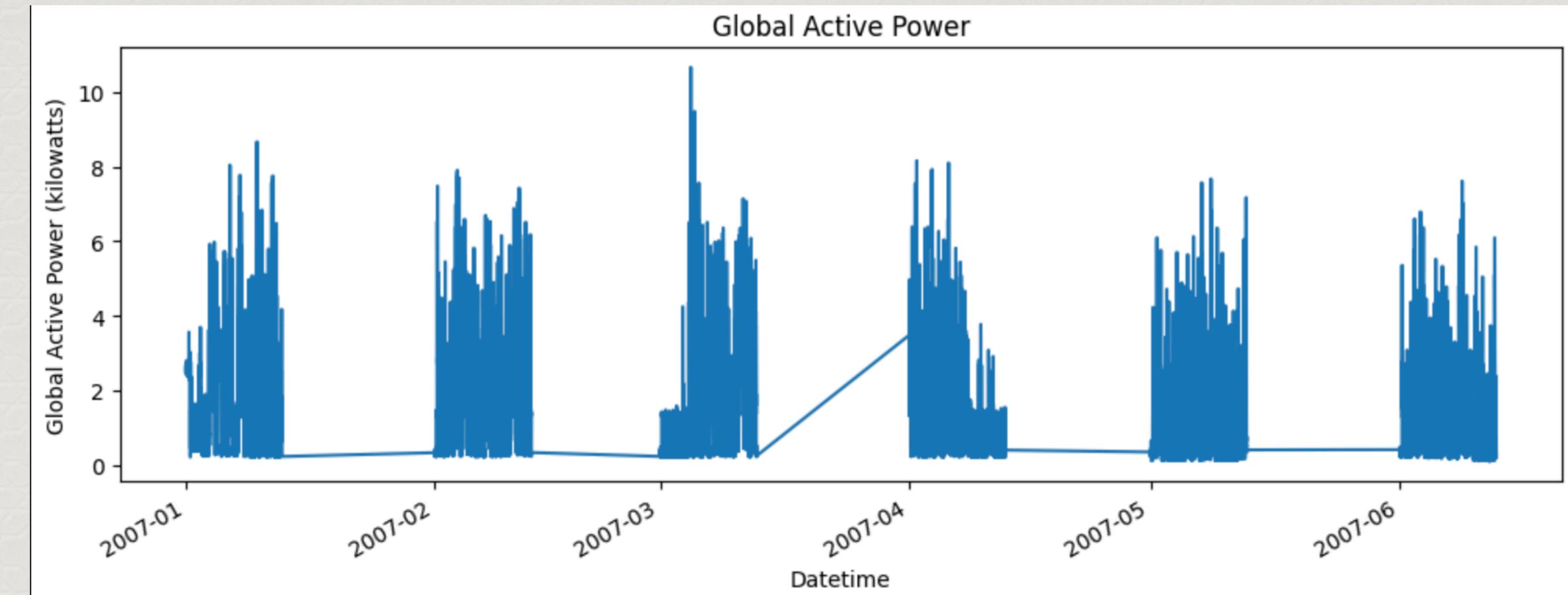
statistically
examining the
data

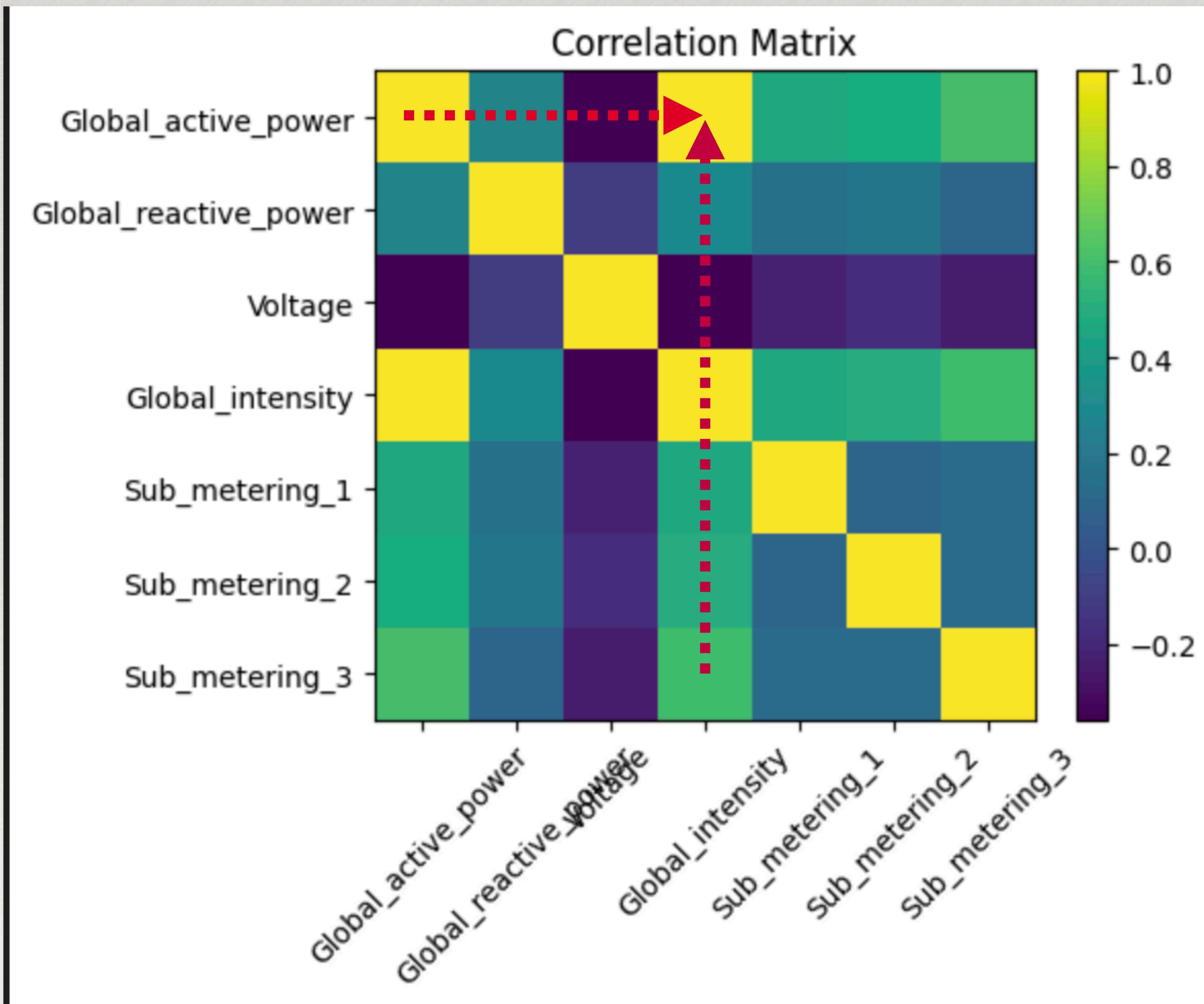
Sanity check

Time series plot

histogram

Correlation matrix



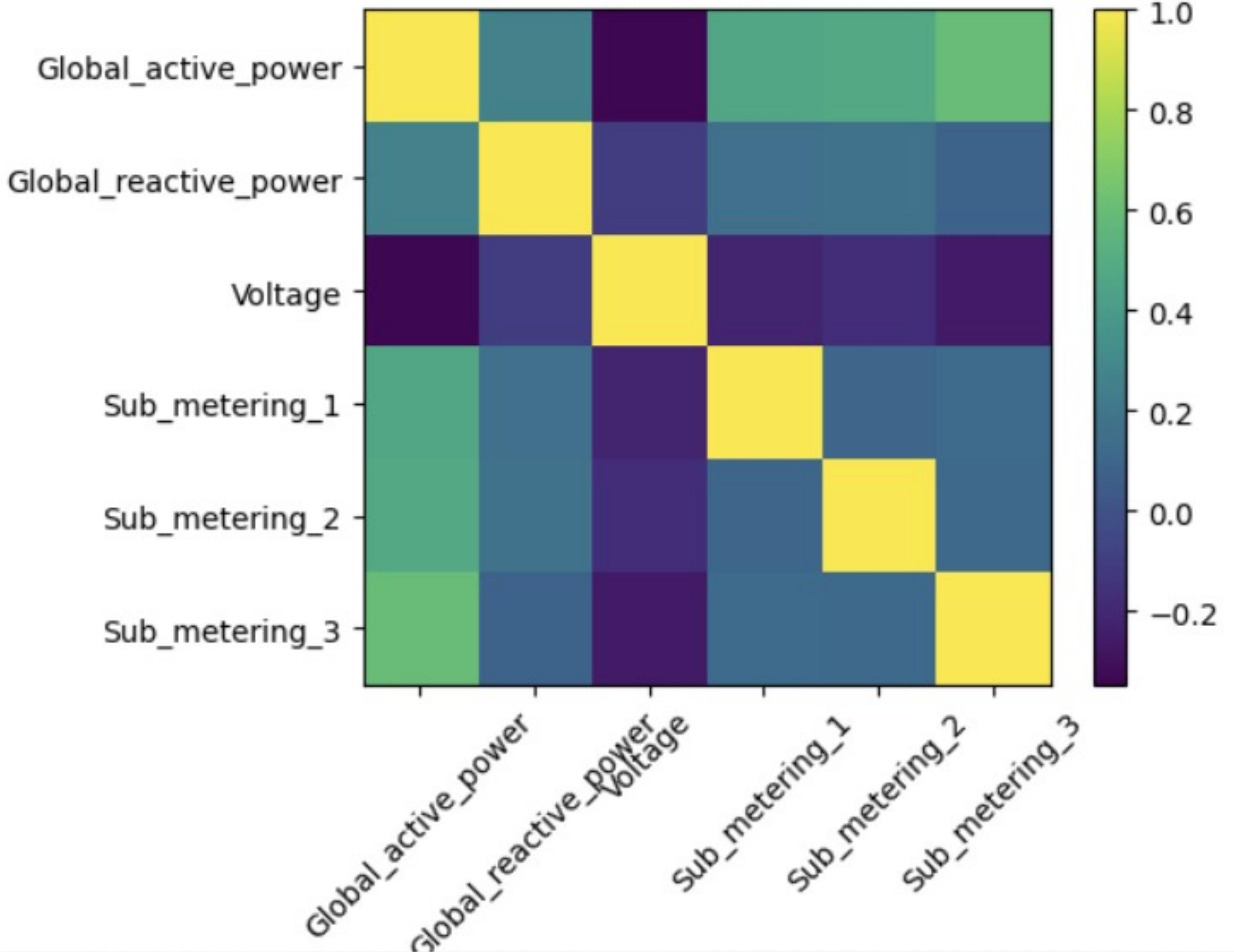


Yellow box : Correlation $\approx +I$
 Greenish box : correlation ≈ 0
 Dark purple box: correlation $\approx -I$

Select
features for
modeling using the
correlation
matrix

Drop the collinear features

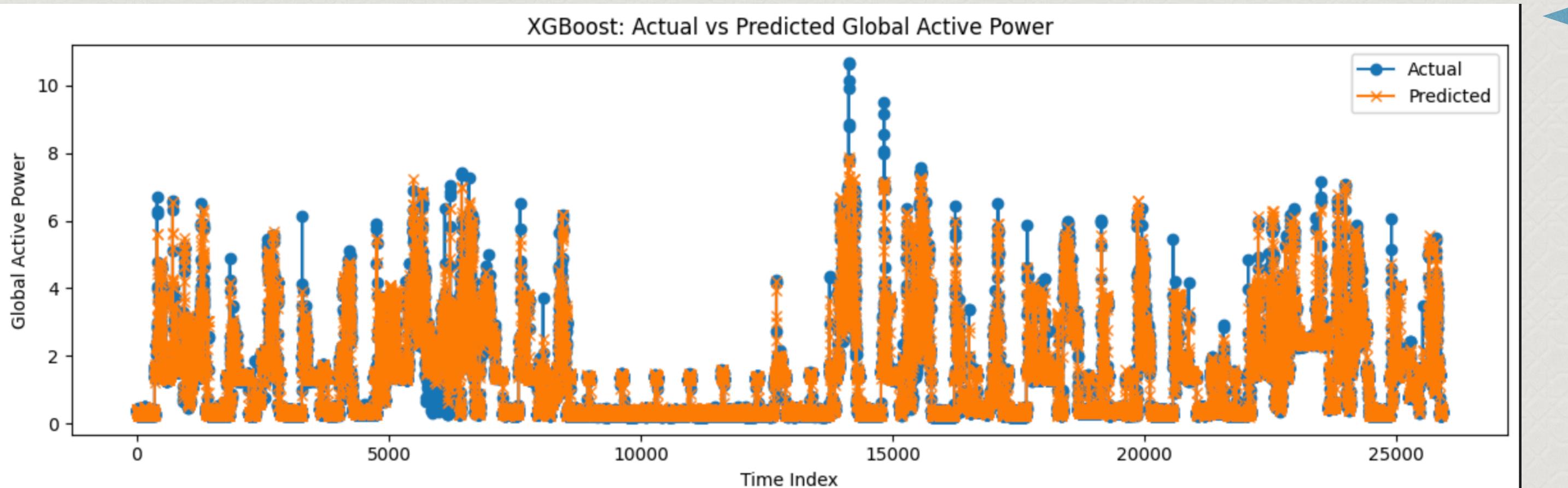
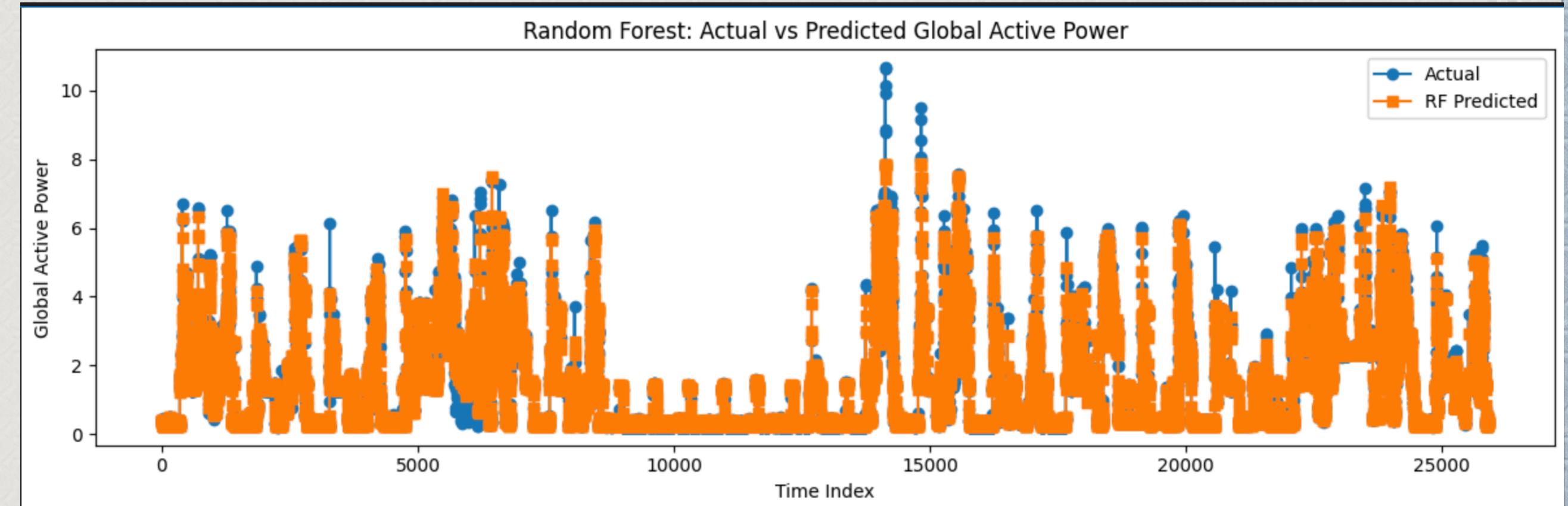
Correlation Matrix



$$R^2 = 0.9672$$

XGBoost algorithm & random forest to learn from training data

Model predicts



$$R^2 = 0.9729$$

- 1. Builds many small decision trees, each tree tries to fix the mistakes of the previous ones
- 2. majority voting/average of their predictions

ETHICAL perspective

Can AI help detect energy inefficiencies or recommend eco-friendly practices without penalizing vulnerable groups?

Eco-friendly recommendations

Running appliances at off-peak times.

switching to led lights



~~Cutting electricity
Making them pay
more~~

AI suggests :
Tell
government ,NGOS
—> help replace old
fridges

What are the risks of using granular energy data for prediction?



To track Personal habits

Misused by landlords ,insurance companies

How might AI models favor certain socio-economic groups or regions?

Biased
training
data



Unfair
predictions

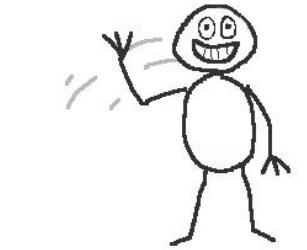
How can we explain AI decisions in a fair, interpretable way?

Interpretable models

LIME

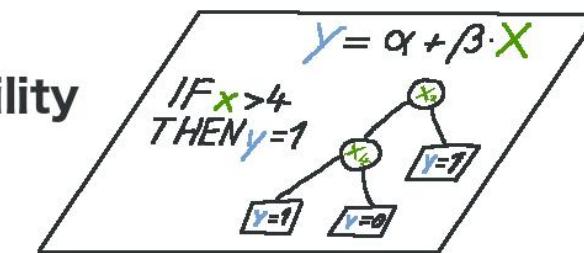
SHAP

Humans



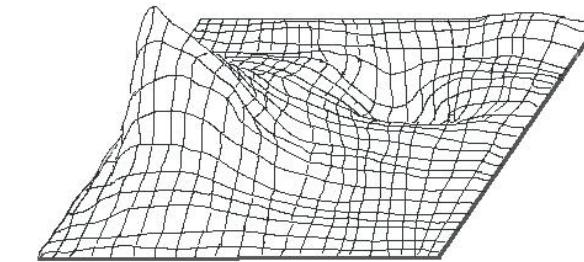
↑ inform

Interpretability Methods



↑ extract

Black Box Model



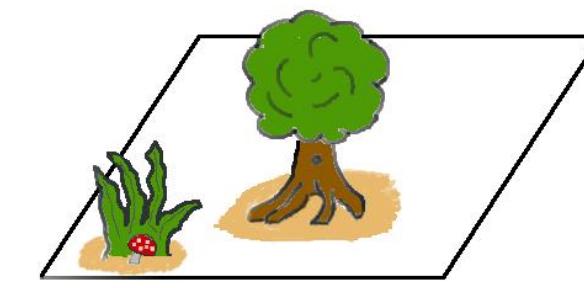
↑ learn

Data

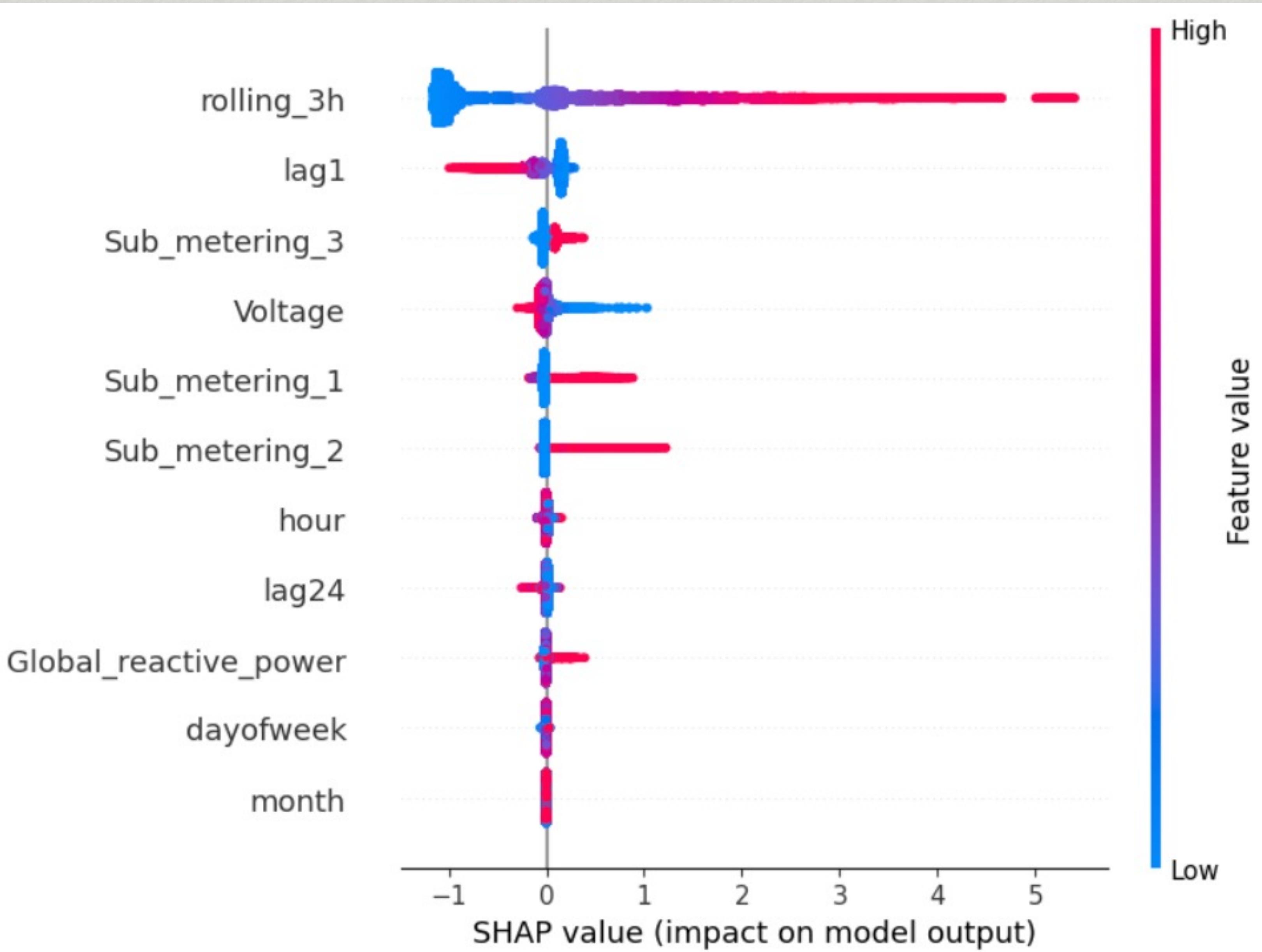
X ₁	X ₂	X ₃	X _n
10	2	0				10
5	1	0				5
7	-1	0				7
1	0	0				1

↑ capture

World



SHAP plot



Thank you