http://www.moderntechno.de/index.php/meit/article/view/meit33-00-008
DOI: 10.30890/2567-5273.2024-33-00-008

UDC 519.23:004.85:621.311

SARIMA(X) AND DECISION TREE MODELS COMPARISON FOR LOAD FORECASTING IN ELECTRICAL GRIDS

Valerii Kyryk
d.t.s., prof.
ORCID: 0000-0003-0419-8934
Yevhen Shatalov
Ph.D. student
ORCID: 0009-0003-5505-5674
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute",
Kyiv, Prospect Beresteiskyi 37, 03056

Abstract. *This work compares load forecasting models in the electrical network based on SARIMA(X) and Decision Tree. Various versions of the above models were trained on load data for 2021 for one of the substations in the Kyiv region. The day 01/01/2022 was selected as test data - as the most difficult part of the load graph to predict. This example shows the imperfection of the models under consideration.*

Key words: *STLF, SARIMA, SARIMAX, Decision Trees, load forecasting, electrical grid.*

Introduction. Forecasting the load in the electrical network is a critical element in the activities of energy companies. It is used to reserve generation capacity, and also determines the planned volumes of electrical energy imports. Short-term load forecasting (STLF) is one of the types of forecasting, the purpose of which is to determine the load for the next day from the current one.

Autoregressive integrated moving average (ARIMA) models were among the first used for STLF [1, 2]. One of their modifications is SARIMA, where S – seasonal component. Also, another complication introduced into the model may be the use of several parameters (SARIMAX) on which the target result will depend. At the same time, these models have a serious drawback – as the seasonal coefficient increases, when sampling data over a significant period of time, their complexity increases too.

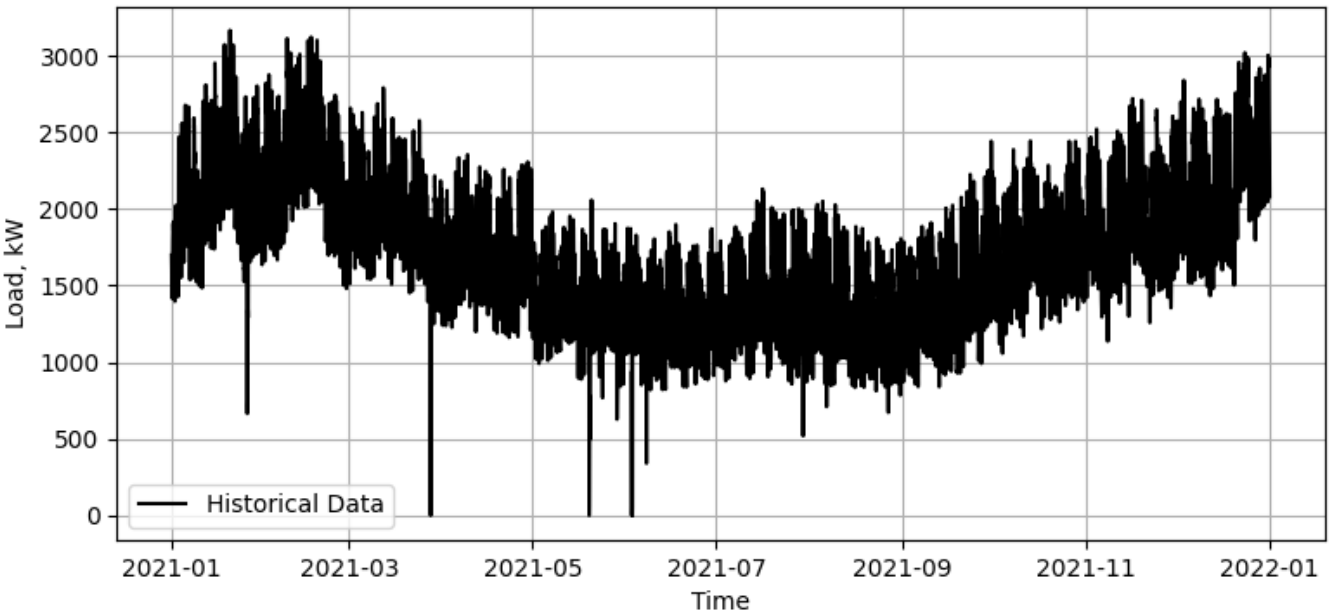

Figure 1 – Load data from 110/10 kV substation from 2021 year

For comparison with the specified statistical models, the Decision Trees (DT) method was chosen. Such models are a nonparametric supervised learning method used for classification and regression. Their goal is to predict the value of a target variable by learning simple decision rules derived from features of the data. DT can be considered as a piecewise constant approximation, which makes it similar to all ARIMA models.

The most difficult periods of time to predict using the selected models are holidays and other events for which the consumption pattern stands out from the overall picture. That is why data for 01/01/2022 is used to compare models.

SARIMA and SARIMAX models fitting.

To create this type of model, a library for the Python programming language was used [4]. With its help, models with different sets of parameters were trained. The test data set is measurements with an interval of 1 hour for 2021 (fig. 1). In addition to active load data, data on weather conditions is also used – Temperature (°C), Dew/Frost Point (°C), Relative Humidity.

Of all the models of this type, 10 were selected with the closest result to the test data. The main criterion for evaluation is root mean square error (RMSE), AIC and BIC – criteria for the quality of model fit and its complexity.

Table 1 – 10 best SARIMA(X) models for load forecast

№	Type	AIC	BIC	LLF	MAE	MSE	RMSE
1	SARIMA (3, 1, 1), (2, 1, 1, 24), 'ct'	108001.6	108072.3	-53990.8	55.73109	13152.64	381.8619
2	SARIMAX (3, 1, 1), (2, 1, 1, 24), 'ct', T, DF, RH	107995.3	108087.2	-53984.6	55.44703	13194.35	401.4878
3	SARIMAX (3, 1, 1), (2, 1, 1, 24), 'ct', T	107993.3	108071.1	-53985.7	55.47397	13135.17	402.22
4	SARIMA (3, 1, 1), (1, 1, 1, 24), 'ct'	108300.6	108364.3	-54141.3	56.40161	13177.25	404.4839
5	SARIMA (3, 1, 1), (1, 1, 2, 24), 'ct'	108012.6	108083.3	-53996.3	55.89954	13106.42	407.6566
6	SARIMA (3, 1, 1), (2, 1, 2, 24), 'ct'	108001.5	108079.3	-53989.8	55.77918	13103.92	408.3428
7	SARIMA (1, 1, 1), (2, 1, 2, 24), 'ct'	107986.7	108050.3	-53984.3	56.29603	13023.12	410.3191
8	SARIMAX (3, 1, 1), (1, 1, 1, 24), 'ct', T, DF, RH	108518.7	108603.6	-54247.4	57.66396	13780.43	414.5345
9	SARIMA (1, 1, 1), (1, 1, 2, 24), 'ct'	107993.1	108049.6	-53988.5	56.45625	13044.73	414.7623
10	SARIMA (3, 1, 1), (2, 1, 1, 24), 't'	108002.4	108066.1	-53992.2	55.58316	13153.86	416.2958

where T, DF, RH – Temperature, Dew/Frost Point, Relative Humidity; ‘t’ indicates a linear trend with time, and ‘ct’ is constant and linear trend.

From the simulation results, the disadvantage of the considered models is clearly visible – the inability to predict the load drop on the selected test day. This is explained by the fact that the selected frequency is 24, that is, one day. To take into account the nature of the load on the 1st day of the year, the frequency should be 8760 – the number

of hours in a year, and the amount of data should be for at least 2 years. Another possible approach is to use a model trained on parameters from the first week of the previous year.

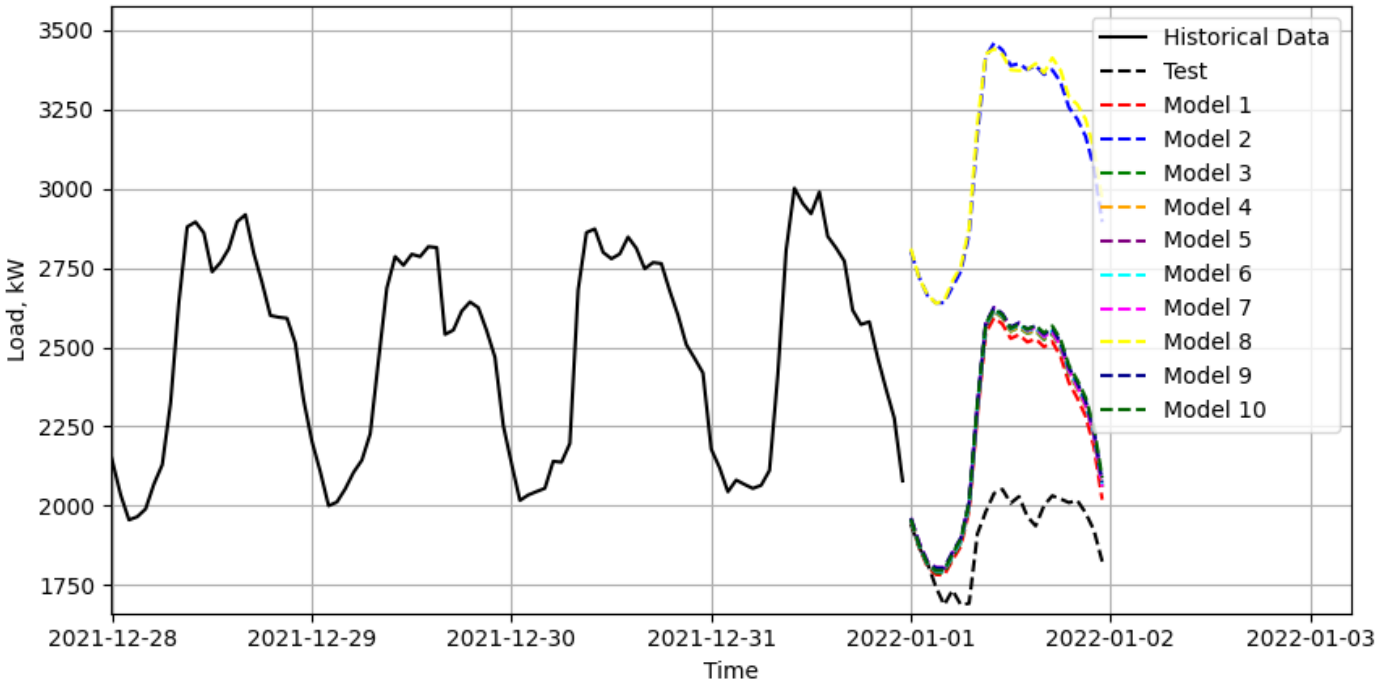

Figure 2 – Forecasting for the 01/01/2022 obtained from 10 SARIMA(X) models

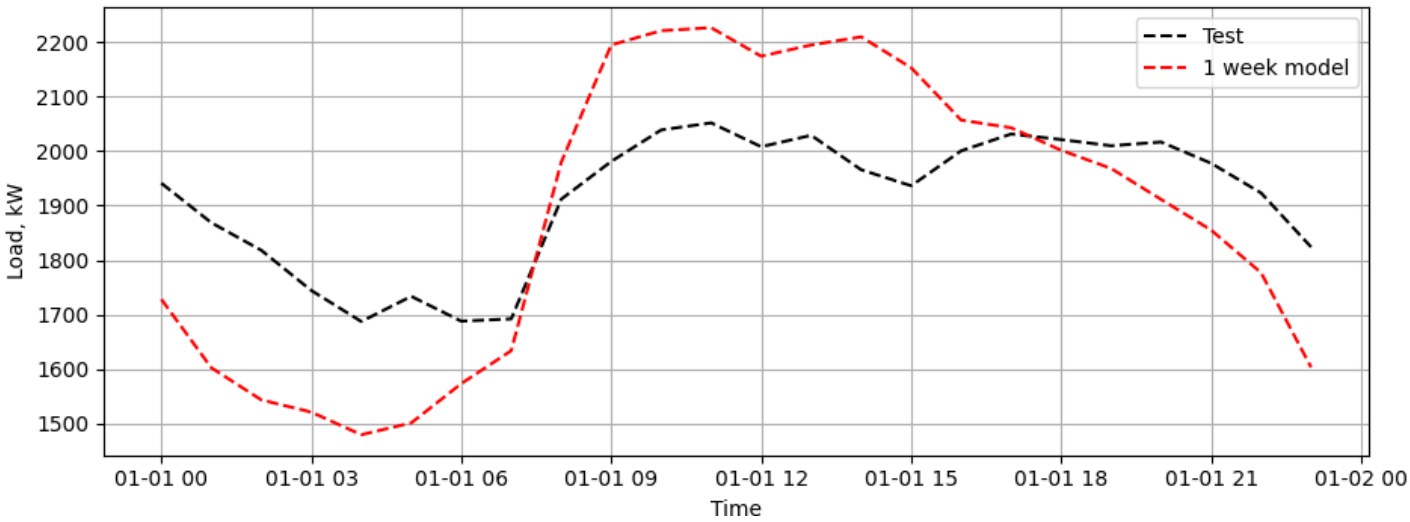

Figure 3 – Forecasting from SARIMA model trained on 01/01/2021 – 07/07/2021 data

When using the SARIMA model trained on data from the first week of the previous year, significantly better results were obtained with an RMSE of 174.86 (Fig. 3). It should be taken into account that such a model still will not be able to take into account monthly seasonality.

Decision Tree models fitting.

The basic concept of decision trees is to compute time instants ahead a response or a class from a set of known inputs and feature values. By growing a binary tree, at each internal node test is applied to an input and based on the outcome, hierarchical decision-making process flows towards left or right branches or sub-trees. When an

eventual leaf node appears, prediction is substantiated which averages or aggregates all the training data points that reach that terminal node of decision-making tree structure. There are two types of partitioning or decision trees applied in machine learning and data mining- classification tree and regression tree. In this work, regression trees were used, built using the library [6] for Python.

For these models, in addition to the load values, the following parameters were used: Temperature (°C); Dew/Frost Point (°C); Relative Humidity; Precipitation (mm/hour); Surface Pressure (kPa); Wind Speed (m/s); All Sky Surface Shortwave Downward Irradiance (Wh/m^2), as well as hourly, daily and monthly data. Using the listed parameters, all possible combinations of models were sorted out and the 10 best were selected, their results are presented in Table 2.

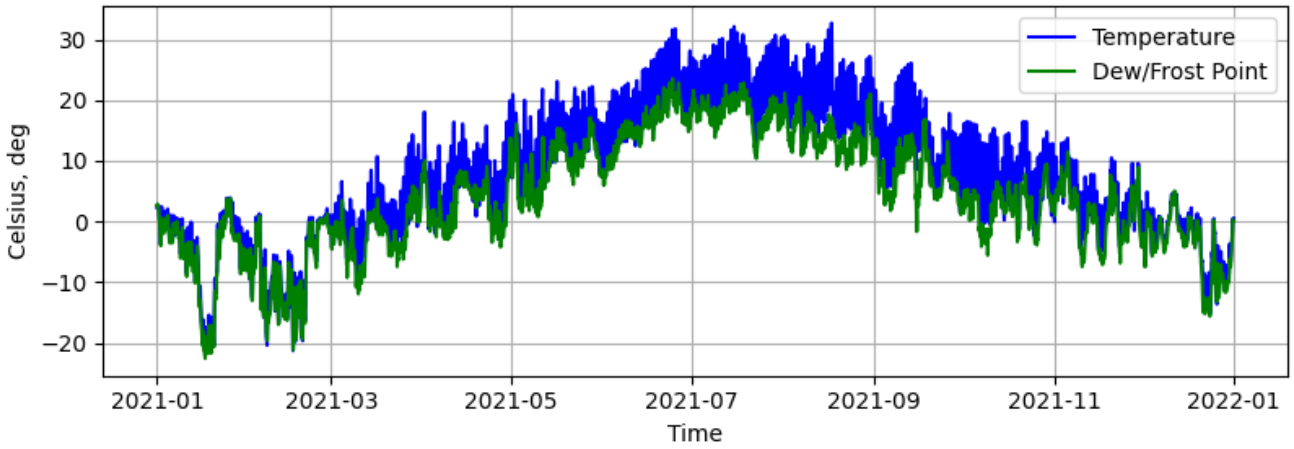

Figure 4 – Temperature and Dew/Frost Point from 2021 year

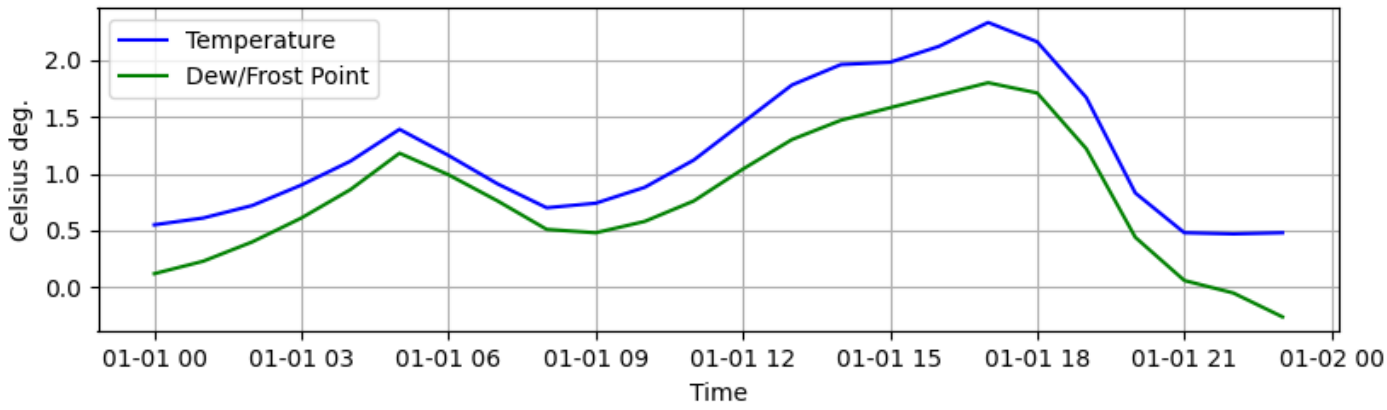

Figure 5 – Temperature and Dew/Frost Point from 01/01/2022

The main feature of Decision Tree models is the ability to take into account various parameters, and as follows from Table 2, they all use data about the day and month. To assess the quality of prediction in [6], the score parameter is used; the closer its value is to 1, the better the model. The best of the considered models does not contain information about the time of day. The result of this is a loss of load curve shape (fig. 7). The only model with data about the time of day is №. 7, while its score is practically zero.

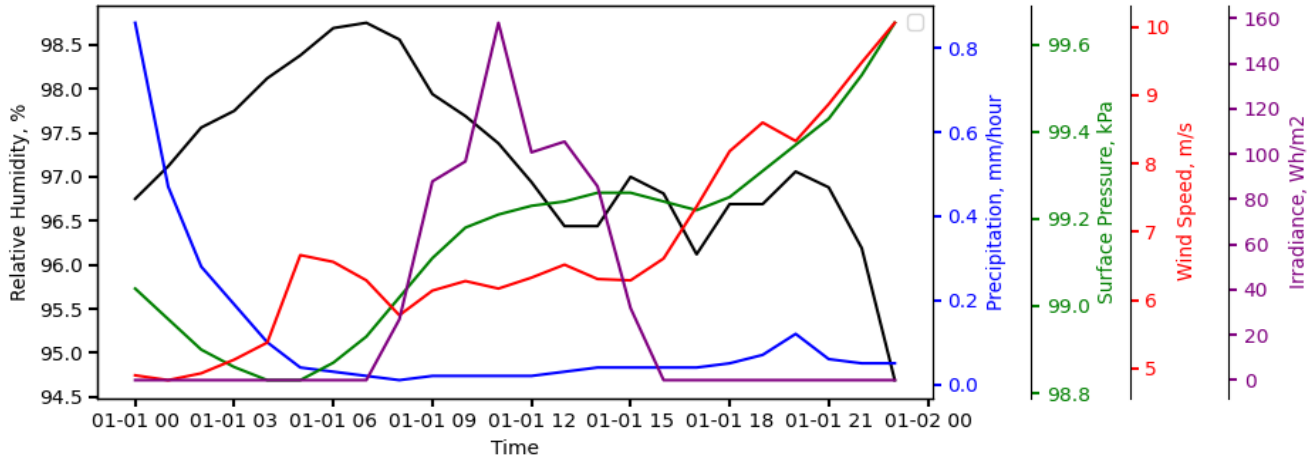

Figure 6 – Rest of the weather parameters from 01/01/2021

Table 2 – 10 best Decision Tree models for load forecast

№	Features used	MSE	RMSE	MAE	Score
1	Temperature, DF Point, Relative Humidity, Surface Pressure, DY, MO	11115.01	105.43	94.79	0.25
2	Temperature, DF Point, Relative Humidity, DY, MO	11787.90	108.57	96.29	0.20
3	DF Point, Relative Humidity, Surface Pressure, DY, MO	11847.16	108.84	95.35	0.20
4	DF Point, Relative Humidity, DY, MO	12513.12	111.86	97.79	0.15
5	Temperature, DF Point, Irradiance, DY, MO	12839.96	113.31	91.26	0.13
6	Temperature, DF Point, Surface Pressure, Irradiance, DY, MO	13654.79	116.85	100.48	0.08
7	Temperature, Relative Humidity, Precipitation, Surface Pressure, Wind Speed, Irradiance, HR, DY, MO	14006.96	118.35	98.95	0.05
8	Temperature, DF Point, Relative Humidity, Surface Pressure Irradiance, DY, MO	14113.56	118.80	105.55	0.05
9	Temperature, Relative Humidity, DY, MO	14119.04	118.82	103.05	0.04
10	Temperature, Relative Humidity, Wind Speed, DY, MO	14119.04	118.82	103.05	0.04

where DY, MO – day and month of year; DF Point – Dew/Frost Point

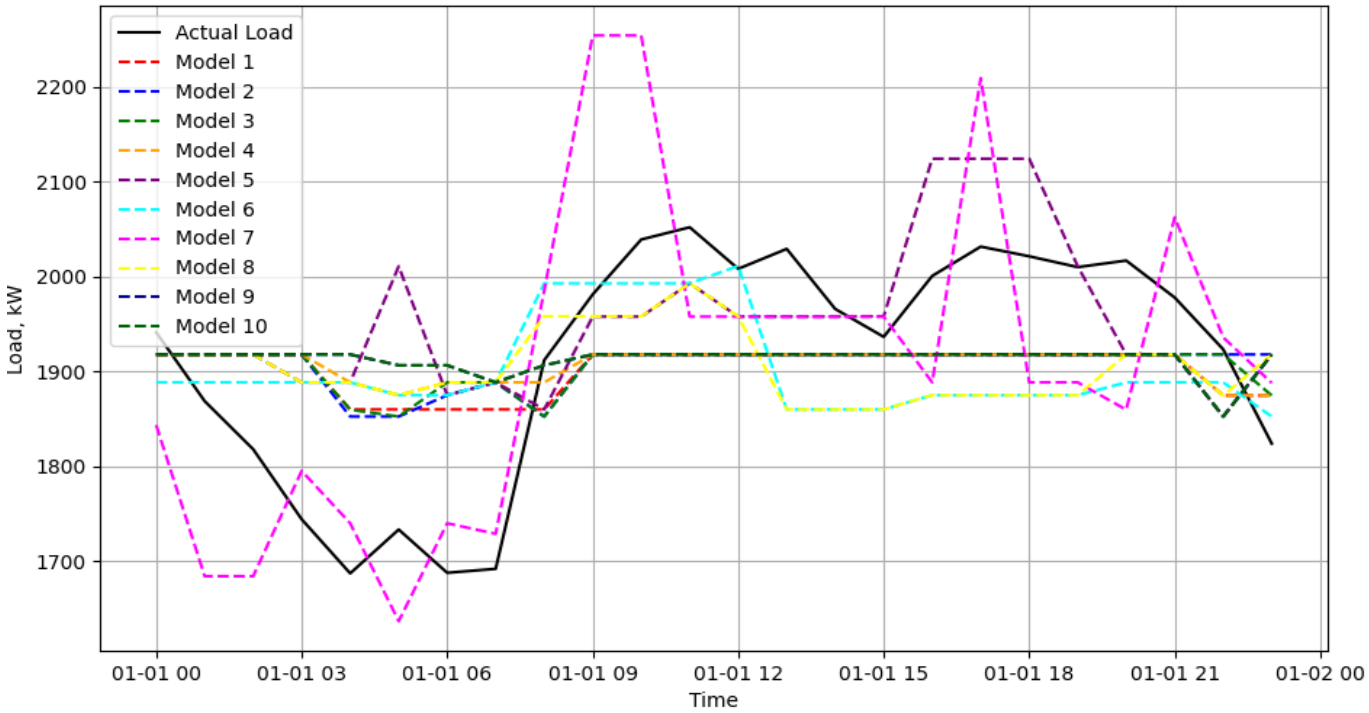

Figure 7 – Forecasting from Decision Tree models

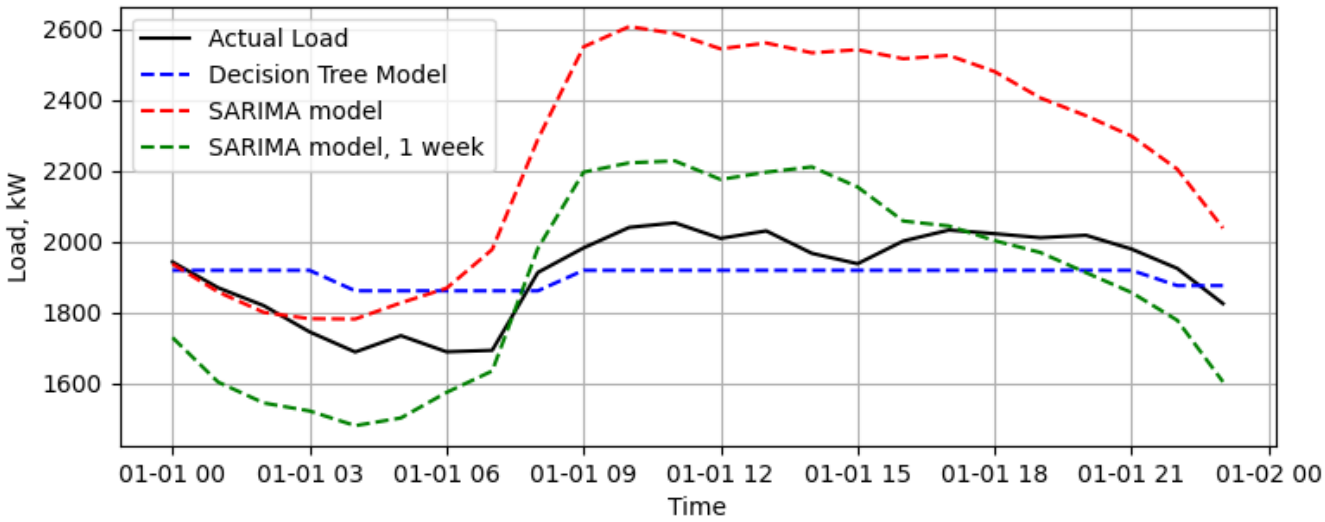

Figure 8 – Forecasting comparisons for used models

Summary and conclusions. Considered models are basically – approximating functions, which allows them to forecast load in the grid. Models like SARIMA, with a low seasonality coefficient, cannot take into account the annual cycle of load changes (red curve in fig. 8), but when the sample is reduced to one week, the accuracy increases. This is explained by the similarity of the load curve at the beginning of each of the years considered. These models are also characterized by reproducing the nature of the load curve, unlike models like Decision Tree.

Decision Tree models allows more accurately forecast grid load with lower computational costs for their training. As follows from the load forecasting plot, their results is more generalized – a smoothed load graph relative to the graphs of the input parameters.

An important difference is the resulting size of the models: SARIMA models on average occupy 6-8 GB (when using the library [4]), while Decision Tree does not exceed 200 MB of memory.

References:

1. Hagan, M. T., & Behr, S. M. (1987). The Time Series Approach to Short Term Load Forecasting. IEEE Transactions on Power Systems, 2(3), 785–791. https://doi.org/10.1109/tpwrs.1987.4335210

2. G. Gross and F. D. Galiana, “Short-term load forecasting,” Proceedings of the IEEE, vol. 75, no. 12, pp. 1558–1573, 1987, doi: https://doi.org/10.1109/proc.1987.13927.

3. J. W. Taylor and P. E. McSharry, “Short-Term Load Forecasting Methods: An Evaluation Based on European Data,” IEEE Transactions on Power Systems, vol. 22, no. 4, pp. 2213–2219, Nov. 2007, doi: https://doi.org/10.1109/tpwrs.2007.907583.

4. Statsmodels.org, 2023. https://www.statsmodels.org

5. S. Nallathambi and K. Ramasamy, “Prediction of electricity consumption based on DT and RF: An application on USA country power consumption,” IEEE Xplore, Apr. 01, 2017. https://ieeexplore.ieee.org/document/8191939

6. “scikit-learn: machine learning in Python — scikit-learn 0.22.2 documentation,” scikit-learn.org. https://scikit-learn.org

Article sent: 22.05.2024