Electricity price forecasting for the day-ahead market in Croatia

Kristijan Kosteski
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University in Skopje
Skopje, Macedonia
kristijan.kosteski@students.finki.ukim.mk

Aleksandra Dedinec
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University in Skopje
Skopje, Macedonia
aleksandra.kanevche@finki.ukim.mk

***Abstract*—This paper presents a Machine Learning model, specifically a Neural Network, to forecast the prices of the Croatian day ahead power exchange. The price forecasting is of great importance for building and maintaining the smart energy grids that are a part of Smart Cities around the globe. Knowing the price of electricity can make a switch to renewable energy sources easier to predict and maintain so that it is available to the citizens of the city.**

The day ahead market requests bids for the next day for all of the 24 hours in that day. From all of the participants' bids, a price is calculated where supply meets the demand and this is the price that the models try to predict without using the bids from the participants. Predicting the price of electricity for markets that produce a big part of its electricity from renewable sources can be unpredictable and cause a lot of problems for researchers. Croatia is trying to increase production of electricity from renewable sources in the following years, which will make the prices more unpredictable.

The data used is collected for the period from 11.02.2016 to 24.08.2021. Multiple Neural Network structures were explored, but the best one was a Neural Network with 2 hidden layers with 11 neurons each. The testing period was the last 20 % of the data as the training and testing data was split 80:20 and some years were dropped for better results.

The results for the model were evaluated using multiple evaluation metrics, most of which being error rates.

Keywords—Artificial Intelligence, Machine Learning, Neural Networks, time-series forecasting, electricity price forecasting, day-ahead market forecasting for Croatia, Smart City, Smart energy grids

I. INTRODUCTION

Forecasting the prices of the day-ahead market is a field of research that holds great importance for the maintenance and implementation of smart energy grids as a key part of Smart Cities around the globe. The predicted prices can help an electricity company optimize its bidding strategy and its own production or consumption schedule in order to reduce the risk or maximize the profits in day-ahead trading [1]. For this to be effective, a decent accuracy has to be achieved so the company can have some confidence in the predicted prices and act accordingly. Unfortunately, the prices can be difficult to predict due to various unforeseen circumstances, especially renewable energy sources [3-6]. As the world is trying to switch to renewable energy sources to achieve The Sustainable Development Goals (SDGs), the need to accurately predict the prices on the day-ahead electricity market becomes an important task.

The problem is formally presented as predicting the prices for the next day or day d+1. The bids to calculate the prices are placed on day d and the prices are calculated sometime on the day d, before the next day or day d+1. The prices are calculated for all of the 24 hours of the day d+1. The forecasting is done for a specific hour the next day, and it requires attributes that are available on day d, so that the prices can be predicted successfully. This problem is presented as a time-series problem where the next item in the series is predicted for a future time. Each item in the series is an hour of the day and the price is measured in Euros per Mega Watt Hour (€/MWh).

This research explored many different Machine Learning models to predict the prices, but the best one seemed to be a Neural Network with 2 layers with 11 neurons each. Deep Learning was also explored, but it didn't produce the same results as Neural Networks with simpler structures. In the results section, there are more details on the models explored and future research directions.

Before we explain how the paper is structured, there is important information to state about this problem that will help the reader to understand the problem into more detail. First of all, the Croatian day-ahead electricity market operated by CROPEX was coupled with the Slovenian day-ahead electricity market operated by BSP South Pool on June 19th 2018. Another information worth to note is that Croatia is trying to increase production of electric energy from renewable sources [8].

The paper is structured as such. First, the introduction presents the problem and important information for future sections. Then, the methodology explains the methodology used in short, because Neural Networks have been explored in thousands of papers before and it would be redundant to go into excessive detail. The third part has interesting visualizations of the data used. The fourth part goes into the data used and the model structure. The fifth part discusses the results of the models tested and touches over what future work can be conducted on this topic.

II. METHODOLOGY

The methodology used for forecasting the prices has been used many times by countless of research papers. This is why I will only touch briefly on the methodology used.

The methodology used for forecasting the prices is the Feed Forward Neural Network, specifically the Multi-layer feed-forward Perceptron (MLP).

The neural networks that are used in this paper are represented as a directed acyclic graph, which has several layers - an input layer, 2 hidden layers and an output layer, which makes the structure a not deep Neural Network. The connection between the input $\boldsymbol{x}$ and the output y are presented with the following equation:

$$y = \sum_{j=0}^{h} \left\{ w_j x\, f\left(\sum_{i=0}^{d} w_{ij} x_i \right) \right\} \tag{1}$$

Where w_j and w_{ij} represent the weight and biases that connect the layers.

Supervised Learning is used and the optimization algorithm used is 'Adam' as presented in [9]. The loss function used for the optimizer was the Mean Absolute Error (MAE) for a regression problem.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{2}$$

Where n is the number of observations, y_i is the actual value, and $\hat{y}_i$ is the predicted value.

For evaluating the model's performance, multiple evaluation metrics were used. One of them is the MAE metric represented with (2). The other evaluation metrics are Mean-Squared Error (MSE), Root Mean-Squared Error (RMSE), and Adjusted R2 Score. MSE is represented with this formula:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{3}$$

The RMSE is represented with this formula, which is just a square root of the MSE:

$$RMSE = \sqrt{MSE} \tag{4}$$

Finally, the formula for the Adjusted R2 Score is given with this formula:

$$Adjusted\ R2 = 1 - \frac{(1-R^2)(N-1)}{N-p-1} \tag{5}$$

where N is the number of samples or observations, R^2 is the R squared score for the observations, and p is the number of predictors, which in our case is 1.

III. Visualizations of the Data

In this section, some interesting data visualizations are presented. The purpose of this section is to describe the data and to explain some facts about the data that will be of use in the later sections and to better understand the problem at hand.

In Fig. 1 below, we can see that the prices range from 300 to around -250 €/MWh, however, this is because there are 2 peaks or outliers in the data that need to be removed so they don't affect the models. This is the reason they are called Old prices. The outliers were replaced by the value above itself, so that the time-series isn't affected too much.

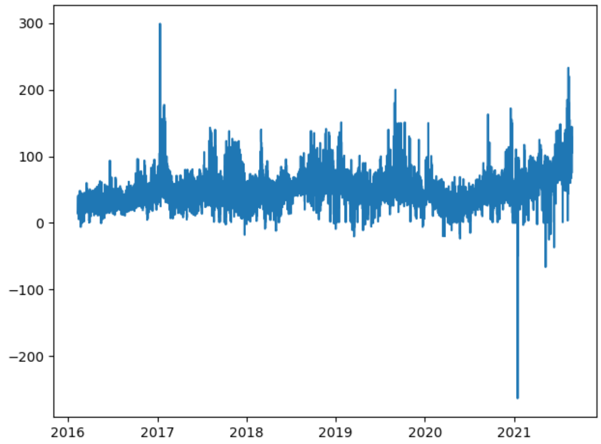

Fig. 1. Old graph for all of the prices in the dataset from 11.02.2016 to 24.08.2021. The x axis presents the time and the y axis presents the prices of electricity measured in €/MWh.

In Fig. 2 the new prices are shown with the 2 outliers removed around the start of 2017 and the start of 2021, positive and negative values respectively. The new graph shows that the prices range from 250 to around -70 €/MWh. We can also see a lot of spikes of the prices and some zero prices and negative prices as well.

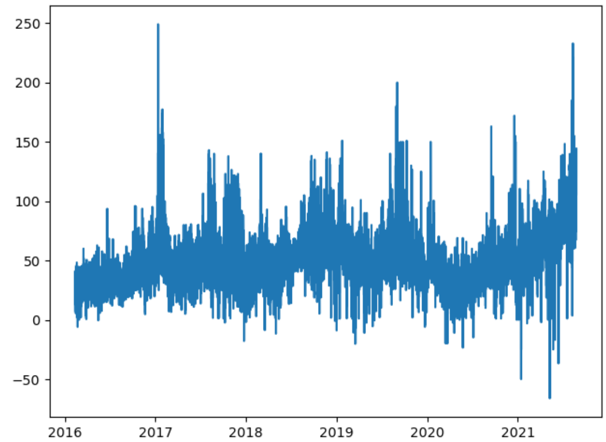

Fig. 2. New prices from 11.02.2016 to 24.08.2021 without the 2 outliers from the previous graph. The x axis represents the time and the y axis represents the price of electricity in €/MWh.

Fig. 3 shows the mean price for every day of the week from Monday to Sunday, represented as numbers from 1 to 7 respectively. We can see a drop in the price as the weekend approaches. The price seems to be the lowest on Sunday and the highest on Wednesday, right in the middle of the week.

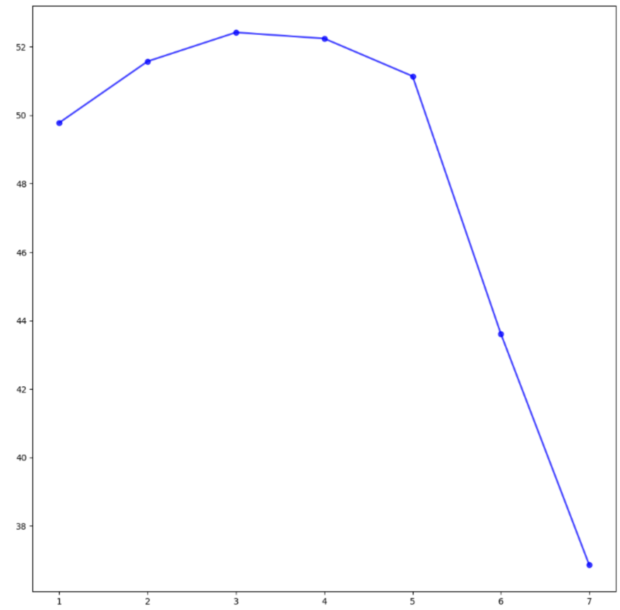

Fig. 3. Mean of the prices for every day of the week from Monday to Sunday, represented as numbers from 1 to 7 respectively. The x axis presents the day of the week from 1 to 7 and the y axis presents the mean price in €/MWh.

In Fig. 4 the mean price for every month is presented. The moths are numbered from 1 to 12 presenting the months from January to December respectively. We can see in Fig. 4 that the prices tend to be higher in the middle of winter and in the middle of summer, when people use electricity to heat or cool their homes. More and more people use electricity for heating or cooling, which makes the prices higher due to the increase in demand.

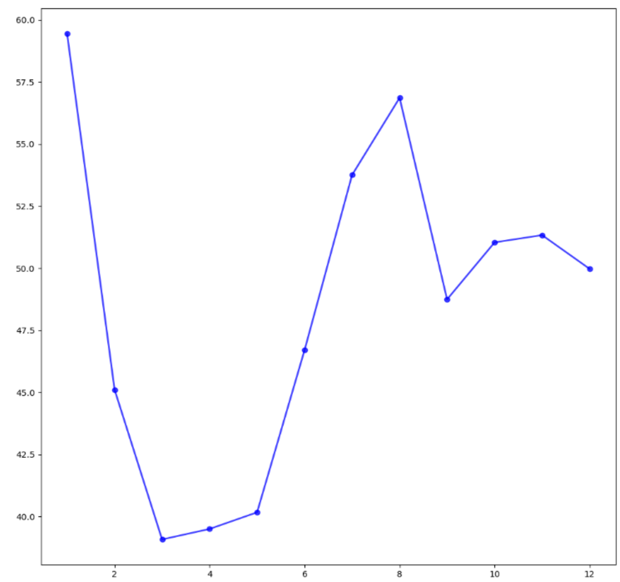

Fig. 4. Mean of the prices for every month of the year from 1 to 12 for January to December respectively. The x axis presents the months from 1 to 12 and the y axis presents the mean of the prices calculated for every month in the dataset.

The last graph, Fig. 5, shows the volume of electricity generated with respect to time. The volume of electricity produced is presented with Megawatt hour (MWh). We can see a sharp increase in the volume produced from around the middle of 2018. This is due to the aforementioned coupling of Croatia's day-ahead electricity market and Slovenia's day-ahead electricity markets on June 19th 2018. We can see that the volume produced (Fig. 5) and the prices of electric energy don't have much similarities and it shows in the results later that the volume of electricity produced isn't a very good predictor or attribute in the Neural Network model.

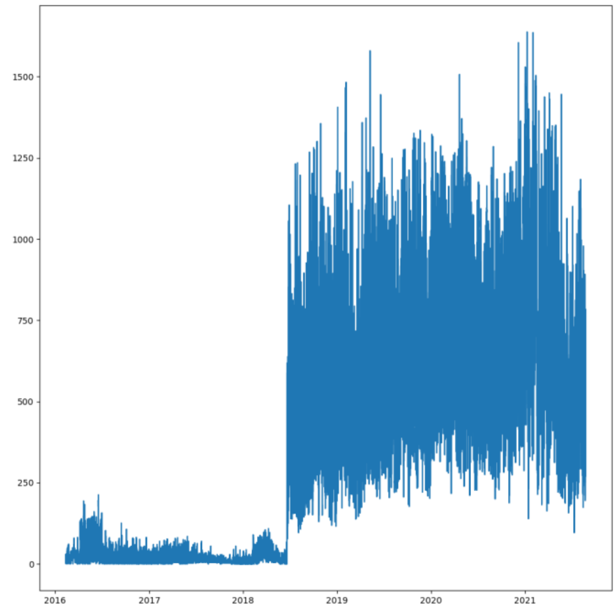

Fig. 5. Volume of electric energy produced from 11.02.2016 to 24.08.2021. The x axis presents the time from 2016 to 2021 and the y axis presents the volume of electricity produced in MWh.

IV. Input Data for the Neural Network

The input data was collected from the official CROPEX website where they post the prices and volumes of electricity generated for every hour from 11.02.2016 to 24.08.2021 [7]. The latter is decided to be August 24th, because the research was conducted in August 2021 and the data was collected up until that point. There is more data to collect for the year 2021 and 2022 to see how the prices were affected from the COVID19 pandemic, if at all.

Additionally, weather data was also collected from the website Weather2Umbrella [10]. The data collected was the wind strength, weather description, and temperature for the city of Zagreb. The problem with this data was its quality, because it had a lot of missing values and incorrect values. This data throttled the potential of the model and it is suspected that better quality weather data would lead to better performances, especially because Croatia is switching to more renewable energy sources.

The last data source was the website Entso-E Transparency program, which had data on the Hourly Load Forecast for the country of Croatia [11]. The data was collected for the same time period from 11.02.2021 to 24.08.2021.

From all of these data sources, these were the attributes or features that were created to forecast the electricity prices:

- Price - The target variable y that is measured in €/MWh.

- Price of day 1 – 7 before – The prices for the same hour, but 1 to 7 days before. There are 7 features for the price for every day from 1 to 7.
- Volume of day 1 – 7 before – The same as above, but for the volumes of electricity produced.
- Average price yesterday – The average price from the day before or day d.
- Average volume yesterday – Same as above, but only for the volumes of electricity produced on day d.
- Hour – The hour in the day from 0 to 23.
- Day – The day in the month from 1 to 31.
- Day of the week – The day of the week from 1 to 7 for every day of the week respectively.
- Month - The month from 1 to 12 for every month of the year.
- Year – The year from 2016 to 2021.
- Hourly Load Forecast – The hourly forecast for the load for every hour on day d+1. This data column is generated and is measured in Megawatts (MW).
- Wind – Wind gust or the wind strength for that hour.
- Weather description – The weather description such as "Sunny" or "Cloudy" etc.
- Temperature – The forecasted temperature on day d+1 measured in degrees Celsius (**°C**).
- Maximum and minimum temperature – The forecasted maximum and minimum temperature of day d+1 measured in **°C.**
- Holiday – True or False if the date is a national holiday in Croatia.

Some of the collected features were recommended by Jesus Lago et al in [2]. Of these input variables, only 14 were used and found to be useful for the model's predictions. These features were the Hour, Day, Day of the week, Month, Year, Holiday, Price of day 1-2 day before, Price of day 6-7 before, Hourly Load Forecast, Maximum and Minimum Temperature, and Weather description.

Not all of the data was used for training and not all of the available data for training was used. The data for the year 2021 was not used, because it caused overfitting in many of the experimented models. The data from the year 2021 is most likely different from the other years, because of the rise of energy prices that started around May 2021 [14]. With the sudden rise of energy prices in Europe and the war that stared in 2022, it is expected that the prices in the years 2021 and 2022 are different from previous years. Another model to accommodate the new changes in the prices can be created. It would make the prices of the electricity more predictable in irregular periods. This model was shown to not accommodate the changes that happened in the year 2021 as it was trained on years without the irregularities mentioned above.

The data was split 80:20 and exactly 1 year was used for the testing data. The testing year was the year 2020. When using the year 2021, the model trained on part of 2020 where it suffered from overfitting and achieved worse results seen from the error rates (MSE, RMSE, and MAE). The training data was the data from 11.02.2016 to 31.12.2019, while the testing data ranged from 01.01.2020 to 31.12.2020. At least one year was used for testing purposes and MAE was used as an evaluation metric as suggested by Jesus Lago et al in [2].

The model structure was a Neural Network with 14 input neurons in the input layer, 2 hidden layers with 11 neurons each, and 1 neuron in the output layer. The model structures that were experimented with were pretty simple and other models such as Linear Regression achieved good results as seen from the Adjusted R2 score of 0.79. The problem with the Linear Regression model was the fact that it performed poorly as seen by the error metrics (MSE, RMSE, and MAE). Very Deep Neural Network structures were not explored due to computational limits, and only some shallow Deep Neural Networks were explored with about 10 layers with 200 neurons each.

V. Results and discussion

The results of the model were evaluated with multiple evaluation metrics to see the performance of the models clearly. For this model, MSE, RMSE, MAE and Adjusted R2 Score were used as evaluation metrics on the testing dataset.

The results for the Neural Network with 2 hidden layers with 11 neurons each, were:

- MSE 78.41822
- RMSE 8.855406
- AR2 0.721466
- MAE 5.83894.

Unfortunately, MAPE couldn't be used as an evaluation metric, because there were multiple zero values or close to zero values, which make the MAPE value very high since it is dividing with zero values or close to zero values. Using MAPE, we could compare this model with the model proposed by Aleksandra et al in [12] where the model used data from only the year 2016.

The main factor for determining the success of the model was the amount of data used for training. If less data is used for training, then that significantly lowers the performances of the model and this is very impactful, because not much data has been collected in Croatia around electric energy. We suspect that as Croatia collects more data around electric energy production that could benefit researchers trying to build Machine Learning models to benefit Smart Cities in Croatia.

The weather data didn't have the best quality which throttled the model. We suspect that better quality data around weather conditions in Croatia, specifically for the weather conditions around power plants that collect energy from renewable sources, would be of great use. As Croatia is switching to more renewable energy sources, we suspect that it could be beneficial for the country to collect as much data as possible as soon as possible to make prices from renewable energy sources more predictable in the long run. Data is a fundamental part of Smart Cities and collecting more data can benefit researchers that are trying to create Machine Learning models.

As previously mentioned, very Deep Neural Networks weren't explored out of computational limitations. These models were very good at predicting the spikes in the price of electric energy, but unfortunately had a higher error rate. When experimenting with different model structures, it seems that simpler structures are favored and they achieve better performances than more complex Neural Network structures. Future work can be done using very Deep architectures for forecasting the prices for the day-ahead electricity market in Croatia.

Another model that wasn't explored was the LSTM proposed in [13] that is made to predict time-series. Future work can be done using this type of model to forecast the prices on the day-ahead electricity market in Croatia.

In Fig. 6, a comparison of the predicted and actual values is shown. As we can see from Fig. 6, the predicted value fits very well under the actual values.

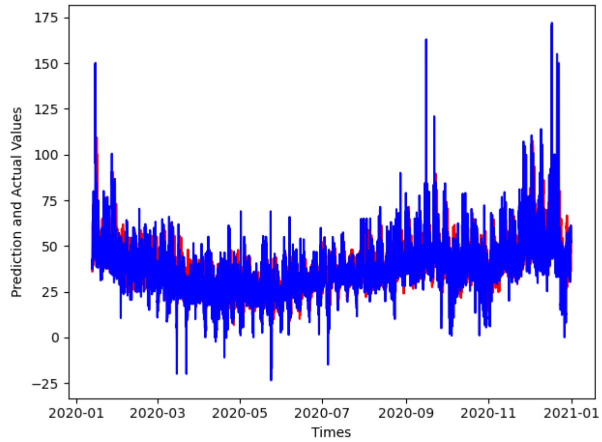

Fig. 6. Comparison between the predicted values from the model and the actual values. The x axis is the time period from the test values and the y axis is the price of electric energy in €/MWh. The blue line is the actual value and the red line represents the predicted values.

Another look at the results is shown on Fig. 7, Fig. 8, and Fig. 9. The graphs show the mean of the actual electricity prices compared with the model's predicted price means for every hour, weekday, and month respectively.

We can see that the model fits all of the lines very well. The only noticeable exception can be found on Fig. 7 for the per hour comparison. We can see that the model makes the biggest mistake from around 7 to about 17, when the people are usually away at work.

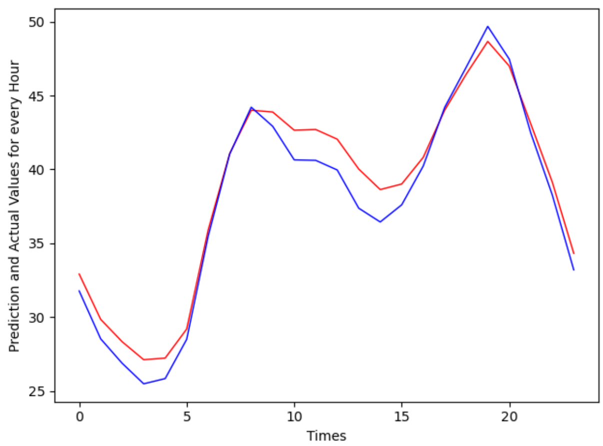

Fig. 7. Comparison between the actual electricity prices (blue line) and model predicted prices (red line) for every hour. The x axis presents the hours from 0 to 23 and the y axis presents the mean of the prices for all of the hours.

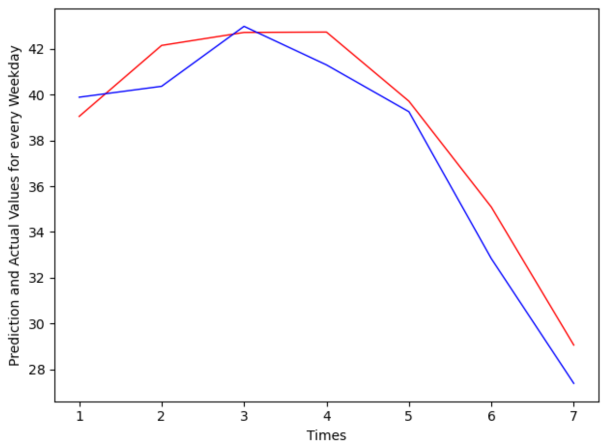

Fig. 8. Comparison between the actual electricity prices (blue line) and model predicted prices (red line) for every weekday. The x axis presents the weekdays from 0 to 7 for Monday to Sunday and the y axis presents the mean of the prices for all of the weekdays.

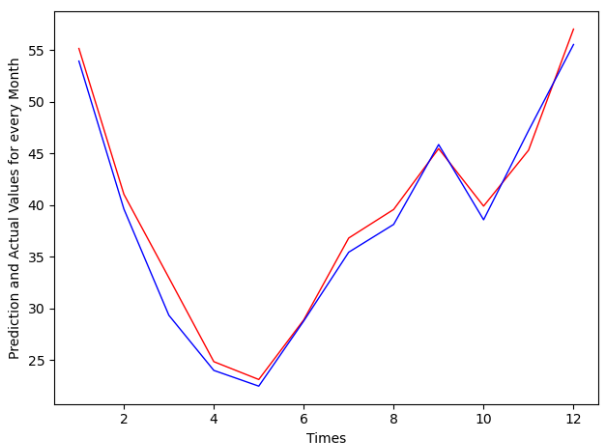

Fig. 9. Comparison between the actual electricity prices (blue line) and model predicted prices (red line) for every month. The x axis presents the months from 1 to 12 for January to December and the y axis presents the mean of the prices for all of the months.

The last thing to note is the fact that the volume of produced electric energy gave the model worse performances and it wasn't of any use at all. This is seen in Fig. 5 were the volume of electric energy produced doesn't seem to have any correlation visually with the prices of electric energy.

VII. Conclusion

In this paper we discussed an approach to forecast the electricity prices on the day-ahead electricity market in Croatia. The data used was from the period of 11.02.2016 to 24.08.2021 and the attributes for the model were features that can help in describing the demand and production of electricity, which in turn, determines the price of electricity. Predicting the electricity prices becomes an important task to predict less predictable sources such as renewable energy sources, which is why the model was built. The results are favorable, however, they cannot be compared to previous models on this topic, because the metrics to evaluate the models differ and the prices have zeros, making it impossible to calculate an accurate MAPE as in the work by Aleksandra et al [12].

Acknowledgment

This work was partially financed by the Faculty of Computer Science and Engineering at the "Ss. Cyril and Methodius" University.

References

[1] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *International Journal of Forecasting*, vol. 30-4, pp. 1030-1081, 2014.

[2] Jesus Lago, Grzegorz Marcjasz, Bart De Schutter, Rafał Weron, "Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark", Applied Energy, Volume 293, 2021, 116983, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2021.116983.

[3] Brancucci Martinez-Anido C, Brinkman G, Hodge B-M. The impact of wind power on electricity prices. Renew Energy 2016;94:474–87. http://dx.doi.org/10.1016/j.renene.2016.03.053.

[4] Gianfreda A, Parisio L, Pelagatti M. The impact of RES in the Italian day-ahead and balancing markets. Energy J 2016;37:161–84. http://dx.doi.org/10.5547/01956574.37.si2.agia.

[5] Grossi L, Nan F. Robust forecasting of electricity prices: Simulations, models and the impact of renewable sources. Technol Forecast Soc Change, 2019, 141:305-18.

http://dx.doi.org/10.1016/j.techfore.2019.01.006.

[6] Maciejowska K. Assessing the impact of renewable energy sources on the electricity price level and variability – A quantile regression approach. Energy Econ 2020;85:104532. http://dx.doi.org/10.1016/j.eneco.2019.104532.

[7] Croatian power exchange (CROPEX) https://www.cropex.hr/hr/trgovanja/dan-unaprijed-trziste/rezultati-dan-unaprijed-trzista.html (accessed August 24, 2021)

[8] Gelo, T.; Šimurina, N.; Šimurina, J. The Economic Impact of Investment in Renewables in Croatia by 2030. Energies 2021, 14, 8215. https://doi.org/10.3390/en14248215.

[9] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).

[10] Weather data, Weather2Umbrella,

https://www.weather2umbrella.com/pr%C3%A9visions-m%C3%A9t%C3%A9orologiques-zagreb-croatia-fr/m%C3%A9teogramme (last accessed 24.08.2021)

[11] Entso-E Transparency program, Hourly Electricity Load forecast, https://transparency.entsoe.eu/ (accessed 30.10.2021)

[12] Dedinec, Aleksandra and Aleksandar Dedinec. "Electricity price forecasting of the South East European power exchanges." (2017).

[13] Hochreiter, Sepp & Schmidhuber, Jürgen. (1997). Long Short-term Memory. Neural computation. 9. 1735-80. 10.1162/neco.1997.9.8.1735.

[14] Eurostat website, Energy prices on the rise in the euro area in 2021, https://ec.europa.eu/eurostat/web/products-eurostat-news/-/edn-20220210-2 (last accessed)