

Short Term Load Forecasting Using TabNet: A Comparative Study with Traditional State-of-the-Art Regression Models [†]

Eugenio Borghini * and Cinzia Giannetti

Faculty of Science and Engineering, Swansea University, Swansea SA1 8EN, UK; c.giannetti@swansea.ac.uk

* Correspondence: eugenio.borghini@swansea.ac.uk

† Presented at the 7th International conference on Time Series and Forecasting, Gran Canaria, Spain, 19–21 July 2021.

Abstract: Electric load forecasting is becoming increasingly challenging due to the growing penetration of decentralised energy generation and power-electronics based loads such as heat pumps and electric vehicles, which adds to a transition to more variable work patterns (accentuated by the COVID-19 pandemic in 2020). In this paper, three different Machine Learning models are analysed to predict the energy load one week ahead for a period of time including the COVID-19 pandemic. It is shown that, by using the recently proposed TabNet model architecture, it is possible to achieve an accuracy comparable to more traditional approaches based on gradient boosting and artificial neural networks without the need of performing complex feature engineering.

Keywords: short-term electricity demand forecasting; neural networks; TabNet



Citation: Borghini, E.; Giannetti, C. Short Term Load Forecasting Using TabNet: A Comparative Study with Traditional State-of-the-Art Regression Models. *Eng. Proc.* **2021**, *5*, 6. <https://doi.org/10.3390/engproc2021005006>

Published: 25 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Electric power load forecasting is widely recognised as a key task for electrical utilities. Accurate predictions in the short time horizon allow to minimise spinning reserve capacity, plan the generation of electric power and configure cost-effective battery charging schedules [1,2]. In the past few years several models based on artificial neural networks have been proposed and shown to be successful for this task [3,4]. Despite this, model selection is not trivial and heavily depends on several aspects of the specific case under study, such as the time resolution of the available data, the type of climate of the location and the required prediction horizon among others. Moreover, the adoption of distributed energy generation, such as wind turbines and solar photovoltaics, the increasing popularity of low carbon technologies (specially, electric vehicles) and even unusual events such as the ongoing COVID-19 pandemic increment the uncertainty and demand levels experienced by distribution networks.

In this context, the recently proposed TabNet model architecture is analysed and compared with two state-of-the-art models such as gradient boosting based on decision trees and deep neural networks (see [3–9]) in the task of predicting the energy load one week ahead at Stentaway primary substation, UK (the choice of forecast horizon is motivated by a Data Science Challenge recently hosted by [Energy Systems Catapult](#)). It was found that the performance achieved by TabNet is comparable with the one exhibited by the more established models, with the advantages of learning directly from the raw data (i.e., no pre-processing is needed) and requiring minimal feature engineering. In addition, given the different nature of TabNet's inductive bias in comparison to more traditional regression algorithms, a further improvement in accuracy was obtained by combining it with the traditional models via ensemble methods.

The article is structured as follows. In Section 2, the description and pre-processing of the employed datasets is given. In Section 3, the three models used for load forecasting are presented. Section 4 is devoted to the analysis of the obtained results. Section 5 contains the summary of the work and some future research lines.

2. Data Description

The historical demand data were collected from the Stentaway Primary substation. They contained average demand power values measured in Megawatts (MW) spanning around 2 1/2 years (between November 2017 and July 2020) and totalling slightly more than 47,000 samples.

Since it is well-known that the weather plays a major role in the energy load, this dataset was complemented with what is known as reanalysis weather data from six sites surrounding the substation extracted using MERRA-2 (the data extraction was based on code available at <https://github.com/emilylaiken/merradownload>, last accessed on 23 June 2021). Reanalysis is a data processing technique that provides a consistent and complete estimation of weather variables over a period of interest. The process consisted of applying modern forecasting techniques to a blend of actual observations with past short-range weather forecasts, thus imitating for historical data the way in which the day-to-day forecasts are generated. In this way, estimations for the averaged hourly irradiance (W/m^2) and instantaneous surface temperature ($^{\circ}C$) were obtained for six locations that could be interpreted as weather forecasts. The sites corresponded to grid points on the numerical weather prediction grid for dates between January 2015 and July 2020.

Both datasets are publicly available at the Western Power Distribution Open Data Hub [site](#) upon login [10].

2.1. Data Pre-Processing

The datasets contained very few erroneous values and gaps (far less than 1% of the samples) which were meaningfully filled. More concretely, the demand dataset presented values that were obviously out of range (both too close to zero and too high) for two weeks in May 2018 and a couple of days in November 2018. All these outliers were replaced by the demand values of the corresponding days from the previous weeks. Regarding the weather data, a few missing values were detected for the temperature at location 4 which were simply filled using the temperature at location 3 since these variables were highly correlated (the correlation coefficient was >0.98).

Finally, the cleaned datasets were merged after linearly interpolating the weather variables to 30 min frequency.

2.2. Feature Extraction

An exploratory data analysis was conducted to unveil patterns and factors that could enhance the predictive value of the original dataset, consisting only of historical demand data and weather reanalysis data.

The most important group of extracted features was derived by studying the auto-correlation of the demand (see Figure 1). As the plot reveals, there were strong daily and weekly patterns in the demand. To account for them, the following features were added to the dataset:

- Hour of the day, day of the week, day of the month, month and year.
- Demand values at the same hour for the whole past week.
- Cyclic versions of hour of the day, day of the month and month, which made explicit the similarity between the end of a period and the beginning of the following one (for instance, the demand around 12:00 PM of a given day tended to be strongly related to the demand around 1:00 AM of the next day) by encoding these features as points in a 2D circle (see [5]).

It was also found that the weather variables produced lagged effects on the demand. After experimenting with different time scales, it was decided to enrich the dataset with the averages of temperature and solar irradiance over periods of 2, 12 and 24 h to capture short-term fluctuations, cyclic day and night patterns and daily trends respectively.

Finally, an ad-hoc strategy was adopted to treat bank holidays and the lockdown period. Specifically, the bank holidays were labelled as a Sunday due to the resemblance of demand patterns between both kind of days, and the lagged demand values were

correspondingly shifted to coincide with that of previous Sunday. Since the behaviour of the demand during lockdown was clearly different from that of regular periods (see Figure 2), it was decided to distinguish lockdown days with a flag.

The resulting dataset contained approximately 100 features.

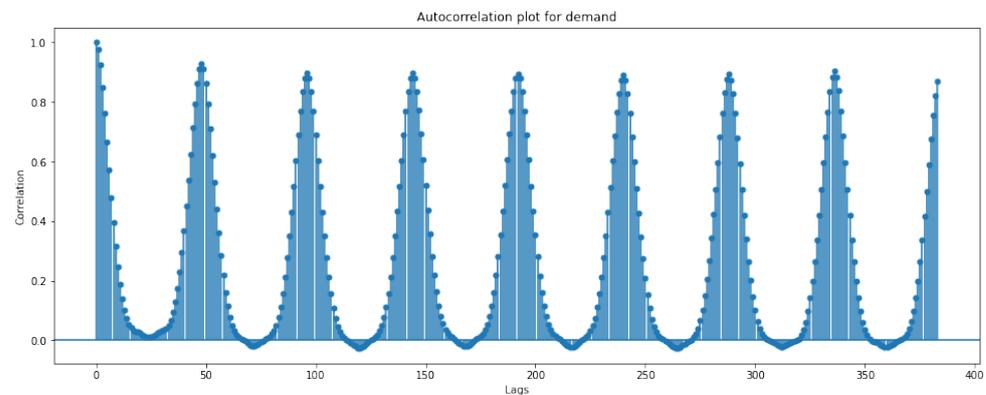


Figure 1. Autocorrelation plot for the demand (the lags are measured at half-hour intervals). There are peaks every 24 h and a slightly higher peak for the same day of the past week.

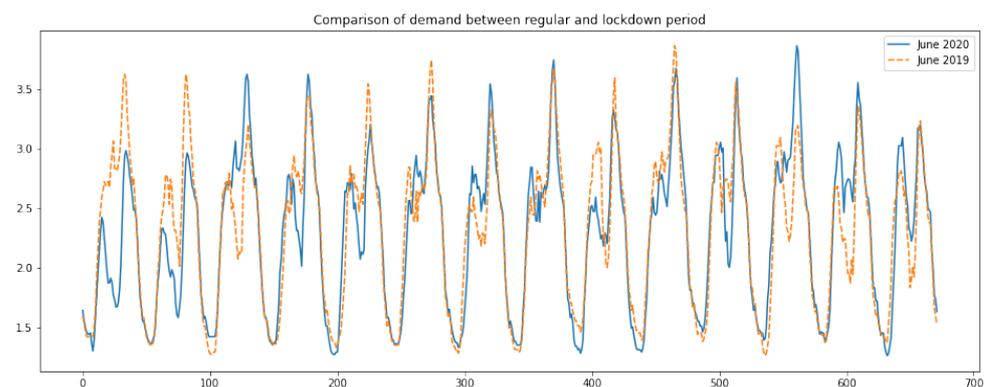


Figure 2. Comparison of demand values between the first two weeks of June 2019 and June 2020 (aligned so that the days of the week coincide).

3. Methodology

The main goal was to forecast one week ahead values of demand (load forecast in MW) using, as model input, its past values in combination with historical and current weather forecast data. As previously stated, the prediction of energy load during the outbreak of the COVID-19 pandemic was one of the main challenges in this study. As it could be expected, the significant change in the energy consumption pattern caused by the various restrictions imposed by the government made it harder to forecast the load for this period. In addition, there is no technique for the short-time load forecasting problem that is known to be superior to all others (see [11]); rather, the best techniques depend heavily on the particular characteristic of the dataset (including factors such as the type of climate and the economic activities at the analysed location, the forecast horizon, etc). For these reasons, three different approaches were contrasted in the present study: gradient boosting tree ensemble model, artificial neural networks and TabNet. The first two techniques are known to achieve state-of-the-art results in several practical tasks and were shown to be successful at short-time load forecasting (see for instance [3–9]). On the other hand, TabNet is a novel deep neural network architecture specially designed to handle tabular data that reportedly outperforms or is on par with standard neural networks and decision trees based variants [12].

All models were trained to minimise the mean squared difference between the predicted and the actual values of demand one week ahead. Roughly 1 year of data was used

(corresponding to the period November 2017–December 2018) as training set, while the remaining weeks (up to July 2020) were used to validate and assess the models' performance using the walk-forward method [2]. Below follows a brief description of each model, together with the specific features and hyperparameters used in each one of them.

CatBoost: CatBoost [13] is an implementation of gradient boosting on decision trees developed by Yandex, which quickly positioned itself as one of the standard methods for learning problems with tabular data, heterogeneous features and complex, non-linear interactions. Gradient boosting is an ensemble method that iteratively improves weak predictors (in the case of CatBoost, decision trees) by performing gradient descent greedily in a certain functional space [14].

All features, both original and extracted, were employed for the CatBoost model. Except for a few relevant hyperparameters that controlled the complexity and regularised the model, the default values were used. These hyperparameters were `n_estimators` (maximum number of trees), `depth` (maximum depth of each decision tree), `max_bin` (number of splits for numerical features) and `rsm` (the proportion of the features considered for each split). Their values were determined by a grid search around initial good values obtained by heuristics and manual experimentation.

Artificial neural network: Artificial neural networks are inspired by a simplified model of how biological neural networks work, and are known to have the capability of learning hidden non-linear and complex pattern in the data. An artificial neural network consists of a directed graph, organised in layers whose nodes are known as neurons. Each neuron applies a non-linear transformation to its input based on learnable parameters and passes the resulting value to neurons in the next layer. These parameters are trained iteratively using stochastic gradient descent with the aim of generating the desired output.

In contrast to the CatBoost model, it was decided to remove several features to reduce multicollinearity issues. Among the time-related features, only the cyclic versions were included and all weather variables were discarded but for the ones corresponding to the two most uncorrelated locations. The total number of neurons was estimated heuristically (proportional to the degrees of freedom of the problem) and it was decided to reduce by a factor of 2 the number of neurons in each hidden layer with the aim of forcing the network to progressively learn more relevant features. The number of neurons in the first hidden layer and the number of layers were determined by a grid search. This resulted in an architecture consisting of four hidden fully connected layers with 64 neurons in the first layer. The non-linear activation ReLU was applied for all layers, while the Adam optimiser was used with the default learning rate 0.001.

TabNet. The new architecture proposed by TabNet learns directly from the raw numerical (not normalised) features of tabular data. The normalisation and feature extraction is somehow embedded in the architecture, since the raw data is filtered by a Batch Normalisation layer and several transformers blocks designed to learn relevant features. One of the salient characteristics of TabNet is the use of a single deep learning block to perform instance-wise feature selection, consisting of a sequential attention mechanism and learnable masks. As a consequence, the accumulated learned weights in this block can be used to interpret the outputs of the model.

For the TabNet model only the cyclic time-related features, the lagged information of the demand and the weather variables of the two most uncorrelated location were employed. The total size of the model was decided by a grid search, following ([12], Guidelines for hyperparameters), to set the values of the hyperparameters `width` and `steps`, which are respectively, the number of hidden neurons in each block and the number of hidden blocks.

4. Discussion and Results

The three models considerably beat naive baselines and achieve a steady accuracy across very dissimilar weeks (see Table 1 below). This is consistent with the existing literature and the common consensus that models based in ensemble of regression trees

and neural networks are the strongest predictors for generic regression tasks. Although in our tests TabNet did not in general outperform the best traditional model, its accuracy was usually close to it. In addition, since TabNet had an inductive bias of different nature to the traditional regression algorithms it allowed us to obtain a further improvement in accuracy by combining it with the traditional models via ensemble methods. Indeed, it was verified that the simple average of the three models achieved an appreciable higher performance than any single model (see Table 2).

Table 1. R^2 scores and root squared errors for the proposed methods. Here the naive baseline consists of predicting the same as the previous week.

Method	R^2 Score	RMSE	R^2 Score (Lockdown)	RMSE (Lockdown)
CatBoost	0.9369	0.2156	0.8562	0.2332
Neural Network	0.9311	0.2254	0.8396	0.2463
TabNet	0.9286	0.2295	0.8424	0.2442
Naive Baseline	0.8740	0.3048	0.7198	0.3256

Table 2. R^2 scores and root mean squared errors for the different averages of the proposed models.

Average	R^2 Score	RMSE
CatBoost+TabNet	0.9477	0.1964
CatBoost+Neural Network	0.9492	0.1936
TabNet+Neural Network	0.9423	0.2062
CatBoost+TabNet+Neural Network	0.9511	0.1898

Regarding the prediction for the lockdown weeks, it was found that reducing the amount of regular samples in the training sets was beneficial for the performance of the predictive models. Concretely, to generate the predictions on lockdown weeks, only samples starting from 2019 were considered for the training set. The rationale behind this decision is that the reduction allows to give more weight to samples corresponding to the lockdown period. The accuracy attained in this way is comparable to the one obtained for normal times (see Figure 3 and Table 1).

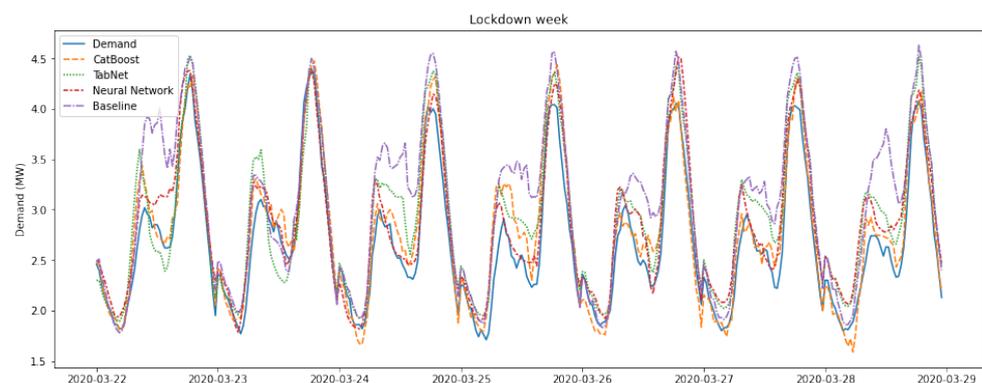


Figure 3. Predictions for the first week of lockdown (from 22 March to 28 March). The consumption pattern is quite different to the one from the previous week.

5. Conclusions

In this study, the performance of the novel TabNet network is compared with two well-established regression models on a short term load forecasting task. It is shown that it is possible to obtain comparable performance to these traditional methods but with little to none feature engineering and data preparation. Moreover, the use of TabNet provides a further boost in the overall accuracy on this task via ensemble methods.

As a future step, it would be interesting to refine the strategy to predict the energy load during the lockdown. As some preliminary evidence suggests, training a strong model on

a regular period and then fine-tuning it using data collected during the lockdown (which can be seen as an application of the transfer learning technique) could lead to further improvements in accuracy.

Author Contributions: Conceptualization, E.B. and C.G.; methodology, E.B. and C.G.; software, E.B.; validation, E.B. and G.C.; formal analysis, E.B. and C.G.; investigation, E.B. and C.G.; resources, C.G.; data curation, E.B.; writing—original draft preparation, E.B.; writing—review and editing, C.G.; visualization, E.B.; supervision, C.G.; project administration, C.G.; funding acquisition, C.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the UK Engineering and Physical Sciences Research Council (EPSRC) project EP/S001387/1 and the European Regional Development Funds projects IMPACT.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets are publicly available at the Western Power Distribution Open Data Hub site [10] upon login. The code required to generate the results referred in the article will be shared upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gross, G.; Galiana, F. Short-term load forecasting. *Proc. IEEE* **1987**, *75*, 1558–1573. [CrossRef]
2. Kaastra, I.; Boyd, M.S. Designing a neural network for forecasting financial and economic time series. *Neurocomputing* **1996**, *10*, 215–236. [CrossRef]
3. Hu, R.; Wen, S.; Zeng, Z.; Huang, T. A short-term power load forecasting model based on the generalized regression neural network with decreasing step fruit fly optimization algorithm. *Neurocomputing* **2017**, *221*, 24–31. [CrossRef]
4. Singh, S.; Hussain, S.; Bazaz, M.A. Short term load forecasting using artificial neural network. In Proceedings of the 2017 Fourth International Conference on Image Information Processing (ICIIP), Shimla, India, 21–23 December 2017; pp. 1–5. [CrossRef]
5. Moon, J.; Park, S.; Rho, S.; Hwang, E. A comparative analysis of artificial neural network architectures for building energy consumption forecasting. *Int. J. Distrib. Sens. Netw.* **2019**, *15*. [CrossRef]
6. Park, D.; El-Sharkawi, M.; Marks, R.; Atlas, L.; Damborg, M. Electric load forecasting using an artificial neural network. *IEEE Trans. Power Syst.* **1991**, *6*, 442–449. [CrossRef]
7. Din, G.M.U.; Marnerides, A.K. Short term power load forecasting using Deep Neural Networks. In Proceedings of the 2017 International Conference on Computing, Networking and Communications (ICNC), Silicon Valley, CA, USA, 26–29 January 2017; pp. 594–598. [CrossRef]
8. Lloyd, J.R. GEFCom2012 hierarchical load forecasting: Gradient boosting machines and Gaussian processes. *Int. J. Forecast.* **2014**, *30*, 369–374. [CrossRef]
9. Hong, T.; Pinson, P.; Fan, S. Global Energy Forecasting Competition 2012. *Int. J. Forecast.* **2014**, *30*, 357–363. 2013.07.001 [CrossRef]
10. Western Power Distribution Open Data Hub Homepage. Available online: <https://www.westernpower.co.uk/innovation/pod> (accessed on 22 April 2021).
11. Hong, T.; Fan, S. Probabilistic electric load forecasting: A tutorial review. *Int. J. Forecast.* **2016**, *32*, 914–938. 2015.11.011. [CrossRef]
12. Arik, S.Ö.; Pfister, T. TabNet: Attentive Interpretable Tabular Learning. *arXiv* **2019**, arXiv:1908.07442.
13. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems*; Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R., Eds.; Curran Associates, Inc.: New York, NY, USA, 2018; Volume 31.
14. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]