



**HAL**  
open science

# Energy Load Forecasting: Investigating Mid-Term Predictions with Ensemble Learners

Charalampos M. Liapis, Aikaterini Karanikola, Sotiris Kotsiantis

► **To cite this version:**

Charalampos M. Liapis, Aikaterini Karanikola, Sotiris Kotsiantis. Energy Load Forecasting: Investigating Mid-Term Predictions with Ensemble Learners. 18th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), Jun 2022, Hersonissos, Greece. pp.343-355, 10.1007/978-3-031-08333-4\_28 . hal-04317183

**HAL Id: hal-04317183**

**<https://inria.hal.science/hal-04317183v1>**

Submitted on 1 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

# Energy Load Forecasting: Investigating Mid-Term Predictions with Ensemble Learners

Charalampos M. Liapis<sup>1</sup>, Aikaterini Karanikola<sup>2</sup>, and Sotiris Kotsiantis<sup>3</sup>

<sup>1</sup> Department of Mathematics, University of Patras, Patras, Rion 265 04, Greece  
`c.liapis@upnet.gr`

<sup>2</sup> Department of Mathematics, University of Patras, Patras, Rion 265 04, Greece  
`karanikola@upatras.gr`

<sup>3</sup> Department of Mathematics, University of Patras, Patras, Rion 265 04, Greece  
`sotos@math.upatras.gr`

**Abstract.** In the structure of the modern world, energy and especially electricity is a prerequisite for regularity. Thus, the requirement for accurate forecasts regarding power system loads seems self-evident. In machine learning, a time series forecasting endeavor can be treated as a regression problem. In such scenarios, ensemble methods are often used for robustness and increased accuracy of the generated predictions. This work is a comparative investigation of the use of ensemble schemes for medium-term forecasting of energy system load. The use of over 300 regression schemes is investigated, in a total of 8 different modifications of the input data, over 5 different time-frames, that is, one day, 7-day, 14-day, 21-day, and 30-day horizons, resulting in a loop of 12000 experiments. Summary tables with representative results from the corresponding Friedman rankings are presented.

**Keywords:** Energy System Load · Power Load Forecasting · Ensemble Learning · Regression Forecasting · Medium-Term Forecasting.

## 1 Introduction

Energy management, from the production stage to storage, sharing and consumption, is a multifactorial issue in which huge interests of diverse parties co-exist, from individuals and production or distribution companies to states and geopolitical formations.

With regard to electricity, and since production and consumption are linked by a synchronicity condition, continuous load monitoring is necessary to regulate the operation of power stations in such a way as to generate the amounts of energy required. Thus, failures in relevant decisions could lead to a variety of problems. Producing more energy than needed may lead to environmental, storing, or supply issues. Less available energy than required may result in sharp increases in production expenditures, given emerging requirements for immediate use of higher-cost units [1]. In addition, black-outs in large areas can have significant consequences on the well-being and well-functioning of society [2]. It

is, therefore, imperative that the best possible forecasts of electricity load are available, as such predictions will help to optimize the power management plan and to make as many sound decisions as possible [3]. Thus, through forecasting, it is possible to ensure the robustness of the entire system.

Predicting energy load, however, is not an easy process. For example, fluctuations in consumer needs depend on many factors. And there is a multitude of such factors than can cause demand to diversify. Meteorological conditions, seasonal and calendar regularities, regional economic standards, the use of new technologies, or the level of industrialization, all tend to alter the energy requirements [2, 4]. There are also random factors that are impossible to foresee. Understandably, the problem of predicting energy system load has been of particular concern to the scientific community with a wealth of relevant research output. On this basis, a variety of methods, ranging from statistical or machine learning to hybrid, have been used in different scenarios with various time horizons and range of applications.

This work aims to apply an extensive comparison of regression models for predicting energy load, investigate whether regression-based ensembles exhibit better performance compared to individual learners and, if so, identify the time-frames where these improvements occur. The work has the following structure: after this short introduction, an indicative part of the literature so far is presented. Next, the methodology and the experimental setup and procedure are introduced. Then, results of the comparison of 24 regression methods for predicting the energy load in 5 different time horizons follow, in contrast with a presentation of outputs of a variety of potential ensemble regressors. Finally, an evaluation of the findings of the work along with a brief report of its possible future extensions are outlined.

## 2 Related Work

In the contemporary social context, the demand for accurate forecasts on energy system load seems more relevant than ever. The growing interest in the subject from the scientific community has led to the publication of numerous papers. This chapter is an attempt to present some of the recent research output. However, this multitude of works, together with the fact that the space is - justifiably - limited make this report just an indicative starting point for the reader.

Regarding regression methods, *multiple linear regression* (MLR) has been used in big data to predict energy load over a 7-day horizon using three-year historical data [5]. The use of *support vector regression* (SVR) in energy problems is also relatively common, mainly due to the fact that these methods identify global minimums rather than local ones, and so, in certain tasks, they appear superior to many, even state-of-the-art, machine learning methods, such as *artificial neural networks* (ANNs) [6]. In [7], the SVR is used for peak load prediction using both historical load values and the corresponding meteorological data. A combination of SVR and *empirical mode decomposition* (EMD) is used in [8] to predict energy load, both in terms of extensive geographic coverage and indi-

vidual buildings, while in [9], the method is utilized in conjunction with *feature selection*, exploiting the national energy load data of Great Britain and France. Deep learning techniques are also quite popular. In [10], *sequence-to-sequence recurrent neural networks* with *attention mechanism* are used to predict the energy load at different time horizons, while in [11], *metaheuristic-search-based* algorithms are used to find the optimal or relatively optimal values for tuning the hyperparameters of *long short-term memory* architectures (LSTMs). *Feed-forward deep neural networks* (FF-DNNs) and *recurrent deep neural networks* (R-DNNs) architectures are compared in [12] on short-term scenarios, while *deep neural network (DNN)* - based load forecasting models are proposed for short-term forecasts in [13]. A combination of *convolutional* and *recurrent neural networks* is presented in [14], comparing it with *linear regression* and SVR in one-day horizon forecasts, using historical and meteorological inputs. LSTM and *gated recurrent unit* (GRU) networks are used with different configurations in [15] for a two-month horizon forecast. Also, a comparison of LSTMs with the classic *auto-regressive integrated moving average* and *exponential smoothing* methods in a small-scale residential scenario of 12 households is presented in [16].

Works comparing classic machine learning methods in terms of their performance in the problem of predicting energy load are presented in [17], in [18], and in [19]. In [20], the *auto-regressive integrated moving average* (ARIMA), SVR and *tree-based* regression, are compared in a household load prediction scenario, while in [21], an LSTM model using *feature selection* and *genetic* algorithm is compared with 7 well-known machine learning algorithms in scenarios of a number of different time horizons. In addition, [22] presents a comparison of eight neural network-based methods in a short-term load forecasting scenario.

Regarding ensemble methods, such schemes combine a finite number of models in order to formulate the final output. They are used due to their robustness and increased accuracy [23], despite the difficulty of their production and the increase in computational costs. Ensembles, given accurate and diverse constituents [24], and bearing in mind the inevitable trade-offs [25], tend to generalize better than individual learners [26]. Thus, the use of ensembles has been suggested for various forecast scenarios, such as pandemic [27], stock market [28], and weather forecasts [29]. In [30], a review of the relevant literature regarding ensemble methods in regression problems is presented. Regarding the use of ensemble methods for energy load forecasting, one can likewise mention a number of related works. Some of them exploit the use of LSTMs in combination with other modern techniques, such as *fully connected cascade* (FCC) networks [31], ARIMA [32] and MLR [33]. SVR-based ensembles using a *sub-sampling* strategy that ensures the diversity of the base learners are presented in [34], while a modified version of *passive aggressive regression* (PAR) for online load forecasting is presented in [35]. Regarding regression tree methods, a scheme based on *multiple regression trees* (CART) is presented in [36], while *gradient boosting* regression tree ensembles are examined in contrast to statistical methods in [37].

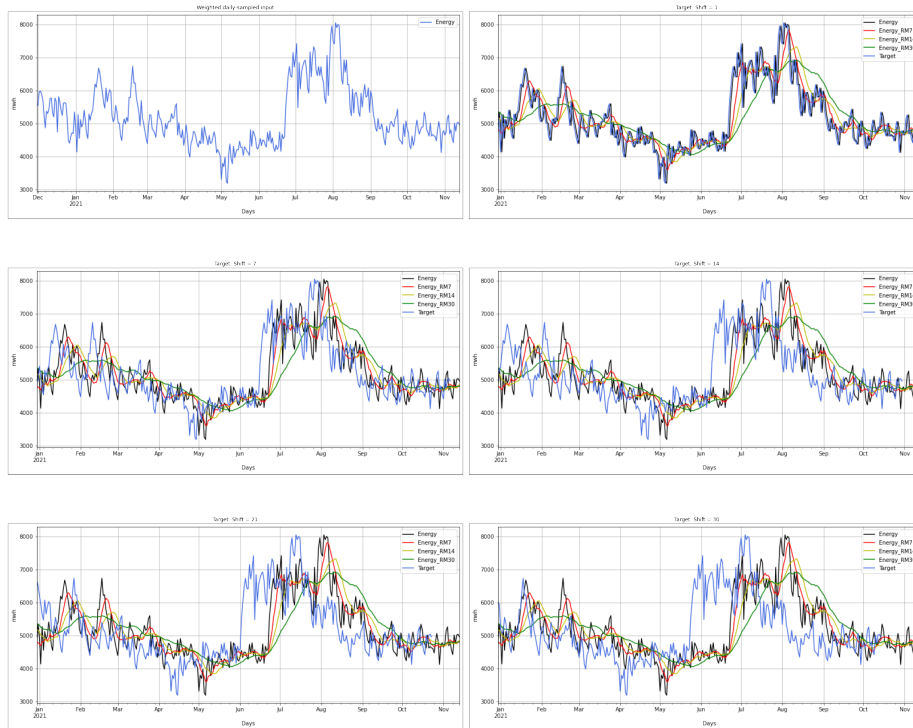
In summary, all of the above show that both a thorough comparison of a large number of regression models over the energy load forecasting problem and the creation of ensembles based on such schemes are both useful and complement the existing literature.

### 3 Experimental Setup

We proceed to sketch the experimental process. Below, the raw input as well as the construction of the final data set used, the base learners tested, the formalization according to which the ensembles were formed and the evaluation metrics will be presented.

#### 3.1 Data

The raw data contains Greek hourly energy system load records ranging from 30 November 2020 to 13 November 2021 [38]. A 24-hour rolling average was



**Fig. 1.** Modified input time series and target data per time-shift

applied to the initial time series, and then a weighted value was exported as

representative of each day. Thus, a new normalized daily time series was created, a component on which the experiments were performed. Then, on the newly formed time series, a rolling mean procedure was applied again for 7-day, 14-day, and 30-day windows.

From the variants of these 4 extracted time series, a variety of input data was created, which consisted of the univariate version and 7 additional multivariate combinations of the individual aforementioned time series. The experiments were performed over 5 time horizons: one-day, 7-day, 14-day, 21-day, and 30-day forecast frames. Based on the prediction time step, works can be characterized as short-term, medium-term or long-term. Thus, the present work is about medium-term predictions. Figure 1 depicts the processed time series used in terms of the corresponding prediction time-shifts.

### 3.2 Algorithms

It has already been mentioned that - in this work - the forecast scenario has been standardized and treated as a regression problem. So, at first, comparisons were made between the individual regression methods and their possible per-2 groupings, the results of which were then used to investigate possible ensembles that exploit triads of the best-performing algorithms. In terms of individual models and base learners, a total of 24 regression methods were utilized. Table 1 contains both the names of the algorithms used together with their respective abbreviations, as well as the reference papers, since it is practically impossible to present them within the limits of the present work.

**Table 1.** Algorithms

	<b>Abbr</b>	<b>Algorithm</b>		<b>Abbr</b>	<b>Algorithm</b>
1	ada	AdaBoostRegressor[39]	13	las	LassoRegression[51]
2	ard	AutoRelevanceDetermination[40]	14	lar	LeastAngleRegression[50]
3	cat	CatBoostRegressor[41]	15	lgbm	LightGradientBoost[52]
4	dt	DecisionTreeRegressor[42]	16	lr	LinearRegression[53]
5	en	ElasticNet[43]	17	mlp	MultiLevelPerceptron[54]
6	et	ExtraTreesRegressor[44]	18	omp	OrthogonalMatchingPursuit[55]
7	xgb	ExtremeGradientBoosting[45]	19	par	PassiveAggressiveRegressor[56]
8	gbr	GradientBoostingRegressor[46]	20	rf	RandomForestRegressor[57]
9	hub	HuberRegressor[47]	21	rsc	RandomSampleConsensus[58]
10	knn	KNeighborsRegressor[48]	22	rid	RidgeRegression[59]
11	kr	KernelRidge[49]	23	svr	SupportVectorRegression[60]
12	llar	LassoLeastAngleRegression[50]	24	tr	TheilSenRegressor[61]

The *pycaret* python library [62] was utilized to conduct the experiments. All examined ensembles were formed under a simple low-cost scheme that produces the final prediction by averaging the outputs of base regressors fit on the training set.

### 3.3 Metrics

The results were evaluated according to the following 6 metrics: *mean squared error* (MSE), *root mean square error* (RMSE), *root mean squared logarithmic error* (RMSLE), *arithmetic mean of the absolute errors* (MAE), *mean absolute percentage error* (MAPE) and the coefficient of determination  $R^2$ . The formalizations could not be presented here due to space constraints. A more detailed description of the structure of the metrics used and how to interpret them can be found in [63].

## 4 Results

In total, more than 8 *data setups* x 300 *algorithmic setups* x 5 *time-shifts* = 12000 *experimental loops* were performed over different data set splits. Due to the size of the results, only a very small part of them was chosen to be presented. Specifically, the results will contain the 5 algorithm schemes with the best performances per metric, regardless of whether they are individual or ensemble methods. This arrangement was formulated after Friedman [64, 65] ranking tests produced over the metric data of each method. Tables 2, 3, 4, 5 and 6 contain the results grouped by prediction time-frame. A much larger portion of the results can be found at [shorturl.at/fvOP9](http://shorturl.at/fvOP9). It should be noted that, in the results presented, the values of the metrics refer to the corresponding scores of their Friedman rankings.

**Table 2.** Friedman Rankings (shift=1)

	<b>Method</b>	<b>MAE</b>	<b>Method</b>	<b>MAPE</b>	<b>Method</b>	<b>MSE</b>
1st	tr	7.625	tr	7.25	tr	7
2nd	tr+gbr+omp	12	tr+gbr+omp	12.25	tr+gbr+omp	14.375
3rd	tr+lgbm+omp	18.125	tr+lgbm+omp	17.75	tr+lgbm+omp	16.625
4th	hub+rsc+gbr	22.125	tr+rsc+gbr	21.75	tr+par	21.625
5th	tr+rsc+gbr	22.375	hub+rsc+gbr	22	par+gbr+omp	24.75
	<b>Method</b>	<b>R2</b>	<b>Method</b>	<b>RMSE</b>	<b>Method</b>	<b>RMSLE</b>
1st	tr	123	tr	7	tr	6.375
2nd	tr+gbr+omp	115.625	tr+gbr+omp	14.375	tr+gbr+omp	14.25
3rd	tr+lgbm+omp	113.375	tr+lgbm+omp	16.625	tr+lgbm+omp	16.125
4th	tr+par	108.375	tr+par	21.625	tr+par	22.25
5th	par+gbr+omp	105.25	par+gbr+omp	24.75	tr+hub	24.5

From the Friedman ranking tables, in each of the time horizons examined, the relative efficiency of the ensemble methods can be observed. Overall they appear to occupy the top positions in the rankings, losing only to *Theil-Sen Regressor* in the first two time-shifts investigated. Regarding longer prediction steps, ensembles prevail. This apparent superiority in scenarios involving longer time-frames



**Table 3.** Friedman Rankings (shift=7)

	Method	MAE	Method	MAPE	Method	MSE
1st	tr	4.5	tr	4.375	tr	4.75
2nd	tr+hub	8.375	tr+hub	8.875	tr+hub	7
3rd	tr+par	9.5	tr+hub+rsc	9.125	tr+par	8.125
4th	tr+hub+rsc	10.5	tr+par	9.75	tr+par+hub	9.125
5th	tr+par+hub	11.125	tr+par+hub	11.375	tr+hub+rsc	11.75
	Method	R2	Method	RMSE	Method	RMSLE
1st	tr	125.25	tr	4.75	tr	4.875
2nd	tr+hub	123	tr+hub	7	tr+hub	6.875
3rd	tr+par	121.875	tr+par	8.125	tr+par	8.5
4th	tr+par+hub	120.875	tr+par+hub	9.125	tr+par+hub	9.25
5th	tr+hub+rsc	118.25	tr+hub+rsc	11.75	tr+hub+rsc	11.25

**Table 4.** Friedman Rankings (shift=14)

	Method	MAE	Method	MAPE	Method	MSE
1st	tr+rsc+gbr	22.125	tr+rsc+gbr	20.625	rsc+omp	23
2nd	rsc+omp	22.625	rsc+omp	22.75	tr+rsc+lgbm	25.125
3rd	tr	26	tr	24	tr+rsc+omp	26.875
4th	tr+rsc+lgbm	26.375	tr+rsc+lgbm	24.5	tr+rsc+gbr	27.25
5th	tr+hub+rsc	28.5	tr+hub+rsc	25.875	hub+rsc+omp	28.5
	Method	R2	Method	RMSE	Method	RMSLE
1st	rsc+omp	107	rsc+omp	23	rsc+omp	22.125
2nd	tr+rsc+lgbm	104.875	tr+rsc+lgbm	25.125	tr+rsc+lgbm	23.75
3rd	tr+rsc+omp	103.125	tr+rsc+omp	26.875	tr+rsc+gbr	26.125
4th	tr+rsc+gbr	102.75	tr+rsc+gbr	27.25	tr+rsc+omp	26.25
5th	hub+rsc+omp	101.5	hub+rsc+omp	28.5	tr	26.75

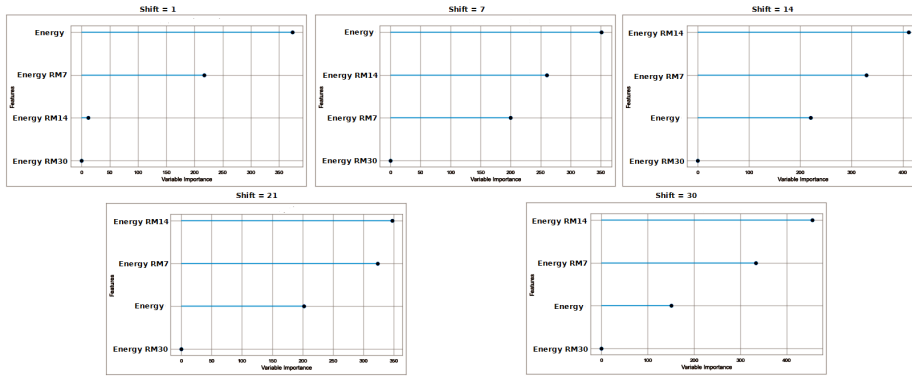
**Table 5.** Friedman Rankings (shift=21)

	Method	MAE	Method	MAPE	Method	MSE
1st	tr+par+gbr	18.625	tr+par+gbr	19	tr+par+gbr	20.5
2nd	par+hub+gbr	18.875	par+hub+gbr	20.125	tr+rsc+omp	20.625
3rd	tr+hub+gbr	20.5	tr+hub+gbr	21.25	par+hub+gbr	22.25
4th	tr+rsc+omp	25.25	tr+rsc+omp	25	tr+hub+gbr	23.875
5th	tr+par	27.625	tr+par	25.875	tr	26.125
	Method	R2	Method	RMSE	Method	RMSLE
1st	tr+par+gbr	109.5	tr+par+gbr	20.5	tr+rsc+omp	20
2nd	tr+rsc+omp	109.375	tr+rsc+omp	20.625	tr+par+gbr	21.625
3rd	par+hub+gbr	107.75	par+hub+gbr	22.25	tr	22.75
4th	tr+hub+gbr	106.125	tr+hub+gbr	23.875	par+hub+gbr	23.625
5th	tr	103.875	tr	26.125	tr+par	24.625

**Table 6.** Friedman Rankings (shift=30)

	Method	MAE	Method	MAPE	Method	MSE
1st	tr+rsc+omp	19	tr+rsc+omp	18.75	tr+rsc+omp	20.25
2nd	hub+rsc+lgbm	22.75	hub+rsc+lgbm	21.75	hub+rsc+lgbm	24.375
3rd	tr+par+rsc	24.5	tr+par+rsc	23.625	tr+rsc+llar	25.625
4th	rsc+omp	25.125	tr+rsc	23.75	tr+par+lgbm	26.625
5th	tr+rsc	26.5	tr+par+gbr	25.125	rsc+omp	27.25
	Method	R2	Method	RMSE	Method	RMSLE
1st	tr+rsc+omp	109.75	tr+rsc+omp	20.25	tr+rsc+omp	20.5
2nd	hub+rsc+lgbm	105.625	hub+rsc+lgbm	24.375	hub+rsc+lgbm	23.25
3rd	tr+rsc+llar	104.375	tr+rsc+llar	25.625	tr+rsc+llar	26.25
4th	tr+par+lgbm	103.375	tr+par+lgbm	26.625	tr+par+lgbm	26.375
5th	rsc+omp	102.75	rsc+omp	27.25	tr	27

has come after an extensive comparison, first, of individual algorithms and, subsequently, of their possible groupings. Although there are ensembles that seem to dominate some of the scenarios, there is still no general predominance of a particular one. However, further investigation of the efficient formations seems to point to a prospect of clear conclusions regarding the proposal of a specific methodology.

**Fig. 2.** Feature Importance per Time-Shift - Data Setup No8

Finally, figure 2 shows indicative representations of the feature importance for each time-frame, regarding the *Theil-Sen* method and over the eighth setup of the ones used, that is, the one that contained as input, in a multivariate layout, all the versions of the modified time series produced. It can be observed that in longer-range forecasts, the predictive importance of the smoothed input time series created by the use of wider rolling averages increases. This indicates that during the forecasting methodology set up, the creation of possible variations of

the data set given, with such appropriate preprocessing, may increase efficiency in the long run.

## 5 Discussion and Future Work

Concluding, in this work, the use of regression ensembles to forecast the Greek energy system load was investigated. Ensembles consisting of 2 or 3 base learners were implemented, forming combinations from a total of 24 regression methods, regarding 5 prediction time-frames. Weighted versions of the raw data were used as input in both univariate and multivariate configurations. Results show that, as the time horizon grows, ensemble methods seem to dominate, implying that, in longer prediction time steps, their use can be beneficial.

The exported results suggest that future extensions of the present work could relate to the use of more complex grouping schemes over even longer forecast time horizons and larger data sets. In addition, the continuation of the experiments towards an extensive presentation of the results creates the possibility of formalizing specific ensemble schemes. Lastly, additional information such as meteorological factors or season regularities could be used in such weighted multivariate settings to improve the predictive effectiveness of the models.

## References

1. Bianchi, F.M., De Santis, E., Rizzi, A., Sadeghian, A.: Short-Term Electric Load Forecasting Using Echo State Networks and PCA Decomposition. *IEEE Access*. 3, 1931–1943 (2015). <https://doi.org/10.1109/ACCESS.2015.2485943>
2. Zhang, J.: Research on power load forecasting based on the improved elman neural network. *Chem. Eng. Trans.* 51, 589–594 (2016). <https://doi.org/10.3303/CET1651099>
3. Filatova, E.S., Filatov, D.M., Stotckaia, A.D., Dubrovskiy, G.: Time series dynamics representation model of power consumption in electric load forecasting system. *Proc. 2015 IEEE North West Russ. Sect. Young Res. Electr. Electron. Eng. Conf. ElConRusNW 2015*. 175–179 (2015). <https://doi.org/10.1109/ElConRusNW.2015.7102256>
4. Khamaira, M.Y., Krzma, A.S., Alnass, A.M.: Long Term Peak Load Forecasting For The Libyan Network. 185–193 (2018). <https://doi.org/10.21467/proceedings.2.23>
5. Saber, A.Y., Alam, A.K.M.R.: Short term load forecasting using multiple linear regression for big data. *2017 IEEE Symp. Ser. Comput. Intell. SSCI 2017 - Proc. 2018-Janua*, 1–6 (2018). <https://doi.org/10.1109/SSCI.2017.8285261>
6. Clarke, S.M., Griebisch, J.H., Simpson, T.W.: Analysis of support vector regression for approximation of complex engineering analyses. *J. Mech. Des. Trans. ASME*. 127, 1077–1087 (2005). <https://doi.org/10.1115/1.1897403>
7. Azad, M.K., Uddin, S., Takruri, M.: Support vector regression based electricity peak load forecasting. *11th Int. Symp. Mechatronics its Appl. ISMA 2018. 2018-Janua*, 1–5 (2018). <https://doi.org/10.1109/ISMA.2018.8330143>
8. Ghelardoni, L., Ghio, A., Anguita, D.: Energy load forecasting using empirical mode decomposition and support vector regression. *IEEE Trans. Smart Grid*. 4, 549–556 (2013). <https://doi.org/10.1109/TSG.2012.2235089>

9. Maldonado, S., González, A., Crone, S.: Automatic time series analysis for electric load forecasting via support vector regression. *Appl. Soft Comput. J.* 83, 105616 (2019). <https://doi.org/10.1016/j.asoc.2019.105616>
10. Sehovac, L., Grolinger, K.: Deep Learning for Load Forecasting: Sequence to Sequence Recurrent Neural Networks with Attention. *IEEE Access.* 8, 36411–36426 (2020). <https://doi.org/10.1109/ACCESS.2020.2975738>
11. Bouktif, S., Fiaz, A., Ouni, A., Serhani, M.A.: Multi-Sequence LSTM-RNN Deep Learning and Metaheuristics for Electric Load Forecasting. *Energies.* 3, 1–21 (2020)
12. Din, G.M.U., Marnerides, A.K.: Short term power load forecasting using Deep Neural Networks. *2017 Int. Conf. Comput. Netw. Commun. ICNC 2017.* 594–598 (2017). <https://doi.org/10.1109/ICCNC.2017.7876196>
13. Ryu, S., Noh, J., Kim, H.: Deep neural network based demand side short term load forecasting. *Energies.* 10, 1–20 (2017). <https://doi.org/10.3390/en10010003>
14. He, W.: Load Forecasting via Deep Neural Networks. *Procedia Comput. Sci.* 122, 308–314 (2017). <https://doi.org/10.1016/j.procs.2017.11.374>
15. Kumar, S., Hussain, L., Banarjee, S., Reza, M.: Energy Load Forecasting using Deep Learning Approach-LSTM and GRU in Spark Cluster. *Proc. 5th Int. Conf. Emerg. Appl. Inf. Technol. EAIT 2018.* 1–4 (2018). <https://doi.org/10.1109/EAIT.2018.8470406>
16. Mubashar, R., Javed Awan, M., Ahsan, M., Yasin, A., Partab Singh, V.: Efficient residential load forecasting using deep learning approach. *Int. J. Comput. Appl. Technol.* (2021)
17. Almalaq, A., Edwards, G.: A review of deep learning methods applied on load forecasting. *Proc. - 16th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2017.* 2017-Decem, 511–516 (2017). <https://doi.org/10.1109/ICMLA.2017.0-110>
18. Aslam, S., Herodotou, H., Mohsin, S.M., Javaid, N., Ashraf, N., Aslam, S.: A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. *Renew. Sustain. Energy Rev.* 144, 110992 (2021). <https://doi.org/10.1016/j.rser.2021.110992>
19. Hammad, M.A., Jereb, B., Rosi, B., Dragan, D.: Methods and Models for Electric Load Forecasting: A Comprehensive Review. *Logist. Sustain. Transp.* 11, 51–76 (2020). <https://doi.org/10.2478/jlst-2020-0004>
20. Shabbir, N., Ahmadiyahangar, R., Kutt, L., Rosin, A.: Comparison of machine learning based methods for residential load forecasting. *2019 Electr. Power Qual. Supply Reliab. Conf. 2019 Symp. Electr. Eng. Mechatronics, PQ SEEM 2019.* 1–4 (2019). <https://doi.org/10.1109/PQ.2019.8818267>
21. Bouktif, S., Fiaz, A., Ouni, A., Serhani, M.A.: Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies.* 11, (2018). <https://doi.org/10.3390/en11071636>
22. Dudek, G.: Neural networks for pattern-based short-term load forecasting: A comparative study. *Neurocomputing.* 205, 64–74 (2016). <https://doi.org/10.1016/j.neucom.2016.04.021>
23. García-pedrajas, N., Hervás-martínez, C., Ortiz-boyer, D.: Cooperative Coevolution of Artificial Neural Network Ensembles for Pattern Classification. 9, 271–302 (2005)
24. Cuncheva, L., Whitaker, C.: Measures of Diversity in Classifier Ensembles. *Mach. Learn.* 51, 181–207 (2003). <https://doi.org/10.1049/ic:20010105>
25. Chandra, A., Chen, H., Yao, X.: Trade-Off Between Diversity and Accuracy in Ensemble Generation. 464, 429–464

26. Tumer, K., Ghosh, J.: Analysis on Decision Boundaries in Linearly Combined Neural Classifiers. *Pattern Recognit.* 29, 341–348 (1996)
27. Liapis, C.M., Karanikola, A., Kotsiantis, S.: An ensemble forecasting method using univariate time series COVID-19 data. *ACM Int. Conf. Proceeding Ser.* 50–52 (2020). <https://doi.org/10.1145/3437120.3437273>
28. Nti, I.K., Adekoya, A.F., Weyori, B.A.: A comprehensive evaluation of ensemble learning for stock - market prediction. *J. Big Data.* (2020). <https://doi.org/10.1186/s40537-020-00299-5>
29. Kankanala, P., Member, S., Das, S., Pahwa, A.: ADABOOST+: An Ensemble Learning Approach for Estimating Weather-Related Outages in Distribution Systems. *IEEE Trans. Power Syst.* 29, 359–367 (2014)
30. Mendes-Moreira, J., Soares, C., Alipio, M.J., De Sousa, J.F.: Ensemble Approaches for Regression: A Survey. *ACM Comput. Surv.* 45, (2012). <https://doi.org/10.1145/2379776.2379786>
31. Wang, L., Mao, S., Wilamowski, B.M., Nelms, R.M.: Ensemble Learning for Load Forecasting. *IEEE Trans. Green Commun. Netw.* 4, 616–628 (2020). <https://doi.org/10.1109/TGCN.2020.2987304>
32. Tang, L., Yi, Y., Peng, Y.: An ensemble deep learning model for short-term load forecasting based on ARIMA and LSTM. 2019 IEEE Int. Conf. Commun. Control. Comput. Technol. Smart Grids, SmartGridComm 2019. 1–6 (2019). <https://doi.org/10.1109/SmartGridComm.2019.8909756>
33. Li, J., Deng, D., Zhao, J., Cai, D., Hu, W., Zhang, M., Huang, Q.: A Novel Hybrid Short-Term Load Forecasting Method of Smart Grid Using MLR and LSTM Neural Network. *IEEE Trans. Ind. Informatics.* 17, 2443–2452 (2021). <https://doi.org/10.1109/TII.2020.3000184>
34. Li, Y., Che, J., Yang, Y.: Subsampled support vector regression ensemble for short term electric load forecasting. *Energy.* 164, 160–170 (2018). <https://doi.org/10.1016/j.energy.2018.08.169>
35. Von Krannichfeldt, L., Wang, Y., Hug, G.: Online Ensemble Learning for Load Forecasting. *IEEE Trans. Power Syst.* 36, 545–548 (2021). <https://doi.org/10.1109/TPWRS.2020.3036230>
36. Dudek, G.: Short-term load forecasting using random forests. *Adv. Intell. Syst. Comput.* 323, 821–828 (2015). [https://doi.org/10.1007/978-3-319-11310-4\\_71](https://doi.org/10.1007/978-3-319-11310-4_71)
37. Papadopoulos, S., Karakatsanis, I.: Short-term electricity load forecasting using time series and ensemble learning methods. 2015 IEEE Power Energy Conf. Illinois, PECEI 2015. 1–6 (2015). <https://doi.org/10.1109/PECEI.2015.7064913>
38. IPTT Energy System Load, [https://www.data.gov.gr/datasets/admie\\_realtimescad\\_asytemload/](https://www.data.gov.gr/datasets/admie_realtimescad_asytemload/).
39. Drucker, H.: Improving regressors using boosting techniques. 14th Int. Conf. Mach. Learn. 107–115 (1997)
40. Wipf, D., Nagarajan, S.: A New View of Automatic Relevance Determination. In: Platt, J., Koller, D., Singer, Y., and Roweis, S. (eds.) *Advances in Neural Information Processing Systems*. Curran Associates, Inc. (2008)
41. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A.: CatBoost: unbiased boosting with categorical features arXiv: 1706 . 09516v5 [ cs . LG ] 20 Jan 2019. 1–23
42. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: *Classification and regression trees*. Routledge (2017)
43. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 67, 301–320 (2005). <https://doi.org/10.1111/j.1467-9868.2005.00503.x>

44. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. *Mach. Learn.* 63, 3–42 (2006). <https://doi.org/10.1007/s10994-006-6226-1>
45. Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., others: Xgboost: extreme gradient boosting. *R Packag. version 0.4-2. 1*, 1–4 (2015)
46. Friedman, J.H.: Greedy function approximation: A gradient boosting machine. *Ann. Stat.* 29, 1189–1232 (2001). <https://doi.org/10.1214/aos/1013203451>
47. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A.: *Robust statistics: the approach based on influence functions*. John Wiley & Sons (2011)
48. Devroye, L., Györfi, L., Krzyżak, A., Lugosi, G.: On the Strong Universal Consistency of Nearest Neighbor Regression Function Estimates. *Ann. Stat.* 22, (2007). <https://doi.org/10.1214/aos/1176325633>
49. Vovk, V.: Kernel Ridge Regression. In: Schölkopf, B., Luo, Z., and Vovk, V. (eds.) *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*. pp. 105–116. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
50. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. *Ann. Stat.* 32, 407–499 (2004). <https://doi.org/10.1214/009053604000000067>
51. Tibshirani, R.: Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B.* 58, 267–288 (1996). <https://doi.org/https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
52. Fan, J., Ma, X., Wu, L., Zhang, F., Yu, X., Zeng, W.: Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. *Agric. Water Manag.* 225, 105758 (2019). <https://doi.org/10.1016/j.agwat.2019.105758>
53. Seber, G.A.F., Lee, A.J.: *Linear regression analysis*. John Wiley & Sons (2012)
54. Murtagh, F.: Multilayer perceptrons for classification and regression. *Neurocomputing.* 2, 183–197 (1991). [https://doi.org/https://doi.org/10.1016/0925-2312\(91\)90023-5](https://doi.org/https://doi.org/10.1016/0925-2312(91)90023-5)
55. Rubinfeld, R., Zibulevsky, M., Elad, M.: Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit. *CS Tech.* 1–15 (2008)
56. Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., Singer, Y.: Online Passive-Aggressive Algorithms. *J. Mach. Learn. Res.* 7, 551–585 (2006)
57. Breiman, L.: Random Forests. *Mach. Learn.* 45, 5–32 (2001). <https://doi.org/10.1017/CBO9781107415324.004>
58. Choi, S., Kim, T., Yu, W.: Performance evaluation of RANSAC family. *Br. Mach. Vis. Conf. BMVC 2009 - Proc.* 1–12 (2009). <https://doi.org/10.5244/C.23.81>
59. Marquardt, D.W., Snee, R.D.: Ridge regression in practice. *Am. Stat.* 29, 3–20 (1975). <https://doi.org/10.1080/00031305.1975.10479105>
60. Smola, A. j., Schölkopf, B.: A tutorial on support vector regression. *Stat. Comput.* 14, 199–222 (2004)
61. Dang, X., Peng, H., Wang, X., Zhang, H.: The Theil-Sen Estimators in a Multiple Linear Regression Model. *Manuscript*, Available URL . . . 1–30 (2009)
62. Ali, M.: PyCaret: An open source, low-code machine learning library in Python, <https://pycaret.org/>.
63. Liapis, C.M., Karanikola, A., Kotsiantis, S.: A Multi-Method Survey on the Use of Sentiment Analysis in Multivariate Financial Time Series Forecasting. *Entropy.* 23, 1603 (2021).
64. Friedman, M.: The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *J. Am. Stat. Assoc.* 32, 675–701 (1937). <https://doi.org/10.1080/01621459.1937.10503522>
65. Dunn, O.J.: Multiple Comparisons Among Means. *J. Am. Stat. Assoc.* 56, 52 (1961). <https://doi.org/10.2307/2282330>