Maksym Oliinyk, Jaroslav Džmura, Jozef Humeník, Peter Havran, Robert Štefko, Daniel Pál

Data analysis and predictive models in Smart Grid

Abstract With the increasing penetration of smart elements into the electric power industry, the problem of processing, analyzing, and using data is becoming increasingly important. The presence of a large number of measuring devices allows you to collect hundreds of gigabytes of data, and every year the number of such devices will increase, allowing power companies to build more accurate predictive models using machine learning. These models can be used for a more accurate forecast of cross-border flows, improvement of control systems for energy storage systems, and improvement of work on the electricity trade market. This article explores several predictive models on the available data.

Keywords: machine learning; smart grids; renewable energy; forecasting; big data; data analysis

I. INTRODUCTION

The power system, which is based on the concept of Smart Grid ("Smart Grid", "Aquive-Adaptive Grids"), is a single complex, within which all elements of the electrical network are interconnected, have the possibility of remote control, control systems allow the networks to operate with the highest possible efficiency. The creation of such networks allows power supply companies to reach a completely new level of power grid management. The transition from classic electrical networks to new ones is only a matter of time. One of the main drivers of the transition will be the widespread adoption of renewable energy sources and electric vehicles. Experts point to a trend towards a rapidly narrowing gap in the cost of conventional and non-traditional energy sources [1]. A stimulating role in the use of alternative energy sources is likely to be played by the growth of oil and gas prices, as well as the tightening of environmental requirements during the construction of traditional generating facilities. However, it should be noted that the generated capacity of wind farms, solar power plants and other alternative energy sources is not constant and depends on natural conditions - the presence of wind activity, solar radiation, etc. In this case, such instability of renewable energy production makes its negative adjustments to stable operation power systems in the fight against these negative factors will help modern control systems, as well as modern forecasting models (artificial intelligence). Analysis of publicly available information shows that the main areas of application of artificial intelligence in the energy sector are currently [2]:

- Forecasting tasks (meteorological information, consumption change, etc.);
- Optimization tasks (consumption, network configuration, etc.);
- Management tasks (renewable energy sources and batteries, energy market)

Every year the number of elements in the power system is growing, which allow collecting various information (quality indicators of electrical energy, the amount of electricity consumed, power flows, weather conditions, etc.). After the liberalization of the energy market, energy supply companies have the opportunity to buy electricity on the exchange at more attractive prices, which also allows them to accumulate market data. There are enough examples of the use of artificial intelligence algorithms in forecasting problems today [3]. The dependence of renewable energy production on weather conditions has greatly increased the need for accurate forecasting. According to analysts, the large-scale use of artificial intelligence algorithms to improve the operation of US wind farms will theoretically allow them to increase production in 2017 by 12 billion kWh and increase the share of wind energy in the total balance [4]. PowerScout received two grants from the US Department of Energy to develop cost reduction programs for grid companies and consumers (smart home), taking into account the integration of renewable sources [5]. The programs also use artificial intelligence algorithms. London-based Green Running Ltd. develops a machine learning-based very application designed to optimize energy consumption in homes. The application works on computers, tablets and smartphones. The German company Schleswig-Holstein Netz AG, which operates electrical networks in the federal state of Schleswig-Holstein, reported an interesting application of artificial intelligence methods. Here, a self-learning network is used to locate suspected damage locations. [6] [7].

II. NEURAL NETWORK

A neural network is a mathematical model in the form of software and hardware implementation, based on the principles of functioning of biological neural networks. Today, such networks are actively used for practical purposes due to the possibility of not only development, but also training. They are used for prediction, pattern recognition, machine translation, audio recognition, etc. There are two large groups of neural networks:

- Conventional neural networks
- Convolutional neural networks

A fully connected neural network is often called common. In it, each node (except for the input and output) acts as both an input and an output, forming a hidden layer of neurons, and each neuron of the next layer is connected to all neurons of the previous one. The inputs are supplied with weights, which are adjusted during the learning process and do not change later. At the same time, each neuron has an activation threshold, after passing which it takes one of two possible values: -1 or 1, or 0 or 1.

The convolutional neural network has a special architecture that allows it to recognize patterns as efficiently as possible. The very idea of a CNN is based on the alternation of convolutional and subsampling layers (pooling), and the structure is unidirectional. SNN got its name from the convolution operation, which assumes that each fragment of the image will be multiplied by the convolution kernel element by element, while the result must be summed and written to a similar position in the output image. Such an architecture ensures the invariance of recognition with respect to the shift of the object, gradually enlarging the "window" at which the convolution "looks", revealing more and more large structures and patterns in the image. Neural networks are used to solve complex problems that require analytical calculations similar to those of the human brain. The most common uses for neural networks are [7]:

- Classification - distribution of data by parameters. For example, a set of people is given at the entrance and it is necessary to decide which of them to give a loan, and who does not. This work can be done by a neural network, analyzing information such as age, ability to pay, credit history, etc.
- Prediction is the ability to predict the next step. For example, the rise or fall of a stock based on the situation in the stock market.
- Recognition is currently the widest application of neural networks. Used on Google when you are looking for a photo or in phone cameras when it detects the position of your face and makes it stand out and much more.

The structure of the simplest neural network is shown in the figure below. The neurons of the input layer are shown in green, the neurons of the hidden layer are blue, and the neuron (s) of the output layer are yellow.

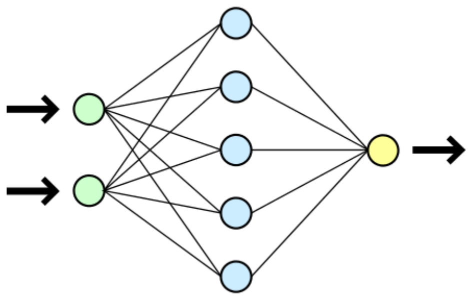

Fig. 1. Magnetizácia v závislosti od magnetickej intenzity.

The neurons of the input layer receive data from the outside (for example, from the sensors of the face recognition system) and, after processing them, transmit signals through the synapses to the neurons of the next layer. The neurons of the second layer (it is called hidden, because it is not directly connected to either the input or the output of the ANN) process the received signals and transmit them to the neurons of the output layer. Since we are talking about imitating neurons, each input-level processor is associated with several latent-level processors, each of which, in turn, is associated with several output-level processors. Such a simple ANN is capable of learning and can find simple relationships in data. An ANN capable of finding not only simple relationships, but also relationships between relationships has a much more complex structure. It can have several hidden layers of neurons, interspersed with layers that perform complex logical transformations. Each subsequent layer of the network looks for relationships in the previous one. Such ANNs are capable of deep (deep) learning [8].

III. ACCURACY ESTIMATION PARAMETERS

To assess the accuracy of the predictive models used in this article, the following parameters will be used:

- Mean absolute error
- Mean squared error
- Coefficient of determination of R^2

Mean absolute error (MAE) is a measure of errors between paired observations expressing the same phenomenon. Examples of comparing Y versus X of the comparative period of the predicted subsequent time and one measurement method versus another measurement method [8].

The Mean Squared Error (MSE) is a measure of how close a fitted line is to data points. For every data point, you take the distance vertically from the point to the corresponding y value on the curve fit (the error), and square the value. Then you add up all those values for all data points, and, in the case of a fit with two parameters such as a linear fit, divide by the number of points minus two. The squaring is done so negative values do not cancel positive values. The smaller the Mean Squared Error, the closer the fit is to the data. The MSE has the units squared of whatever is plotted on the vertical axis [8].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|$$

Fig. 2. Formulas for calculating MAE and MSE, where Y_i - is the actual sales volume for the analyzed period; $\hat{Y}_i$ - is the value of the forecast model for the analyzed period; N - is the number of periods.

Evaluation of the quality of the regression equation is carried out based on a set of criteria that verify the adequacy of the model to actual conditions and the statistical reliability of the regression. One of the most effective estimates of the adequacy of the model is the coefficient of determination of R^2. The true coefficient of determination of the model of the dependence of the random variable y on the factors x is determined as follows [9]:

$$R^2 = 1 - \frac{\sigma^2}{\sigma_y^2}$$

Fig. 3. Formulas for calculating R^2, where σ_y is the variance of the random variable **y**, σ — conditional (in terms of **x** factors) variance of the dependent variable (variance of model error).

R^2 describes the proportion of variation in the dependent variable due to regression or variability of the explanatory variables. The closer R^2

is to unity, the better the constructed regression model describes the relationship between the explanatory and the dependent variable. In the case of $R^2 = 1$, the studied relationship can be interpreted as functional (rather than statistical), which requires additional qualitative and quantitative information and changes in the research process [9] [10].

IV. MODEL CREATION AND DATA USED

To create forecast models, the Python programming language was used, as well as several TensorFlow and Scikit-learn libraries [11] [12]. The forecast model was based on neural networks that were created using TensorFlow. The list of variables that were predicted within the framework of this article are indicated below:

- Electricity price - knowing the future price, utilities will be able to better handle their electricity purchases, operators will be better able to use spare capacity.
- Solar power generation - renewable energy production is an unreliable parameter that introduces instability to the grid. With a more accurate forecast, operators can better manage the power system.
- Load - having an accurate forecast for consumption, you can make more accurate purchases, as well as reduce the amount of reserved power
- Cross-border power flows DE>CZ, DE>PL, DE>AT - in the past few years, the neighboring countries of Germany have had problems with overloading high voltage electric lines due to the unregulated capacity of renewable energy sources. Having a more accurate forecast allows operators to more accurately plan the transmission of energy.

The data for this article was taken from the Open Power System Data and ENTSOE Transparency Platform [13] [14]. Two datasets were generated. The first dataset was used to predict the first 3 parameters (Prices for electricity on the stock exchange in the Czech Republic, Power generated by solar power plants in the Slovak Republic, Load of the power system of the Slovak Republic). The second dataset was created to predict power flows between Germany and 3 countries (Poland, Austria, Czech Republic). Dataset parameters are presented in the table below.

TABLE I
First dataset data

Name features	Variable size	Units
utc_timestamp	43822	Date,time
SK_load_actual_entsoe_transparency	43822	MW
SK_load_forecast_entsoe_transparency	43822	MW
SK_solar_generation_actual	43822	MW
Price_day_ahead	43822	€/MWh
SK_temperature	43822	°C
SK_radiation_direct_horizontal	43822	W/m^2
SK_radiation_diffuse_horizontal	43822	W/m^2

TABLE II
Second dataset data

Name features	Variable size	Units
DE_load_actual_entsoe_power_statistics	8760	MW
DE_price_day_ahead	8760	€/MWh
DE_solar_capacity	8760	MW
DE_solar_generation_actual	8760	MW
DE_wind_capacity	8760	MW
DE_wind_generation_actual	8760	MW
DE > AT	8760	MW
DE > PL	8760	MW
DE>CZ	8760	MW
DE > NL	8760	MW
DE > CH	8760	MW
DE > FR	8760	MW
DE > DK	8760	MW

After creating the datasets, an exploratory analysis was performed, within which some data injections were identified that had to be eliminated in order to improve the accuracy of the model. The load graph of the Slovak Republic is presented as an example. The initial schedule is shown in Figure 4.

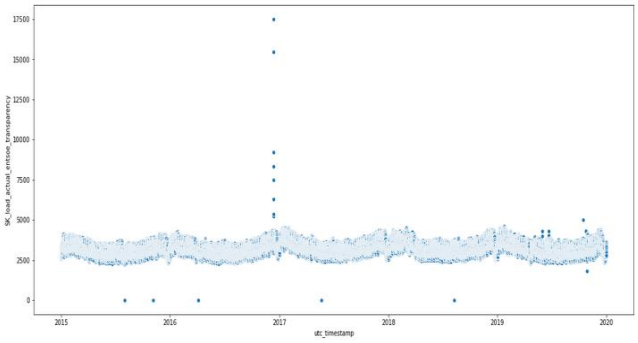

Fig. 4. SK_load_actual_entsoe_transparency

As you can see in the figure, hours with zero load periodically appear in the country's power system, which cannot be. Also, one day it can be seen through a high load, which at its peak reached 17,500, which also cannot be, since the Slovak Republic does not have such an established power. This stuffing’s have been replaced with averages. The corrected graph is shown in Figure 5.

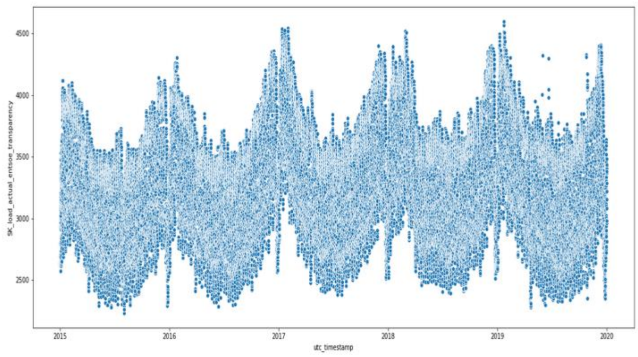

Fig. 5. SK_load_actual_entsoe_transparency

V. ANALYSIS OF THE OBTAINED RESULTS

To build a model and evaluate the results obtained, the original dataset was divided into two samples, data for training the model and checking the result. Statistical metrics of the obtained results are presented in Table 3. There are also graphs that show how much the prediction results correspond to real data. The lowest accuracy turned out to be for the price, this is explained by the fact that the price is influenced by players from different countries, in order to increase the accuracy of the model, it is necessary to use the data of all countries that work on this exchange. The best results are the predictions of cross-border power flows between Germany and Austria, this can be explained by good data, as well as by the stability of electricity exchange between these countries. The quality of load predictions exceeded the forecast provided by the operators of the Slovak Republic for ENTSOE, which indicates the high quality of the model.

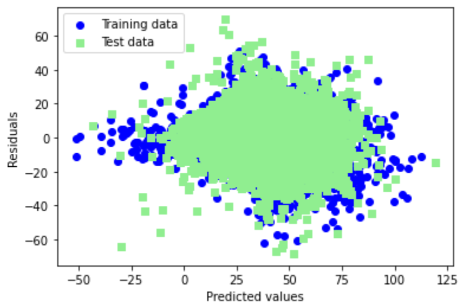

Fig. 4. Real and Forecast Price Prediction Data Chart

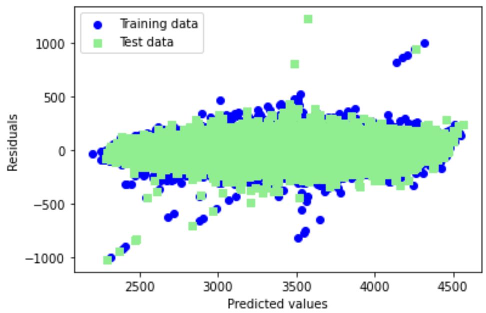

Fig. 4. Plot of reaD l and predicted load prediction data

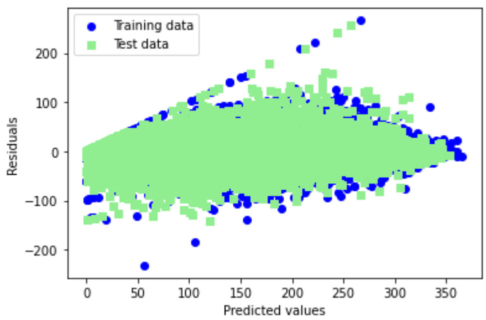

Fig. 4. Plot of real and predicted data on the prediction of power generated by solar power plants

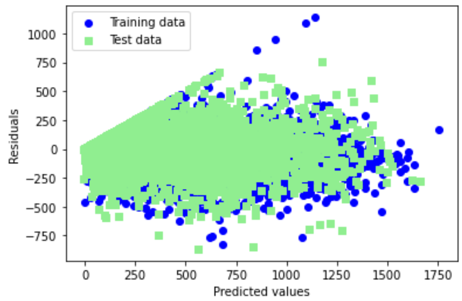

Fig. 4. Plot of real and predicted data cross-border power flows DE> CZ

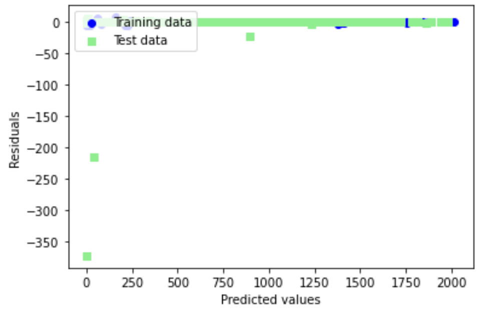

Fig. 4. Plot of real and predicted data cross-border power flows DE>AT

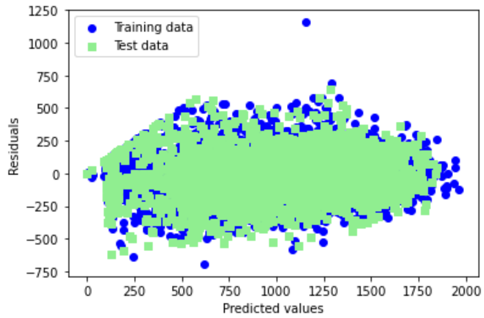

Fig. 4. Plot of real and predicted data cross-border power flows DE>AT

TABLE III
Prediction accuracy assessment

Prediction	MAE	MSE	R^2
Electricity price	6.419	9.266	0.669
Solar power generation	11.041	21.947	0.941
Load	62.188	83.815	0.964
Cross-border power flows DE>CZ	119.534	174.012	0.824
Cross-border power flows DE>PL	124.798	161.167	0.836
Cross-border power flows DE>AT	0.41	4.527	0.999

CONCLUSION

The accuracy of the model predictions is quite high, especially for the energy generated by solar power plants; this is primarily due to the constancy of solar radiation. electricity price forecasts are not very accurate. this is mainly due to the fact that not only the Czech Republic, but also a number of neighboring countries have access to the exchange (where the price is set), to improve the accuracy of forecasts, we need data on energy consumption, as well as on energy produced by renewable energy sources in countries which participate in the auctions. the article confirmed that the amount of power passing through Poland and the Czech Republic has a strong dependence on the amount of energy that is generated in wind farms. with the increase in the number of renewable energy sources, as well as the number of smart meters, the relevance of machine learning in the electric power industry will only increase.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-19-0576.

REFERENCES

[1] S. Borlase, “Smart Grids: Infrastructure, Technology and Solutions”, 577, (CRC, 2013) ISBN: 978-1-4398-2905-9

[2] A. Mogilenko, “Application of artificial intelligence algorithms in the global energy industry”, in Energy and Industry of Russia, 2018, no 7, p. 23.

[3] A. Losev, “Artificial Intelligence in Nuclear Power”, in Atomic Expert, 2018, no 3-4, pp. 35-43.

[4] E. Kamchatova, A. Vasilieva, N. Lyasnikov, M. Dudin, N. Vysotskaya, “Energy saving management in urban economy and industry” in International Journal of Civil Engineering and Technology, 2018, vol. 9, no 6, pp. 1423-1429.

[5] American Wind Energy Association, “U.S. wind industry annual market report year ending 2012,” Amer. Wind Energy Assoc., Washington, DC, USA, 2013 [Online]. Available: http://awea.files.cmsplus.com/images/AWEA_USWindIndustryAnnualMarketReport2012_ Executive Summary.pdf.

[6] “How does the domestic business connect to digitalization projects in the power industry”, 24 RosInfo, (Unpublished work style),” unpublished

[7] Claesen, Marc, and Bart De Moor. "Hyperparameter Search in Machine Learning." arXiv preprint arXiv:1502.02127 (2015)".

[8] M. Monfared, H. Rastegar, and H.M. Kojabadi, “A new strategy for wind speed forecasting using artificial intelligent methods”, Renew. Energy, vol. 34, pp. 845–848, 2009

[9] D. A. Freedman, Statistical Models: Theory and Practice, Cambridge University Press. p. 26.

[10] Warne, T. Russell (2011). "Beyond multiple regression: Using commonality analysis to better understand R2 results". Gifted Child Quarterly. 55 (4): 313–318. doi:10.1177/0016986211422217

[11] Martín Abadi, Ashish Agarwal, Paul Barham and others, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[12] Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.

[13] ENTSOE Transparency Platform. [Online]. Available: https://transparency.entsoe.eu/dashboard/show

[14] Open Power System Data. [Online]. Available: https://data.open-power-system-data.org

ADDRESSES OF AUTHORS

Ing. Maksym Oliinyk, Technická Univerzita Košice, Katedra elektroenergetiky, Mäsiarska 74, Košice, SK 04210, Slovenská Republika, Maksym.oliinyk@tuke.sk

Doc. Ing. Jaroslav Džmura, PhD, Technická Univerzita Košice, Katedra elektroenergetiky, Mäsiarska 74, Košice, SK 04210, Slovenská Republika, Jaroslav.dzmura@tuke.sk

Ing. Jozef Humeník, Technická Univerzita Košice, Katedra elektroenergetiky, Mäsiarska 74, Košice, SK 04210, Slovenská Republika, Jozef.humenik@tuke.sk

Ing. Peter Havran, Technická Univerzita Košice, Katedra elektroenergetiky, Mäsiarska 74, Košice, SK 04210, Slovenská Republika, Peter.havran@tuke.sk

Ing. Robert Štefko, Technická Univerzita Košice, Katedra elektroenergetiky, Mäsiarska 74, Košice, SK 04210, Slovenská Republika, Robert.stefko@tuke.sk

Ing. Daniel Pál, Technická Univerzita Košice, Katedra elektroenergetiky, Mäsiarska 74, Košice, SK 04210, Slovenská Republika, Daniel.Pal@tuke.sk