Predicting Toronto's Energy Demand

Tanishq Maheshwari
School of Computer Science
University of Waterloo
Waterloo, ON, N2L 3G1
tmaheshw@uwaterloo.ca

Abstract

This project used hybrid deep learning architectures to predict energy demand for Toronto in the form of a week-long hourly load forecast. We were able to achieve a mean absolute percentage error (MAPE) of 3.97%.

Project data
Project code

1 Introduction

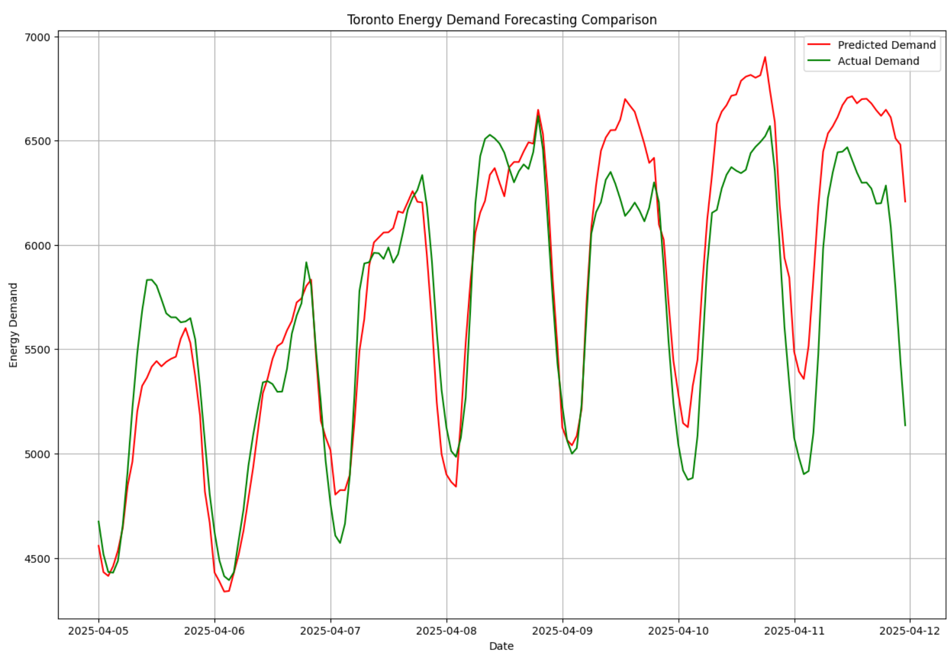


Figure 1: Week long hourly energy demand forecast using LSTM-attention hybrid architecture vs. actual demand values obtained after April 12.

Predicting electricity energy demand—also referred to as energy load forecasting—is extremely important in order to effectively optimize and balance the electricity grid. This helps prevent blackouts by ensuring demand is met, energy dumping in cases of excess supply, and improves energy production efficiency with respect to sustainability and cost. However, accurate forecasting using machine learning models is challenging because the demand for electricity is dependent on many variables: historical demand, weather, grid status information, socioeconomic factors, and more. These factors are all important to consider as the relationship between them is often non-linear and interdependent.

Furthermore, there is significant opportunity to improve energy demand prediction as the last few years have seen rapid development in the application of novel machine learning and deep learning model architectures. As such, this research explored different hybrid architectures in order to create week-long hourly load forecasts for Toronto (which is defined by Ontario's energy operator as the area here) and all of Ontario.

For a week-long hourly forecast from April 5th to April 12th we were able to achieve a MAPE of 3.97% for the Toronto region and a MAPE of 4.71% for Ontario.

The Independent Electricity System Operator (IESO)—the crown corporation responsible for operating the electricity market—provides load forecasts for Ontario but not Toronto. IESO's predictions for the week of April 5th achieve a MAPE of 2.59% which means we were unsuccessful in achieving our goal of surpassing the state of the art.

2 Background

2.1 Related Work In Academia

There has been significant research in applying machine learning to energy markets and there are a number of papers that serve as motivation for our attempt at improving short-term load forecasting. In terms of applications to energy markets, there have been attempts to predict the energy demand of single households [12], the energy mix required to meet demand (i.e. how much energy comes from renewable sources vs. fossil fuels) [10], energy pricing [8], etc. Over the last ten years, techniques employed for these various problems have evolved from rule-based systems and statistical optimization methods to machine learning and deep learning algorithms. We now provide a brief overview of some papers that motivated our project.

[13] leverages a hybrid architecture that employs convolutional neural networks (CNNs) and long short-term memory (LSTM) together in order to predict energy demand. This first illustrated the potential of hybrid architectures and led us to further research the concept.

[12] leverages dual-stage attention-based recurrent neural networks (DA-RNNs) to predict electrical load consumption for a single household. This paper demonstrated the promise of using hybrid architectures that leverage attention specifically. We opted to use LSTMs over DNNs as LSTMs are able to better learn long-term dependencies that would be valuable in making week-long predictions [5].

[10] leveraged different forms of data such as weather data, energy consumption from different sectors, and grid data regarding energy flow, load distribution, and grid efficiency in order to predict energy generation and demand patterns. They were able to achieve strong results in spite of using traditional machine learning models rather than deep learning techniques. This demonstrated the importance of having good data when training ML models.

Finally, [8] was the most novel approach towards predicting energy pricing and was a large source of inspiration in our initial project proposal. This paper implemented soft voting and stacking ensemble models that predicted energy pricing based on weighted output from different models. Unfortunately, in our implementation, we were ultimately unable to implement ensemble methods as a result of time and compute constraints.

Given that this is an applied machine learning problem, application in industry is likely closer to state of the art. However, as a result, less information can be found online about machine learning techniques employed. That being said, the next section discusses the importance of various data and considers what data IESO might use to predict energy demand in Ontario.

2.2 Application In Industry

Independent System Operators employ algorithms much more complex than most people realize. Algorithmic advancement requires strong market knowledge, real-time optimization capabilities, understanding of renewable energy production, consideration of distributed energy resources such as privately owned energy storage systems, and more.

Ontario's IESO is already quite advanced and shares a number of publicly accessible system and market reports [7]. Notably, IESO has a real-time energy market that serves as a platform for matching supply and demand of electricity in Ontario by providing a market clearing price for energy every five minutes based on bids and offers in the wholesale electricity market [6]. Note that we have opted not to apply our machine learning models to real-time energy predictions like this as it requires access to data and data engineering at a scale that is beyond our resources.

Based on IESO's publicly-shared data, there are a number of unique variables and factors that IESO could be using in their demand and cost forecasts. These include energy output per generator and maximum generation capabilities, day-ahead operating reserve (the amount of backup electricity supply available), industrial load by sector, and imports/exports of electricity. These all have significant implications on energy demand. For example, the Niagara region produces much more energy than they consume due to energy production from Niagara Falls, and thus consistently exports energy to New York. As mentioned before, these different sources of data and information are all interdependent and nonlinear, so feeding this information into machine learning models can be valuable in understanding how it might affect demand in ways that we might not otherwise understand. Above all, the most important factors remain historical demand and weather, and this was confirmed by speaking to an IESO data engineer.

3 Main Results

3.1 Approach

At a high level, we predict electricity demand by examining historical demand, weather information, and temporal features. We had 70 different features with information such as current demand, demand 24 and 168 hours ago, temperature, cloud coverage, wind speed, day of the week, whether or not it was a holiday, etc. All data was gathered from IESO and Open-Meteo, an open-source weather API. Unfortunately, the weather data is only from the Toronto region and does not comprise weather data as accurately as we would like due to API limits. Thus, it represents Toronto well but not necessarily all of Ontario, this is especially important for a province like Ontario that is so vast and has such intense climate up north.

We use a hybrid deep learning architecture that combines a LSTM (long short-term memory) model with attention mechanisms in order to capture both short-term patterns and long-range dependencies in the data. The model learns through a sliding window approach where we look at the past week's data (168 hours) in order to predict future demand. During training, the window slides through the historical dataset such that hours 1-168 predict hour 169, hours 2-169 predict hour 170, etc. When we are creating our week-long hourly forecast, the sliding window is modified to incorporate forecasted weather data and demand predictions as they are made.

Expressed mathematically, the goal is to learn a function f that predicts future energy demand:

$$\hat{y}_{t+h} = f(x_{t-w+1}, x_{t-w+2}, ..., x_t)$$

Where:

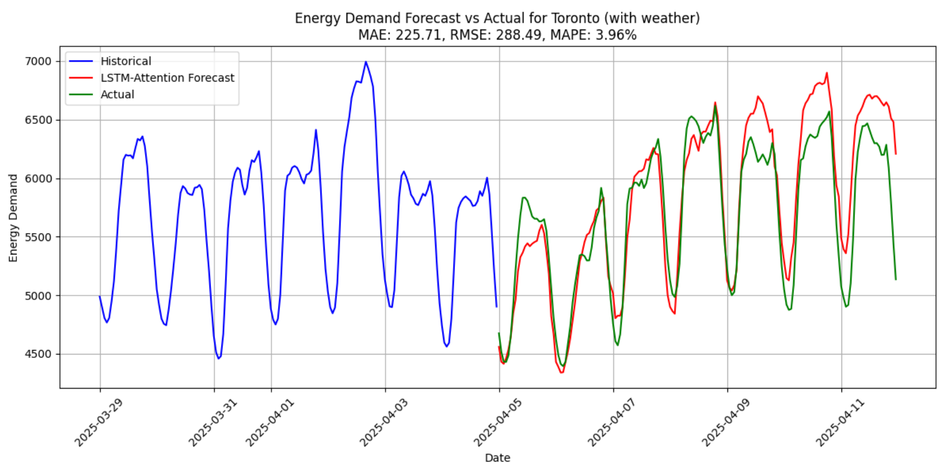
- $\hat{y}_{t+h}$ is the predicted energy demand at time $t + h$.
- $x_t \in \mathbb{R}^n$ is the feature vector at time t containing historical energy demand, weather variables, and temporal features. If the date is outside of the existing dataset, we use predicted energy demand and forecasted weather variables.
- w is the window size (168 hours).
- h is the forecast horizon (1 hour ahead).

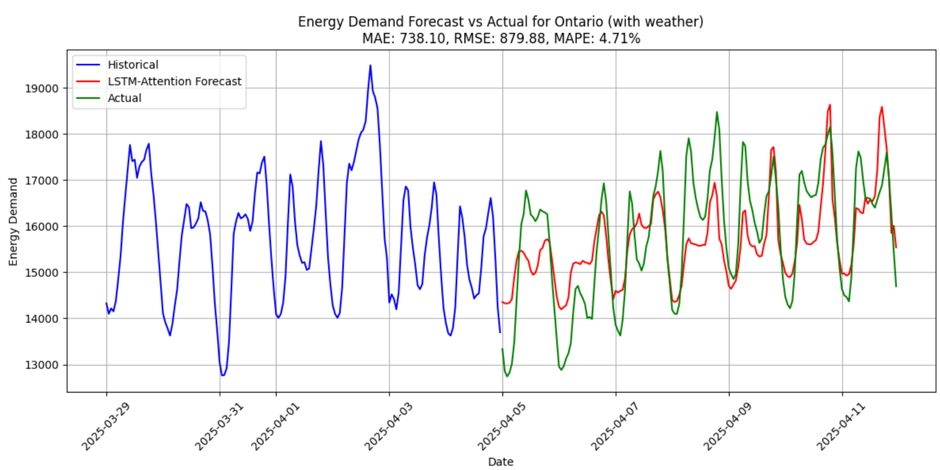
3.2 Findings

Comparing Forecasted Predictions Against Actual Energy Demand

Testing Conditions	MAE	RMSE	MAE
Toronto with forecast weather	225.71	288.49	3.97%
Toronto without forecast weather	271.35	352.48	4.77%
Ontario with forecast weather	738.10	879.88	4.71%
Ontario without forecast weather	854.49	1004.43	5.37%
Ontario forecasted by IESO	412.99	504.17	2.59%

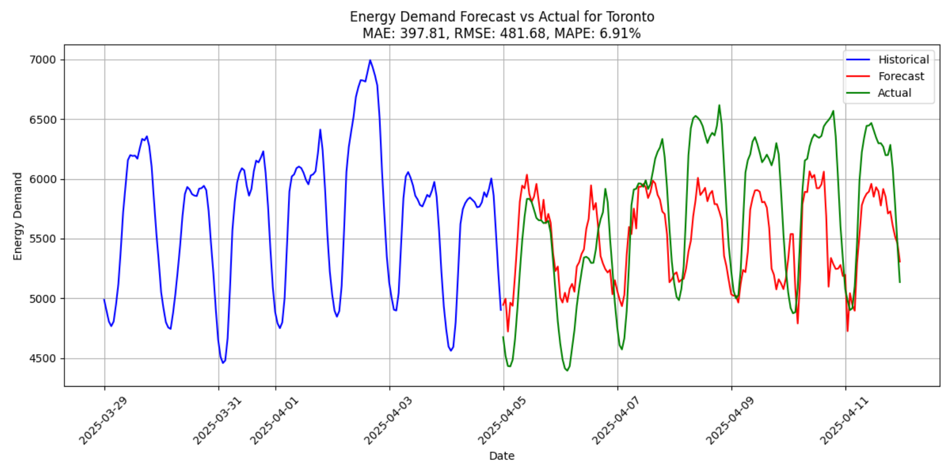
We tested our model with week-old data so that real demand data would be available to compare to. As mentioned before, our model performed best when making predictions for the Toronto region and when incorporating forecasted weather from Open-Meteo. We wanted to incorporate interpretability into our model using methods such as SHAP but lacked the ability to do so due to time constraints. While testing the model with and without the forecasted weather data doesn't explain how each feature contributes to the model, it does demonstrate that weather data was valuable in better predicting energy demand.

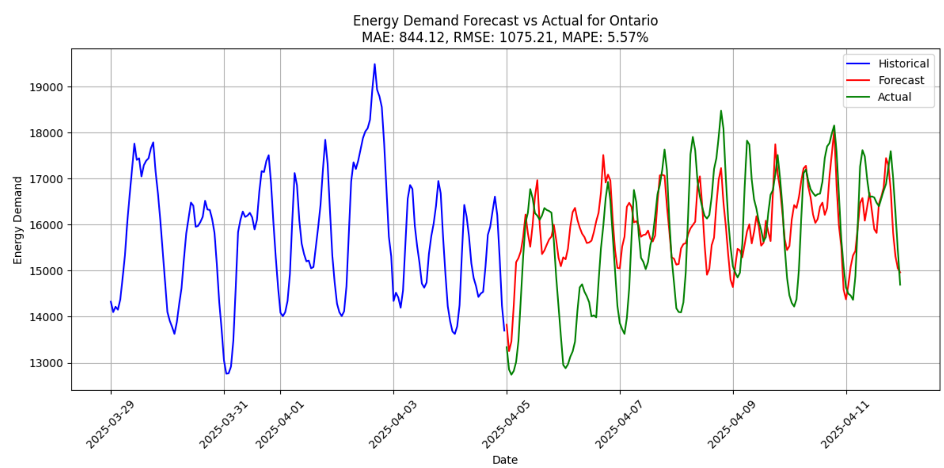




3.3 Experiments

Based on preliminary research, our hypothesis was that a CNN-attention hybrid architecture would be best suited to learn from weather-related features and have the best performance. Even though we did incorporate different weather regions into our data, it was still structured as a time-series, and consequently the LSTM-attention architecture ended up performing better. When creating a week-long hourly load forecast we had a MAPE of 6.91% for Toronto and we had a MAPE of 5.57% for Ontario. Curiously, Ontario fared better than Toronto in this instance in spite of the data being a poor representation of weather across Ontario. This demonstrates how different models identify different patterns and trends within the data.





Throughout the project, we experimented with different hyperparameters by modifying window sizes, epochs, batch size, as well as the number of attention heads and number of hidden dimensions in the model. In the end, we found that the following combination of hyperparameters gave us the best results. That being said, there was definitely room for more experimentation and it could be the case that some other permutation of hyperparameters would have led to better results. This is especially true in instances where increasing window size or the number of hidden dimensions was beyond our compute capabilities.

Optimal Hyperparameters

Window Size	168
Epochs	50
Batch size	16
Hidden dimensions	128
Number of attention heads	4

4 Conclusion

While we failed to achieve state of the art results for energy demand forecasting in Ontario, partly due to time, compute, and data constraints, completing this project was a great learning experience.

I had the opportunity to create a machine learning model and go through the stages of planning, data preparation, model engineering, and model evaluation. Through this work, we were able to achieve a MAPE of 3.97% when creating a week-long energy load forecast for the Toronto region.

With more time and compute resources, we would have experimented with more models such as the temporal fusion transformer [9], N-BEATS [**<empty citation>**], and Prophet [4]. Also, we would have likely benefited from incorporating ensemble methods, which are techniques that look to combine the outputs of multiple models into a single final prediction with each model contributing weighted according to its expected accuracy. This would have been valuable as ensemble methods are less likely to overfit to specific patterns in data and would be more robust to variance in different features (https://www.ibm.com/think/topics/ensemble-learning).

It was already difficult enough to determine what features were of the most importance when working with a single model, and this would become more problematic as we add more data or incorporate more models. As such, I will work to develop machine learning models with interpretability from the start next time, this would also be valuable in the experimentation process so that we can perform feature engineering with some precision.

Finally, it would have been valuable to incorporate more data, such as the aforementioned import and export of energy through Ontario. It is almost always the case that training on more high-quality data will result in a better model. Above all, completing this project truly showed me just how necessary good compute and data are for producing good machine learning models.

Acknowledgement

Thank you Professor Yu and the TAs for running a challenging but fun course, I particularly enjoyed the lecture on fairness! (Yet didn't include interpretability in my model)...

References

[1] U.S. Energy Information Administration. "API Dashboard". URL: https://www.eia.gov/opendata/browser/electricity/rto/region-data. (accessed: 27.02.2025).

[2] U.S. Energy Information Administration. "ERCOT Reported Outages". URL: https://www.eia.gov/opendata/browser/electricity/rto/region-data. (accessed: 27.02.2025).

[3] National Centers for Environmental Information. "Global Hourly - Integrated Surface Database (ISD)". URL: https://www.ncei.noaa.gov/products/land-based-station/integrated-surface-database. (accessed: 27.02.2025).

[4] Facebook. "Prophet Model". URL: https://facebook.github.io/prophet/.

[5] Hassaan Idrees. "RNN vs. LSTM vs. GRU". URL: https://medium.com/@hassaanidrees7/rnn-vs-lstm-vs-gru-a-comprehensive-guide-to-sequential-data-modeling-03aab16647bb.

[6] IESO. "Real-time Energy Market". URL: https://www.ieso.ca/sector-participants/market-operations/markets-and-related-programs/real-time-energy-market.

[7] IESO. "System and Market Reports". URL: https://reports-public.ieso.ca/.

[8] V. Laitsos et al. "Data-Driven Techniques for Short-Term Electricity Price Forecasting through Novel Deep Learning Approaches with Attention Mechanisms". *Energies* (2024). URL: https://www.mdpi.com/1996-1073/17/7/1625#sec5dot1-energies-17-01625.

[9] B. Lim et al. "Temporal Fusion Transformer". 2019. URL: https://arxiv.org/abs/1912.09363.

[10] B. Oladapo, M. Olawumi, and F. Omigbodun. "Machine Learning for Optimising Renewable Energy and Grid Efficiency". *Atmosphere* (2024). URL: https://www.mdpi.com/2073-4433/15/10/1250.

[11] B. Oreshkin et al. "N-BEATS: Neural basis expansion analysis for interpretable time series forecasting". 2020. URL: https://arxiv.org/abs/1912.09363.

[12] A. Ozcan, C. Catal, and A. Kasif. "Energy Load Forecasting Using a Dual-Stage Attention-Based Recurrent Neural Network". *Sensors* (2021). URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC8587894/.

[13] C. Tian, J. Ma, and P. Zhan. "A Deep Neural Network Model for Short-Term Load Forecast Based on Long Short-Term Memory Network and Convolutional Neural Network". *Energies* (2018). URL: https://www.mdpi.com/1996-1073/11/12/3493.