Electricity Consumption Forecasting Using an Explainable AI Framework

Tyra Wodén (021215-7440)

id21twn@cs.umu.se

February 2, 2026
Semester: VT 26
External partner: Umeå Energi

1   Introduction

In recent years, the Swedish electricity market has been affected by large variations in price and periodically very high electricity prices[1][2]. This has increased the need for customers to understand and influence their electricity consumption and future electricity costs. To address this need, Umeå Energi has worked on developing and improving their interfaces for customers, for example by developing an application for private customers. According to customer analytics from Umeå Energi, a highly requested feature is to receive an estimation of both the electricity consumption and the electricity cost before the invoice arrives. Currently, Umeå Energi has developed a linear regression model to forecast the consumption for the current month. The model is based on historical consumption and temperature, aggregated by month. For certain customer groups, Umeå Energi has access to high definition measurement data where electricity consumption is registered on a 15-minute level, which creates opportunities for more advanced modeling. There may also be other factors, apart from historical consumption, that affect the electricity consumption, for example weather data, calendar information (holidays etc.), and energy prices.

The goal with this Master's thesis is to develop and evaluate a forecasting model that shows the predicted energy consumption for the current month. Previous research has identified several models that are able to

---

[1] Allhorn, J., Tidningen Näringslivet: Chockvändningen: Nu har norra Sverige högst elpriser - "Har förändrats ganska snabbt" (January 2026). https://www.tn.se/inrikes/46369/chockvandningen-nu-har-norra-sverige-hogst-elpriser-har-forandrats-ganska-snabbt/

[2] Rydegran, E., Energiföretagen: Dramatik och rekord sammanfattar Elåret 2022 (December 2022). https://www.energiforetagen.se/pressrum/pressmeddelanden/2022/Dramatik-och-rekord-sammanfattar-Elaret-2022/

accurately predict energy consumption for different time intervals, utilizing various combinations of data sources [1, 2]. In recent years, increasing interest has been shown for Machine Learning (ML) and AI solutions to energy consumption forecasting, due to their strength in handling nonlinear problems [3, 2]. The black-box property of many popular AI models has increased the demand of asserting confidence and transparency through means of Explainable Artificial Intelligence (XAI) [3]. This study will address these topics by using XAI to evaluate the contributions of different variables to the energy forecast. Additionally, efforts will be made to continuously update and improve the forecast during the month, as actual observed values of electricity consumption become available.

2   Problem Formulation

This project aims solve the problem of accurately predicting customers' monthly electricity consumption based on consumption data and external data sources. The consumption data is in form of time series data from 16000 customers over roughly three years on a 15-minute level (some customers may not have data for all three years). Other variables such as weather data, calendar information, and electricity prices will be retrieved from Open APIs. The main focus of the project will be to quantify and explain the contribution of each variable in the forecast. Forecasting will be performed using a state-of-the-art ensemble learning model, Gradient Boosting, which has been proven to be efficient for electricity consumption prediction [4]. XAI methods such as SHAP and Feature Importance will be used explain the variable contributions. Previous studies have used ensemble learning for short-term (hourly-daily) forecasting [5, 6, 4], but this study will address the novel task of monthly forecasts using ensemble learning. Additionally, XAI methods will be applied for the monthly forecasts, which previously has been done for the short-term task [7, 8, 3, 4]. This research will also explore the possibility of improving the accuracy of the forecast as the month progresses, by using observed values of electricity consumption. Two research questions are posed, where RQ1 has the highest priority:

1. Using an ensemble learning approach, which forecasting variables contribute significantly to the accuracy of the monthly forecast, and which sources provide little to no improvement?

2. As the month progresses, observed values of electricity consumption become available. How can the observations be used to increase the accuracy of the forecast during the month?

3   Method

To answer the research questions, the following main steps will be performed:

- **Literature study**: Review existing work in the field of electricity consumption forecasting, in order to gain valuable theoretical background and insights on suitable ML models and XAI techniques.

- **Data preparation**: Retrieve 15-minute level electricity consumption data, as well as relevant external data sources (such as weather data, calendar data, and electricity prices). Perform data analysis and pre-processing.

- **Model implementation**: Implement a Gradient Boosting forecasting model based on the consumption data and the external sources.

- **Evaluation of data sources**: Analyze and evaluate how the different data sources contribute to the accuracy of the forecasts using Explainable AI. The goal is to identify data sources that add significant value, and sources that are not able to motivate the increased complexity.

- **Continuous forecast updates**: Develop a method for improving the forecasts during the current month by continuously replacing predicted values with actual observations.

- **Integration perspectives**: Develop a prototype that demonstrates how the forecasting models can be integrated into Umeå Energi's existing environments (Microsoft SQL Server). The prototype includes reading of input data from database tables, execution of forecasting models in Python, and storing of results.

- **Quality assurance**: Develop a suggestion on how to continuously follow up the accuracy of the models after implementing them in production.

3.1   Evaluation Methods

Evaluation will mainly be performed on the data sources' contributions, the forecast improvement over time, and in case of time, the different models' performance. As a baseline comparison, an existing linear regression electricity consumption model (based on historical monthly consumption and temperature data) will be utilized. XAI methods such as feature importance and Shapley values will be used to evaluate the contributions of the different data sources. Different combinations of data sources may also be evaluated. All results will be evaluated for statistical significance, using metrics such as MAPE, MSE, RMSE, NMAE.

4   Literature

Two comprehensive reviews on electricity consumption forecasting have been identified:

- An extensive review and comparison of both statistical and ML/DL techniques for forecasting is found in [1]. They also review combinations of different techniques, i.e. hybrid models. Claims that ANN has more advantages than statistical models, and has better performance for nonlinear problems. Highlights that hybrid models can be beneficial to capture complexities in building energy and operational data.

- Another review of statistical, AI, and hybrid methods for forecasting is found in [2]. They also highlight the strength in AI models for dealing with nonlinear patterns. Claims that hybrids between AI and Swarm Intelligence (SI) methods show potential for increased accuracy. Provides a clear overview of different studies regarding prediction time intervals, included features, building types etc.

Papers where experiments have been performed:

- Support Vector Machine (SVM) for forecasting energy consumption: [9].

- Monthly electricity consumption forecasting based on decomposition methods and ARIMA: [10].

- Forecasting cooling energy using ANN (for three university buildings, weekly/monthly): [11].

- Forecasting high voltage consumers' electricity consumption using LSTM, GRU, TCN: [12].

- SVR and fruit fly optimization with seasonal indexing to address the fact that electricity consumption has a seasonal component: [13]. Results show that the proposed model is a reliable forecasting tool.

- Ensemble learning for short-term electricity consumption forecasting: [5].

- Ensemble learning for short-term electricity consumption forecasting of office buildings: [6].

Papers specifically on XAI and energy consumption forecasting:

- Forecasted hourly energy consumption for the steel sector using three different LSTM models [8]. Used SHAP to interpret the decision-making, and found that leading current reactive power and the number of seconds from midnight contributed significantly to the model output.

- Ensemble learning for electricity consumption forecasting. Evaluated several decision tree-based ensemble learning techniques using SHAP [4]. Found that ensemble learning models outperform deep learning models. Also found that temperature-humidity index and wind chill temperature has a greater impact on short-term forecasts than more traditional parameters such as temperature. Released the code at https://github.com/sodayeong/PLOS-ONE_Github.

- Predicted electricity consumption for residential buildings based on hourly data with information about consumption for different household areas (such as kitchen and appliances) [3]. Used LSTM as prediction model, and LIME and SHAP to provide comprehensible explanations of the predictions.

- Proposed a methodology for selecting input variables for energy consumption prediction using XAI (SHAP) [7]. Used Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), Light Gradient Boosting Model (LightGBM), and LSTM for prediction. Found that variables with strong impact on the forecast include year, hour, energy consumption difference, temperature, and surface-temperature.

5 Implementation

The implementation of the project consists of a model for predicting electricity consumption, as well as integration of resulting model predictions in the existing Umeå Energi database. Part of the implementation is also to use Explainable AI methods to evaluate the contributions of each variable in the prediction. The implementation will be performed on an Umeå Energi computer with access to their database holding electricity consumption data. Python and open-source ML/DL/XAI libraries will be used for the model implementation and variable interpretation. The integration part of the project will utilize SQL.

6 Work Structure and Time Plan

The work structure of the project will follow agile principles, in order to allow flexibility and adjustments of plans as the project develops. Writing the paper will be performed continuously throughout the project. Peer

review sessions will be organized with other students taking the course, to get more feedback in addition to the feedback from supervisors. A time plan in form of a GANTT chart can be found via this link: [https://docs.goo gle.com/spreadsheets/d/10SVX8NpNxUo1fw0eReo7ZZRdi8UZkltd-iCj r12EO6o/edit?usp=sharing](https://docs.google.com/spreadsheets/d/10SVX8NpNxUo1fw0eReo7ZZRdi8UZkltd-iCjr12EO6o/edit?usp=sharing). Week 17 has been left intentionally blank to make time for catching up in any area needed, but also because I will be quite busy outside the thesis that week.

7   Risk Analysis

Below follows a list of potential risks for the project, and what can be done to circumvent them.

- **Risk 1: Difficulties with finding data sources.** Possibly the largest risk for the project is the availability of high-quality external data sources for the forecasts. Previous studies have, apart from historical electricity consumption, focused on numerous variables, often pertaining to weather and building specific factors. The historical consumption data will in this project be provided by Umeå Energi, but retrieval of external parameters relies on open APIs. To begin with, two external APIs have been identified: SMHI[3] which provides both historical and forecasted weather data, and Dagsmart[4] which provides Swedish calendar information. Ideally, other variables will be retrieved from other open APIs, but availability will be determined during the data preparation phase of the project.

- **Risk 2: Problems with assembling training data.** Since several data sources will be used, there may be difficulties in assembling all data into a format that is useable with the prediction model(s). If this turns out to be the case, the methodologies of similar experiments will be surveyed further to find out how they solved the problem. This risk is also mitigated by Gradient Boosting models' lower need for data preprocessing.

- **Risk 3: Limited time.** The limited time frame of the project means that all project goals may not be fulfilled. To address this, the research questions have been ordered in priority, with the XAI part of the project being ranked highest. The other goals will be worked on in case of time. Writing the paper will be done continuously to avoid any big chunks of writing being left at the end of the project.

---

[3]SMHI Open APIs: https://opendata.smhi.se
[4]Dagsmart API: https://dagsmart.se/api/

8 Supervision

8.1 UmU Supervision

Supervision with my internal supervisor will be arranged as needed, roughly once a week or once every two weeks. Email will be used to book time slots, and meetings can happen either in person on campus or digitally via zoom.

Table 1: Contact information for internal supervisor.

Name	Esteban Guerrero Rosero
Email	esteban.guerrero@umu.se

8.2 External Supervision

I will be spending most of my time at the Umeå Energi office, so arranging supervision with my external supervisor will be easy. If needed, I am also able to get help from other team members at Umeå Energi.

Table 2: Contact information for external supervisor.

Name	Elin Eriksson
Email	elin.eriksson@umeaenergi.se

9 Project Journal

The project journal can be found via this link: https://docs.google.co m/document/d/1z1Nd_QyX17tCN-93EjTk_8BK8745s6sru_kyz7s-aGM/edi t?usp=sharing.

References

[1] Deb, C., Zhang, F., Yang, J., Lee, S.E., Shah, K.W.: A review on time series forecasting techniques for building energy consumption. Renewable and Sustainable Energy Reviews **74** (July 2017) 902–924

[2] Mat Daut, M.A., Hassan, M.Y., Abdullah, H., Rahman, H.A., Abdullah, M.P., Hussin, F.: Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review. Renewable and Sustainable Energy Reviews **70** (April 2017) 1108–1118

[3] Janjua, J.I., Ahmad, R., Abbas, S., Mohammed, A.S., Khan, M.S., Daud, A., Abbas, T., Khan, M.A.: Enhancing smart grid electricity prediction with the fusion of intelligent modeling and XAI integration. International Journal of Advanced and Applied Sciences **11**(5) (May 2024) 230–248

[4] Moon, J., Maqsood, M., So, D., Baik, S.W., Rho, S., Nam, Y.: Advancing ensemble learning techniques for residential building electricity consumption forecasting: Insight from explainable artificial intelligence. PLOS ONE **19**(11) (November 2024) e0307654 Publisher: Public Library of Science.

[5] Divina, F., Gilson, A., Goméz-Vela, F., Torres, M.G., Torres, J.F.: Stacking Ensemble Learning for Short-Term Electricity Consumption Forecasting. Energies **11**(4) (April 2018) Company: Multidisciplinary Digital Publishing Institute Distributor: Multidisciplinary Digital Publishing Institute Institution: Multidisciplinary Digital Publishing Institute Label: Multidisciplinary Digital Publishing Institute Publisher: publisher.

[6] Pinto, T., Praça, I., Vale, Z., Silva, J.: Ensemble learning for electricity consumption forecasting in office buildings. Neurocomputing **423** (January 2021) 747–755

[7] Sim, T., Choi, S., Kim, Y., Youn, S.H., Jang, D.J., Lee, S., Chun, C.J.: eXplainable AI (XAI)-Based Input Variable Selection Methodology for Forecasting Energy Consumption. Electronics **11**(18) (September 2022) Multidisciplinary Digital Publishing Institute.

[8] Maarif, M.R., Saleh, A.R., Habibi, M., Fitriyani, N.L., Syafrudin, M.: Energy Usage Forecasting Model Based on Long Short-Term Memory (LSTM) and eXplainable Artificial Intelligence (XAI). Information **14**(5) (April 2023) Multidisciplinary Digital Publishing Institute.

[9] Dong, B., Cao, C., Lee, S.E.: Applying support vector machines to predict building energy consumption in tropical region. Energy and Buildings **37**(5) (May 2005) 545–553

[10] Sun, T., Zhang, T., Teng, Y., Chen, Z., Fang, J.: Monthly Electricity Consumption Forecasting Method Based on X12 and STL Decomposition Model in an Integrated Energy System. Mathematical Problems in Engineering **2019**(1) (2019) 9012543

[11] Deb, C., Eang, L.S., Yang, J., Santamouris, M.: Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks. Energy and Buildings **121** (June 2016) 284–297

[12] Hadjout, D., Torres, J., Troncoso, A., Sebaa, A., Martínez-Álvarez, F.: Electricity consumption forecasting based on ensemble deep learning with application to the Algerian market. Energy **243** (March 2022) 123060

[13] Cao, G., Wu, L.: Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting. Energy **115** (November 2016) 734–745