

*Н. Г. Акопов<sup>1\*</sup>, Е. А. Кузнецова<sup>1</sup>*

## **Особенности применения машинного обучения при прогнозировании потребления энергетических ресурсов**

<sup>1</sup>Новосибирский государственный университет, г. Новосибирск,  
Российская Федерация  
\* e-mail: n.akopov@g.nsu.ru

**Аннотация.** В данной статье был спрогнозирован спрос на природный газ в Китае с помощью методов машинного обучения. Были выбраны наиболее подходящие модели, способные справиться с проблемой мультиколлинеарности, такие как случайный лес, градиентный бустинг и гребневая регрессия, а также отобраны факторы, наиболее полно отражающие спрос на газ с различных сторон. Для каждой из построенных регрессий были оценены ошибки и влияние каждого фактора на объясняемую переменную. Лучшие результаты показала модель с применением градиентного бустинга. Итогом исследования стало понимание значимости природного газа как ресурса для стран и возможности использования методов машинного обучения для дальнейшего исследования спроса на данный ресурс. Подтверждено предположение авторов о значимости численности населения и площади городов для спроса на газ.

**Ключевые слова:** природный газ, машинное обучение, спрос на природный газ, прогнозирование

*N. G. Akopov<sup>1\*</sup>, E. A. Kuznetsova<sup>1</sup>*

## **Features of the use of machine learning in forecasting energy consumption**

<sup>1</sup> Novosibirsk State University, Novosibirsk, the Russian Federation  
\* e-mail: n.akopov@g.nsu.ru

**Abstract.** In this article, the demand for natural gas in China was predicted using machine learning methods. The most suitable models capable of coping with the problem of multicollinearity, such as random forest, gradient boosting and ridge regression, were selected, as well as factors that most fully reflect the demand for gas from various sides. For each of the constructed regressions, the errors and the influence of each factor on the explained variable were evaluated. The best results were shown by a model using gradient boosting. The result of the study was an understanding of the importance of natural gas as a resource for countries and the possibility of using machine learning methods to further study the demand for this resource. The authors' assumption about the importance of the population and the area of cities for gas demand is confirmed.

**Keywords:** natural gas, machine learning, natural gas demand, forecasting

### ***Введение***

В современном мире машинное обучение становится всё более востребованным инструментом в совершенно различных областях. При этом спрос на энергоресурсы постоянно растёт. Имея инструмент для моделирования спроса на

природный газ, компании могут более точно оценивать риски, моделируя различные сценарии. При этом и страны, используя похожие модели [1,2,3], смогут не допускать дефицита ресурса, что очень важно как для обычных граждан страны, так и для промышленности.

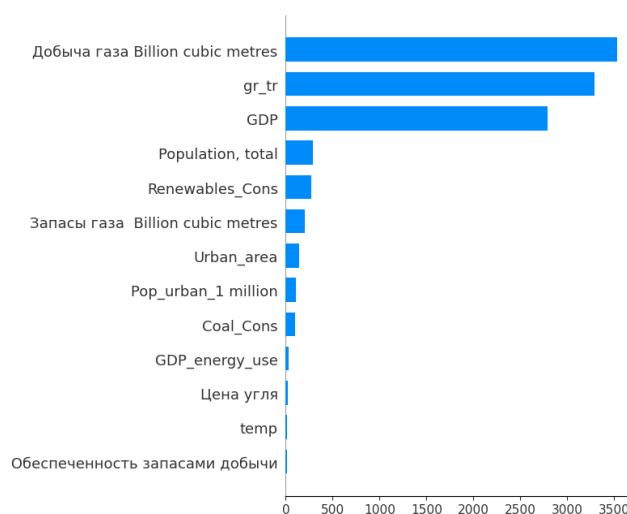
### ***Методы и материалы***

Для анализа спроса на газ были выбраны следующие факторы: валовый ВВП, ВВП на единицу потребляемой энергии, среднегодовая температура, городская территория, численность населения, население в городских агломерациях численностью более миллиона человек, годовое потребление каменного угля, годовое потребление энергии их возобновляемых источников, количество гражданского транспорта, добыча газа, запасы газа, обеспеченность запасами добычи, цена угля [4,5,6,7]. Информационной базой исследования послужил Energy Institute Statistical Review of World Energy [8].

Также достаточно важным моментом в данном исследовании являлся выбор моделей для построения регрессии. Так как при предварительном анализе данные показали достаточно большую зависимость между друг другом, то выбор моделей сократился до наиболее устойчивых к проблеме мультиколлинеарности. Из всего множества была выбрана гребневая регрессия в качестве линейной модели, а также модели случайный лес и градиентный бустинг в качестве нелинейных.

### ***Результаты и обсуждение***

Начнем анализ результатов с линейных моделей, а именно с гребневой регрессии. Значение  $R^2$  и MAPE получились равными 0,93 и 6,1% соответственно. Если же говорить о значимости коэффициентов, то она была оценена с помощью SHAP метода, ниже приведены результаты (рис. 1). По нему видно, что наибольший вклад в спрос на природный газ вносят его добыча, гражданский транспорт и ВВП. Интересным здесь кажется тот факт, что площадь городской территории (Urban\_area) вносит не такой весомый вклад, как ожидалось до получения результатов.



**Рис. 1. Значимость факторов по значениям Шэпли (гребневая регрессия)**

Далее, переходя к нелинейным моделям, хочется начать со случайного леса. Показатель  $R^2$  у данной модели получился равным 0,95, а ошибка MAPE 5,8%. Оба этих показателя лучше, чем у первоначально рассмотренной линейной модели. При анализе второй нелинейной модели, а именно градиентного бустинга, результаты получились следующими.  $R^2$  равен 0,95, MAPE 5,73%. Получилось так, что данная модель стала лучшей среди всех моделей, рассмотренных в данном исследовании. Результаты значений SHAP приведены ниже (рис. 2). Их интерпретация кажется более уместной и логичной, если сравнивать с гребневой регрессией. Тут виден явный отрыв у численности населения и добычи газа, а на третьем месте расположилась площадь городской территории.

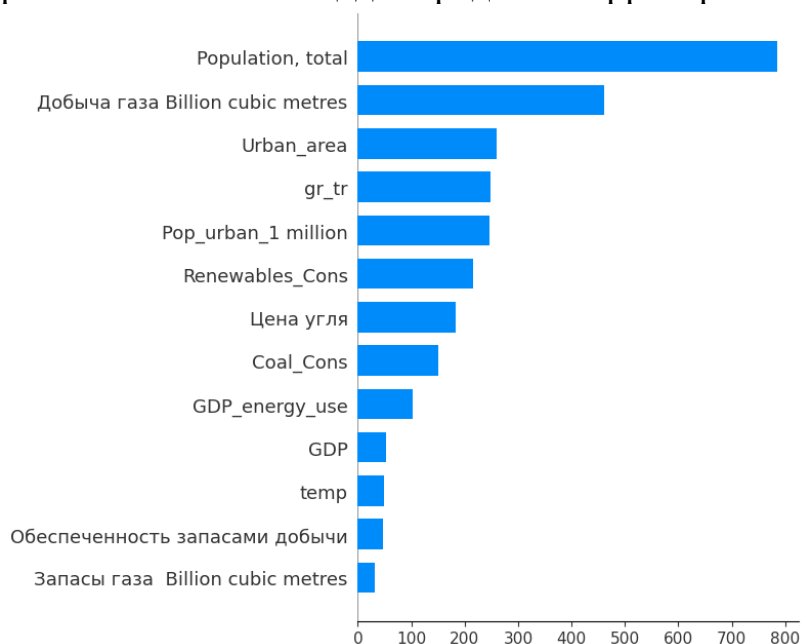


Рис. 2. Значимость факторов по значениям Шэпли (градиентный бустинг)

### ***Заключение***

Анализируя данное исследование, можно сделать вывод, что нелинейные регрессионные модели машинного обучения лучше справляются с прогнозированием спроса на природный газ. Полученные с помощью градиентного бустинга результаты подтвердили предположения авторов о том, что спрос на газ сильно зависит от численности населения, площади городов. Ведь одно из основных направлений, куда тратится данный ресурс, это газификация домов. Именно с помощью газа люди готовят еду и получают тепло в дом.

Можно сказать, что методы машинного обучения показали свою эффективность и в дальнейшем их можно использовать для прогнозирования различных макроэкономических показателей.

### ***Благодарности***

Исследование выполнено за счет гранта Российского научного фонда № 23-78-10157, <https://rscf.ru/project/23-78-10157/>.

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Ahmed S. et al. Prediction of Natural Gas Consumption in Bahçeşehir Using Machine Learning Models /S. Ahmed, S. Madanian, F. Mirza, S. Zain //Association for Information Systems, 2021.
2. Gao F., Shao X. Forecasting annual natural gas consumption via the application of a novel hybrid model //Environmental Science and Pollution Research. – 2021. – vol. 28. – no. 17. – P. 21411-21424.
3. Zhang W., Yang J. Forecasting natural gas consumption in China by Bayesian model averaging //Energy Reports. – 2015. – vol. 1. – P. 216-220.
4. Богоявленский В. И. и др. Газовая революция в Китае /Богоявленский В. И., Баринов П. С., Богоявленский И. В., Якубсон К. И. //Бурение и нефть. – 2016. – №. 11. – С. 3-14.
5. Ершова Е. В. Природный газ в энергопотреблении Китая: основные тенденции развития //Известия Байкальского государственного университета. – 2016. – Т. 26. – №. 1. – С. 80-89.
6. Резникова О. Б. Нефть и газ Центральной Азии: разворот на рынок КНР //Серия—Библиотека Института мировой экономики международных отношений – 2014. – С. 6.
7. Халова Г. О., Сычева А. М., Спивак В. Ю. Деятельность китайских нефтегазовых компаний в государствах Центрально-Азиатского региона //Нефть, газ и бизнес. – 2013. – №. 12. – С. 39-43.
8. BP: Energy Institute Statistical Review of World Energy. 2023. URL: <https://www.energyinst.org/statistical-review> (дата обращения 16.03.2024)

© Н. Г. Акопов, Е. А. Кузнецова, 2024