MINI PROJECT CANVAS

Title (preliminary): Electricity Price Estimator

Group members: Pontus Hedlund, Sanna Korpi, Topi Ranta

Workshop # : 3

MOTIVATION

As the prices of electricity are rising, it is more important for a consumer with an hourly rate to optimize their electricity usage. Our Electricity Price Estimator provides information about expected changes in electricity prices.

With our Estimator a consumer can evaluate when it is cheaper to run consuming household devices or, for example, charge their electric car. By targeting their costly consumption to periods of potentially lower prices, a consumer could manage their future electricity bills.

DATA COLLECTION

We use following historical data:

- Fingrid Wind Power Production Data;
- Helen District Heating Power Data;
- Entsoe Day-ahead Market Prices History Data;
- Finnish Meteorological Institute Weather Observation Data.

Data is mostly in CSV format and publicly available. The data is downloaded to a shared cloud location with version control.

PREPROCESSING

The goals of the preprocessing pipeline are to:

- Discard erroneous data;
- Handle missing values;
- Merge the data so that a ML model can be trained.

To reach the goals, we:

- Find potential errors with EDA and discard datasets with errors;
- Evaluate missing data rates and apply imputation method(s) if missing rates are high, or discard datasets with missing data if rates are low;
- Transform data by normalization, up- or downsampling;
- Merge data frames;
- Double check units, date formats and time zones to be compatible.

EXPLORATORY DATA ANALYSIS (EDA)

To understand the data and to find potentially incorrect data we plan to calculate statistical summaries and visualize data distributions. These can be for example minimum and maximum values, middle values and variances.

To determine whether to discard datasets with missing values or to impute such values we have to count missing data occurrences.

Also, cross-tabulations could be beneficial to further understand missing data mechanisms and the data itself.

VISUALIZATIONS

To explain the mechanism behind the market price, visualizing energy consumption on different weekdays and hours could be beneficial.

Also showing how wind affects the electricity production or how temperature affects consumption would be useful to the end user.

We will probably show the future price prediction via some kind of web application. Future price can be visualized for example with charts and confidence intervals and color schemes.

LEARNING TASK

(focus on problem definition)

The task is to learn how different weekdays and weather conditions affect electricity production and consumption and thus the price. It is a supervised task with time series regression.

The target is to evaluate the cost of electricity in the near future. To determine that we use past electricity costs, electricity consumption, electricity production and weather conditions.

LEARNING APPROACH

(focus on solution implementation)

Since this is a regression problem, we are planning on using the XGBoost regressor for this task. It is fairly easy to implement and even comes with built-in cross-validation.

COMMUNICATION OF RESULTS

We find an interactive web app is the easiest for the end user. Our target group is all the electricity users who are interested in saving on the energy bills and are capable of using web browsers.

Further reporting of interesting findings in the data and explaining the components affecting the market price can be done with a combination of extra visualizations and explanatory text.

DATA PRIVACY AND ETHICAL CONSIDERATIONS

(if applicable)

Since we use aggregated open data with no links to individual persons or their private data, there are no special privacy considerations in data collection.

Our web application does not need to store user-specific data either.

ADDED VALUE

We aim to provide a better estimate by combining different sources. Thus it would be possible to make a better prediction taking more underlying factors into account.

Prediction data enables the customer to better estimate the forthcoming electricity bill. Best- and worst-case scenarios give the customer a possibility to prepare financially.

In an optimum scenario the customer can manage their future electricity bill by adapting their consumption and by using electricity when it is cheaper.

LEGEND

WEEK 1: Data collection/preprocessing

WEEK 2: EDA & visualizations

WEEKS 3-4: Machine/deep learning

WEEK 5: Fairness & data privacy