

MASTER'S THESIS PROJECT SPECIFICATION

ELECTRICITY CONSUMPTION FORECASTING USING HIGH-RESOLUTION CUSTOMER DATA AND EXTERNAL DATA SOURCES

Tyra Wodén (021215-7440)

id21tn@cs.umu.se

January 20, 2026

Semester: VT 26

External partner: Umeå Energi

Introduction

In recent years, the electricity market has been affected by large variations in price and periodically very high electricity prices. 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. Customers wish to, before the invoice arrives, receive an estimation of the electricity cost. A first step towards a cost forecast is forecasting the electricity consumption. 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 forecasting models that show the predicted energy consumption for the current month. The forecast will be updated continuously during the month, and as actual observed values of electricity consumption become available, these values will be used to improve the forecast. The main subject of the thesis is data analysis, with a focus on Machine Learning (ML) and statistical analysis.

Problem formulation

This project aims solve the problem of accurately predicting customers' electricity consumption based on consumption data and external data sources. The forecasts will have a monthly scope. Several research questions are posed:

- What type of forecasting model gives the best results for predicting electricity usage of the current month?
- Which external sources of data contribute to the accuracy of the forecast, and which sources provide little to no improvement?
- How can the accuracy of the forecasts be validated over time after the models are implemented in production?

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 statistical/ML models.
- **Data gathering:** Assemble 15-minute level electricity consumption data, as well as relevant external data sources (such as weather data, calendar data, and electricity prices).
- **Development and evaluation of models:** Implement and evaluate several different statistical and ML (possibly DL) forecasting models.
- **Evaluation of data sources:** Analyze and evaluate how different combinations of the data sources contribute to the accuracy of the forecasts. 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.

Evaluation methods

Evaluation will mainly be performed on the data sources' contributions and 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. The different statistical/ML/DL models will also be evaluated and compared against each other. For the data sources, different combinations will be investigated, and their contribution to the forecast accuracy will be evaluated both individually and in the different combinations. Results will be evaluated for statistical significance.

Self assessment

I am suitable for this project since I have taken fundamental courses in statistics and Machine Learning, as well as having an interest in how raw data can be transformed into useful insights. On advanced level, I have taken a course in database management which will help me in the integration part of the project. During the Student Conference in Computing Science course, I worked on a project where the main focus was Machine Learning/Deep Learning. I have also taken an advanced course in applications and methods of Artificial Intelligence, which gives me good insights in how to reason about knowledge. My background in Human-Computer Interaction helps me keep end-users in mind when developing, focusing on insights that are valuable for the customers.

Resources

For completing this project, the following resources are needed:

- **Internal data:** Electricity consumption data from Umeå Energi on a 15-minute level.
- **External data:** For example, weather data, calendar information, and energy prices. The external data sources are the largest uncertainty of the project, since the availability and quality of API's are not thoroughly explored before starting the project. At the very least, SMHI¹ has an open API for weather forecasts, and Dagsmart² pro-

¹SMHI Weather Forecast API: <https://opendata.smhi.se/metfcst/snow1gv1>

²Dagsmart API: <https://dagsmart.se/api/>

vides an open API for Swedish holidays, bridge-days (klämdagar), and half-days (halvdagar), which will allow for analysis extended beyond Umeå Energi's internal data.

- **Software and environments:** Standard Machine Learning models and tools for statistical analysis. For the integration part of the project, access to a database instance where forecast data can be stored. No unusual hardware is required.
- **Competence from external supervisor:** Knowledge about the current forecasting model and data at Umeå Energi, as well as general competence in data analysis. Also, communication with the Web Development team that will potentially be implementing the forecasts in an existing application for customers.