

S. V. SHEVCHENKO, D. O. DRUPPOV, M. I. BEZMENOV

SITUATIONAL FORECASTING OF ELECTRICITY DEMAND IN THE REGION

The process of forecasting volumes of electricity sales on the wholesale market is considered. To improve the quality of the forecast, it is proposed to use the method of machine learning Random Forests as part of the solution of the task of situational forecasting of electricity consumption. A comparison of the Random Forests with a simple linear regression is performed. The forecast is based on historical data on electricity consumption in Ukraine, as well as changes in cost per hour of consumption and a number of key factors. Forecasting takes into account weather conditions, macro – financial and economic characteristics. When the software was implemented, the library used includes the implementation of the prognostic algorithms Spark MLlib, which specializes in machine learning methods. Training samples were created based on historical data found in different open sources. In the introduction section the justification of the problem of forecasting demand for electricity and the impossibility of taking into account all factors affecting the environment when using standard approaches is made. In the results section, a number of indicators have been calculated capable of determining the accuracy of the forecast: the mean square error, the mean relative error and the absolute error. In the mathematics section, a description and analysis of the Random Forests algorithm was given. The graphs were built showing the results of the forecast in different time periods: one day, one week, one year. The results were compared with the original historical data. Added tables that show the input data and the results obtained using linear regression and the algorithm of machine learning Random Forests. In conclusion, conclusions were drawn about the effectiveness of the algorithm Random Forests, as well as a possible problem when working with machine learning algorithms.

Keywords: Forecasting, Electric Power System, Smart Grid, Machine Learning, Random Forests, Algorithm.

С. В. ШЕВЧЕНКО, Д. О. ДРУПPOВ, М. І. БЕЗМЕНОВ

СИТУАЦІЙНЕ ПРОГНОЗУВАННЯ ПОПИТУ НА ЕЛЕКТРОЕНЕРГІЮ В РЕГІОНІ

Розглядається процес прогнозування обсягів продажу електроенергії на оптовому ринку. Для поліпшення якості прогнозу пропонується використання методу машинного навчання Random Forests у складі вирішення задачі ситуаційного прогнозування споживання електроенергії. Робиться порівняння методу Random Forests з звичайною лінійною регресією. Прогноз виконується на базі історичних даних споживання електроенергії в Україні, а також змін вартості за одну годину споживання електроенергії і ряду ключових факторів. При прогнозуванні враховувалися показники погодних умов, макрофінансові та економічні характеристики. При реалізації програмного забезпечення була використана бібліотека Spark MLlib, що містить реалізацію прогностичних алгоритмів та спеціалізується на методах машинного навчання. Для алгоритмів були створені навчальні вибірки на базі історичних даних, знайдених у різних відкритих джерелах. У розділі Вступ робиться обґрунтування актуальності задачі прогнозування попиту електроенергії і неможливість врахування всіх факторів, що впливають на середовище при використанні стандартних підходів. Обчислено ряд показників, здатних визначити точність прогнозу: середня квадратична помилка, середня відносна і абсолютна помилки. Побудовано графіки, що відображають результати прогнозу в різних часових періодах: один день, один тиждень, один рік. Було виконано порівняння отриманих результатів із вихідними історичними даними. У розділі математичного обґрунтування наводиться детальний опис і аналіз алгоритму Random Forests. Додано таблиці, в яких наведені вхідні дані і отримані результати за допомогою лінійної регресії і алгоритму машинного навчання Random Forests. На завершення було зроблено висновки щодо ефективності алгоритму Random Forests, а також можливі проблеми при роботі з алгоритмами машинного навчання.

Ключові слова: прогнозування, електроенергетична система, Smart Grid, Машинне навчання, Random Forests, алгоритм.

С. В. ШЕВЧЕНКО, Д. А. ДРУПPOВ, Н. И. БЕЗМЕНОВ

СИТУАЦИОННОЕ ПРОГНОЗИРОВАНИЕ СПРОСА НА ЭЛЕКТРОЭНЕРГИЮ В РЕГИОНЕ

Рассматривается процесс прогнозирования объемов продаж электроэнергии на оптовом рынке. Для улучшения качества прогноза предлагается использование метода машинного обучения Random Forests в составе решения задачи ситуационного прогнозирования потребления электроэнергии. Делается сравнения метода Random Forests с обычной линейной регрессией. Прогноз выполняется на базе исторических данных потребления электроэнергии в Украине, а также измененной стоимости электроэнергии за один час потребления и ряда ключевых факторов. При прогнозировании учитывались показатели погодных условий, макрофинансовые и экономические характеристики. При реализации программного обеспечения была использована библиотека Spark MLlib, которая содержит реализацию прогностических алгоритмов и специализируется на методах машинного обучения. Для алгоритмов были созданы обучающие выборки на основании исторических данных, найденных в различных открытых источниках. В разделе Введение делается обоснование и актуальность задачи прогнозирования спроса электроэнергии и невозможность учета всех факторов, оказывающих влияние на среду при использовании стандартных подходов. Вычислено ряд показателей способных определить точность прогноза: средняя квадратическая ошибка, средняя относительная и абсолютная ошибки. В разделе математического обоснования приводится детальное описание и разбор алгоритма Random Forests. Построены графики, отображающие результаты прогноза в различных временных периодах: один день, одна неделя, один год. Было выполнено сравнение полученных результатов с исходными историческими данными. Добавлены таблицы, в которых приведены входные данные и полученные результаты с помощью линейной регрессии и алгоритма машинного обучения Random Forests. В заключении были сделаны выводы об эффективности алгоритма Random Forests, а также возможной проблеме при работе с алгоритмами машинного обучения.

Ключевые слова: прогнозирование, электроэнергетическая система, Smart Grid, Машинное обучение, Random Forests, алгоритм.

Introduction. Operational planning and effective management of the electricity system (ES) are impossible without reliable forecast of the load in the nodes of the calculation scheme, conducted in advance within a few minutes to several days. The forecast of nodal loads is necessary for optimization and correction of current modes, consideration of operational dispatch requests related to the withdrawal of electric equipment for repair,

for examinations, and others. A particular importance is given to this problem in connection with the transition to a competitive electricity market. Load nodes is considered consumption of active and reactive power in the substitution circuit of the electrical network. The load at the node for each time period is determined by the loadings of the set of receivers of the electrical network connected to this EC node, and the loss of power when

transmitting electricity to this node [1]. The need for qualitative prediction is a result of technological and economic reasons [2]. Technological reasons are connected with the key role of forecasting in the processes of planning energy balances and power system capacity, determined by the modes of parameters: technical and economic indicators of electricity consuming objects and calculations of electrical loads in power nodes and sections.

Forecasting of future indicators of electric consumption allows achieving the most important principle of forming a reliable and efficient operation of the Unified Energy System ensuring a clear system balance of production and consumption of electric energy under the conditions of simultaneous momentum of these processes. Balance electricity production and consumption is the basis of technological stability of the power system, it determines the system reliability and continuity of its operation. In case of violations of the balance conditions, the quality of the electricity deteriorates (changing the parameters of the frequency and voltage of the network), which reduces the efficiency of the work of the consumer equipment of end users.

There is a number of economic reasons that necessitate qualitative forecasting. Accurate calculations provide an optimal distribution of the load between power plants in the grid and improve the quality of electricity. Forecasting and planning of electricity consumption indicators for large consumers allows managing the cost of purchasing electricity through the regulation of equipment loading with the help of production processes management, transforming the main volumes of electricity consumption in the hours with the lowest cost, thereby reducing the cost of production and the amount of payments to energy supplying organizations.

A particular relevance the forecasting task was acquired after the emergence of the wholesale electricity market, the rules of which require the need for accurate forecasting of the electricity volume in order to make its purchase in the wholesale market [3].

In accordance with these rules, market participants should make a forecast of their planned hourly consumption and the day before the operational one, to submit an hourly request for electricity consumption to the system administrator for every hour of operating days. It means that all market players should fulfill their electricity consumption forecasts, as well as the hourly discretion for a few days ahead. Substandard submission by the enterprise of a participant in applications for electricity can lead to significant economic losses. Mistakes in forecasting reduce the quality of management of the electricity supply and worsen the cost-effectiveness of its complex regimes. In case of deviation of the actual hourly rates from the forecast, the participants are imposed with penalties that increase the cost of purchased electricity. This is due to the feature of electricity as a commodity. Under estimation of the forecast leads to the need to use emergency power. Exaggeration of the forecast leads to an increase in the cost of maintaining reserve power in the operating state. For the consumer that acquires electricity

on the wholesale and (or) retail market, there is the problem of accurate application power consumption for some time ahead. This is due to the fact that excessive or inadequate electricity consumption booked by the previous applications of the enterprise-user leads to unscheduled costs of the supplier at the point of generation. Therefore, increasing the accuracy of the forecast even by a tenth of a percent can lead to a significant reduction in the cost of paying deviations from the plan for the supply of electricity [4]. It turns out that the task of forecasting is highly relevant for a large number of players operating in the wholesale electricity market: guaranteeing suppliers serving the consumers of entire regions, independent energy sales organizations serving individual industrial enterprises in different regions, and for large economic entities that buy electricity on the wholesale market for their own consumption needs. Therefore, qualitative forecasting of electricity consumption for the subjects of the wholesale market is economically feasible and in today's highly competitive conditions, it receives an increasing relevance. The article describes a detailed description of the Random Forests algorithm and based on historical data for previous years makes a forecast that is compared with the forecast of linear regression and real historical data.

Task setting. One of the distinguishing features of the technological process of making decisions on the planning of electricity supply is the cyclicity (repetition) and the relationship of problems being solved to input and output data. As a result on of the accuracy and reliability of the electricity consumption forecast depends on the accuracy of the solution of the load optimal distribution tasks between the generating capacities, the efficiency of the management of the combined energy system and the management of power consumption and, that is especially important in market conditions, the wholesale price of purchases from the generating companies and, hence, the wholesale price of sales to electricity suppliers at the wholesale electricity market.

When moving to the free trade sector, in addition to winning from participation in competitive bidding, the wholesale market entity takes some risks associated with the impossibility of an exact planning of the application for electricity consumption. Deviations of actual consumption from declared values above a certain percentage values lead to the purchase of electricity from a balancing market at a higher price. Deviation to the smaller side is also punishable by payment of non-delivered electricity, determined by the difference between the declared and actual consumption at the established rates. The forecast is especially responsible for those power systems that do not have their own generating capacities, and there is no possibility to influence the electricity loads of consumers. The complexity of the forecast of electricity consumption is due to the presence of a large number of consumers and the need to take into account many factors affecting the consumption of electricity. This is a ambient temperature, degree of illumination, longitude of the day, day of the week, transitions from winter to summer time and back, the

presence of extraordinary events (catastrophes, mass actions), weather forecasts, the state of other factors affecting the change in consumption in accordance with the data obtained from the processing of consumption statistics, planned inclusion/disconnection of energy-intensive industries. For the solution of the problem of forecasting power consumption, traditional statistical models (regression models and time series models), models based on expert systems, artificial neural networks (ANN), and machine-driver algorithms can be used.

Ways of solution. The choice of the optimal method for solving a particular practical problem is a separate and rather difficult task. Up to nowadays in many power-systems in the world, when creating forecasting models, statistical methods of analysis are used: dynamic (time) rows, that are ordered sequences of observations of a process that changes over time. A sufficient number of mathematical stochastic methods of forecasting electric load schedules is known, their practical implementation in order to reduce the error of calculations requires the collection and subsequent use of significant volumes of output retrospective data of real hourly production of electricity. Determination and correction of errors that is maintaining a certain level of account correctness affecting the forecast result requires additional amount of data.

In connection with the transition to the balancing market and the market of bilateral agreements between direct producers and suppliers of electricity to solve the problem of prediction of electrical load in the ES, including one day ahead, can find and use new information technologies that are developed on the basis of the device expert systems, artificial neural networks, cellular automata, machine learning and others.

Mathematics. The decision trees are the greedy algorithm for building the chain if then else for constructing data forecasting models. Allows getting stable solutions that are comparable to SVM [5] and Neurons networks [6], while not using the high computing power required by the previous ones.

The most effective method of improving any algorithms based on deciduous trees was proposed by Leo Breiman and Adele Cutler in 2001 [7, 8]. They proposed an algorithm called Random Forests. Its main idea is use one tree instead of a whole ensemble of deciduous trees built with a few modified algorithms. At the same time, the essence of the struggle with the problem solving of ineffective selection of features was to use in the process of constructing a tree of some random samples that removes the determinism of tree construction and makes this process stochastic. Let us proceed directly to the description of the algorithm. We give a statement of the problem in the general case. Let us give a set of A power objects N , a component of the learning sample and has a set of $(M + 1)$ attributes. The set of first M attributes is denoted by Q . For a given set A all $(M + 1)$, the attributes are known. For other (new) known elements of the first M attributes, the target $(M + 1)$ attribute is needed to be found. At the same time, the parameter N ,

the number $M \geq 1$, parameters $m \leq M, mq \leq M$, some parameter $r (0 < r \leq 1)$ and the parameter number of trees in the ensemble $NTrees \geq 1$ are given to the input. A generalized prediction algorithm based on Random Forests looks like:

Basing the original training set A , a random sample with rN repetitions is generated. Basing on the generated sample, a decision tree is constructed (in the general case, any algorithm for constructing deciduous trees can be used for this task). Moreover, during the construction of the next node of a tree with M of existing features, on the basis of which it is possible to divide the tree, we choose mq random. The partition decision is made on the basis of the best of $mq \leq M$ selected characteristics (ie, based on the application of the branching criterion to mq signs).

The procedure is repeated $NTrees$ times. The resulting $NTrees$ random trees combine into an ensemble. The classes are defined for the next new element, (its value $(M + 1)$ parameter) using all $NTrees$ -constructed trees and choose the resulting class for the one that "voted" by the most trees. Each tree is constructed using the following algorithm in fig. 1:

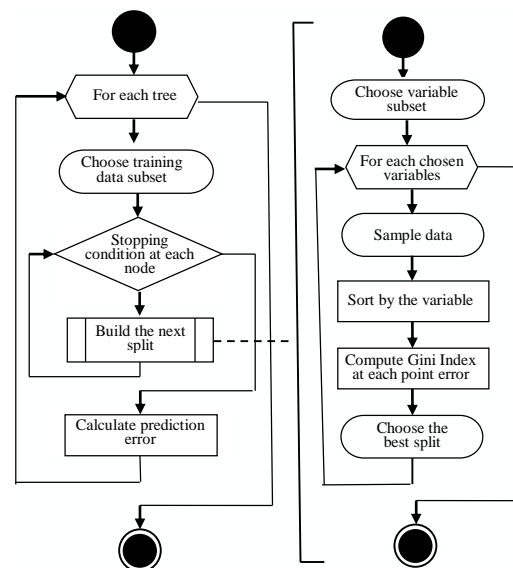


Figure 1 – Flow chart of Random Forests algorithm

Let the number of training cases be N and the number of variables in the classifier be M .

- It is given the number m of input variables to be used to determine the decision at a node of the tree; m should be much less than M .
- Choose a training set for this tree by choosing n times with replacement from all N available training cases (i.e. take a bootstrap sample). Use the rest of the cases to estimate the error of the tree, by predicting their classes.
- For each node of the tree, randomly choose m variables on which to base the decision at that node. Calculate the best split [9] based on these m variables in the training set.

- Each tree is fully grown and not pruned (as may be done in constructing a normal tree classifier).

For prediction, a new sample is shifted down the tree. It is assigned the label of the training sample in the terminal node it ends up in. This procedure is iterated over all trees in the ensemble and the average vote of all trees is reported as Random Forests prediction.

Consider the complexity of the proposed algorithm based on the random forests. We will consider the variant of the algorithm's operation with the following parameters: $r = 1, mq = M$. With these values of parameters, the computational complexity of the algorithm will be greatest. Then the complexity of constructing a single tree will coincide with the complexity of constructing the original deciduous tree without modifications. In this case the complexity of the entire algorithm for constructing a random forests will be determined as follows:

$$Trf = O(T \cdot NTrees) = O(NTrees \cdot t^2 \cdot n^m).$$

This complexity can be reduced if you do not build the entire forest at once and engage in the construction of only those branches in each tree from the forest, which are needed to predict this particular element. Taking as a basis the proposed adaptive algorithm for constructing trees in a random forests, we asymptotically reduce the complexity to the following value:

$$Trf = O(T \cdot NTrees) = O(NTrees \cdot k \cdot t^2),$$

where k is the number of elements that need to be predicted. We call the received modification of the "Random Forests" algorithm in application to the task of forecasting the time series of the adaptive algorithm "Random Forests".

Methods. The method described here is simple, a learning sample was created, based on historical data that was taken from different public sources. That training sample was used by the Random Forests algorithm, in our case, we used the Spark MLlib [10] software library to solve the regression problem, it supports different algorithms and Random Forests in particular. The training sample was presented in LIBSVM [11] format. In this format, each line represents a single set of data with the result and attributes values.

The first value is the result, others represent the input data. Example lines in LIBSVM format using formula:

$$y(\text{result})1: x1: 2: x2: 3: x3 \dots n: xn.$$

Then simple software program was implemented that able to load training sample and make a prediction using Spark MLlib [10].

After the training sample was loaded, the training was completed, and algorithm could make a prediction based on new LIBSVM data in vector format. The complexity of working with Spark MLlib comes down to find a relevant training sample.

Results. Overall the results presented below show that forecasting of electric energy consumption includes a number of factors among that a considerable importance takes weather conditions: temperature, humidity, wind speed, and others. Weather indicators are essential factors that influence the consumption of electricity, during the analysis of historical data, it was concluded that at low temperatures and high wind speed, electricity consumption figures are increasing. Thus, indicators for winter periods exceed the rates for the warmer ones: in the fall or in the summer season. The main task when working with the algorithms of machine learning is the collection of historical data. During the computation of the results the following data were collected. The weather data was taken from the site [12], and data on electricity consumption were obtained from site [13] in the interval of 3 years 2006–2008, based on which the learning algorithm was made. It is also worth mentioning that weather conditions are far from all indicators that affect consumption, but weather is a significant criterion in forecasting demand for electricity in the region. An important factor is the large amount of historical data, so that the algorithm can do the training.

To test the Random Forests algorithm a sample in the LIBSVM data format was created.

In tab. 1 contains data for the weekly chart. For a monthly forecast, the data has a similar structure to which month and year are added. Also, the table shows the forecast data using linear regression to compare the Random Forests algorithm with standard forecasting approaches.

From the graphs in fig. 2 and fig. 3 the results of the forecast have rather good performance. To ensure this, the mean squared error (MSE) for the given formula is calculated:

$$m = \sqrt{\frac{\Delta_1^2 + \Delta_2^2 + \dots + \Delta_n^2}{n}} = \sqrt{\frac{\Delta^2}{n}}, \quad (1)$$

where n – the number of forecasts;

Δ – the difference between historical data and predicted.

The average absolute value of the error is calculated as:

$$\Delta'_{\text{av. abs. err.}} = \frac{1}{n} \sum_{t=1}^n |y_t - y_t^*|, \quad (2)$$

Average relative error (approximation error) is calculated as:

$$\epsilon_{\text{av. rel. err.}} = \frac{1}{n} \sum_{t=1}^n \frac{|y_t - y_t^*|}{y_t}, \quad (3)$$

where y_t – historical value,

y_t^* – predicted value

n – the number of forecasts.

Table 1 – Input data and Results for forecasting electricity demand

Historical data	Forecast		Weather indicators		
Demand for electricity (MW*hour)	Demand for electricity RF	Demand for electricity REG	Air Temperature (C)	Air humidity, %	Wind speed, ms
590	586.4	579	-15	70	4
580	574.3	567.3	-12.9	65	5
570	566.7	564.3	-15.9	75	7
590	584.4	577.1	-16.5	65	4
600	591.1	582.5	-14.9	80	4
590	584.4	581.3	-15.2	80	1
600	593.7	588.3	-14.2	80	4

From the graphs in fig. 2 and fig. 3 the results of the forecast have rather good performance.

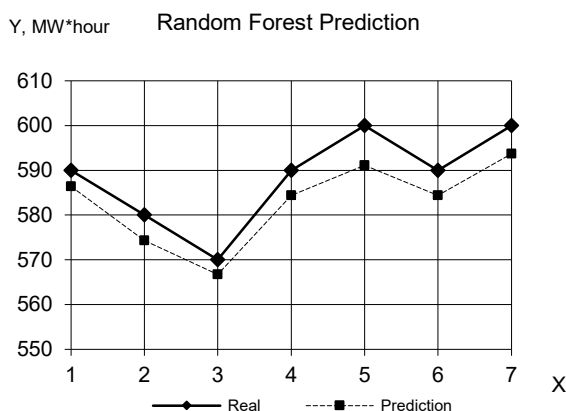


Figure 2– Results of the Random Forests algorithm for a weekly forecast

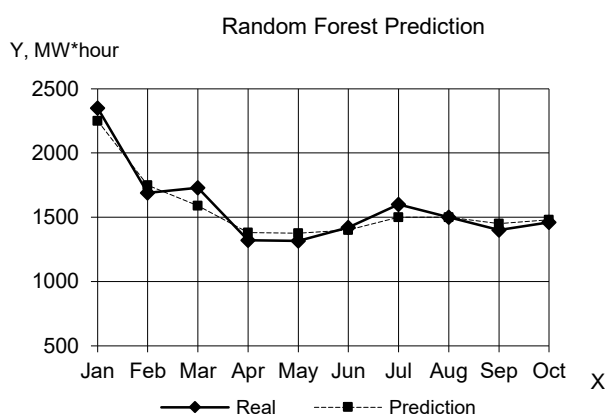


Figure 3 – Prediction of Electricity Demand Random Forests method

To ensure this, the mean squared error (MSE) for the given formula is calculated: forecast data using linear regression to compare the Random Forests (RF) algorithm with standard forecasting approaches (REG) using formula (1).

To estimate the accuracy of the Random Forests algorithm, we calculate MSE for regression from tab. 1:

$$m_{REG} = \sqrt{\frac{(590 - 579)_1^2 + \dots + (600 - 588.3)_7^2}{7}} = 11.98.$$

In the same way

$$m_{RF} = \sqrt{\frac{(590 - 586.4)_1^2 + \dots + (600 - 593.7)_7^2}{7}} = 5.83.$$

We have

$$m_{RF} < m_{REG}.$$

The relative average and absolute mean errors for Linear Regression by formulas (2)–(3) were calculated:

$$\epsilon_{av. rel. err. REG} = \frac{16.26}{7} \approx 2.32,$$

$$\Delta'_{av. abs. err. REG} = \frac{96.2}{7} \approx 13.74.$$

The relative average and absolute mean errors for Random Forests were calculated

$$\epsilon_{av. rel. err. RF} = \frac{6.72}{7} \approx 0.96,$$

$$\Delta'_{av. abs. err. RF} = \frac{39.1}{7} \approx 5.57.$$

In tab. 2 the test data for forecasting the cost of electricity in the wholesale market is presented.

For the forecast a number of indicators influencing the price characteristics were selected: the cost of raw materials: natural gas and coal were taken from the site [14, 15], the dollar exchange rate for the current month was obtained from the site [16] as well as the wholesale average monthly price per hourly segments for one month was taken from the site [13]. Also, there are a few essential global macroeconomic parameters as Gross

Table 2 – Results and input data for the forecast of the wholesale price of the electricity power per day

Forecast wholesale price RF (kop / kW*hour)	Forecast wholesale price REG (kop/ kW*hour)	Historical wholesale prices (kop / kW*hour)	Cost of natural gas (Dollars for 1 MMBTU)	Demand for coal (million tons)	Hours	Dollar exchange rate (UAH)	Gross domestic product (UAH million)	GDP Per capita income (USD)	Size of investments (USD million)
Year 2006									
13.50	16.50	13.55	7.34	40.6	1	5.05	544153	7070	5737
13.45	14.30	12.91	7.34	40.6	2	5.05	544153	7070	5737
13.21	17.40	12.87	7.34	40.6	3	5.05	544153	7070	5737
12.79	18.79	12.86	7.34	40.6	4	5.05	544153	7070	5737
...
13.06	15.05	12.86	7.34	40.6	24	5.05	544153	7070	5737
Year 2007									
14.11	17.09	13.90	7.36	40.7	1	5.03	720731	7 890	9218
11.23	16.04	10.70	7.36	40.7	2	5.03	720731	7 890	9218
10.91	17.49	10.47	7.36	40.7	3	5.03	720731	7 890	9218
11.23	19.91	10.36	7.36	40.7	4	5.03	720731	7 890	9218
...
13.54	19.45	14.05	7.36	40.7	24	5.03	720731	7 890	9218
Year 2008									
14.24	19.89	15.60	10.16	41.2	1	8.50	948056	8 320	9903
13.01	18.15	12.86	10.16	41.2	2	8.50	948056	8 320	9903
13.75	20.56	13.28	10.16	41.2	3	8.50	948056	8 320	9903
13.80	18.09	13.29	10.16	41.2	4	8.50	948056	8 320	9903
...
16.01	28.67	16.94	10.16	41.2	24	8.50	948056	8 320	9903

domestic product [17], GDP per capita [18] and foreign investments [19].

Figure 3 shows the forecast made by the Random Forests algorithm based on the data obtained from years 2006–2008. The forecast was drawn up for a period of 10 months. Axis X – months of the year. Axis Y – demand for energy in MW*hour.

Figure 4 shows the forecast of the calculation of the wholesale price for each hour (kop / kW*hour), predicted by the Random Forests algorithm based on the data obtained in the period 2006–2008 years. Axis X – hours of day. Axis Y – the amount on the wholesale market.

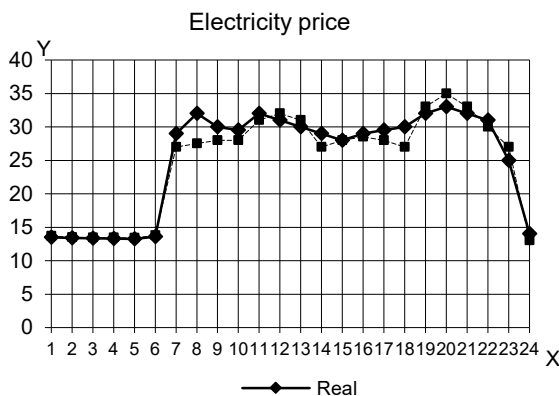


Figure 4 – Forecast of the Random Forests algorithm to determine the price per day

List of acronyms used in the tab. 2:

- RF – The Random Forests algorithm;
- REG – Algorithm of linear regression;
- MMBTU – British Thermal Unit;
- GDP – Gross domestic product.

Consequently, the mean square error, relative average error and absolute average error in the Random Forests algorithm is less than in the linear regression algorithm and can be verified that the Random Forests algorithm has given more precise metrics.

The MSE of the regression algorithm for tab. 2 using formula (1):

$$m_{REG} \approx \sqrt{\frac{874.45}{24}} \approx \sqrt{36.43} \approx 6.03.$$

The MSE of the Random Forests algorithm for tab. 2

$$m_{RF} \approx \sqrt{\frac{117.44}{24}} \approx \sqrt{4.89} \approx 2.2.$$

We have again

$$m_{RF} < m_{REG}.$$

The relative average and absolute mean errors for Linear Regression were calculated by formulas (2)–(3):

$$\varepsilon_{\text{av. rel. err. REG}} = \frac{16.26}{7} \approx 15.3,$$

$$\Delta'_{\text{av. abs. err. REG}} = \frac{96.2}{7} \approx 4.3.$$

The relative average and absolute mean errors for Random Forests were calculated:

$$\varepsilon_{\text{av. rel. err. RF}} = \frac{6.72}{7} \approx 6.46,$$

$$\Delta'_{\text{av. abs. err. RF}} = \frac{39.1}{7} \approx 1.84.$$

According to the results, the Random Forests algorithm and in the forecast for the wholesale price gave better performance than regression.

Conclusions. This work has been attempted to analyze of the one of the of Machine learning algorithms, which is used to predict the demand for electricity.

Prediction of Electricity demand is a complex task, a lot of factors play a key role, which are very difficult to consider. The Random Forests algorithm is really well suited for solving prediction problems. The main difficulty when working with the algorithm is to search for relevant historical data that will affect the course of the forecast. The calculated results showed that the standard prediction algorithms based on regression gave less accuracy in comparison with Random Forests algorithm. However, in order to improve the accuracy of the algorithm, a large statistic of historical data were required, based on which the algorithm will carry out training. It should also be noted that all algorithms of machine learning have the so-called shortcoming as a retraining, when some data may become irrelevant to reality and will only worsen the accuracy of the forecast. Here you should promptly detect such data and remove them from a training set.

References

1. Шумилова Г. П., Готман Н. Е., Старцева Т. Б. Прогнозирование активной и реактивной нагрузки узлов ЭЭС с использованием инверсии искусственной нейронной сети. *Управление электроэнергетическими системами – новые технологии и рынок: Сб. науч. тр.* Сыктывкар: Изд-во Коми НЦ УрО РАН, 2004. С. 115–122.
2. Поляхов Н. Д., Приходько И. А., Ван Ефэн. Прогнозирование электропотребления на основе метода опорных векторов с использованием эволюционных алгоритмов оптимизации. *Электронный научный журнал. Современные проблемы науки и образования.* 2013, № 2. URL: <http://vestnik.kpi.kharkov.ua> (accessed: 16.04.2018).
3. Шумилова Г. П., Готман Н. Е., Старцева Т. Б. Прогнозирование электрических нагрузок ЭЭС с использованием методов искусственного интеллекта. *РНЭЭ, 10–14 сентября 2001: Материалы докладов. Т. II.* Казань: Казан. гос. энерг. ун-т, 2001. С.103–106.
4. Breiman Leo, Friedman Jerome, Stone Charles J. Olshen R. A. *Classification and regression trees.* Wadsworth and Brooks, CA, 1984. 368 p.
5. *Support Vector Machines (SVM) Introductory Overview.* URL: <http://www.statsoft.com/Textbook/Support-Vector-Machines> (accessed: 16.04.2018).
6. Новотарський М. А., Нестеренко Б. Б. *Штучні нейронні мережі: обчислення.* Київ: Інститут математики НАН України, 2004.

- 408 с. URL: http://www.immsp.kiev.ua/postgraduate/Biblioteka_trudy/ShtuchnN_ejronMeregNester2004.pdf (accessed: 16.04.2018).
7. Breiman Leo. Random Forests. *Machine Learning.* 2001. Т. 45, № 1. P. 5–32.
8. Tin Kam Ho. Random Decision Forest. *Proceedings of the 3rd International Conference on Document 8. Analysis and Recognition.* Montreal, QC, 1995. P. 278–282.
9. Ishwaran H. *The Effect of Splitting on Random Forests.* URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC559918/> (accessed: 16.04.2018).
10. *MLlib is Apache Spark's scalable machine learning library.* URL: <https://spark.apache.org/mllib/> (accessed: 16.04.2018).
11. *Data Types – RDD-based API.* URL: <https://spark.apache.org/docs/2.0.2/mllib-data-types.html> (accessed: 16.04.2018).
12. *Weather archive in Kharkiv.* URL: [https://rp5.ru/Weather_archive_in_Kharkiv_\(airport\)](https://rp5.ru/Weather_archive_in_Kharkiv_(airport)) (accessed: 16.04.2018).
13. *Аналіз роботи ОЕС України по місяцям.* URL: <http://www.er.gov.ua/doc.php?c=24> (accessed: 16.04.2018).
14. *Динамика цены на природный газ.* URL: <https://www.calc.ru/dinamika-Gas.html> (accessed: 16.04.2018).
15. *Отраслевой обзор. Уголь Украины.* URL: http://www.bakertilly.ua/media/Baker%20Tilly%20-%20Report_coal_industry_rus.pdf (accessed: 16.04.2018).
16. *Курсы валют (доллар, евро) в 2008 году.* URL: <https://www.kurs.metrinfo.ru/kurs/> (accessed: 16.04.2018).
17. *Валовой внутренний продукт.* URL: <https://index.minfin.com.ua/economy/gdp/> (accessed: 16.04.2018).
18. *Ukraine – Gross domestic product per capita in current prices.* URL: https://knoema.com/atlas/Ukraine/GDP-per-capita?origin=knoema.ru&_ga=2.86031384.1897352500.1528129995-423093515.1528129995 (accessed: 16.04.2018).
19. *Прямые иностранные инвестиции.* URL: <https://index.minfin.com.ua/index/fdi/> (accessed: 16.04.2018).

References (transliterated)

1. Shumilova G. P., Gotman N. E., Startseva T. B. Prediction of active and reactive load nodes EPS using an artificial neural network inversion [Prognozirovaniye aktivnoy i reaktivnoy nagruzki uzlov EES s ispolzovaniyem inversii iskusstvennoy neyronnoy seti]. *Upravleniye elektroenergeticheskimi sistemami – novyye tekhnologii i rynek: Sb. nauch. tr.* [Power systems management – new technologies and market: Collection of scientific papers]. Syktyvkar, Komi NTs UrO RAN, 2004, pp. 115–122.
2. Polyakhov N. D., Prikhodko I. A., Van Ye. Forecasting electricity consumption based on support vector using evolutionary optimization algorithms [Prognozirovaniye elektropotrebleniya na osnove metoda opornykh vektorov s ispolzovaniyem evolyutsionnykh algoritmov optimizatsii]. *Elektronnyy nauchnyy zhurnal. Sovremennyye problemy nauki i obrazovaniya* [Electronic scientific journal. Modern problems of science and education], 2013, no.2.
3. Shumilova G. P., Gotman N. E., Startseva T. B. Prediction of electrical loads using artificial intelligence techniques [Prognozirovaniye elektricheskikh nagruzok s primeneniym metodov iskusstvennogo intellekta]. *RNSE, 10–14 sentyabrya 2001: Materialy докладов* [RNEE, September 10–14, 2001: Material of the reports]. Available at: <http://www.energy.komisc.ru/seminar/StShum1.pdf> (accessed 16.04.2018).
4. Breiman Leo, Friedman Jerome, Stone Charles J. Olshen R. A. *Classification and regression trees.* Monterey, CA: Wadsworth & Brooks, 1984.368 p.
5. *Support Vector Machines (SVM) Introductory Overview.* Available at: <http://www.statsoft.com/Textbook/Support-Vector-Machines>, (accessed: 16.04.2018).
6. Novatorskiy M. A., Nesterenko B. B. *Shtuchni neyronni merezhi: obchislennya* [Artificial Neural Networks: calculation]. Kiev, Instytut matematyky NAN Ukrai'ny Publ., 2004.408 p. Available at: http://www.immsp.kiev.ua/postgraduate/Biblioteka_trudy/ShtuchnN_ejronMeregNester2004.pdf (accessed: 16.04.2018).
7. Breiman Leo. Random Forests. *Machine Learning.* 2001,vol. 45, no 1, pp. 5–32.

8. Tin Kam Ho. Random Decision Forest. *Proceedings of the 3rd International Conference on Document 8. Analysis and Recognition*. Montreal, QC, 1995, pp. 278–2826.
9. Ishwaran Hemant. *The effect of splitting Random Forest*. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5599182/> (accessed 16.04.2018).
10. *MLlib is Apache Spark's scalable machine learning library*. Available at: <https://spark.apache.org/mllib/> (accessed: 16.04.2018).
11. *Data Types – RDD-based API*. Available at: <https://spark.apache.org/docs/2.0.2/mllib-data-types.html> (accessed: 16.04.2018).
12. *Weather archive in Kharkiv*. Available at: [https://tp5.ru/Weather_archive_in_Kharkiv_\(airport\)](https://tp5.ru/Weather_archive_in_Kharkiv_(airport)) (accessed: 16.04.2018).
13. *Analiz roboty OES Ukrainy po misiatsiam* [Analysis of the work of the UES of Ukraine by months]. Available at: <http://www.er.gov.ua/doc.php?c=24> (accessed: 16.04.2018).
14. *Dinamika tseny na prirodnyy gaz* [The dynamics of prices for natural gas]. Available at: <https://www.calc.ru/dinamika-Gas.html> (accessed: 16.04.2018).
15. *Otraslevoy obzor. Ugol Ukrainyi* [Industry review. Coal of Ukraine]. Available at: http://www.bakertilly.ua/media/Baker%20Tilly%20-%20Report_coal_industry_rus.pdf (accessed: 16.04.2018).
16. *Kursyi valyut (dollar, evro) v 2008 godu* [Exchange rates (dollar, euro) in 2008]. Available at: <https://www.kurs.metinfo.ru/kurs/2008-1-3/#start> (accessed: 16.04.2018).
17. *Valovoy vnutrenniy produkt* [Gross domestic product]. Available at: <https://index.minfin.com.ua/economy/gdp/> (accessed: 16.04.2018).
18. *Ukraine – Gross domestic product per capita in current prices*. Available at: https://knoema.com/atlas/Ukraine/GDP-per-capita?origin=knoema.ru&_ga=2.86031384.1897352500.1528129995-423093515.1528129995 (accessed: 16.04.2018).
19. *Pryamyie inostrannyye investitsii* [Direct Foreign investments]. Available at: <https://index.minfin.com.ua/index/fdi/> (accessed: 16.04.2018).

Received 27.04.2018

Відомості про авторів / Сведения об авторах / About the Authors

Шевченко Сергій Васильович (Шевченко Сергей Васильевич, Shevchenko Serhiy Vasylyovych) – кандидат технічних наук, професор НТУ «ХПІ», Національний технічний університет «Харківський політехнічний інститут», професор кафедри програмної інженерії та інформаційних технологій управління; ORCID: <https://orcid.org/0000-0002-3831-5425>; e-mail: sv-shevchenko@ukr.net.

Друппов Дмитро Олександрович (Друппов Дмитрий Александрович, Drupnov Dmytro Oleksandrovych) – бакалавр, Національний технічний університет «Харківський політехнічний інститут», студент кафедри «Програмна інженерія та інформаційні технології управління»; м. Харків, Україна; ORCID: <https://orcid.org/0000-0002-1318-0325>; e-mail: dmitry.drupnov@gmail.com

Безменов Микола Іванович (Безменов Николай Иванович, Bezmenov Mykola Ivanovych) – кандидат технічних наук, доцент, професор НТУ «ХПІ», Національний технічний університет «Харківський політехнічний інститут», професор кафедри системного аналізу та інформаційно-аналітичних технологій; ORCID: <https://orcid.org/0000-0002-2995-2350>; e-mail: bezmenov@kpi.kharkov.ua