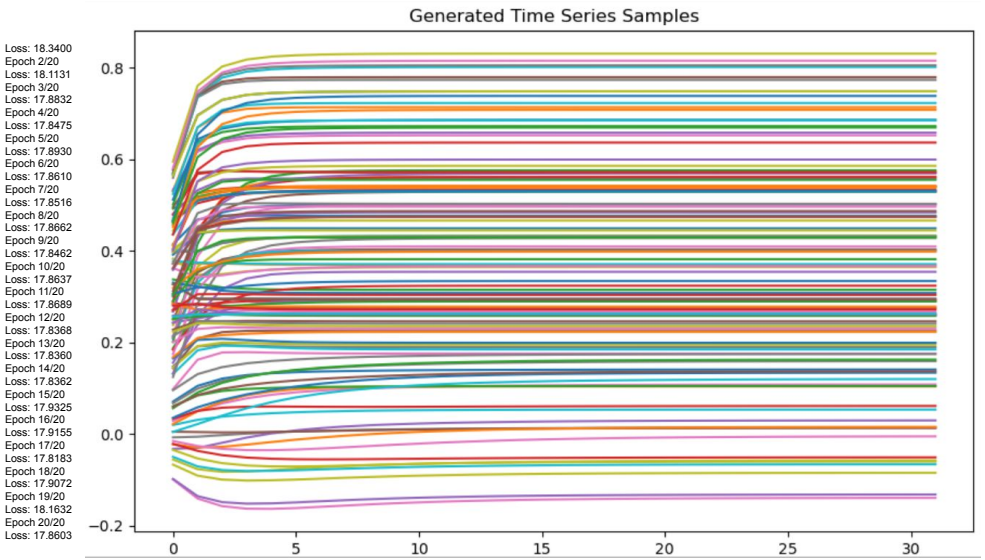
Energy Encryptors

Method for synthetic data generation

Part 1: Synthetic data generator

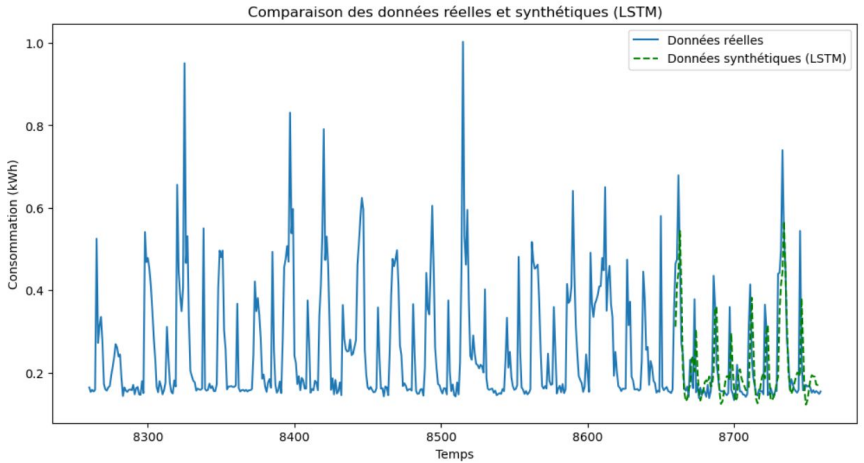
An **Encoder (LSTM)** processes sequential data and encodes it into a fixed-length hidden state representation.

Trained with all daily data households



KS Test Statistic: 0.858675799086758, p-value: 0.0
Cosine Similarity: 0.5016493949046211

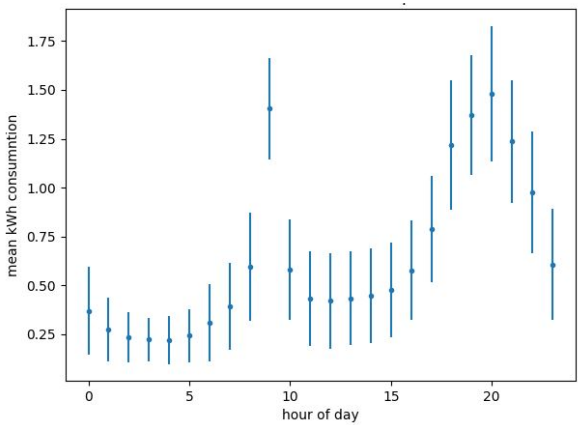
Trained with only one daily data households



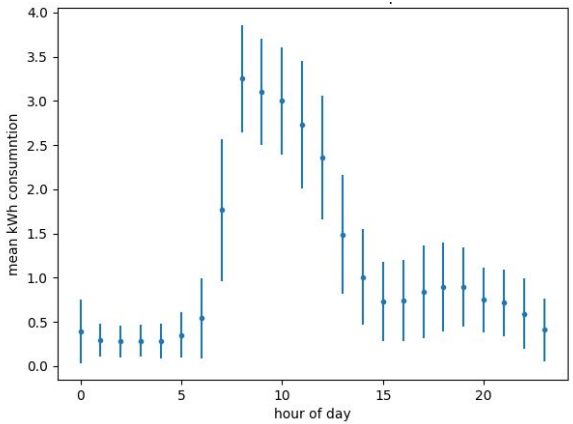
KS Test Statistic: 0.7195205479452055, p-value: 0.0
Cosine Similarity: 0.8555294497145418

all are goods, but we needed more time to train better our models and generate a synthetic dataset with those

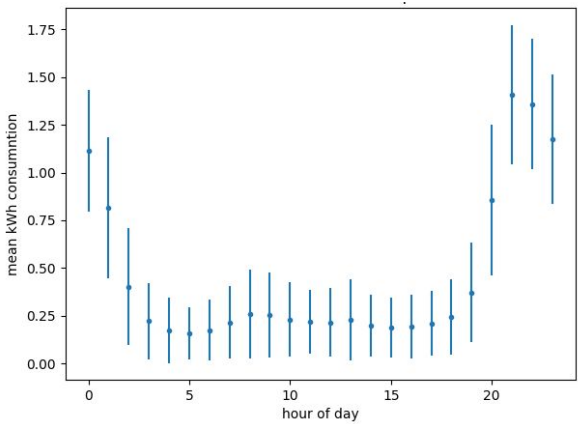
Examples of a typical daily consumption



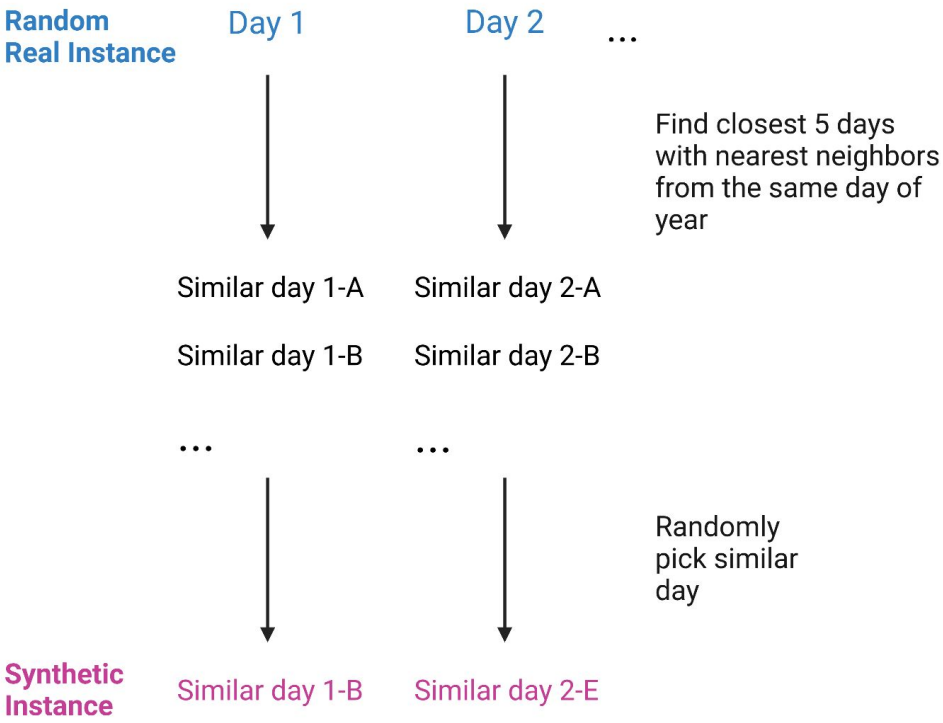
Working in the office



Taking care of household



Let’s party!

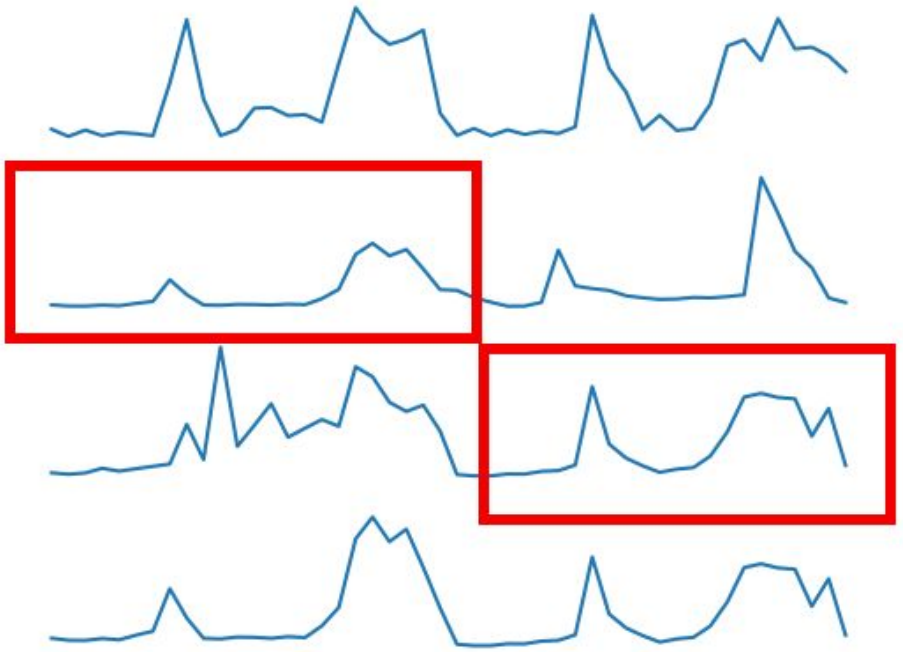

Random Real Instance
Day 1
Day 2
...
Find closest 5 days with nearest neighbors from the same day of year
Similar day 1-A
Similar day 2-A
Similar day 1-B
Similar day 2-B
...
...
Randomly pick similar day
Synthetic Instance
Similar day 1-B
Similar day 2-E

Real sequence

First day

Second day

Synthetic sequence (inc. noise)

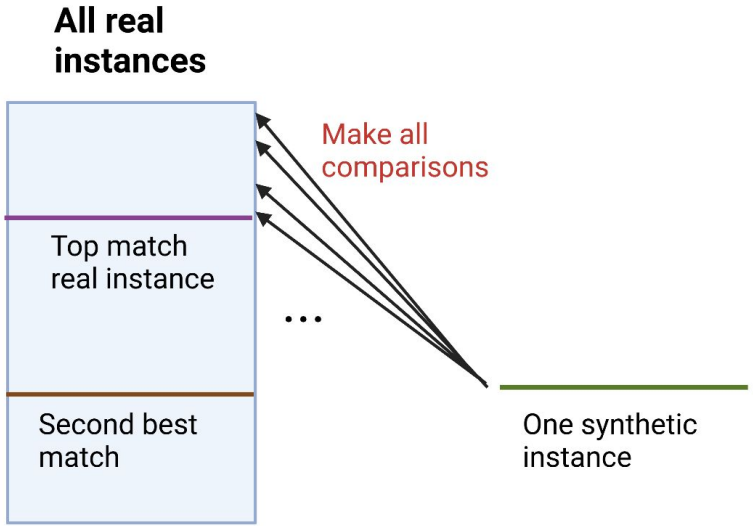


Statistics score: 1.15

Test accuracy XGBoost: 0.84

Feature	Importance
value__has_duplicate_min	0.293615
value__longest_strike_above_mean	0.158698
value__percentage_of_reoccurring_values_to_all...	0.077412
value__median	0.076222
value__absolute_sum_of_changes	0.066261
value__count_above_mean	0.064588
value__cid_ce__normalize_True	0.059581
value__skewness	0.055487
value__variance	0.052959
value__root_mean_square	0.049815
value__mean	0.045362
value__has_duplicate	0.000000
value__has_duplicate_max	0.000000
value__mean_abs_change	0.000000
value__sum_values	0.000000

Part 2: Privacy metric



Privacy score:
(Similarity top match)/
(Similarity second best match)

Typical ratio:
0.73/0.68=1.01


Thanks