Short-term Electricity Consumption and Demand Prediction (AI4Demand)

This project was funded by I-NERGY project

Project Intro/Objective

The purpose of AI4Demand project was the development of short-term (daily or hourly) energy demand and consumption forecasting module using novel AI-based approaches where the cause-and-effect relations between energy data and other externally influencing factors are considered and analyzed.

The project aims to develop a prediction module that:

- Can be used to optimize building energy management systems (BEMS).
- Enables energy suppliers to optimize the production procedure and anticipate peak demands.
- Allows end-users to have a better understanding of the insights of their consumption habits and scenarios.
- Helps consumers to reduce the energy consumption and lower energy bills.
- Aims supplier companies to manage energy waste which leads in reducing the carbon footprint and financial loses.

Developed Services

The developed solution includes the following assets and services:

1. A small code in Python that can be used to collect the necessary data from the API.
2. The prediction module that can be used for exploratory data analysis and forecasting services. The provided codes are configurable and integrable in terms of:
    - **Pre-processing stage:** The whole data analysis layer is fully modifiable and can be changed and added new functionalities depending on the user's preferences.
    - **Prediction horizons:** The user can set a preferable forecast horizon. It should be noted that our model is trained with 70% of the data and 30% is used for the validation of predictions. Any modification of the forecast horizon can lead to different results.

- **A number of external variables:** The user can set the number of necessary variables used for the forecasting of the target variable.
- **Integrability:** The developed source code can be integrated into the 3rd party applications. It should be noted that the provided code is adapted to the specific database and integration should be done with the modification of learning parameters.

3. There are two datasets with hourly and daily periodicity provided with the source code which can be used for multivariate analysis and prediction purposes. The user can update the existing model with the desired datasets and the input file should follow the .csv format.

Methods Used

- Data Collection (NUUKA API and manual collection)
- Data Cleaning (Elimination of missing values, MinMax scaling)
- Feature Selection (Random Forests)
- Noise Reduction (Discrete Wavelet Transform)
- Deep Learning (LSTM - Autoencoders)
- Evaluation (MSE, POCID)

Libraries

- Pandas
- Numpy
- Scikit Learn
- Pwt
- Math
- Keras
- Tensorflow

Project Description

The project began with the investigation of the current state-of-the-art of energy data analytics. Several methods have been examined for short-term energy demand prediction systems suitable for buildings and the main architecture of the multi-level prediction system was developed. Next, a data collection procedure has been implemented to collect the necessary datasets related to energy consumption, IoT, and weather data. Hourly and daily electricity consumption data were collected from the City of Helsinki's Utility and Service Properties through the NUUKA API service, and the meteorological data were gathered from the Finnish Meteorological Institute's open data portal. For the building of the prediction module, the collected datasets were divided into training and testing sets with a ratio of ~70% to ~30%. For the evaluation of the results, a simple Random Forest-based prediction algorithm and the hybridized LSTM-based learning algorithm combining cross-correlation-based feature selection methods are also applied to the same dataset. The results suggest that the use of the newly developed hybridized module has successfully improved the prediction accuracy, and the project achieved its main objectives.

Data description

The dataset part includes two .csv files at hourly and daily resolution.

1. The first dataset contains the hourly electricity and heating consumption values from the office building located in Malmi area together with 9 meteorological features obtained from the nearest meteorological station called Helsinki Malmi lentokenttä. The weather-related variables include cloud amount, pressure, relative humidity, air temperature, dew-point temperature, horizontal visibility, wind direction, gust speed, and wind speed. The observation period is 01/01/2020 – 17/03/2020.

2. The second dataset contains daily time series data from the period of 01/01/2019 – 31/12/2021. The data variables include the daily electricity and heating consumption values from the school building located in Kaisaniemi area, as well as 5 meteorological variables (precipitation amount, snow depth, air temperature, maximum temperature, minimum temperature) collected from the nearest weather observation station called Helsinki Kaisaniemis.

For the collection of other necessary datasets, we suggest using the small script provided as a service for the collection of electricity, heating, water, and cooling-related datasets. We provide also an example of how to use the code to obtain the data through the NUUKA API service.

System Architecture

The architecture of the proposed demand forecasting module follows the hybrid-type methodology integrating different components or layers based on data mining, machine learning, and signal processing techniques. The whole project is implemented in Python programming language using different data science and machine learning libraries. The architecture of the prototype is given in the next diagram.

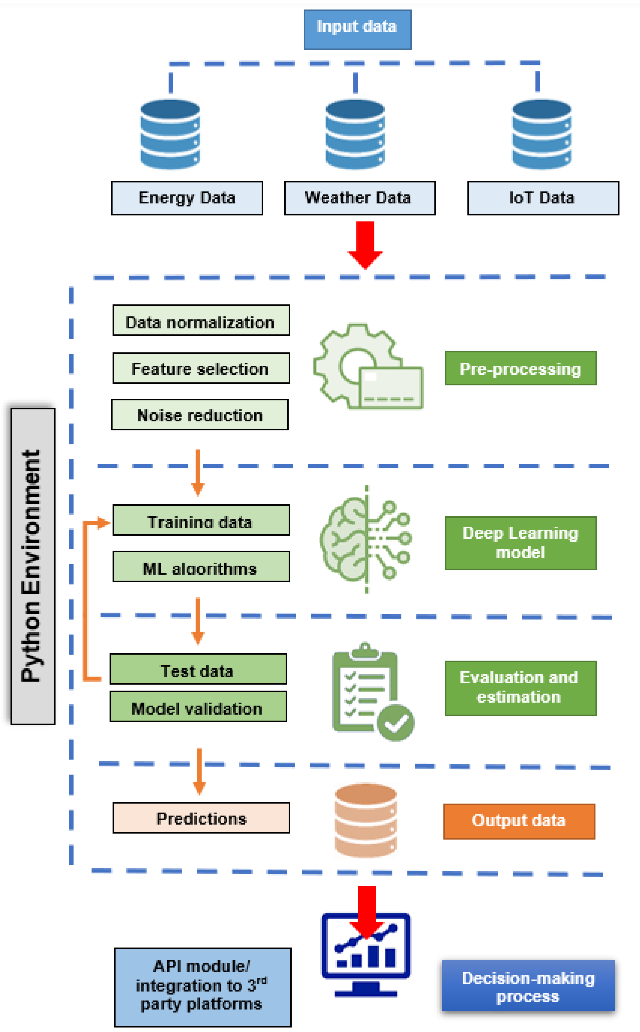

Figure 1. The functionality of the developed prototype

As a first stage, all necessary datasets are collected and then converted to the .csv standardized format. Raw datasets are checked for missing values and normalized for the data pre-processing procedures.

Next, a feature selection algorithm based on the ensemble learning method called Random Forest is implemented to filter out irrelevant and redundant features and keep only important ones for the learning algorithms.

Further, the signal processing technique called Discrete Wavelet Transform is used for cleaning the selected features from the existing noise.

The prediction stage is performed using a neural network architecture based on long short-term memory (LSTM) combined with Auto-Encoder networks.

The evaluation procedure is done with the help of two precision metrics, namely, a mean square error (MSE) and the prediction of change in direction (POCID).