Quebec Energy Consumption Prediction

Philippe Joly – 261205161
pjoly3@mail.mcgill.ca

2024-02-09

Summary of Proposal

The goal of this project is to use past energy usage data supplied by Hydro-Québec [5] to estimate future needs in energy. To predict energy demand, energy consumption data since 2019 [5], coupled with weather data [4], population metrics [3], and economic indicators [2] [1] will be used as features for a support vector regression, deep neural network, Long short-term memory network, and Prophet [6] model. The performance of each algorithm will then be compared using the root-mean square and coefficient of determination metrics. This project attempts to predict something complex like energy usage, using easier values to estimate like weather, population size, and economic growth.

Models

- Support Vector Regression (SVR): Here, SVR can be a good choice of model as it combines the power of linear regression and the kernel representation of SVM. This permits the features' projection to a higher-dimensional space (even infinite depending on the choice of kernel), without the associated time complexity. This could allow the model to capture the data's complexity and variance.
- Deep Neural Network (DNN): A DNN can do a similar job to SVR. However, the advantage of DNN is that it removes the choice of kernel from the equation. A DNN, through multiple layers, will, in essence, find the optimal "kernel" of the size of its ultimate hidden layer. This allows the model to pick up on complex trends like the ones in energy usage.
- Long Short-Term Memory Network (LSTM): LSTM is a good choice of model as it can capture complex patterns in time-series data. The main advantage of using a LSTM over a recurrent neural network (RNN) is that unlike LSTMs, RNNs can struggle with long-term dependencies due to the vanishing and exploding gradient problems, where gradients either diminish or grow exponentially as they propagate through time.
- Prophet [6]: Prophet is a model designed by *Facebook* is constructed to modelize time-series data with an emphasis on capturing seasonal and recurrent trends, fitting perfectly for the task at hand.

Performance Metrics

- Root Mean Squared Error (RMSE): RMSE is a good metric to measure the performance of the model in fitting the data. Other advantages of RMSE is that it is highly interpretable as it is in the same units as the output. One disadvantage of RMSE that must be taken into account is that it can be highly influenced by aberrant data points.
- Coefficient of Determination (R^2): R^2 on the other hand will quantify how the models' capacity to capture the data's variance. It will give insight into whether the models' performance is attributable to its capacity to analyze variance in the input data or simply because of chance.

Features

- Year [5]: The year will only be used for the LSTM and Prophet as they deal with time-series data.
- Month [5]: Considering months as a feature opens the way to capturing monthly and seasonal trends in energy consumption.
- day [5]: Using the day as a feature can help capture trends in energy usage linked to specific days in the year like holidays.
- Weekday [5]: Taking into account the weekday could allow models to use weekly consumer energy usage trends.
- Time [5]: Time is a logical feature as household energy usage vary hourly depending on the time of day.
- Temperature [4]: Temperature could give insight into energy usage as it is most-likely linked to household heating and cooling energy consumption.
- Population of Quebec [3]: It is probable that the province energy consumption is highly correlated to its population size.
- Quebec Gross Domestic Product (GDP) [2]: This economical index captures the industrial growth and activity in the province which are to be considered in electrical energy usage.
- Household Disposable Income [1]: This feature quantifies the financial situation of consumers, which could influence their energy consumption.

Keep in mind that the models will forecast slightly varying outcomes. The Prophet and LSTM models will forecast the progression of the time-series, while the SVR and DNN will aim to predict energy consumption based on particular inputs. Nevertheless, their performance can be compared.

Data Processing

Most of the data is already cleaned and in readable formats (`.csv`, `.xlsx`, and `.json`). The date and time data will be encoded numerically. Temperature, GDP, Disposable Income, and Population will all be normalized for all four models. The LSTM and Prohphet models can handle time-series data as is. For the SVR and DNN models, date and time (and year will be disregarded) will be normalized like other inputs.

Web Integration

The performance of the models could be displayed on a web app by contrasting the model prediction with the real time data on energy usage offered Hydro-Québec [5]. The user could alternate between the different models and compare the predicted and actual chart of energy usage.

References

[1] Statistics Canada. Adjusted household disposable income, canada, provinces and territories, annual (x 1,000,000), 2023. URL: https://doi.org/10.25318/3610061201-eng.

[2] Statistics Canada. Gross domestic product (gdp) at basic prices, by industry, monthly, growth rates (x 1,000,000), 2023. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610043402.

[3] Statistics Canada. Population estimates, quarterly, 2023. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000901.

[4] Environment and Natural Resources Canada. Past weather and climate historical data, 2024. URL: https://climate.weather.gc.ca/historical_data/search_historic_data_e.html.

[5] Hydro-Québec. Electricity demand in québec, 2024. URL: https://www.hydroquebec.com/documents-data/open-data/electricity-demand-quebec/.

[6] Letham B. Taylor SJ. Forcasting at scale. *PeerJ Preprints*, 2017. URL: https://doi.org/10.7287/peerj.preprints.3190v2.