IEEE-CIS TECHNICAL CHALLENGE ON ENERGY PREDICTION FROM SMART METER DATA: DATA DESCRIPTION AND SUBMISSION

https://ieee-dataport.org/competitions/ieee-cis-technical-challenge-energy-prediction-smart-meter-data

Data Description

The goal of this competition is to predict the monthly electricity consumption for 3248 households in a coming year (January to December).

You are provided with historical half-hourly energy readings for the 3248 smart meters. To simulate a realistic use case, we take the 1st of January of a given year as the day we want to make predictions. Thus, different smart-meters will have available a range of months' worth of consumption, ranging from only last month (i.e. December) to the entire last year (January to December), acknowledging that customers may have joined at different times during the previous year. For example, we may have a few customers for which we only have data from last December, and we aim to predict January to December of the coming year, whilst for others, we may have the entire January-to-December time series. We would like to see how well we can predict the coming year depending on the amount of data that is available.

For each (fully anonymised) meter_id, we provide half-hourly consumption data, but for the months that are not available a nil value is registered. Note that consumption data may contain (a low number of) *missing values* due to errors in sending/receiving from smart meters. In addition to consumption data, we provide some additional information for each meter_id, which is typically available in this scenario, and thus, participants could come up with smart ways to exploit it.

- **Weather data**: we provide the full-year temperature for each meter_id at a daily resolution, including the average temperature, minimum and maximum of each day. To guarantee the privacy of the customers, we are not allowed to provide the location or the postcode/zip code.

- **Additional attributes**: Through a number of voluntary surveys, we have collected some additional information for a subset of the smart meters. Some customers replied voluntarily to this survey, and for those participating in the survey, they did not reply consistently to all questions. You will find information about the dwelling type (for 1702 meters), number of occupants (74 meters), number of bedrooms (1859 meters), etc. You can find all available information below.

Format

All the data is provided in CSV format.

What am I predicting?

You are predicting the expected energy consumption in the next 12 months for each meter_id.

Files

- *consumption.csv* – half hourly consumption data for 3248 smart meters.
- *weather-avg.csv* – average daily temperature associated to the 3248 smart meters.
- *weather-min.csv* – Minimum daily temperature associated to the 3248 smart meters.
- *weather-max.csv* – Maximum daily temperature associated to the 3248 smart meters.
- *addInfo.csv* – Additional information available for the 1859 smart meters.
- *sample_submission.csv* – a sample submission file in the correct format

Columns

consumption.csv

Column name	Meaning	Value
meter_id	Anonymised smart meter ID	String
2017-01-01 00:00:00	Electricity consumption in kWh between 2017-01-01 00:00:00 and 2017-01-01 00:30:00	float64
2017-01-01 00:30:00	Electricity consumption in kWh between 2017-01-01 00:30:00 and 2017-01-01 01:00:00	float64
......	......	......
2017-12-31 23:00:00	Electricity consumption in kWh between 2017-12-31 23:00:00 and 2017-12-31 23:30:00	float64
2017-12-31 23:30:00	Electricity consumption in kWh between 2017-12-31 23:30:00 and 2019-01-01 00:00:00	float64

weather-[avg,min,max].csv

There are three files with the same shape; one for the averages, one for the minimum temperature and another one for the maximum temperature of each day.

Column name	Meaning	Value
meter_id	Anonymised smart meter ID	String
2017-01-01	Avg/Min/Max temperature on 2017-01-01	float64
2017-01-02	Avg/Min/Max temperature at 2017-01-02	float64
......	......	......
2017-12-31	Avg/Min/Max temperature at 2017-12-31	float64

addInfo.csv

Additional information is not available for all meters. Only 2143 contain some information. In the following table we provide the count of meters with information for each specific attribute.

Column name	Meaning	#meters	Value
meter_id	Anonymised smart meter ID	2143	String
Dwelling type	Type of building (categorical variable, 5 categories)	1702	String
Number of occupants	Number of occupants (1,2,3,4)	74	int64

Number of bedrooms	Number of bedrooms in the property (1,2,3,4,5)	1859	int64
Heating fuel	Heating fuel type used for space heating of the property (categorical variable, 4 categories)	78	String
Hot water fuel	Heating fuel type used for water heating of the property (categorical variable, 3 categories)	76	String
Boiler age	If the boiler is new or old (categorical variable, 2 categories)	74	String
Loft insulation	If the loft is insulated (categorical variable, 2 categories)	75	String
Wall insulation	If the wall is insulated (categorical variable, 5 categories)	75	String
Heating temperature	Indoor temperature (categorical variable, 4 categories)	74	String
Efficient lighting percentage	Efficient lighting percentage (categorical variable, 4 categories)	73	String
Dishwasher	Number of dishwashers (0,1,2)	76	int64
Freezer	Number of freezer (0,1,2)	70	int64
Fridge freezer	Number of fridge freezer (0,1,2)	70	int64
Refrigerator	Number of refrigerator (0,1,2)	73	int64
Tumble dryer	Number of tumble dryer (0,1,2)	76	int64
Washing machine	Number of washing machine (0,1,2)	76	int64
Game console	Number of game console (0,1,2,3)	72	int64
Laptop	Number of laptop (0,1,2,3,4)	70	int64
PC	Number of pc (0,1,2,3)	70	int64
Router	Number of router (0,1,2)	69	int64
Set top box	Number of set top box (0,1,2,3)	70	int64
Tablet	Number of tablet (0,1,2,3,4)	70	int64
TV	Number of tv (0,1,2,3,4)	75	int64

Submission

For each meter_id in the provided dataset, you must predict the consumption in the following 12 months in kWh. The file should contain a header and have the following format:

sample_submission.csv:

```
meter_id,Jan,Feb,Mar,Apr,May,Jun,July,Aug,Sep,Oct,Nov,Dec
0xd50,140.089,140.089,140.089,140.089,140.089,140.089,140.089,140.089,140.089,140.089,140.089,140.089
0x47a,475.265,475.265,475.265,475.265,475.265,475.265,475.265,475.265,475.265,475.265,475.265,475.265
0xdc,91.221,91.221,91.221,91.221,91.221,91.221,91.221,91.221,91.221,91.221,91.221,91.221
```

You should save it as a **CSV file** and submit it as the **Analysis document**.

Important: In addition to their predictions, participants are requested to submit a draft description their methodology *in their final submission* (up to 1000 words). This is compulsory as this preliminary description will also be evaluated by the committee to shortlist the top 5 submissions. Please use the '*Abstract*' field to include this information when submitting your final analysis. If your submission is shortlisted you will be asked to provide a final description within a week after the final deadline.
Detailed instructions to register and submit your solutions can be found here.