

УДК 004.89

Т. В. Ермоленко, В. Н. Котенко, А. О. Винник

Государственное образовательное учреждение высшего профессионального образования
«Донецкий национальный университет»

283001, г. Донецк, ул. Университетская, 24

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ПРЕДСКАЗАТЕЛЬНЫХ МОДЕЛЕЙ ДЛЯ СИСТЕМЫ АНАЛИЗА И МОНИТОРИНГА ЭНЕРГОПОРЕБЛЕНИЯ НА ПРЕДПРИЯТИЯХ УГОЛЬНОЙ ПРОМЫШЛЕННОСТИ

T. V. Yermolenko, V. N. Kotenko, A. O. Vinnik

State Educational Institution of Higher Professional Education «Donetsk National Technical University»
283001, Donetsk, University st, 24

STUDY OF THE EFFICIENCY OF PREDICTIONAL MODELS FOR THE SYSTEM OF ANALYSIS AND MONITORING OF ENERGY CONSUMPTION AT COAL INDUSTRY ENTERPRISES

Статья посвящена актуальной на сегодняшний день задаче построения предсказательных моделей потребления объемов электроэнергии такими крупными предприятиями как предприятия угольной промышленности. Модели обучались и тестировались на данных шахты имени академика А. А. Скочинского за 2012-2019 гг. Проведенное исследование временного ряда показателей энергопотребления шахтой показало, что ряд является стационарным, в нем присутствуют тренд и сезонность. Это делает возможным применение регрессионных и авторегрессионных моделей прогнозирования. Для исследования использовались модели ARIMA и экспоненциального сглаживания. Кроме этих моделей исследовалась эффективность нейросетевых методов, в частности, моделей на основе многослойного персептрона и двунаправленной сети долгой краткосрочной памяти. Проведен сравнительный анализ эффективности этих моделей на тестовой выборке, который показал, что наиболее эффективной моделью является ARIMA.

Ключевые слова: временной ряд, стационарность, тренд, сезонность, ARIMA, модуль экспоненциального сглаживания, модель Хольта-Винтерса, многослойный персептрон, двунаправленная сеть долгой краткосрочной памяти.

The article is devoted to the task of constructing predictive models for the consumption of electricity volumes by such large enterprises as coal mining enterprises, which is relevant today. The models were trained and tested on the data of the mine named after Academician A.A. Skochinsky for 2012-2019 years. The study of the time series of indicators of energy consumption by the mine showed that the series is stationary, it has a trend and seasonality. This makes it possible to use regression and autoregressive forecasting models. For the study, ARIMA and exponential smoothing models were used. In addition to these models, the effectiveness of neural network methods was studied, in particular, models based on a multilayer perceptron and a bidirectional network of long short-term memory. A comparative analysis of the effectiveness of these models on a test sample was carried out, which showed that the most effective model is ARIMA.

Key words: financial time series, bars, Japanese candlesticks, tick chart, fractional differentiation, triple barrier method, Lopez de Prado two-stage labeling.

Введение

Методы искусственного интеллекта прочно вошли практически системы управления, внедренные в разные сферы производственной деятельности человека, заменив методы экспертных оценок. Не остались в стороне промышленные технологии электрического хозяйства предприятия, прикладные интеллектуальные системы, ориентированные на обработку и использование знаний, находят все большее применение в задачах производства и эффективного использования электроэнергии. Качественный прогноз обеспечивает баланс электроэнергии, необходимый для устойчивой работы энергосистемы, поскольку потребляемый объем электроэнергии влияет на выбор режимных параметров и расчетных электрических нагрузок, т.е. своевременное получение информации о предстоящей нагрузке позволяет выбрать оптимальный режим работы системы. В свою очередь, фактическое отклонение потребления электроэнергии от спрогнозированного объема приводит к тому, что предприятие либо закупает недостающий объем, либо продает излишки по заведомо невыгодным ценам.

В частности, для нашего региона, где угольная промышленность является основным экономическим стержнем и является одним из основных источников дохода, проведены исследования эффективности интеграции на предприятия добывающей отрасли системы анализа и мониторинга энергозатрат и установлено, что при ее внедрении энергозатраты сокращаются до 5% [1].

Таким образом, задача построения качественной модели прогнозирования потребляемых предприятием объемов электроэнергии является актуальной, а ее реализация способствует снижению затрат предприятия.

Следует учесть, что современные условия существования субъектов хозяйственной деятельности характеризуются нестабильностью социальных и экономических процессов, что делает методы анализа и прогнозирования параметров социально-экономических объектов на основе регрессионного анализа, ориентированные, в основном, на стационарные условия их развития, малоэффективными. Необходимость учета большого количества факторов, влияющих на показатели энергопотребления, представляет собой основную проблему прогнозирования электропотребления.

В условиях наличия детерминированной и недетерминированной информации одним из перспективных направлений исследования являются анализ и прогнозирование параметров электропотребления предприятиями с помощью нейросетевых моделей. Внедрение в практику управления моделей, сформированных на базе нейронных сетей, позволяет наиболее четко видеть взаимодействие возмущающих и управляющих воздействий для достижения желаемого результата.

Цель настоящего исследования – провести сравнительный анализ различных моделей прогнозирования потребления электроэнергии предприятием угольной промышленности, используя для обучения и тестирования моделей данные предприятия «Шахта имени академика А.А. Скочинского».

Для чего необходимо решить следующие задачи:

- осуществить визуальный анализ имеющихся данных;
- исследовать полученные временные ряды показателей электропотребления на стационарность, наличие сезонности и тренда;
- в результате исследования рядов выбрать методы прогнозирования и построить модели прогноза;
- оценить качество прогноза, полученного с помощью разработанных моделей.

Исследование временного ряда объемов потребления электроэнергии предприятием угольной промышленности

Целями разработки системы анализа и мониторинга потребления электроэнергии предприятием являются:

- наблюдение за изменением мощности и потребления энергии в режиме реального времени каждого устройства, на котором установлен датчик;
- вычисление доли энергопотребления каждого устройства от общего потребления предприятием;
- ведение базы данных о потреблении электроэнергии устройствами, установленными на предприятии;
- получение достоверного прогноза энергопотребления предприятием на период до одного года на основе ретроспективных данных, хранящихся в базе данных.

Последовательность накопленных за несколько предшествующих лет ретроспективных данных нагрузки представляет собой временной ряд. Как известно, временной ряд $x_0, x_1, \dots, x_t, \dots, x_n$ ($x_t \in R$) – последовательность измерений, упорядоченных в неслучайные моменты времени.

В общем виде при исследовании экономического временного ряда x_t выделяются несколько составляющих:

$$x_t = f_t + v_t + c_t + \varepsilon_t, t=1, 2, \dots, n,$$

где f_t – *тренд*, плавно меняющаяся компонента, описывающая чистое влияние долгосрочных факторов, т.е. длительную тенденцию изменения потребления;

v_t – *сезонная компонента*, отражающая повторяемость исследуемого показателя в течение не очень длительного периода;

c_t – *циклическая компонента*, отражающая повторяемость исследуемого показателя в течение длительных периодов;

ε_t – *случайная компонента*, отражающая влияние не поддающихся учету и регистрации случайных факторов.

В данном исследовании использованы данные шахты имени А.А. Скочинского за месяцы 2018 – 2019 гг. Визуализация полученных данных представлена на рис. 1.

По графику на рис. 1 видно, что можно выделить чёткую сезонность. Энергопотребление растёт к зимнему периоду и падает к летнему. Циклическая компонента отсутствует. Для более детального анализа исследуем коррелограмму, т.е. график выборочной автокорреляционной функции (АКФ).

На рис. 2 первое значение много больше нуля, что говорит о наличии тренда, который является доминирующим в этом временном ряде. Отличающиеся от нуля значения возникают через определенные промежутки, что говорит о наличии сезонности.

Найти сезонную компоненту в исходном ряде можно, применив сглаживание ряда. Наиболее простым случаем является сглаживание методом скользящего среднего с длиной сглаживающего участка k , в качестве которого выбирают период сбора данных [2]. Сглаживая ряд, избавляются от сезонности периода k . Для имеющегося временного ряда применен метод Хольта-Винтерса (модель тройного экспоненциального сглаживания) [2] на период десяти месяцев. В результате чего, можно убедиться в том, что тренд ряда – это прямая, около которой совершаются колебания значений в рамках доверительных интервалов (рис. 3).

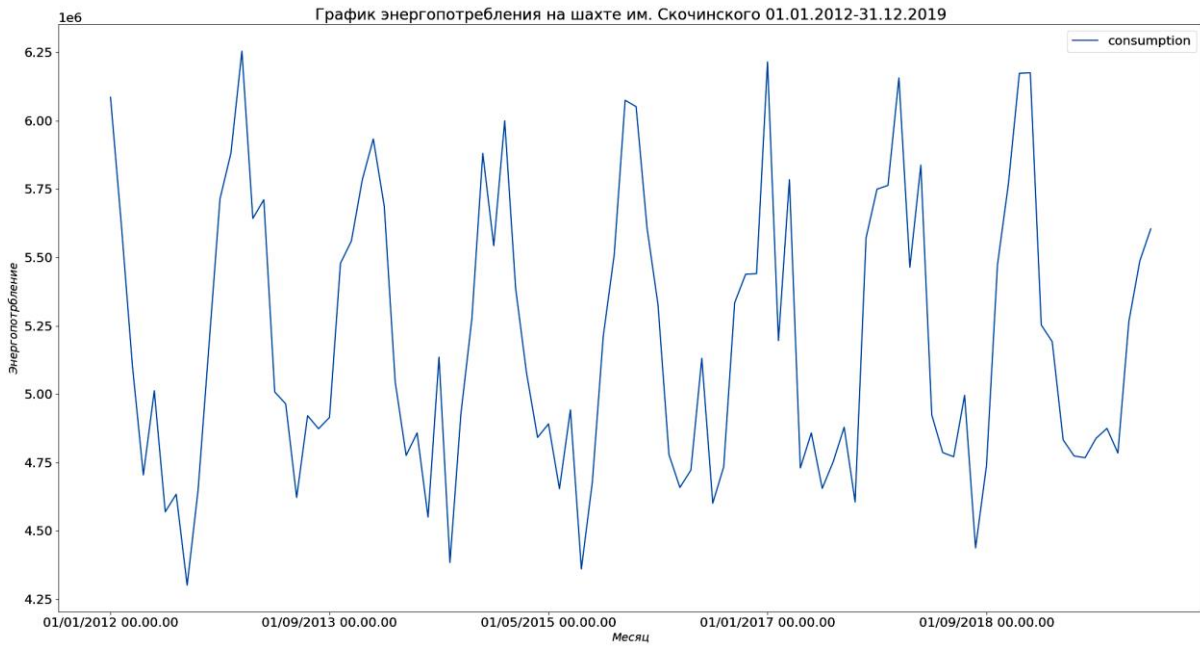


Рисунок 1 – График энергопотребления за 2018-2019 гг.

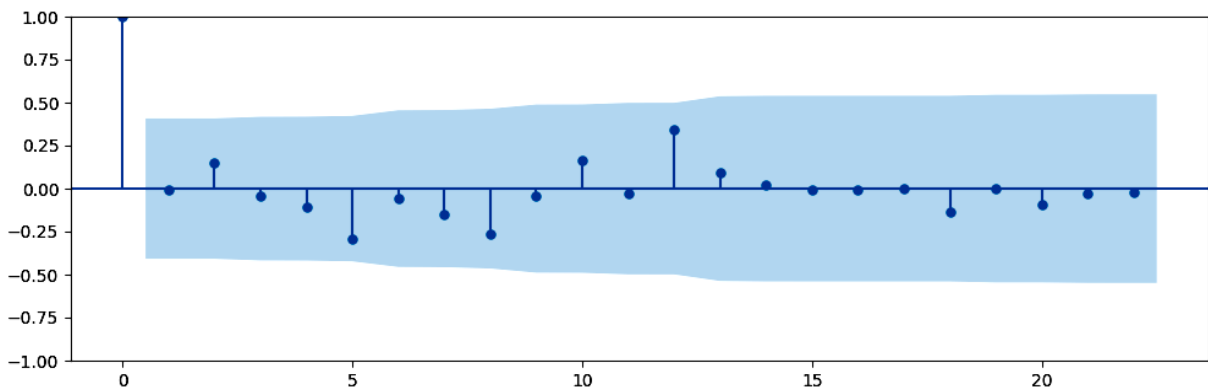


Рисунок 2 – Коррелограмма исследуемого временного ряда показателей энергопотребления

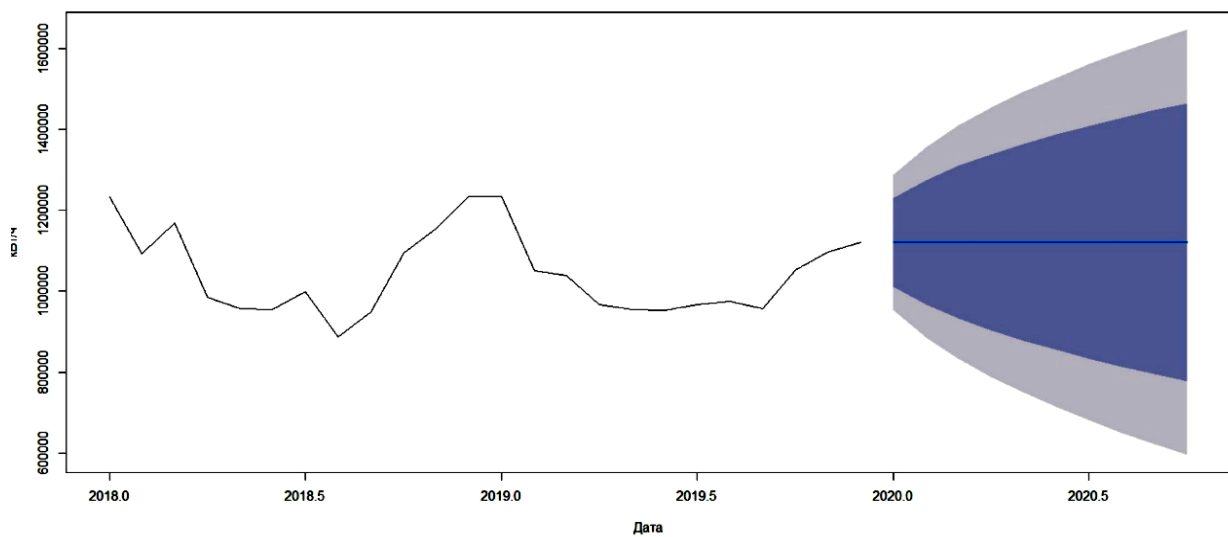


Рисунок 3 – Сглаженный ряд показателей энергопотребления на десять месяцев по методу Хольта-Винтерса

Регрессионные методы анализа применимы только для стационарных временных рядов, т.е. рядов, вероятностные свойства которых не изменяются во времени [3]. Для стационарного временного ряда с увеличением лага τ взаимосвязь членов временного ряда x_t и $x_{t+\tau}$ ослабевает и АКФ по абсолютной величине должна убывать.

На рис. 2 видно, что абсолютная величина убывает с ростом лага, следовательно, ряд является стационарным. Исследование временного ряда показателей энергопотребления показало, что применимыми моделями прогноза являются регрессионные и авторегрессионные модели прогнозирования, модели экспоненциального сглаживания и модели, основанные на нейронных сетях.

Нейросетевые модели позволяют строить прогноз, как основываясь на фактических значениях временного ряда, так и на внешних факторах, влияющих на значения временного ряда. Нейронная модель совмещает в себе возможности регрессионных и авторегрессионных моделей.

Описание используемых моделей прогнозирования

Для обучения и тестирования моделей собран датасет, содержащий статистические данные месячных объемов электропотребления, накопленные за период с января 2012 года по декабрь 2019 года. Полученная выборка делилась в отношении 87/13 на обучающую/тестовую.

Среди статистических моделей были выбраны модели: авторегрессионная интегрированная скользящая средняя (*Autoregressive Integrated Moving Average, ARIMA*) и модель экспоненциального сглаживания (*Exponential Smoothing, ES*). Обе модели строили прогноз на основе обучающей выборки.

Модель ARIMA использует три основных целочисленных параметра (p , d , q), которые учитывают сезонность, тренд и шумовую составляющую в наборах данных:

- p – порядок авторегрессии, который позволяет добавить предыдущие значения временного ряда;
- d – порядок разности, добавляет в модель понятия разности временных рядов, определяя количество прошлых временных точек, которые нужно вычесть из текущего значения;
- q – порядок скользящего среднего, который позволяет установить погрешность модели как линейную комбинацию наблюдавшихся ранее значений ошибок.

Для подбора оптимальных значений параметров модели временного ряда ARIMA (p , d , q) использовался сеточный поиск. При оценке и сравнении статистических моделей, соответствующих различным параметрам p , d , q использовался критерий AIC (*Akaike Information Criterion*). AIC оценивает, насколько хорошо модель соответствует данным, принимая во внимание общую сложность модели: чем меньше функций использует модель, чтобы достичь соответствия данным, тем выше её показатель AIC, поэтому оптимальной является модель с наименьшим значением AIC. В результате проведенного сеточного поиска получены следующие значения параметров модели ARIMA: $p = 20$, $d = 0$, $q = 3$.

Входные данные имеют выраженную сезонность с периодичностью в 12 месяцев, в качестве параметров модели ES передаются: аддитивная сезонность и длина периода выявленной сезонности. В модели ES учитывается степень старения данных: чем старше информация, тем с меньшим весом она входит в формулу для расчета

сглаженного значения уровня ряда. Наряду со стационарными рядами ES применяется и для прогнозирования нестационарных временных рядов, имеющих случайные изменения уровней ряда.

В качестве нейросетевых моделей использовались многослойный перцептрон (*Multilayer perceptron, MLP*) [4] и двунаправленная рекуррентная сеть с долгой краткосрочной памятью (*Bidirectional Long short-term memory, BLSTM*) [5].

Однонаправленные LSTM имеют ограничения: слои этих сетей имеют доступ к прошедшему контексту, и не имеют доступа к следующему контексту. Для этого и используется BLSTM. В этой архитектуре два разных внутренних слоя оперируют с данными в двух направлениях (вперёд и назад). Оба этих слоя соединены с одним выходным слоем, что позволяет использовать контекст из двух направлений. Недостатком подобной архитектуры по сравнению с однонаправленной является большее время обучения.

MLP имеет 20 входных нейронов. На вход перцептрону подаются 11 ретроспективных значений временного ряда, а также девять значений, полученных с помощью авторегрессионной модели. MLP имеет один скрытый слой с пятью нейронами с функцией активации сигмоид. На выходе один нейрон показывает предсказанное значение.

При обучении BLSTM модели использовались следующие гиперпараметры:

- две модели LSTM направленные в разные стороны;
- размер пакета: 256;
- размер выходного слоя каждой модели LSTM: 12;
- размер буфера: 10000
- количество блоков в каждом направлении прохода: 32;
- количество эпох обучения: 10;
- функция активации скрытых слоёв: hyperbolic tangent (tanh);
- функция рекуррентной активации: sigmoid;
- метод оптимизации: Adam.

Для оценки качества прогноза в рамках данного исследования в качестве метрики использовались относительная погрешность прогноза d и среднеквадратическая ошибка S прогноза потребления электроэнергии:

$$d = \frac{|y_t - y_t^*|}{y_t}, S = \sqrt{\frac{\sum_{t=1}^n (y_t - y_t^*)^2}{n}}$$

где y , y^* – модельное и наблюдаемое значение соответственно.

Результаты расчета прогноза потребления электроэнергии на 2019 год, полученные используемыми моделями по метрикам d и S , сведены в табл. 1 и 2 соответственно.

Таблица 1 – Качество исследуемых моделей прогнозирования потребления электроэнергии по относительной погрешности прогноза

Модель	Прогноз потребления электроэнергии, кВтч	Фактическое потребление электроэнергии, кВтч	относительная погрешность прогноза, %
Январь			
MLP	6194610	6175236	0.314
BLSTM	6219020		0.71
ARIMA	5955007		3.566
ES	6075293		1.618
Февраль			
MLP	5693352	5253244	8.378
BLSTM	5202147		0.973
ARIMA	5246838		0.122
ES	5587983		6.372
Март			
MLP	5647805	5192153	8.776
BLSTM	5785513		11.428
ARIMA	5389828		3.807
ES	5293012		1.943
Апрель			
MLP	4976016	4833005	2.959
BLSTM	4731155		2.107
ARIMA	4919116		1.781
ES	4851972		0.392
Май			
MLP	5057698	4774217	5.938
BLSTM	4860079		1.798
ARIMA	4648480		2.634
ES	4845304		1.489
Июнь			
MLP	4729243	4767830	0.809
BLSTM	4652426		2.42
ARIMA	4772586		0.1
ES	4646601		2.542
Июль			
MLP	5212777	4838769	7.729
BLSTM	4747141		1.894
ARIMA	4732947		2.187
ES	4930541		1.897
Август			
MLP	4665783	4875070	4.293
BLSTM	4870421		0.095
ARIMA	4788677		1.772
ES	4502642		7.639
Сентябрь			
MLP	5020041	4784897	4.914
BLSTM	4593488		4
ARIMA	5076705		6.099
ES	4812734		0.582

Продолж. Табл. 1

Октябрь			
MLP	5426444	5265802	3.05
BLSTM	5555872		5.509
ARIMA	5379656		2.162
ES	5303928		0.724
Ноябрь			
MLP	5883597	5487333	7.221
BLSTM	5747533		4.742
ARIMA	5526374		0.711
ES	5686345		3.626
Декабрь			
MLP	5946268	5604557	6.097
BLSTM	5733762		2.305
ARIMA	5721072		2.079
ES	5843207		4.258

Таблица 2 – Значения среднеквадратической ошибки прогноза S , полученного различными моделями

Модель	Среднеквадратическая ошибка прогноза, %
MLP	5.722
BLSTM	4.314
ARIMA	2.768
ES	3.535

Как видно из приведенных в таблицах данных, наиболее качественная модель прогнозирования – ARIMA.

Выводы

На основе проведенного анализа временного ряда выявлено, что ряд является стационарным, имеет выраженный тренд и сезонность. Свойство стационарности делает возможным использование методов на основе автокорреляционной функции, в частности, модели ARIMA, оптимальные параметры которой подобраны с помощью сетчатого поиска. Кроме того, исследовалась модель экспоненциального сглаживания, позволяющая выявить тенденции развития ряда, не являющегося стационарным. Экспоненциальное среднее является примером асимметрической скользящей средней, в которой учитывается степень старения данных.

Помимо этих статистических моделей реализованы нейросетевые методы прогнозирования и проведено исследование эффективности всех используемых моделей. Критериями качества моделей в рамках данного исследования являлись относительная погрешность прогноза и среднеквадратическая ошибка прогноза.

Как показали численные исследования, статистические модели заняли лидирующее место по сравнению с нейросетевыми моделями. Очевидно, это связано с тем, что обучающая выборка крайне мала для того, чтобы использовать аппарат нейросетей для предсказания. При достаточном объеме обучающей выборки, можно предположить, что модель BLSTM будет давать наилучшие среднесрочный и долгосрочный прогнозы. Модель ARIMA отлично справится с краткосрочным прогнозом, как при маленькой, так и при большой обучающей выборке, но будет показывать далеко не самый хороший результат при долгосрочном прогнозе, по сравнению с нейросетевыми моделями.

Список литературы

1. Винник А. О. Программный комплекс мониторинга энергопотребления на предприятиях угольной промышленности [Текст] / Винник А.О., Мартыненко А-А.М., Бондаренко В.И. // Вестник ДонНУ. Серия Г: Технические науки. – Донецк. 2021. – № 2. – С. 33-38.
2. Prajakta S. K. Time series Forecasting using Holt-Winters Exponential Smoothing [Текст] – 2004.
3. Афанасьев, В. Н. Анализ временных рядов и прогнозирование: учебник [Текст] // В. Н. Афанасьев. – Ай Пи Ар Медиа, Саратов, Оренбургский гос. ун-т. – Оренбург, 2020. – 286 с.
4. Хайкин, Саймон Нейронные сети. Полный курс: моногр. [Текст] // Саймон Хайкин. – М.: Вильямс, 2017. – 788 с.
5. Zhiyong Cui, Ruimin Ke, Ziyuan Pu, Yin Hai Wang. Stacked Bidirectional and Unidirectional LSTM Recurrent Neural Network for Networkwide Traffic Speed Prediction [Электронный ресурс] – Режим доступа: <https://arxiv.org/ftp/arxiv/papers/1801/1801.02143.pdf>
6. Математическое моделирование и анализ параметров устройства вихревой пароконденсации в прямоточной системе обессоливания шахтных вод [Текст] / Павлыш В. Н., Доценко Г. В., Овсянников В. П., Малеев В. Б. // Международный рецензируемый научно-теоретический журнал «Проблемы искусственного интеллекта». – 2020. – № 4(19). – С. 30–45.

References

1. Vinnik A.O. Programmyj kompleks monitoringa energopotrebleniya na predpriyatiyah ugol'noj promyshlennosti / Vinnik A.O., Martynenko A-A.M., Bondarenko V.I. // Vestnik DonNU. Seriya G: Tekhnicheskie nauki. – Doneck. 2021. – № 2. – S. 33-38.
2. Prajakta S.K. Time series Forecasting using Holt-Winters Exponential Smoothing. // 2004.
3. Afanas'ev, V. N. Analiz vremennyh ryadov i prognozirovanie: uchebnyk / V. N. Afanas'ev. – Aj Pi Ar Media, Saratov, Orenburgskij gos. un-t. – Orenburg, 2020. – 286 s.
4. Hajkin, Sajmon Nejrornyje seti. Polnyj kurs: monogr. / Sajmon Hajkin. - M.: Vil'yams, 2017. - 788 s.
5. Zhiyong Cui, Ruimin Ke, Ziyuan Pu, Yin Hai Wang. Stacked Bidirectional and Unidirectional LSTM Recurrent Neural Network for Networkwide Traffic Speed Prediction <https://arxiv.org/ftp/arxiv/papers/1801/1801.02143.pdf>
6. Pavlysh V.N., Dotsenko G.V., Ovsyannikov V.P., Maleev V.B. Matematicheskoye modelirovaniye i analiz parametrov ustroystva vikhrevoj parokondensatsii v pryamotochnoy sisteme obessolivaniya shakhtnykh vod [Mathematical modeling and analysis of the parameters of a vortex steam condensation device in a once-through mine water demineralization system]. *Mezhdunarodnyy retsenziruyemy nauchno-teoreticheskiy zhurnal «Problemy iskusstvennogo intellekta»* [International peer-reviewed scientific and theoretical journal "Problems of Artificial Intelligence"], 2020, No. 4 (19), pp. 30-45.

RESUME

T. V. Yermolenko, V.N. Kotenko, A. O. Vinnik

Study of the efficiency of predictional models for the system of analysis and monitoring of energy consumption at coal industry enterprises

Studies of the effectiveness of integrating the system of analysis and monitoring of energy costs into extractive industry enterprises have shown that when it is implemented, energy costs are reduced by up to 5%, which makes the task of building a high-quality model for predicting the amount of electricity consumed by an enterprise relevant and in demand. Its solution is complicated by the need to take into account a large number of factors affecting energy consumption, especially in the context of unstable social and economic processes.

This study is devoted to a comparative analysis of various models for predicting electricity consumption by a coal industry enterprise, for training and testing of which data from the enterprise «Mine named after academician A.A. Skochinsky » for 2012-2019 years.

Visual analysis of the data of the time series of energy consumption indicators, as well as the correlogram, made it possible to identify a clear seasonality and a trend, which is a straight line around which values fluctuate within confidence intervals. In addition, the stationarity of the time series is established.

The stationarity of the series makes it possible to analyze it using autoregressive methods. In this work, the ARIMA method was used, the parameters of which were selected using a grid search. Another method belonging to the statistical group, which was used for forecasting in this study, is the exponential smoothing model.

In addition, neural network models were used, which, due to their properties, make it possible to build a forecast, both based on the actual values of the time series, and on external factors that affect the values of the time series. MLP and BLSTM were used as neural network models.

The efficiency of the implemented models was evaluated by the relative error and the root mean square error of the electricity consumption forecast. The ARIMA model showed the best result. This can be explained by the fact that the volume of training data for the qualitative work of neural network models was extremely small, but quite enough to obtain a short-term forecast using ARIMA.

РЕЗЮМЕ

Т. В. Ермоленко, В. Н. Котенко, А. О. Винник

Исследование эффективности предсказательных моделей для системы анализа и мониторинга энергопотребления на предприятиях угольной промышленности

Исследования эффективности интеграции на предприятия добывающей отрасли системы анализа и мониторинга энергозатрат показали, что при ее внедрении энергозатраты сокращаются до 5%, что делает задачу построения качественной модели прогнозирования потребляемых предприятием объемов электроэнергии актуальной и востребованной. Ее решение осложняется необходимостью учета большого количества факторов, влияющих на показатели энергопотребления, особенно в условиях нестабильности социальных и экономических процессов.

Настоящее исследование посвящено сравнительному анализу различных моделей прогнозирования потребления электроэнергии предприятием угольной промышленности, для обучения и тестирования которых использовались данные предприятия «Шахта имени академика А.А. Скочинского» за 2012-2019 гг.

Визуальный анализ данных временного ряда показателей энергопотребления, а также коррелограмма позволили выделить четкую сезонность и тренд, представляющий собой прямую, около которой совершаются колебания значений в рамках доверительных интервалов. Кроме того, установлена стационарность временного ряда.

Стационарность ряда делает возможным его анализ с помощью авторегрессионных методов. В данной работе использовался метод ARIMA, параметры которого подобраны с помощью сеточного поиска. Еще один метод, относящимся к группе статистических, который применялся для прогнозирования в рамках данного исследования, – модель экспоненциального сглаживания.

Кроме того, использовались нейросетевые модели, которые в силу своих свойств позволяют строить прогноз, как основываясь на фактических значениях временного ряда, так и на внешних факторах, влияющих на значения временного ряда. В качестве нейросетевых моделей использовались MLP и BLSTM.

Эффективность реализованных моделей оценивалась по относительной погрешности и среднеквадратической ошибке прогноза потребления электроэнергии. Наилучший результат показала модель ARIMA. Это можно объяснить тем фактом, что объема обучающих данных для качественной работы нейросетевых моделей было крайне мало, но вполне достаточно для получения краткосрочного прогноза с помощью ARIMA.

Статья поступила в редакцию