Eppu Heilimo, Janne Alatalo*, and Mika Rantonen

Predicting the electricity consumption of Finland

Abstract: This paper evaluates the performance and computational requirements of seven different machine learning (ML) algorithms to predict the electricity consumption of Finland. The forecasted period is 24 hours into the future using 24 hours of historical data as an input. The tested ML algorithms were linear regression, random forest (RF), gradient decent regression, support vector regression, multilayer perceptron, convolutional neural network (CNN), and WaveNet. A dataset was constructed by combining three data sources containing historical data about the electricity usage, weather, and industry turnover. The CNN model achieved the best results with both RF and WaveNet in the second place with comparable performance to each other.

Keywords: electricity consumption prediction, machine learning, benchmark

Eppu Heilimo: Jamk University of Applied Sciences, E-mail: eppu.heilimo@jamk.fi
***Corresponding Author: Janne Alatalo:** Jamk University of Applied Sciences, E-mail: janne.alatalo@jamk.fi
Mika Rantonen: Jamk University of Applied Sciences, E-mail: mika.rantonen@jamk.fi

1 Introduction

Knowing the country's next day electricity consumption in advance would be advantageous. To keep the power infrastructure running a balance needs to be kept between the electricity production and consumption. If the electricity consumption is known in advance, the production can be planned ahead of time with optimal resources. Electricity consumption prediction is not a new concept. Prior art can be found, such as [5], where the authors tested different machine learning (ML) algorithms to predict the power consumption of a university campus. In this study we benchmark different ML algorithms to predict the electricity usage of Finland for the next 24 hours.

2 Materials and Methods

2.1 Dataset

The dataset was generated by combining three open access dataset sources. The electricity consumption of Finland was obtained from the open access download portal hosted by the Finnish national grid operator Fingrid [3]. The data is recorded with one hour interval and includes the total electricity consumption in Finland in MW h/h.

This data was enriched with weather data from Helsinki-Vantaa Airport weather station. Weather changes the heating and cooling requirements of buildings; therefore, it is an important variable to convey for the models. The data was obtained from an open access portal hosted by the Finnish Meteorological Institute [4].

The industry accounts 45% of total electricity usage in Finland [12]. The assumption was that the industry uses less electricity during economic depression, thus the industry turnover was used to convey this information for the models. The data was obtained from Statistics Finland open access portal [11].

Finally, the dataset was supplemented with the information about holidays, time of the week, and the time of the year. Time of the week and time of the year were encoded using the cyclical feature sin/cos encoding technique [8]. The encoding aids the models to understand cyclical features such as the time of the week and time of the year. With naive encoding method, where the week times are converted to numbers ranging from 1 to 168 (168 hours in a week), Monday 1 hour past the midnight is encoded to 1, and Sunday 1 hour to midnight is encoded to 167. The numbers are far apart although in the weekly cycle the times are right next to each other. Converting the time of the week to vectors using the sin/cos method solves this problem.

The dataset consists of a set of input/target samples. The objective was to predict the electricity consumption for the next 24 hours, thus the targets were vectors with the length of 24. The length of the input was likewise set to 24 hours, producing input matrices with a size of 24×19 where the final features were:

- Electricity consumption
- Air temperature
- Air temperature for the next day
- Wind speed

- Wind speed for the next day
- Direct solar radiation
- Snow depth
- Snow depth for the next day
- Industry turnover
- Time of the year (sin/cos encoded)
- Time of the year for the next day (sin/cos encoded)
- Time of the week (sin/cos encoded)
- Time of the week for the next day (sin/cos encoded)
- Days until the next holiday
- Days until the next holiday for the next day

For the models that do not support multidimensional input, the input matrix was squashed to a vector with a length of 456. For the CNN model, the input matrix was constructed without separating the next day's features (air temperature, wind speed, snow depth, day of the week, days until the next holiday) to a separate feature. Instead, the input was constructed for 48 hours making the input matrix to be 48×11. The features that did not have known forecasts for the next day (electricity consumption, industry turnover and direct solar radiation) were set to zero for the hours from 25 to 48.

The generated dataset spanned the period from the 1st of January 2013 to the 10th of December 2019. The data from the year of 2019 was reserved for testing and all other data was added to the training dataset. The samples were generated using rolling window method where the first 24 hours were used as the input and the following 24 hours were used as the target. The next sample was generated by shifting the window one hour over and this was repeated until reaching the end of the dataset. The final dataset consisted of $43,800$ and $8,760$ samples in train and test datasets respectively.

2.2 Models

The tested ML models were linear regression (LR), random forest (RF), gradient descent regression (GRR), support vector regression (SVR), multilayer perceptron (MLP), convolutional neural network (CNN), and WaveNet [9]. These are all well known ML models except for the WaveNet that was added to the benchmark because we had previous experience with the model. The neural network models were implemented using TensorFlow deep learning framework [1], and for the other models an existing implementation was used from the scikit-learn ML library [10].

The random forest model was initialized with 100 estimators while all other parameters used the default values. Likewise, all other scikit-learn models were initialized with the default values.

The MLP model was implemented using five fully connected layers with batch normalization [6] and ReLU [2] activation functions. The unit sizes for the layers were: 8, 16, 32, 64, and 24. The batch normalization layer was located immediately after each dense layer before the activation function was applied. Final layer used linear activation (no activation). Moreover, the model included a dropout layer before the first dense layer. The total number of trainable parameters was $8,832$.

The CNN model was implemented by stacking four convolutional blocks. Each of the blocks were constructed from a convolutional layer, batch normalization layer, and ReLU activation layer. The convolutional layers used 5×5 kernel size with stride 2×1. The number of output filters in the convolutional layers in the different blocks were 64, 32, 16, and 8. The total number of trainable parameters was $114,552$.

Finally, the WaveNet model implementation was the same as what was used in our earlier paper [7]. The total number of trainable parameters was $1,039,872$.

2.3 Experiment Setup

The models were trained using the train dataset and the test dataset was used to compute the mean absolute error, median error, and maximum 97th percentile error metrics. Moreover, the training and inference times were measured. The error metrics were used to compare the model performances, and the measured training and inference times were used to compare the computational requirements of the models.

3 Results

Figure 1 illustrates the accuracies of the different algorithms. The CNN model achieves the highest performance with the lowest scores in mean absolute error and maximum 97th percentile error metrics. The CNN model performance is substantially better in the maximum 97th percentile error metric when compared to the other algorithms. The difference is not as large in the mean absolute error metric with the WaveNet and RF algorithms performing nearly as well. Likewise, the median error metric of the CNN model is comparable to the WaveNet and RF performance which outperform the CNN model with only a small margin.

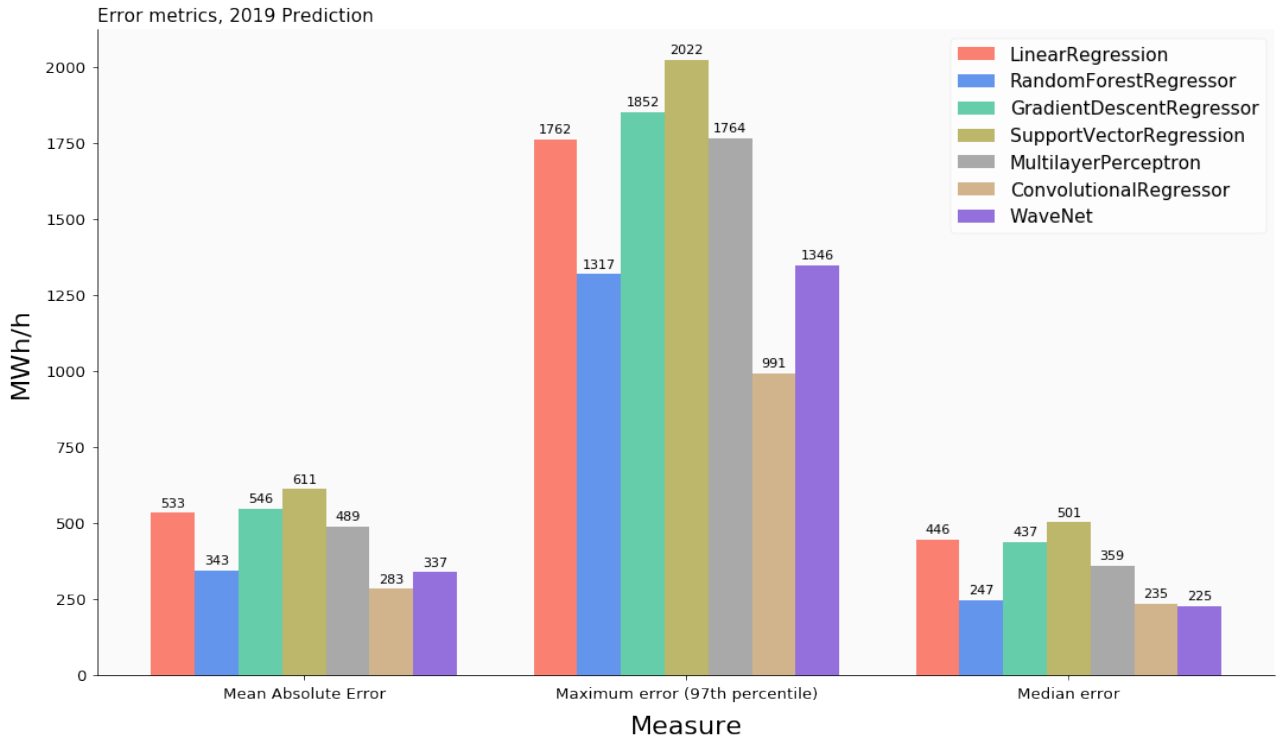


Fig. 1. Accuracy comparison of the different algorithms using three metrics.

The LR, GRR, SVR, and MLP models have clearly worse performance in all three metrics when comparing to the three best performing models. However, even the best performing models have more than 200 MW h/h mean absolute error and median error, which is a large error.

Figure 2 illustrates the predictions for the four best performing models to the test dataset in the year 2019. The plot is smoothed using a 48-hour sliding window that is defined in the Equation 3 where X is the original signal, n is the length of the signal, and $\hat{X}$ is the smoothed signal.

$$X = [x_1, x_2, ..., x_n] \tag{1}$$

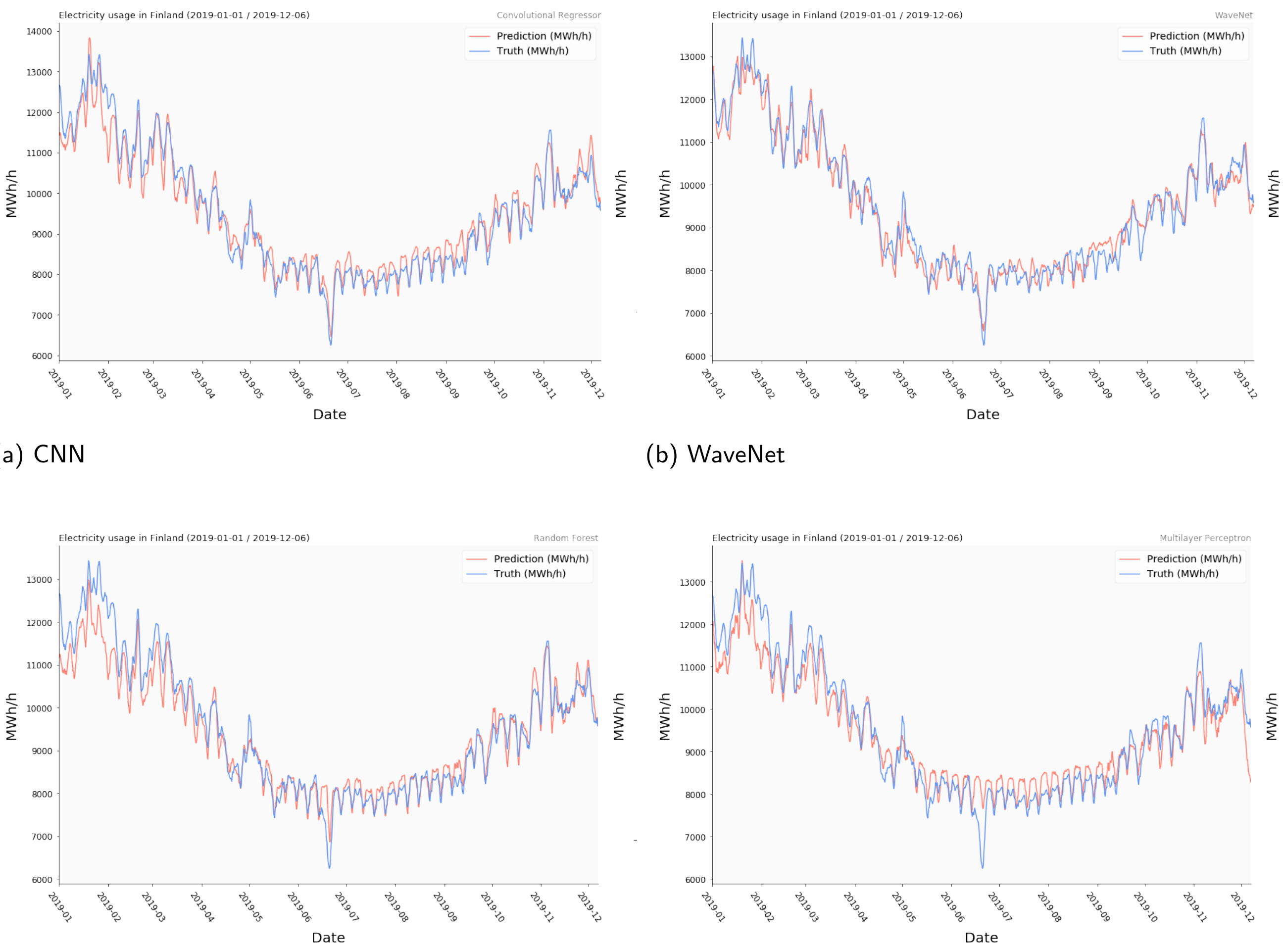
$$\hat{X} = [\hat{x}_1, \hat{x}_2, ..., \hat{x}_n] \tag{2}$$

$$\hat{x}_t = \begin{cases} \frac{x_1 + \ldots + x_{t+25}}{t+25}, & \text{if } t < 24 \\ \frac{x_{t-23} + \ldots + x_{t+25}}{48}, & \text{if } t >= 24 \text{ and } t < n-25 \\ \frac{x_{t-23} + \ldots + x_n}{23+n-t}, & \text{if } t >= n-25 \end{cases} \tag{3}$$

The smoothed signal illustrates better the model accuracy when it is inspected at the span of a full year. Normally the electricity usage includes large changes during the day, thus at the year scale the signal would appear very noisy. However, by smoothing the real and predicted electricity usage signals, the daily accuracies are better highlighted thus creating more readable plot.

Interesting phenomenon to note in the plots are the deviations at the middle of the summer and large variations during the winter. The large dip in the real electricity usage at the center of the plot is positioned around the midsummer festival which is an important holiday in Finland. The large variations in the winter are caused by weather changes where cold temperatures cause raise in heating requirements, thus raising the electricity usage.

The plots verify the observation that was determined from the collected metrics. The three best performing models in Figures 2a, 2b, and 2c predict the next day electricity usage better when comparing to the fourth best performing model in Figure 2d. The difference is most obvious at the midsummer festival deviation, where the MLP model fails completely to consider the holiday and instead predicts the normal electricity usage pattern. Likewise, the RF model predicts similar usage pattern to continue over the holidays with small dip at the end. However, the two neural network models create excellent predictions for the holiday. The WaveNet model creates the best predictions for the winter variations with good predictions in the beginning of February where all other models struggle. All plotted



(a) CNN (b) WaveNet

(c) RF (d) MLP

Fig. 2. Smoothed model predictions for the four best performing models to the year 2019 test data.

models have a clear tendency to overshoot the predictions during summer and autumn months, however all of them have learned to predict the weekly consumption variation patterns.

In addition of computing the model performance metrics, the training and inference times were also measured from all models. Table 1 lists the times for the different algorithms. Computationally the neural network models are more demanding than the more traditional ML models. Even though the neural networks can efficiently utilize GPUs, the training and inference times are longer when comparing to RF model that runs on CPU and has comparable performance.

Table 1. Computational Requirements.

Model	Train time	Inference time
CNN	5 min (GPU)	0.5 ms (GPU)
WaveNet	30 min (GPU)	85 ms (GPU)
RF	52 sec	0.35 ms
MLP	1 min (GPU)	0.5 ms (GPU)
LR	1 sec	0.04 ms
GRR	10 sec	4 ms
SVR	90 sec	3 ms

4 Conclusions

The CNN model worked the best for predicting the electricity usage for the next day. Comparable performances were measured from the WaveNet and RF models. The RF model is computationally less demanding when comparing to the two other well performing neural network models. Therefore, if the algorithm performance is an important aspect, the RF model is a good option to use. However, the more complex neural network models show better performance in predicting the electricity usage on days with large deviations, such as the midsummer festival.

The goal of this research was to benchmark the ML algorithms. Therefore, the study used a simple dataset that contained only one weather station from the Southern part of Finland. However, Finland is a long country, therefore only one weather station cannot convey the information about weather elsewhere in Finland. This is not a problem in this study since all models were trained on the same dataset, thus the results between the models are comparable and the best performing models can be identified.

However, further study is needed to implement a model that is more accurate and therefore usable in real use-cases where electricity usage prediction is needed. The number of weather stations should be increased to convey more information about the weather in other parts of Finland. The CNN model is the most promising of the tested models and it is likely that it can utilize more input data to make better predictions. The model has a good balance between performance and computational requirements, and the benchmark shows that the model can learn complex information from the input data to make more accurate predictions. Therefore, improving the CNN model would be the recommended choice for further studies in this topic.

Acknowledgements

This research was conducted in the Jamk University of Applied Sciences with funding from two projects.

Data for Utilisation – Leveraging digitalisation through modern artificial intelligence solutions and cybersecurity project which is funded by the Regional Council of Central Finland/Council of Tampere Region and European Regional Development Fund.

coADDVA - ADDing VAlue by Computing in Manufacturing project which is funded by REACT-EU Instrument as part of the European Union's response to the COVID-19 pandemic.

References

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org. URL: https://www.tensorflow.org/.

[2] A. F. Agarap. Deep learning using rectified linear units (ReLU), 2019. arXiv:1803.08375.

[3] Fingrid. Electricity consumption in Finland. URL: https://data.fingrid.fi/en/dataset/electricity-consumption-in-finland.

[4] Finnish Meteorological Institute. Download observations. URL: https://en.ilmatieteenlaitos.fi/download-observations.

[5] I. Hajjaji, H. E. Alami, M. R. El-Fenni, and H. Dahmouni. Evaluation of artificial intelligence algorithms for predicting power consumption in university campus microgrid. In *2021 International Wireless Communications and Mobile Computing (IWCMC)*, pages 2121–2126, 2021. doi:10.1109/IWCMC51323.2021.9498891.

[6] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015. arXiv:1502.03167.

[7] T. Kokkonen, S. Puuska, J. Alatalo, E. Heilimo, and A. Mäkelä. Network anomaly detection based on wavenet. In O. Galinina, S. Andreev, S. Balandin, and Y. Koucheryavy, editors, *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*, pages 424–433, Cham, 2019. Springer International Publishing.

[8] T. Mahajan, G. Singh, and G. Bruns. An Experimental Assessment of Treatments for Cyclical Data. In *CSCSU 2021*. 2021. URL: http://hdl.handle.net/20.500.12680/th83m446n.

[9] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. Wavenet: A generative model for raw audio, 2016. doi:10.48550/ARXIV.1609.03499.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[11] Statistics Finland. 112c – index of turnover in industry monthly. URL: https://statfin.stat.fi/PxWeb/pxweb/en/StatFin/StatFin___tlv/statfin_tlv_pxt_112c.px/.

[12] Statistics Finland. Energy in Finland. Technical report, 2022. URL: https://www.doria.fi/handle/10024/185778.