



Hierarchical Clustering and Deep Learning for Short-Term Load Forecasting with Influenced Factors

Rio Indralaksono¹, M. Abdul Wakhid², Novemi Uki A³, Galih Hendra Wibowo⁴,
 M. Abdillah⁵, Agus Budi Rahardjo⁶, Diana Purwitasari^{7*}

^{1,2,3}PLN Indonesia

^{1,2,3}Magister Management Technology, Institut Teknologi Sepuluh Nopember, Indonesia

⁴Department of Informatics Engineering, Polytechnic State Banyuwangi, Indonesia

⁵Departement of Electrical Engineering Pertamina University Jakarta, Indonesia

^{6,7}Department of Informatics, Institut Teknologi Sepuluh Nopember, Indonesia

¹rio@indralaksono.net, ³galih@poliwangi.ac.id, ⁴m.abdillah@universitaspertamina.ac.id, ⁵agus.budi@its.ac.id,

⁵diana@if.its.ac.id (* corresponding author)

Abstract

Stable and reliable electricity is one of the essential things that must be maintained by the transmission system operator (TSO). That can be achieved when the TSO is able to set the balance between demand and production. To maintain the balance between production and demand, TSO should estimate how much demand must be served. In order to do that, the next day short-term load forecasting is an essential step that TSO should be done. Generally, load forecasting can be done through conventional techniques such as least square, time series, etc. However, this method has been sought over time as the electricity demand is increasing significantly over the years. Hence, this paper proposed another approach for short-term load forecasting using Deep Neural Networks, widely known as Long Short-Term Memory (LSTM). In addition, this paper clusters historical electrical loads to obtain similar patterns into several clusters before forecasting. We also explored other influence factors in the observed days, such as weather conditions and the human activity cycle represented by holidays, in a neural network-based classification model to predict the targeted clusters of electrical loads. East Java sub-system is used as the test system to investigate the efficacy of the proposed load forecasting method. From the simulation results, it is found that the proposed method could provide a better forecast on all indicators compared to the conventional method, as indicated by MaxAPE and MAPE are around 4,91% and 2,02%, while the RMSE is 112,08 MW.

Keywords: forecasting, electricity, load, Long Short-term Memory, Analytical Hierarchical Clustering

1. Introduction

In the 4.0 industrial revolution, power systems should be able to provide both stable and reliable electricity. To get stable and reliable electricity Transmission System Operator (TSO) should maintain the balance between electricity production and demand [1]. If TSO is failed to maintain the balance between production and demand, the possibility of the system experiencing unstable conditions is high [1]. In addition, the penalties from the consumers could have happened if the unstable condition emerged in the system. Hence, it is important to forecast the load demand from the customer side so that the TSO can produce the optimal electricity [2]–[4].

In power systems, load forecasting can be divided into several categories depending on how long the prediction takes. Long-term load forecasting is used to

predict the peak load for the next ten years. Medium-term load forecasting is used to predict the load for a month or year period. Short-term load forecasting is used when TSO wants to predict the load demand between half an hour to weekly. While very short-term load forecasting is used when TSO wants to predict the load demand in less than half an hour [5], [6]. Among them, short-term load forecasting is more essential as they play an important role in the energy management system's real-time control and security functions. The accurate short-term load forecasting (STLF) could provide operational cost savings and safe conditions that allow utilities to process production resources to optimize energy prices and trade-offs with producers and consumers. Short-term load forecasts for the next 0.5 - 24 hours are important for the day-to-day operation of electric utilities.

Generally, short-term load forecasting can be solved by using a conventional method such as extrapolation, correlations, least squares, stochastic time series (ARMA and ARIMA), and an expert system. The application of the extrapolation method for short-term load forecasting is reported in [7]. From the results, it was found that extrapolation can be used to predict how much load demand must be handled by TSO. Srinivasan et al. proposed an application of the multiple correlations model for solving short-term load forecasting problems, as reported in [8]. It was found that short-term load forecasting problems can be handled properly by using correlations approaches. The research effort in [9] proposed applications of least square for short-term load forecasting. It was observed that using the concept proposed in [9] the TSO could maintain the balance between production and electricity demand. The application of ARMA model identification for short-term load forecasting is reported in [10]. In [11], non-gaussian process consideration is included in the ARMA model for optimally predicting the load demand of the power systems. From the results, it was found that the method can be used to forecast the load demand in power systems. ARIMA model approach for load forecasting is used in the following paper [12]. While the application of an expert system for short-term load forecasting is reported in [13]. Although the conventional method can be used to solve the short-term load forecasting problem, the concept is out of date in the nowadays condition [14].

Electricity load demand has been increasing significantly over the past few years. In addition, the power system is also undergone significant changes over the years with high penetration of intermittent inverter-based power plants (photovoltaic generation and wind generation [1]). Hence, conventional approaches are not enough to predict the load demand in uncertain conditions. To handle this problem, artificial intelligence approaches are essential to solve short-term load forecasting in modern power systems. Among numerous types of artificial intelligence methods Deep Neural Network or widely known as Long Short-Term Memory (LSTM) model, is becoming favorable. The application of LSTM for predicting the circulation water pump bearing condition is reported in [15]. In [15], a coal fire power plant is used as the test system of water pump bearing conditions. From the results, it is found that LSTM can be used to predict the condition of water pump bearing optimally. Dalgkitsis et al. show the successful application of LSTM for cellular network traffic forecasting, as reported in [16]. The research effort in [17] shows the application of LSTM for predicting the load demand in the residential area. It was noticeable that LSTM can be used as a method for short-term residential load forecasting optimally.

From the literature review, it is noticeable that LSTM is a highly recommended method for short-term load forecasting. However, to make the LSTM perform more accurately, additional approach such as grouping similar load profiles is essential. The application of daily temperature (weather condition) grouping in load forecasting is reported in [18]. From the results, it is noticeable that by grouping the load based on the daily weather condition profiles, the forecasting is more accurate than without grouping. Quilumba et al. proposed a grouping method based on customer behavior for load forecasting, as reported in [19]. To understand the behavior of the customer, the smart meter is used to capture all the important data. From the result, it is observed that by using this concept, load forecasting can be done more accurately. From the above research, it can be concluded that to increase the accuracy of load forecasting, clustering the load is important. Researchers in [20] proposed an optimal method for clustering data using Analytical Hierarchy Clustering (AHC) method. It is noticeable that by using this method, the data can be optimally clustered.

This paper proposed an application of long short-term memory for short-term load forecasting. To get optimal and accurate results analytical hierarchy clustering method is also considered in the research. The rest of the paper is organized as follows: Chapter 2 focuses on the research method of the paper. Results and discussions are thoroughly presented in Chapter 3. Chapter 4 highlighted the conclusions and future works of this research.

2. Research Methods

2.1 Analytical Hierarchical Clustering

Agglomerative AHC is a clustering method that is carried out on a bottom-up basis by combining a number of scattered data into a cluster. The AHC method uses several choices of algorithms in performing clustering, namely single linkage, complete linkage, and average linkage [20], [21].

Single linkage is an algorithm to combine two clusters (e.g., clusters A and B) into one cluster (AB) by considering the smallest distance between clusters, denoted by $D = \{d_{AB}\}$. Calculation of the distance between clusters can use a variety of methods but generally use the excluded distance. Next, the distance between clusters is calculated, including the new cluster AB and chooses the shortest distance. The distance calculation process is repeated until the desired number of clusters or the highest hierarchical tree is obtained, namely one cluster. Single linkage can be denoted as in Eq. 1.

A complete linkage is a grouping approach that is carried out by calculating the largest distance between data or clusters. The process begins by determining the

distance between the largest clusters $D = \{ij\}$, then they are combined into one cluster (e.g., clusters A and B become clusters AB). The distance calculation process is repeated with a new cluster (cluster AB), and the largest distance is recalculated to be a candidate for a new cluster. This process is repeated until the desired number of clusters or the highest hierarchical tree is obtained, namely one cluster. Complete linkage can be denoted as in Eq.2.

$$d(AB)C = \min(dAC, dBC) \quad (1)$$

$$d(AB)C = \max(dAC, dBC) \quad (2)$$

In determining distance, the commonly used calculations include Euclidian, Minkowski/Manhattan, Mahalanobis, etc. In this study, the determination of distance using the Euclidian method was chosen so that the mapping of an object was carried out for homogeneity.

2.2 Neural Network Classification

Neural Network is one approach to classifying data. The basic idea of this method comes from the ability of the human brain to determine or make a decision. Based on this, the architectural form in this method consists of an input layer, process layer, and output layer. The use of this method varies greatly, as the number of layers used will be greatly influenced by the level of complexity of the data or the problem to be solved. The higher the level of complexity of the problem, the more layers used. This layer is supported by several processes in it, such as the initialization process, activation, weight training, and the number of iterations. The initialization process is the process of determining the data into neurons in the input layer (X_n). This process also defines the number of iterations and the initial weight assigned to each data or neuron to another layer ($W_1, W_2, \dots W_n$). The activation process will play a role in connecting neurons and layers to the maximum limit of iterations or the problem is solved, where the activation function can be written $Y=f(\text{net})$ and $\text{net} = X_1W_1 + X_2W_2 + \dots + X_nW_n$ [22], [23].

2.3 Forecasting Modelling

Long Short-Term Memory (LSTM) is a Deep Learning Recurrent Neural Network (DLRNN) development. In general, this method is used to predict Sequential Values or forecasting. This method is applied in many fields such as cellular network [16], power generation [24], bearing conditions [15], Load Forecasting [6], [25], etc. LSTM was developed to solve the RNN problem [26], namely the loss of important information at the beginning of knowledge; if the sequence processed is long enough (forward propagation) and the gradient vanishing problem (backward propagation), the gradient value is very small, so it does not contribute to changes in weight.

The basic concept of LSTM is to combine new knowledge (Short Term Memory) with old knowledge (Long Term Memory) by deleting or updating that knowledge. Which consists of a cell state and three main gates: forget gate, input gate, and output gate. Cell state provides comprehensive information (from the beginning of the data to the end) at the stage of model formation. The information on the cell state will be influenced by the existing gates, such as forget gates and input gates. Forget gate aims to eliminate information that is less relevant to the new input data. The checking process uses a sigmoid function from the previous output data with the current data to produce values ranging from zero to one.

This process is continued by performing multiplication operations of new knowledge with old knowledge. The closer to zero, the information will be forgotten and vice versa. Input Gate aims to update the state of the Cell, whether the existing information values need to be updated or maintained. This gate is started by performing sigmoid and tan-h operations on the previous and current output data. The results of the two operations are then processed by multiplying them. The results of the sigmoid operation on this gate will determine which new information from the operation is important to store. After the input gate processing results are obtained, it is added to the Forget Gate processing results to get the latest information from the inputs. This result will be used to get the Output Gate value by operating the input data with the latest knowledge. Output Gate results will be stored and used for processing to the next Cell as data from the previous output [6], [27].

3. Results and Discussions

Section 1 explains that this study uses two reference datasets to obtain the predicted value of the Electrical Load to increase the efficiency of the use of the required resources. There are two types of these datasets, the electric Load dataset (Conventional) that has been obtained and the electrical Load dataset supported by influence factors such as weather conditions and human activity that is preceded by the clustering process (Proposed). In general, both approaches will be processed using LSTM.

The data used in this study consists of Electricity Load data, which is load data for three years every 30 minutes, as shown in Fig. 1, and Influenced Factor data consisting of weather data and human activity data represented by day-off / holiday data. The electricity load data is data on the East Java sub-system obtained from the PLN Group. Weather data used as features in this study include Dew Point, Temperature, Humidity, Wind Speed, and Pressure. The weather data was obtained from 6 Amateur Weather Stations (AWS) spread over the province, which were obtained from the

website wunderground.com. While the human activity cycle data that represent by weekday/weekend, national holidays, or day-off between holidays. These data were obtained from the ministry's circular. The data used in this study is historical data from January 1, 2019, to February 14, 2022. From the data, it will be divided into two parts, January 1, 2019, to February 7, 2022, as testing data, and February 8 – 14, 2022, as testing data.

3.1. Conventional Method

The first form of prediction is done by using electrical load data that has been collected. The dataset has numerical features and has a very different minimum and maximum range. This condition makes the modeling process cannot be carried out optimally. In this study, min-max normalization was carried out. Normalization is done by setting a min-max limit with values that have occurred in each type of feature and rounded up for max and down for a min. Based on this normalization processing, the Minimum value in the Electrical Load data is set to 2500, and the Maximum value is set to 6100. Furthermore, the data is used to predict the Electricity Load for the next seven days using LSTM.

In this study, two types of datasets are used, so it is necessary to determine the architecture used as a measurable form of comparison. In the first dataset, 18 types of LSTM architectural modeling were carried out to get the best LSTM architecture by testing ten times in each architectural model with 1000 iterations. The 18 types of architectural models used are presented in Table 1. there is the combination of the LSTM layer with various numbers of hidden units dan the Dropout Layer coefficient. The best architectural modeling value is determined by calculating the Maximum Absolute Percentage Error (MaxAPE) value from each test and accumulated by the average value approach. Forecasting is done using limited training data, namely four days or 4 x 48 data, to predict the next one day or 1 x 48 data, which is sufficient according to the author's previous research [6], [25]

The results of the LSTM architectural modeling test are presented in Table 2. on these results, it was found that the best LSTM architecture model lies in model number 9, with the smallest MaxAPE average value of 7,061. This is supported by the relatively small standard deviation value so that when combined with the MaxAPE average value obtained, model number 9 still shows the best MaxAPE average performance value. The selected architectural model is presented in Table 1. with a number of layers of 7.

The following process is to test using data testing. Testing data is obtained by taking part of the last data from the dataset, which is seven days, and the rest as training data. Tests are carried out using a predetermined architectural model and calculate the

performance of the predictions through the values of MaxAPE, MAPE, and RMSE. The test results are presented in Table 3. in the Conventional section. From the test results, it was found that the first dataset from Electrical Load alone can produce a MaxAPE value of 19.97%, MAPE of 9%, and RMSE of 492.38. These results will then be compared with the second approach.

3.2. Proposed Method

In the second dataset, the electrical load's value prediction is compiled with influenced factors such as weather data and holiday data. Each data is pre-processed according to the characteristics of the data. Similar to the conventional method, the electrical load data is normalized using Min-Max. As for the weather data, it was found that the data that had been obtained still contained missing values. The missing value is solved by using the average value from 6 different AWS datasets. Then normalization is performed using the Min-Max method. While the human activity data, normalization is not carried out because the data is categorical.

The Training dataset will be processed according to the proposed method as described in Fig. 2 Electrical load data clustering is needed because the value of the electrical load has a different pattern on the types of the date that exist. The training dataset for electricity load that has been normalized will be clustered using AHC. The parameter used is Complete Linkage with cluster selection based on the height of the hierarchy. In this study, the cluster selection was determined between 5 and 10, as shown in Fig. 3. From the Clustering process, label cluster data will be obtained for each row in the Date feature and the Total Demand feature, and then this dataset will be called electricity load with a label. Then the dataset is modeled based on each label using the LSTM method to create a Prediction Model for each label. Forecasting using the Prediction model is carried out to estimate the load of 48 steps every day for a week.

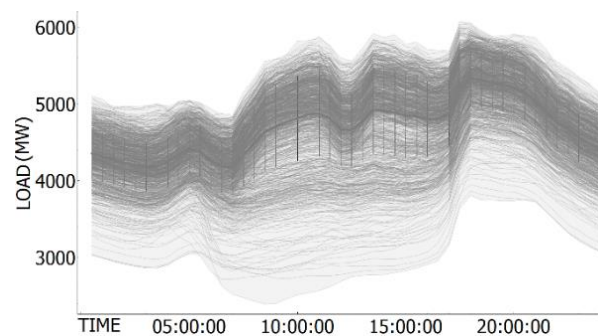


Fig. 1. Electricity Load From 2019 to Feb 2022

Table 1. Architectural Models for Mini Forecasting

MODEL	1	2	3	4	5	6	7	8	9	
LAYER ARCHITECTURE	LSTM : 250 DROPOUT : 0.25 LSTM : 300	LSTM : 250 DROPOUT : 0.1 LSTM : 300	LSTM : 25 DROPOUT : 0.25 LSTM : 30	LSTM : 25 DROPOUT : 0.1 LSTM : 30	LSTM : 25 DROPOUT : 0.25 LSTM : 30 DROPOUT : 0.25 LSTM : 30 DROPOUT : 0.25 LSTM : 35 DROPOUT : 0.25 LSTM : 40	LSTM : 250 DROPOUT : 0.25 LSTM : 300 DROPOUT : 0.25 LSTM : 350 DROPOUT : 0.25	LSTM : 250 DROPOUT : 0.25 LSTM : 300 DROPOUT : 0.25 LSTM : 350	LSTM : 250 LSTM : 300 LSTM : 50	LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50	
MODEL	10	11	12	13	14	17	15	16	17	18
LAYER ARCHITECTURE	DROPOUT : 0.25 LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50	LSTM : 250 LSTM : 250 LSTM : 250 LSTM : 250 LSTM : 250 LSTM : 250 LSTM : 250 LSTM : 250	LSTM : 250 LSTM : 300 LSTM : 350 LSTM : 300 LSTM : 250 LSTM : 250 LSTM : 50 LSTM : 25	LSTM : 400 LSTM : 350 LSTM : 300 LSTM : 250 LSTM : 150 LSTM : 100 LSTM : 50	LSTM : 250 LSTM : 300 LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50	LSTM : 250 LSTM : 300 LSTM : 350 LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50	LSTM : 100 LSTM : 100 LSTM : 100 LSTM : 100 LSTM : 100 LSTM : 100 LSTM : 100	LSTM : 75 LSTM : 75 LSTM : 75 LSTM : 75 LSTM : 75 LSTM : 75 LSTM : 75	LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50	LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50 LSTM : 50

Table 2. Simulation Results Based on MaxAPE for 10 Trial Each Model

Model	Average	Min	Max	Std Dev
1	11.4512	8.5338	13.7522	1.6445
2	11.4491	8.8863	13.8878	1.6664
3	11.5866	7.9803	14.2476	2.2212
4	12.9731	10.0216	17.9174	2.3707
5	17.4180	10.3920	24.9135	4.7248
6	13.9639	12.6002	15.3541	0.7706
7	12.2990	9.1160	14.2172	1.6730
8	12.9774	11.5877	13.9090	0.6462
9	7.0610	4.2605	14.0007	2.7399
10	14.0726	7.8484	21.5523	4.1185
10	11.9959	5.3182	21.7351	5.6537
12	12.5055	5.2569	21.8209	5.4497
13	14.0392	4.0765	27.5914	8.5502
14	10.9924	4.1521	20.9886	6.5873
15	21.0353	7.8932	43.0616	9.6519
16	12.4159	5.0299	24.0205	7.0228
17	8.9724	5.1829	18.9552	4.3161
18	8.7731	5.0759	15.4611	3.5577

Table 3. Simulation Results Based on Performance Indicators

Day -		1	2	3	4	5	6	7	Mean	Min	Max
COVEN TIONAL	MaxAPE	13.79	11.64	8.29	23.41	18.64	39.29	24.71	19.97	8.29	39.29
	MAPE	6.65	4.62	2.76	3.43	10.70	18.15	16.67	9.00	2.76	18.15
	RMSE	378.11	300.51	160.50	260.44	558.82	907.43	880.82	492.38	160.50	907.43
PROPO SED	MaxAPE	5.35	4.52	4.09	4.60	5.31	5.9333	4.56	4.91	4.09	5.93
	MAPE	2.10	2.16	1.62	1.78	1.67	3.3367	1.49	2.02	1.49	3.34
	RMSE	123.58	122.02	90.78	103.70	99.51	155.59	89.39	112.08	89.39	155.59

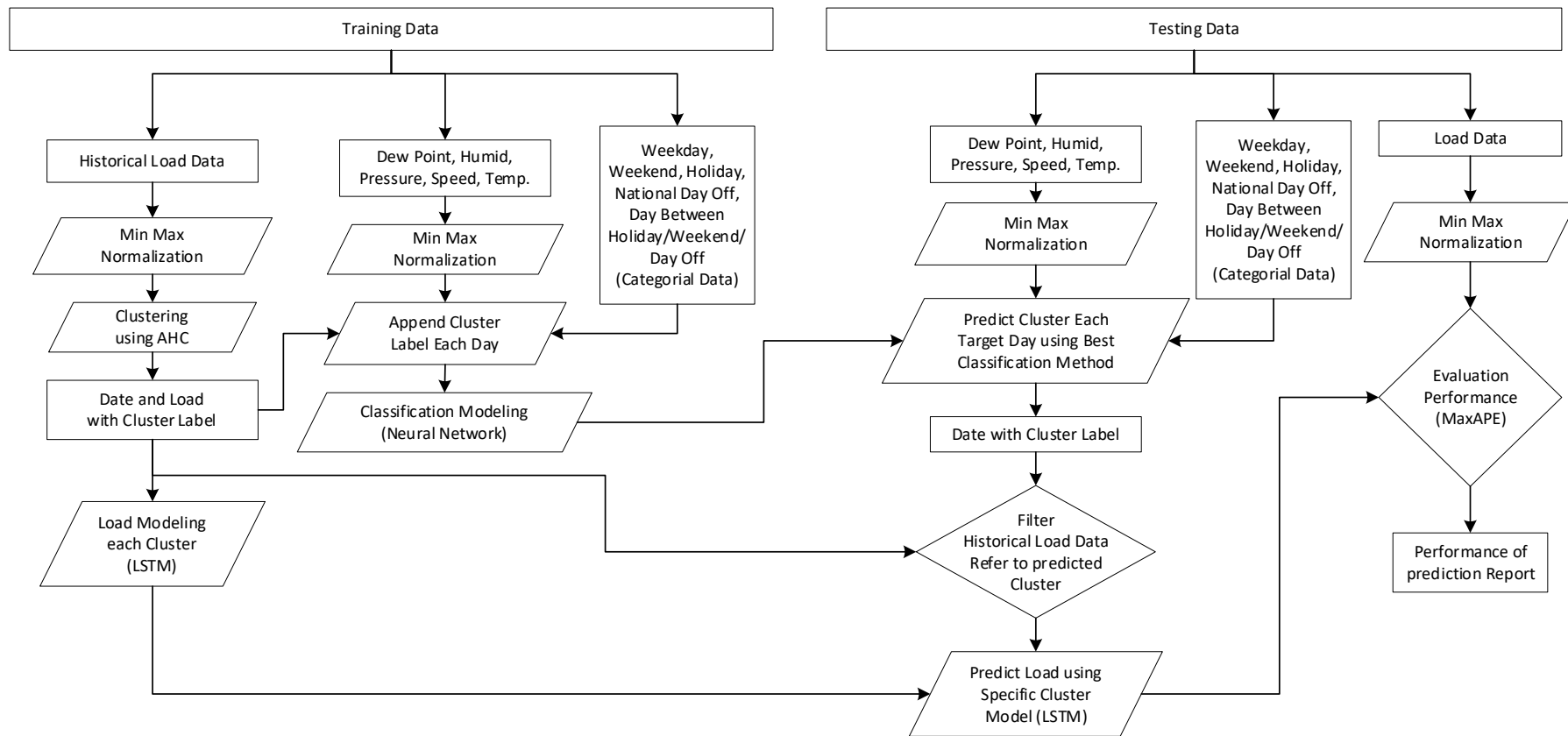


Fig. 2. Proposed Method Flowchart

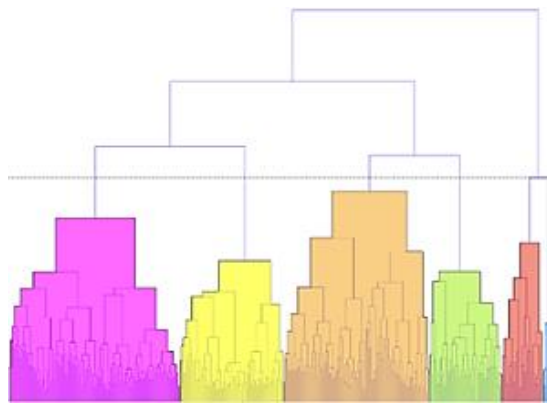


Fig. 3. AHC using Orange Data Mining

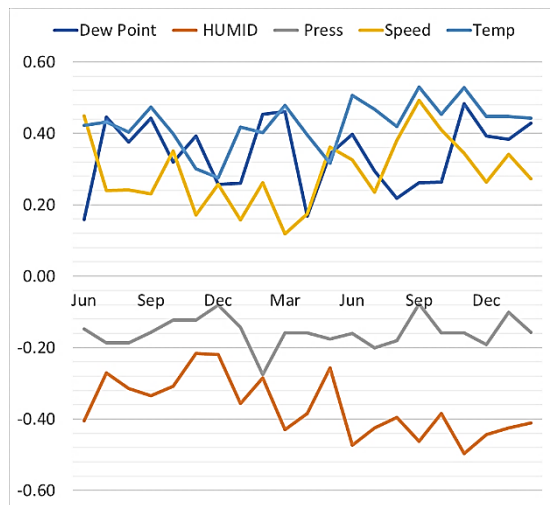


Fig. 4. Rank Spearman Correlation of Weather Dataset

The cluster label for each date will be appended to Influenced factor dataset. The added label is the Target Feature of the following classification process. Because the data used are data from different fields, further analysis is needed regarding the correlation value to the electrical load data. Calculation of correlation in this study is done using the Rank Spearman method because the data distribution does not follow the normal distribution. Based on Fig. 4 and Fig. 5, it is known that the results of the merging of these features have a correlation value that is sufficient to affect the Electrical Load but not strong enough to have a direct influence, so the authors assume that these features still cannot be used as a direct predictor of Electricity Load, so further analysis or processing was needed.

For the stage of processing the forecasting, using influenced factor dataset, the classification process of the testing dataset is carried out using the classification model that will be done in advance to get a label on each date in the testing dataset. As discussed previously, the expected target group or cluster is 5 to 10 clusters. Determination of the optimal cluster is done by comparing the results of the classification of the testing data with the actual data.

Table 4. Clustering Results Based on Influenced Factor

CLUSTER	Model	AUC	Precision
C05	kNN	0.583	0.714
	Neural Network	0.999	0.999
	SVM	0.999	0.857
C06	kNN	0.583	0.714
	Neural Network	0.999	0.999
	SVM	0.999	0.857
C07	kNN	0.501	0.735
	Neural Network	0.917	0.857
	SVM	0.917	0.857
C08	kNN	0.545	0.51
	Neural Network	0.999	0.857
	SVM	0.999	0.595
C09	kNN	0.545	0.51
	Neural Network	0.999	0.857
	SVM	0.999	0.595
C10	kNN	0.545	0.51
	Neural Network	0.909	0.857
	SVM	0.999	0.595

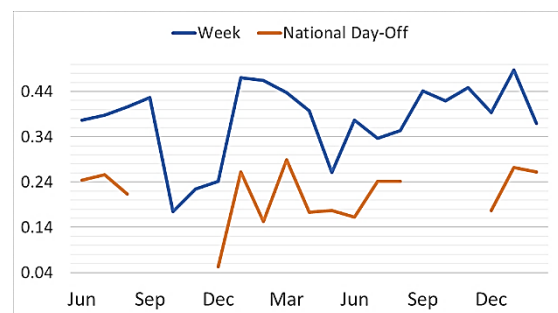


Fig. 5. Rank Spearman Correlation of Human Activity Dataset

Furthermore, the ability of the Influenced factor as a predictor in the classification process was tested using several methods, as shown in Table 4. This table shows performance measurements from the kNN, Neural Network, and SVM methods. In this study, the classification simulation was carried out using Orange Data Mining software using the default parameters for each method. This aims to obtain the most optimal model for determining the electrical load cluster. The test results in Table 1 show that the highest number of the cluster that can be classified is cluster number 6 as illustrated in Fig. 6. It is also shown that the best classification method is Neural Network, according to AUC and Precision performance.

The forecasting is carried out using the LSTM method, assuming that the predicted day's Influenced Factor data is accurate. Based on this data, cluster classification is carried out on the predicted day. Then forecasting is done using historical data from the previous 100 days, which have the same type of cluster. The simulation results by comparing the conventional method and the proposed method are shown in Fig. 7 and Fig. 8, and the details of the comparison of performance indicators are shown in Table 3. Based on these results, the MaxAPE average value is 4.56%, MAPE 2.02%, and RMSE 112.08 MW.

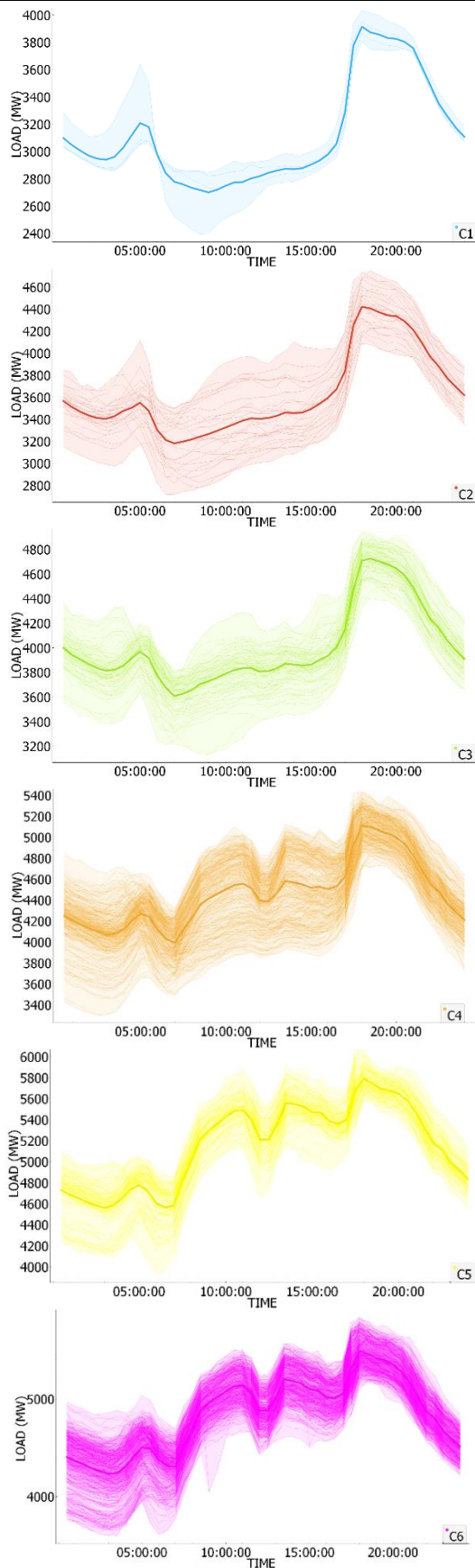


Fig. 6. Line Plot of Clustering Results

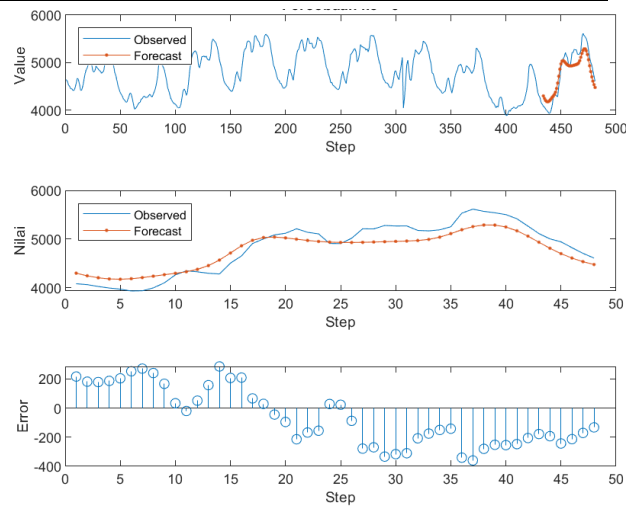


Fig. 7. Forecasting Without Clustering

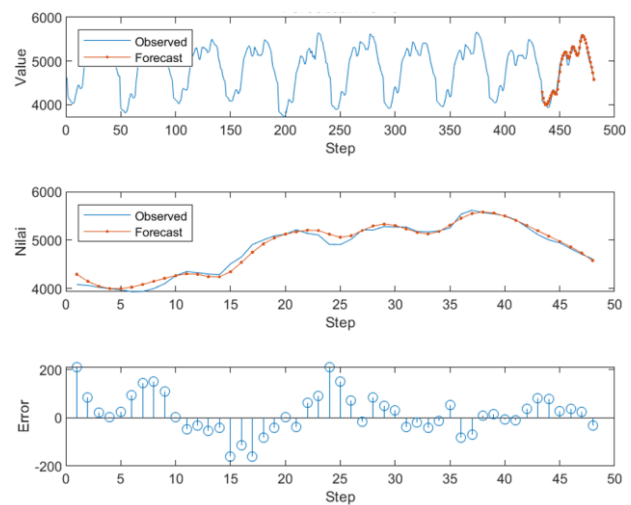


Fig. 8. Forecasting With Clustering

These data indicate that the average MaxAPE value in the testing data is better than using the conventional method containing only electrical loads data. It can be seen that the value of the difference, which is quite significant with the proposed method, is $\pm 15\%$, while the MAPE value is $\pm 7\%$. This difference will certainly significantly affect the efficiency of supplying the amount of material and power that needs to be prepared to accurately produce electricity regarding customer's demand. In addition, based on the test results obtained the fact that the electrical load data will be more optimal to predict by involving several influenced factors such as weather and holidays. This is evidenced by the proposed method, combining clustering as a reference medium for prediction, which is able to show more optimal prediction results with an optimization value of $\pm 15\%$.

4. Conclusion

This paper proposed another approach method for short-term load forecasting based on the Deep Neural

Network or widely known as the Long Short-Term Memory approach. To get optimal and accurate research, the Analytical Hierarchy Clustering method is used to cluster the load data. East-Java sub-system is used as the test system to observe the efficacy of the proposed method. From the simulation results, it is found that by using the proposed method, the load demand can be predicted more accurately than using

conventional approaches. This can be a seed by the average value of MaxAPE up to 4.91%. While the MAPE and RMSE are around 2.02% and 112.08%. Further research can be carried out by implementing the method to predict solar irradiance and wind speed for PV and wind power systems. A hybrid fuzzy and deep learning method can also be considered for future research.

Reference

- [1] H. Setiadi *et al.*, "Multi-Mode Damping Control Approach for the Optimal Resilience of Renewable-Rich Power Systems," *Energies (Basel)*, vol. 15, no. 9, May 2022, doi: <https://doi.org/10.3390/en15092972>.
- [2] R. M. Bamatraf, R. Indralaksono, and I. Robandi, "Application of Craziness Particle Swarm Optimization for Optimal Unit Commitment," in *International Student Conference on Advanced Science and Technology (ICAST)*, 2009, pp. 9–16.
- [3] R. Indralaksono, R. M. Bamatraf, and I. Robandi, "Craziness Particle Swarm Optimization (CRPSO) Solution for Optimal Economic Dispatch," in *Proceedings of National Seminar on Applied Technology, Science, and Arts (APTECS)*, 2009, pp. 2086–2091.
- [4] R. Indralaksono, S. Dwiratha, A. M. Benie Zakariya, F. Ulfi, M. A. Fassamasi, and I. Robandi, "Optimal Management Decision for Large Scale Economic Dispatch using Artificial Immune System Clonal Selection Algorithm (AISCAS)," in *International Student Conference on Advanced Science and Technology (ICAST)*, 2010, pp. 1–1.
- [5] A. Sallam and O. Malik, "Load Forecasting," in *Electric Distribution Systems, Second Edition*, Second., Hoboken, NJ, USA: John Wiley & Sons, Inc., 2018, pp. 41–71. doi: <https://doi.org/10.1002/9781119509332.ch4>.
- [6] R. Indralaksono, A. M. Firdaus, M. Abdul Wakhid, N. U. Andreas, G. H. Wibowo, and M. Abdillah, "Short-Term Load Forecasting using Long Short-Term Memory Network on Various Sub System Load in PLN Indonesia," in *ICT-PEP 2021 - International Conference on Technology and Policy in Energy and Electric Power: Emerging Energy Sustainability, Smart Grid, and Microgrid Technologies for Future Power System, Proceedings*, 2021, pp. 175–179. doi: <https://doi.org/10.1109/ICT-PEP53949.2021.9601005>.
- [7] J. Wang, R. Jia, W. Zhao, J. Wu, and Y. Dong, "Application of the largest Lyapunov exponent and non-linear fractal extrapolation algorithm to short-term load forecasting," *Chaos, Solitons and Fractals*, vol. 45, no. 9–10, pp. 1277–1287, 2012, doi: <https://doi.org/10.1016/j.chaos.2012.06.009>.
- [8] K. Srinivasan and R. Pronovost, "Short term load forecasting using multiple correlation models," *IEEE Transactions on Power Apparatus and Systems*, vol. 94, no. 5, pp. 1854–1858, 1975, doi: <https://doi.org/10.1109/T-PAS.1975.32031>.
- [9] J. Wu and D. Niu, "Short-term power load forecasting using Least Squares Support Vector Machines (LS-SVM)," *2nd International Workshop on Computer Science and Engineering, WCSE 2009*, vol. 1, pp. 246–250, 2009, doi: <https://doi.org/10.1109/WCSE.2009.663>.
- [10] S. J. Huang and K. R. Shih, "Short-term load forecasting via ARMA model identification including non-Gaussian process considerations," *IEEE Transactions on Power Systems*, vol. 18, no. 2, pp. 673–679, May 2003, doi: <https://doi.org/10.1109/TPWRS.2003.811010>.
- [11] Shyh-Jier Huang and Kuang-Rong Shih, "Short-term load forecasting via ARMA model identification including non-gaussian process considerations," *IEEE Transactions on Power Systems*, vol. 18, no. 2, pp. 673–679, May 2003, doi: <https://doi.org/10.1109/TPWRS.2003.811010>.
- [12] S. Karthika, V. Margaret, and K. Balaraman, "Hybrid short term load forecasting using ARIMA-SVM," *2017 Innovations in Power and Advanced Computing Technologies, i-PACT 2017*, vol. 2017-January, pp. 1–7, 2017, doi: <https://doi.org/10.1109/IPACT.2017.8245060>.
- [13] K.-L. Ho, Y.-Y. Hsu, C.-F. Chen, and T.-E. Lee, "Short-term Load Forecasting of Taiwan Power System using a Knowledge-based Expert System," *IEEE Transactions on Power Systems*, vol. 5, no. 4, 1990, doi: <https://doi.org/10.1109/59.99372>.
- [14] E. E. Elattar, N. A. Sabiha, M. Alsharef, M. K. Metwally, A. M. Abd-Elhady, and I. B. M. Taha, "Short term electric load forecasting using hybrid algorithm for smart cities," *Applied Intelligence*, vol. 50, no. 10, pp. 3379–3399, Oct. 2020, doi: <https://doi.org/10.1007/s10489-020-01728-x>.
- [15] M. K. Wisyaldin, G. M. Luciana, and H. Pariaman, "Using LSTM network to predict circulating water pump bearing condition on coal fired power plant," in *Proceeding - 2nd International Conference on Technology and Policy in Electric Power and Energy, ICT-PEP 2020*, Sep. 2020, pp. 54–59. doi: <https://doi.org/10.1109/ICT-PEP50916.2020.9249905>.
- [16] A. Dalgitsis, M. Louta, and G. T. Karetos, "Traffic forecasting in cellular networks using the LSTM RNN," in *ACM International Conference Proceeding Series*, Nov. 2018, pp. 28–33. doi: <https://doi.org/10.1145/3291533.3291540>.
- [17] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019, doi: <https://doi.org/10.1109/TSG.2017.2753802>.
- [18] K. E. Farfar and M. T. Khadir, "A two-stage short-term load forecasting approach using temperature daily profiles estimation," *Neural Computing and Applications*, vol. 31, no. 8, pp. 3909–3919, Aug. 2019, doi: <https://doi.org/10.1007/s00521-017-3324-x>.
- [19] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 911–918, Mar. 2015, doi: <https://doi.org/10.1109/TSG.2014.2364233>.
- [20] A. T. R. Dani, S. Wahyuningsih, and N. A. Rizki, "Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu," *Jambura Journal of Mathematics*, vol. 1, no. 2, pp. 64–78, Jul. 2019, doi: <https://doi.org/10.34312/jjom.v1i2.2354>.
- [21] D. T. Utari and D. S. Hanun, "Hierarchical Clustering Approach for Region Analysis of Contraceptive Users," *EKSAKTA: Journal of Sciences and Data Analysis*, pp. 99–108, Sep. 2021, doi: <https://doi.org/10.20885/EKSAKTA.vol2.iss2.art3>.
- [22] A. G. Farizawani, M. Puteh, Y. Marina, and A. Rivaie, "A review of artificial neural network learning rule based on multiple variant of conjugate gradient approaches," in *Journal of Physics: Conference Series*, Jun. 2020, vol. 1529, no. 2. doi: https://ui.adsabs.harvard.edu/link_gateway/2020JPhCS1529b2040F/doi:10.1088/1742-6596/1529/2/022040.
- [23] B. Handoko, R. M. Bamatraf, R. Indralaksono, M. Abdillah, and I. Robandi, "Short-term Load Forecasting using Hybrid Artificial Neural Networks and Particle Swarm Optimization," in *Proceedings of National Seminar on Applied Technology, Science, and Arts (1st APTECS)*, 2009, pp. 2086–2091.
- [24] D. Yu, W. Choi, M. Kim, and L. Liu, "Forecasting day-ahead hourly photovoltaic power generation using convolutional self-attention based long short-term memory," *Energies (Basel)*, vol. 13, no. 15, Aug. 2020, doi: <https://doi.org/10.3390/en13154017>.

- [25] F. Liu, T. Dong, T. Hou, and Y. Liu, "A Hybrid Short-Term Load Forecasting Model Based on Improved Fuzzy C-Means Clustering, Random Forest and Deep Neural Networks," *IEEE Access*, vol. 9, pp. 59754–59765, 2021, doi: <https://doi.org/10.1109/ACCESS.2021.3063123>.
- [26] K. Cho *et al.*, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Jun. 2014, pp. 1724–1734. doi: <https://doi.org/10.48550/arXiv.1406.1078>.
- [27] S. V. Oprea and A. Bara, "Machine Learning Algorithms for Short-Term Load Forecast in Residential Buildings Using Smart Meters, Sensors and Big Data Solutions," *IEEE Access*, vol. 7, pp. 177874–177889, 2019, doi: <https://doi.org/10.1109/ACCESS.2019.2958383>.