МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ЧЕРНІГІВСЬКА ПОЛІТЕХНІКА»
WIN-WIN EDIH



Co-funded by
the European Union



Співфінансується
Європейським Союзом



ВІЛЬНЕ ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ В ОСВІТІ, НАУЦІ ТА БІЗНЕСІ

FREE/OPEN SOURCE SOFTWARE IN EDUCATION, SCIENCE & BUSINESS (FOSS 2026)

Матеріали II Міжнародної науково-практичної конференції
01–02 жовтня 2026 р.
м. Чернігів

Чернігів 2026

УДК 004.49:[37+001+334.722]
В41

Рекомендовано до друку вченою радою Національного університету «Чернігівська політехніка» (протокол № 12 від 02.11.2026 р.)

Вільне програмне забезпечення в освіті, науці та бізнесі : матеріали ІІ Міжнар. В41 наук.-практ. конф. (м. Чернігів, 01–02 жовтня 2026 р.) / відп. за вип. Д. Б. Мехед. Чернігів : НУ «Чернігівська політехніка», 2026. 156 с.

ISBN 978-617-8584-09-2

Збірник містить матеріали ІІ Міжнародної науково-практичної конференції «Вільне програмне забезпечення в освіті, науці та бізнесі» (FOSS 2026), що відбулася онлайн 01–02 жовтня 2026 року в Національному університеті «Чернігівська політехніка». Розглянуто питання використання вільного та відкритого програмного забезпечення в освіті, науці, бізнесі, ІТ-інфраструктурі, кібербезпеці, аналізі даних і штучному інтелекті. Матеріали подано українською та англійською мовами.

Для науковців, викладачів, здобувачів освіти, фахівців ІТ-галузі та всіх, хто цікавиться вільним програмним забезпеченням.

Матеріали подано в авторській редакції. Відповідальність за зміст, достовірність фактів, цитат і посилань несуть автори.

Конференцію проведено в межах проєкту WIN-WIN EDIH (грантова угода № 101191647), співфінансованого Європейським Союзом у межах програми Digital Europe. Висловлені погляди та думки належать виключно автору(ам) і не обов'язково відображають позицію Європейського Союзу чи Європейської Комісії.

Відповідальний за випуск: Мехед Дмитро Борисович

ISBN 978-617-8584-09-2

© НУ «Чернігівська політехніка», 2026

Чопорова О. В., Чопоров С. В. ВИКОРИСТАННЯ 3D-МОДЕЛЮВАННЯ ТА АЛГОРИТМІВ НЕЙРОННИХ МЕРЕЖ В НАВЧАЛЬНИХ ПРОЄКТАХ ІНЖЕНЕРНО-ТЕХНІЧНОГО СПРЯМУВАННЯ / USE OF 3D MODELING AND NEURAL NETWORK ALGORITHMS IN EDUCATIONAL PROJECTS OF ENGINEERING AND TECHNICAL DIRECTION 82

Чуб М. В. ВИЯВЛЕННЯ ТА ПРОТИДІЯ АДАПТИВНИМ ВЕБ-АТАКАМ ЗА ДОПОМОГОЮ МОДЕЛЕЙ ГЛИБОКОГО НАВЧАННЯ 86

Шевченко Ю. О. ПОРІВНЯЛЬНЕ ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ POSTGRESQL ТА INFLUXDB ДЛЯ ЗБЕРІГАННЯ ДАНИХ ФІЗИЧНОЇ АКТИВНОСТІ / COMPARATIVE STUDY OF POSTGRESQL AND INFLUXDB PERFORMANCE FOR STORING PHYSICAL ACTIVITY DATA 89

Al-Akwaa K. H. M. A. OPEN-SOURCE DATA SCIENCE FOR SUSTAINABLE DEVELOPMENT: A REPRODUCIBLE ANALYTICAL FRAMEWORK 93

Borysenko O. A. AN ARCHITECTURAL APPROACH TO LLM-BASED ROOT CAUSE ANALYSIS FOR CONTAINER ORCHESTRATION INCIDENTS 100

Burmaka I. LOCAL, OPEN, AND ENERGY-EFFICIENT: HARNESSING X86 NPUS FOR EVERYDAY AI TASKS 102

Byvoino T. P., Byvoino P. G. SIMULATION MODELING AND STATISTICAL ANALYSIS OF THE EFFICIENCY OF AN EMERGENCY MEDICAL SERVICE STATION 106

Didusenko O. A., Bula Y. I., Shymchuk G. V. REPRODUCIBLE FOSS FRAMEWORK FOR RESILIENCE EVALUATION OF CONTAINERIZED DISTRIBUTED SYSTEMS 109

Didusenko O. A., Bula Y. I., Shymchuk G. V. REPRODUCIBLE FOSS TESTBED FOR FAULT-TOLERANT MULTIPATH ROUTING IN IP NETWORKS 112

Ebirim V., Jadama A. F. R. RESILIENT DIGITAL EDUCATION IN DISRUPTED ENVIRONMENTS: THE CCDS APPROACH 115

Fraya DIGITAL TRANSFORMATION OF BUSINESS THROUGH FREE AND OPEN SOURCE SOFTWARE: OPPORTUNITIES, BARRIERS, AND STRATEGIC PATHWAYS 119

Kala P. RUN MONGODB YOUR WAY: A PRACTICAL OPEN-SOURCE STACK FOR SELF-MANAGED DEPLOYMENTS 121

Lukin O. A MODERN C++ KMIP CLIENT BASED ON A REUSABLE PROTOCOL LIBRARY 124

Mekhed D. B. DEVELOPING STUDENTS' CYBER HYGIENE SKILLS THROUGH THE USE OF FREE SPECIALIZED SOFTWARE 127

Nahornyi B. I., Lysenok D. V., Savenko A. O. AN OPEN BROWSER-BASED CO-SIMULATOR OF STM32 FIRMWARE AND ELECTRICAL CIRCUITS AS A FREE ALTERNATIVE TO PROTEUS 130

Romanenko M. V. FOSS APPROACH IN MOBILE DEVELOPMENT: CREATING AN OPEN-SOURCE ANDROID APPLICATION WITH THE INTEGRATION OF READY-MADE LIBRARIES FOR STYLISTIC RECOMMENDATIONS 134

Shevchenko N. A., Kibitkin D. S., Shymchuk G. V. REPRODUCIBLE OPEN-SOURCE BENCHMARK OF CONCEPT-DRIFT DETECTORS FOR STREAMING REGRESSION 137

Shevchenko N. A., Kibitkin D. S., Shymchuk G. V. INTERPRETABLE OPEN-SOURCE MACHINE LEARNING FOR NATURAL GAS CONSUMPTION FORECASTING WITH WEATHER AND LAG FEATURES 140

Shivashankar C. ARCHITECTING DATABASE DEPLOYMENTS ON KUBERNETES: PATTERNS, TRADE-OFFS AND EXPERIENCES 143

Shymchuk G. V., Yatskiv P. O. DETECTION OF ANOMALOUS STATES IN IP NETWORKS USING MACHINE LEARNING AND OPEN-SOURCE SOFTWARE TOOLS 146

Shymchuk G. V., Yatskiv P. O. OPEN-SOURCE SOFTWARE ENVIRONMENT FOR MODELING FAULT-TOLERANT MULTIPATH ROUTING IN IP NETWORKS 149

Silva F. HOSTING SOMEONE ELSE'S APP: IFRAME CHOREOGRAPHY IN PMM 152

Šupić A. ZERO-DOWNTIME MIGRATION WITH PERCONA CLUSTERSYNC FOR MONGODB 154

Shevchenko N. A., Kibitkin D. S., Shymchuk G. V.
Ternopil Ivan Puluj National Technical University, Department of Computer Science, Ternopil, Ukraine
Corresponding author: Grigorii Shymchuk, e-mail: gorych@gmail.com, ORCID: 0000-0003-2362-7386

INTERPRETABLE OPEN-SOURCE MACHINE LEARNING FOR NATURAL GAS CONSUMPTION FORECASTING WITH WEATHER AND LAG FEATURES

Abstract. An interpretable open-source machine-learning workflow for short-term natural gas consumption forecasting is investigated using historical gas-demand and meteorological observations from 2007-2019. Hourly gas volumes were aggregated into non-overlapping 3-hour intervals and aligned with 3-hour weather records. Models were trained on 2007-2018 and evaluated chronologically on the unseen year 2019. Decision Tree, Random Forest, and Histogram-based Gradient Boosting regressors were compared using R^2, MAE, RMSE, and WAPE. Random Forest achieved the best test performance with $R^2 = 0.9656$, MAE = 4749.6 m^3, RMSE = 7955.9 m^3, and WAPE = 9.52%. Permutation importance showed that the strongest predictive signal was the daily lag Q(t-24 h), followed by Q(t-3 h), Q(t-7 d), outdoor temperature, and cyclic hour features. A reduced eight-feature Random Forest preserved performance (WAPE = 9.46%), while excluding weather increased WAPE to 9.67%. The results demonstrate that an open-source Python/scikit-learn pipeline can combine accurate forecasting with transparent feature-level interpretation.

Keywords: open-source software; natural gas consumption; time-series forecasting; machine learning; Random Forest; explainable AI; permutation importance; weather features.

Introduction

Accurate short-term forecasting of natural gas consumption is important for operational planning, balancing, and demand management. Gas demand is strongly seasonal and depends simultaneously on recent consumption, daily and weekly cycles, and meteorological conditions. Machine-learning methods can represent nonlinear relations between these variables, but operational use also requires an understanding of why a forecast changes. This has increased interest in explainable and interpretable forecasting approaches, especially for energy time series [1-3].

Recent energy-forecasting studies combine predictive models with feature-attribution methods. Eskandari et al. integrated feature selection and interpretable machine learning for energy-consumption forecasting and emphasized the value of a compact feature subset [4]. Neubauer et al. showed that lagged load variables and weather features are central to explainable heating-load forecasting [5]. Mohammadi et al. reported that exogenous variables can improve energy time-series forecasts and that feature perturbation methods provide robust interpretability signals [6]. These findings motivate a reproducible workflow in which forecast accuracy and feature relevance are evaluated together.

The objective of this work is to assess an open-source tree-based forecasting pipeline on a real multiyear gas-consumption dataset and to quantify the contribution of lagged demand, calendar variables, and weather observations. In contrast to a purely predictive comparison, the analysis includes permutation-based feature importance and an ablation test to determine whether a reduced set of variables can preserve forecast quality.

Materials and Methods

The dataset covers 1 January 2007 through 31 December 2019. Hourly gas-consumption values Q were aggregated by summing each non-overlapping block of three consecutive hours, producing 37,984 target observations measured in cubic metres per 3-hour interval. These records were aligned with 3-hour meteorological measurements containing air temperature T, station pressure Po, sea-level pressure P, pressure tendency Pa, relative humidity U, wind speed Ff, precipitation RRR, dew-point temperature Td, and visibility VV. Calendar variables were encoded by sine/cosine transformations of hour-of-day and day-of-year, together with a weekend indicator.

Table 1 – Forecasting feature groups

Group	Variables	Role
Demand lags	Q_lag_1, Q_lag_8, Q_lag_56	3 h, 24 h, 7 d memory
Weather	T, Po, P, Pa, U, Ff, RRR, Td, VV	exogenous conditions
Daily cycle	hour_sin, hour_cos	time-of-day periodicity

Annual / week	doy_sin, doy_cos, weekend	seasonality and weekly effect

Three open-source regressors were evaluated: a CART Decision Tree, Random Forest, and Histogram-based Gradient Boosting, all implemented with scikit-learn. The training interval was 2007-2018, and 2019 was kept completely unseen during fitting. This chronological split avoids information leakage from future observations. The primary relative error measure was weighted absolute percentage error (WAPE),

$$WAPE = \left[\frac{\Sigma|Q_i - \hat{Q}_i|}{\Sigma|Q_i|}\right] \cdot 100\%. \quad (1)$$

where Q_i is the measured 3-hour gas volume and $\hat{Q}_i$ is the corresponding model forecast. R^2, MAE, and RMSE were also calculated. For interpretability, permutation importance was evaluated on a fixed 2019 subset: one feature at a time was randomly permuted and the increase in MAE was recorded. A large increase indicates that the model depends strongly on that variable. This method is model-agnostic and is implemented directly in the scikit-learn inspection module [7].

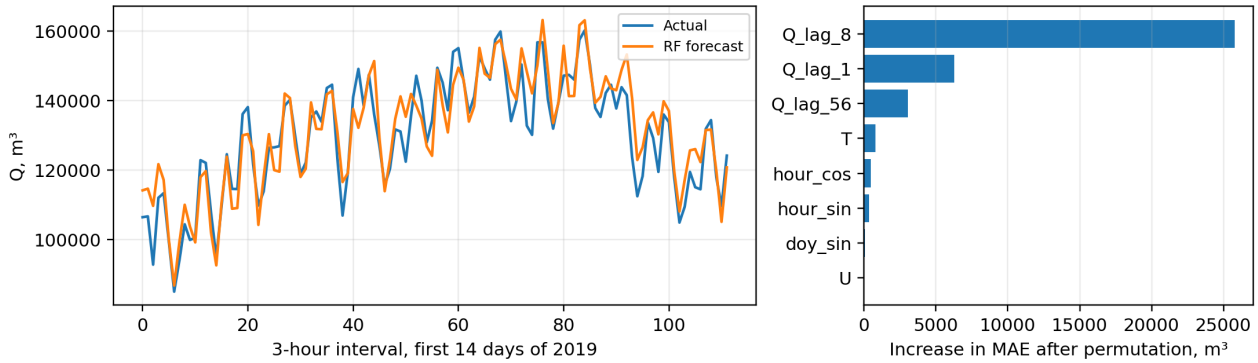

Figure 1 – Example 2019 Random Forest forecast and permutation feature importance

Results

Table 2 summarizes the chronological 2019 test results. Random Forest provided the best overall balance, achieving $R^2 = 0.9656$ and WAPE = 9.52%. Decision Tree remained competitive but produced larger absolute errors, while Histogram-based Gradient Boosting in the tested configuration yielded WAPE = 10.62%. The result confirms that ensemble averaging can reduce the instability of a single tree while preserving the ability to model nonlinear lag-weather interactions.

Table 2 – Forecasting results for the unseen year 2019

Model	R^2	MAE, m³	RMSE, m³	WAPE, %
Decision Tree	0.9564	5447.6	8960.3	10.92
Random Forest	0.9656	4749.6	7955.9	9.52
HistGradientBoosting	0.9537	5298.2	9237.8	10.62

The feature-importance analysis was strongly dominated by Q_lag_8, corresponding to the same 3-hour position one day earlier. Permuting this feature increased MAE by approximately 25,795 m³ on the evaluation subset. Q_lag_1 and Q_lag_56 followed with increases of about 6,288 m³ and 3,048 m³. Among exogenous variables, temperature was the most influential: its permutation increased MAE by about 795 m³. Hour-of-day sine/cosine features also had a clear contribution, whereas most pressure and visibility variables had small marginal influence after lag variables were present.

A reduced Random Forest trained on the eight most important variables achieved $R^2 = 0.9655$, RMSE = 7971.9 m³, and WAPE = 9.46%, practically preserving the full-model accuracy. By contrast, a model that removed all meteorological variables obtained $R^2 = 0.9653$, RMSE = 7988.8 m³, and WAPE = 9.67%. Thus, the weather block produced a modest but reproducible improvement, while the main predictive information remained concentrated in autoregressive lags and temporal periodicity.

Discussion

The dominance of Q_lag_8 is consistent with the operational structure of gas demand: on a 3-hour grid, an eight-step lag represents the same time of day on the previous day. The seven-day lag

captures the weekly schedule, while Q_lag_1 reflects immediate inertia. Temperature adds independent information because heating demand changes with outdoor conditions even when recent gas consumption is known. Similar interactions between load history and meteorological factors have been reported for heating and electricity forecasting [5,6].

The most practically useful result is the near-equivalence between the full 17-feature Random Forest and the reduced eight-feature model. Reducing the feature set can simplify data collection and model maintenance while maintaining predictive quality. At the same time, permutation importance must be interpreted carefully in the presence of correlated predictors: if two variables contain similar information, the importance of each one can be underestimated because the other remains available to the model. Therefore, the reported ranking should be viewed as predictive dependence rather than a causal explanation.

The workflow is fully based on FOSS components and is reproducible without proprietary analytical software. Its main limitation is that the experiment evaluates one regional historical series and one chronological test year. Further work should repeat the analysis on additional regions, use rolling-origin validation, compare local explanation methods such as SHAP, and analyze the stability of importance rankings across seasons and temperature ranges.

Conclusions

An interpretable open-source workflow for natural gas consumption forecasting was evaluated on 2007-2019 data. Random Forest achieved the best 2019 test performance with $R^2 = 0.9656$ and WAPE = 9.52%. The 24-hour lag was the dominant predictor, followed by the 3-hour and 7-day lags; temperature was the most relevant weather variable. An eight-feature model preserved the accuracy of the full feature set, demonstrating that transparent feature analysis can support model simplification without materially degrading forecast quality.

References

1. Breiman L. Random Forests. Machine Learning, 2001, 45, 5-32. https://doi.org/10.1023/A:1010933404324.
2. Friedman J. H. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics, 2001, 29(5), 1189-1232. https://doi.org/10.1214/aos/1013203451.
3. Lundberg S. M., Lee S.-I. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 2017, 30, 4765-4774.
4. Eskandari H., Saadatmand H., Ramzan M., Mousapour M. Innovative framework for accurate and transparent forecasting of energy consumption: A fusion of feature selection and interpretable machine learning. Applied Energy, 2024, 366, 123314. https://doi.org/10.1016/j.apenergy.2024.123314.
5. Neubauer A., Brandt S., Kriegel M. Explainable multi-step heating load forecasting: Using SHAP values and temporal attention mechanisms for enhanced interpretability. Energy and AI, 2025, 20, 100480. https://doi.org/10.1016/j.egyai.2025.100480.
6. Mohammadi A., Napoles G., Salgueiro Y. Deep Learning for Time-Series Forecasting With Exogenous Variables in Energy Consumption: A Performance and Interpretability Analysis. IEEE Access, 2025, 13, 86746-86767. https://doi.org/10.1109/ACCESS.2025.3570618.
7. Pedregosa F., Varoquaux G., Gramfort A. et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 2011, 12, 2825-2830.
8. scikit-learn Developers. Version 1.9.1 release notes, September 2026. https://scikit-learn.org/stable/whats_new/v1.9.html (accessed 01.10.2026).