https://doi.org/10.31891/csit-2024-4-7
UDC 004.89: 004.4: 551.509.331

Tetiana HOVORUSHCHENKO, Vitalii ALEKSEIKO,
Valeriia SHVAIKO, Juliia ILCHYSHYNA, Andrii KUZMIN
Khmelnytskyi National University

INFORMATION SYSTEM FOR EARTH'S SURFACE TEMPERATURE FORECASTING USING MACHINE LEARNING TECHNOLOGIES

Temperature forecasting is a topical issue in many areas of human life. In particular, climate change directly affects agriculture, energy, infrastructure, health care, logistics, and tourism. Anticipating future changes allows you to better prepare for challenges and minimize risks. The paper presents an information system for forecasting the temperature of the Earth's surface using machine learning technologies. The forecast is formed by a model adapted to the region, by learning on the basis of historical data and tracking the most inherent patterns. The selection and training of the model was carried out on the basis of the analysis of the characteristics of climatic zones, according to the Köppen classification. A comparison of the performance of models for forecasting the average monthly temperatures of the earth's surface in different climatic zones was carried out.

The analysis of scientific publications confirmed the relevance of the chosen research topic. Modern approaches to forecasting climatic indicators are considered. Methods and approaches to temperature forecasting, their advantages and disadvantages are analyzed.

The peculiarities of the application of machine learning methods for temperature forecasting are considered, and the criteria for choosing the most accurate and least energy-consuming methods are determined. The research results made it possible to identify machine learning methods that best adapt to temperature patterns and allow accurate short-term forecasting. An approach for long-term forecasting using recurrent neural networks is proposed.

An information system has been developed for forecasting future temperatures depending on the climatic features of the studied territories based on the proposed methods. A concept for further research for the development and improvement of the developed information system has been formed.

Keywords: machine learning (ML), forecasting, Earth's surface temperature, climate zone, information system.

Тетяна ГОВОРУЩЕНКО, Віталій АЛЕКСЕЙКО,
Валерія ШВАЙКО, Юлія ІЛЬЧИШИНА, Андрій КУЗЬМІН
Хмельницький національний університет

ІНФОРМАЦІЙНА СИСТЕМА ПРОГНОЗУВАННЯ ТЕМПЕРАТУРИ ПОВЕРХНІ ЗЕМЛІ З ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ МАШИННОГО НАВЧАННЯ

Прогнозування температури є актуальною темою в багатьох сферах життя людини. Зокрема, зміна клімату безпосередньо впливає на сільське господарство, енергетику, інфраструктуру, охорону здоров'я, логістику та туризм. Передбачення майбутніх змін дозволяє краще підготуватися до викликів і мінімізувати ризики. У роботі представлено інформаційну систему прогнозування температури земної поверхні з використанням технологій машинного навчання. Прогноз формується з використанням моделі, адаптованої до конкретного регіону, шляхом навчання на основі історичних даних і відстеження найбільш притаманних закономірностей. Вибір і навчання моделі проводили на основі аналізу характеристик кліматичних зон, згідно з класифікацією Кеппена. Проведено порівняння ефективності моделей прогнозування середньомісячних температур земної поверхні в різних кліматичних зонах.

Аналіз наукових публікацій підтвердив актуальність обраної теми дослідження. Розглянуто сучасні підходи до прогнозування кліматичних показників. Проаналізовано методи та підходи до прогнозування температури, їх переваги та недоліки.

Розглянуто особливості застосування методів машинного навчання для прогнозування температури та визначено критерії вибору найбільш точних і найменш енергозатратних методів. Результати дослідження дозволили виявити методи машинного навчання, які найкраще адаптуються до температурних закономірностей і дозволяють здійснити точний короткостроковий прогноз. Запропоновано підхід для довгострокового прогнозування з використанням рекурентних нейронних мереж.

Розроблено інформаційну систему прогнозування майбутніх температур залежно від кліматичних особливостей досліджуваних територій на основі запропонованих методів. Сформовано концепцію подальших досліджень щодо розвитку та вдосконалення розробленої інформаційної системи.

Ключові слова: машинне навчання (ML), прогнозування, температура поверхні Землі, кліматичний пояс, інформаційна система.

Introduction

Climate changes have a significant impact on various aspects of human life, covering, in particular, the spheres of health care, industry, agriculture, logistics, tourism, etc. Forecasting of such changes allows timely development and implementation of strategies to minimize negative consequences and reduce their impact on society. Thus, the fight against climate change is one of the defining goals on the way to sustainable development [1, 2].

The climate is characterized by a significant number of parameters that are interconnected to one degree or another. One of the main indicators is the temperature of the earth's surface.

Forecasting the temperature of the Earth's surface is an urgent issue, as it allows to monitor the main trends of climate change. Based on the assessment of such parameters, further studies can be conducted, which will

contribute to the prediction of dangerous phenomena and allow the necessary measures to be taken to avoid them in a timely manner.

The use of machine learning algorithms capable of capturing seasonal trends is extremely important for a number of tasks, including forecasting climate indicators [3, 4, 5].

Domain analysis

The majority of scientific publications related to the forecasting of climate parameters and the temperature of the earth's surface in particular, are devoted to the study of general trends in temperature changes, considering climate patterns from the point of view of global warming on a planetary scale [6], without taking into account regional features, or, on the contrary, focus on the studies of a narrow region [7]. Thus, the proposed approaches are not universal and solve only a narrow range of problems, which creates the need for further study of the issue of forecasting climate indicators.

Analysis of changes in the temperature of the Earth's surface in megacities since the middle of the 20th century showed that the vast majority of cities are not ready for climate change, and measures aimed at reducing the negative effects caused by global warming are insufficient [8].

It should be noted that some geographic areas are actually more prone to extreme weather conditions than others, and therefore these extreme signals should appear in the long-term forecast of climate variables in these areas. Accordingly, no one-size-fits-all model can predict well for all geographic areas [9].

Thus, we consider it expedient to conduct model research separately for each climate zone, according to the Köppen classification [10], to determine the most suitable temperature forecasting methods [11].

In the previous work [12] it was analyzed software development life cycle, the method of system analysis and determined requirements for the proposed information system.

Models evaluation

Prediction accuracy is the most important factor for forecasting model. The method must be able to make accurate predictions on new data. Also, there are some other important factors.

Energy efficiency, environmental friendliness and speed of operation are important factors in choosing machine learning methods for the following reasons.

First of all machine learning methods often require large amounts of computing resources. This can result in significant power consumption, especially during working with large datasets or complex models. Energy-efficient algorithms can reduce energy costs and reduce environmental impact.

Secondly, high energy consumption can lead to increased carbon dioxide emissions, especially if the energy comes from non-renewable sources. Green machine learning techniques can help reduce environmental impact.

Finally, fast machine learning algorithms can provide timely and accurate results.

In addition to it, there are many other factors that play an important role in choosing machine learning methods:

– model complexity;
– interpretability;
– robustness;
– scalability;
– versatility;
– matching task requirements.

Complex models can be more accurate, but they can also be more susceptible to overfitting and take longer to train and predict. In some cases, it is important that the model is not only accurate, but also interpretable. This means human can understand how the model makes its predictions. Also, the model must be robust to small changes in the data. The method must be able to work efficiently with large volumes of data and must be able to work with various data types and tasks. Moreover, some methods may be more suitable for certain types of tasks than others.

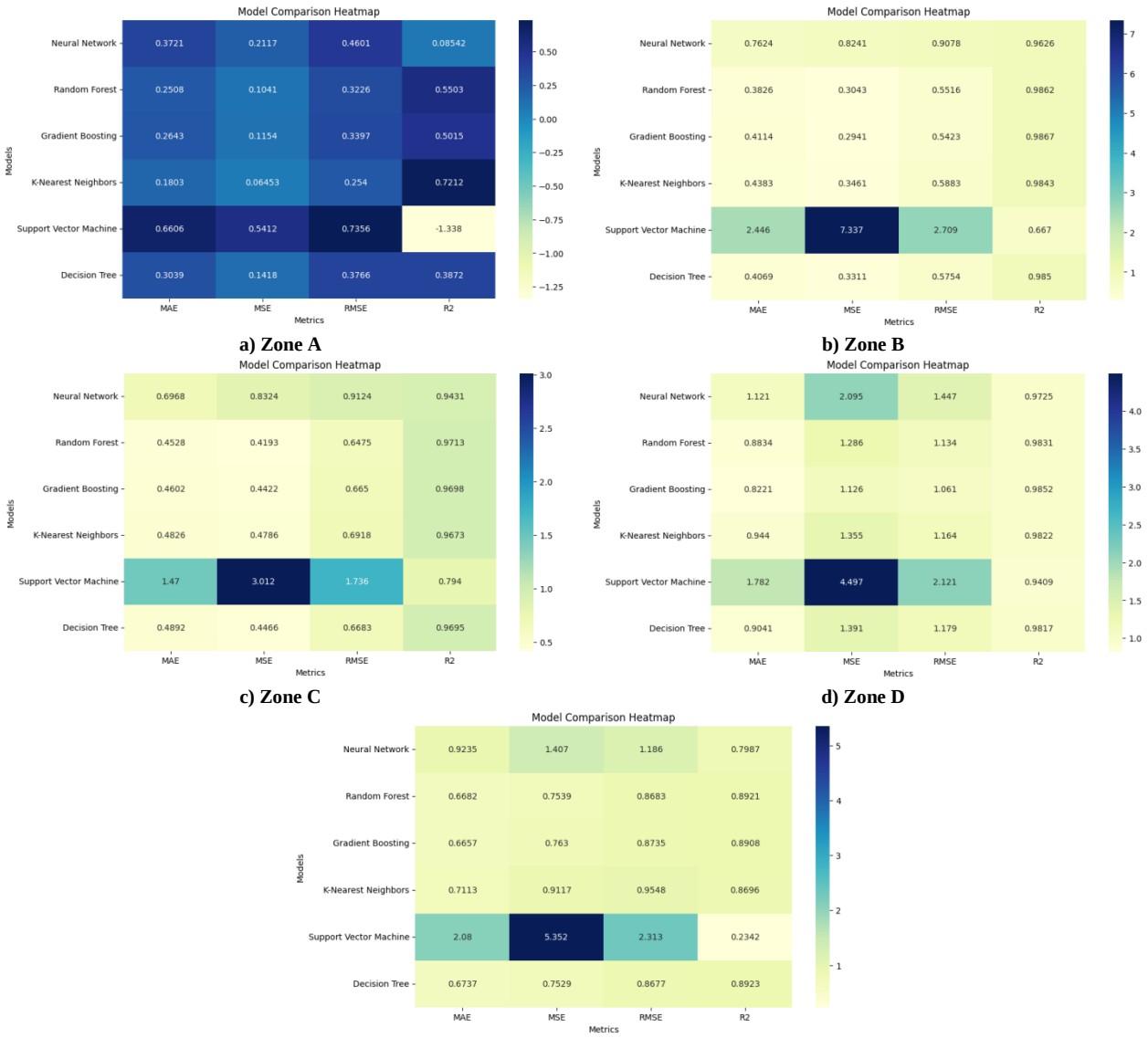
The models were evaluated based on metrics:

– Mean Absolute Error (MAE);
– Mean Squared Error (MSE);
– Root Mean Squared Error (RMSE);
– R2 score (R2).

In the research was used dataset GlobalLandTemperatures [13] with Creative Commons License (CC0: Public Domain) from Kaggle. Due to needs of research the dataset was modified according to the World Climate Data [14].

Figure 1 demonstrates comparison heatmaps for each model in each climate zone.

According to the conducted research, different methods demonstrate different accuracy on the data of different climate zones. Unfortunately, considered methods are not suitable for long-term forecasting because they can not follow up trends of temperature rise due to Global Warming. Table 1 shows the most suitable methods for short-term forecasting for each climate zone.



a) Zone A **b) Zone B**

c) Zone C **d) Zone D**

e) Zone E

Fig. 1. Model Comparison Heatmaps for climate zones

Table 1

The most suitable methods for short-term forecasting for each climate zone

Climate Zone	Method
A	K-Nearest Neighbors (KNN)
B	Gradient Boosting (GB)
C	Random Forest (RF)
D	Gradient Boosting (GB)
E	Decision Tree (DT)

The study of the work of the most suitable models for each climate zone is presented as a charts, where the predicted temperature is indicated in red, and the actual temperature observed in the studied region is indicated in blue. The forecasting results are shown in the figures 2 – 6 for the following climate zones:

– tropical (figure 2);
– arid (figure 3);
– temperate (figure 4);
– continental (figure 5);
– polar (figure 6).

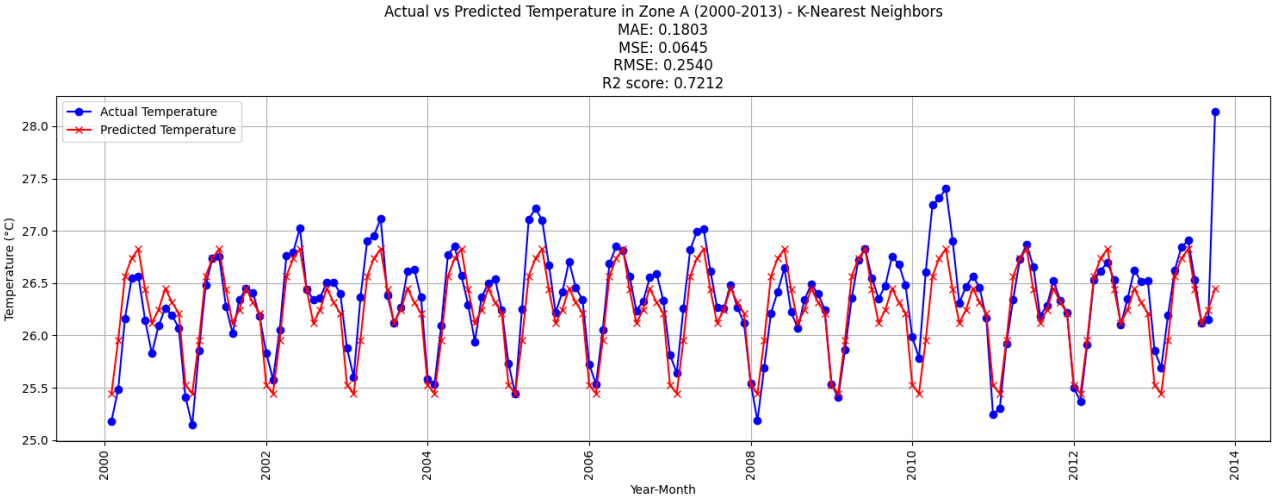


Fig. 2. K-Nearest Neighbors forecast for zone A

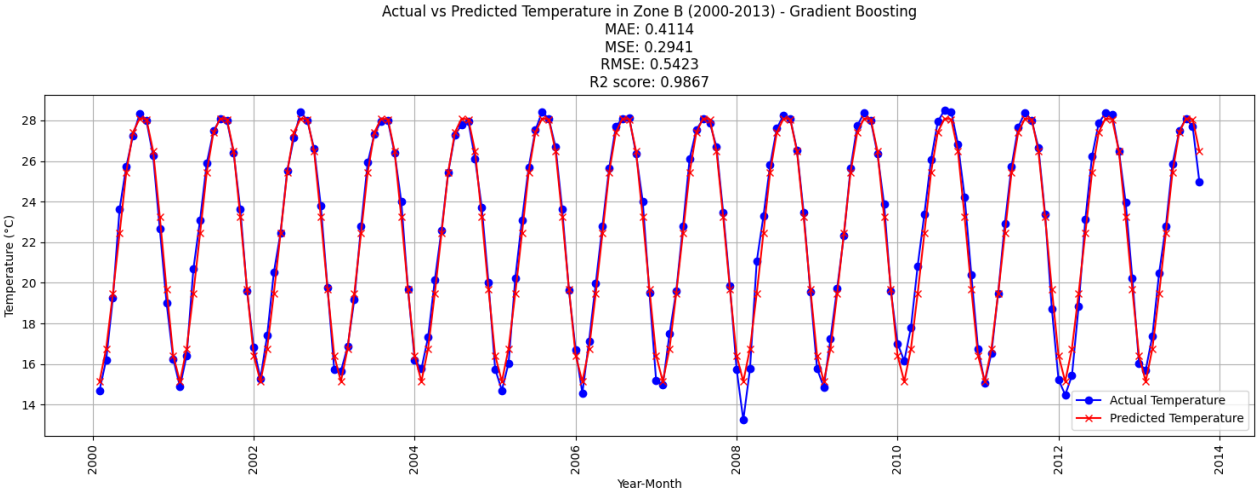


Fig. 3. Gradient Boosting forecast for zone B

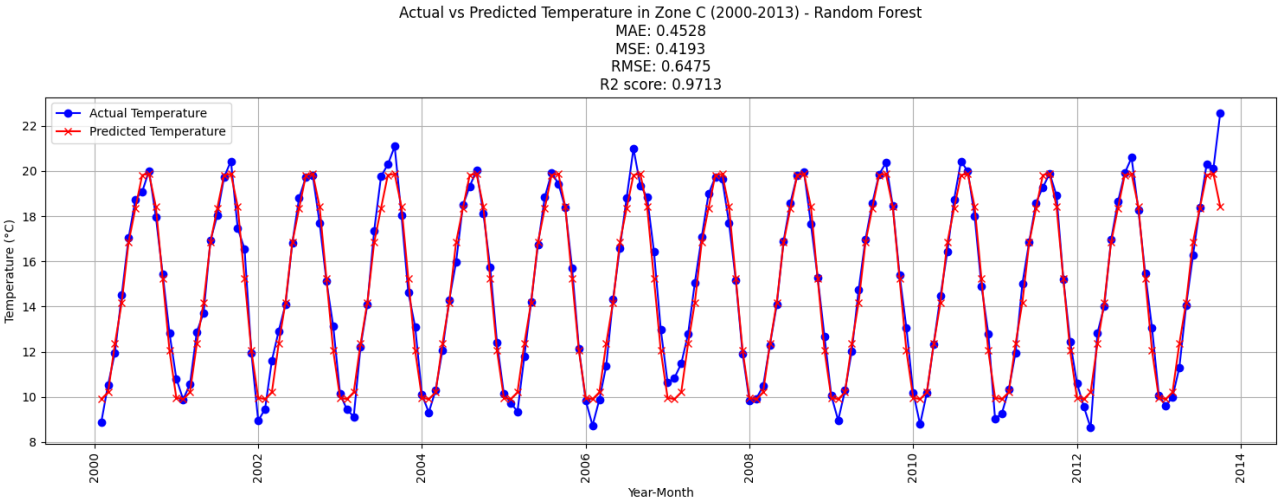


Fig. 4. Random Forest forecast for zone C

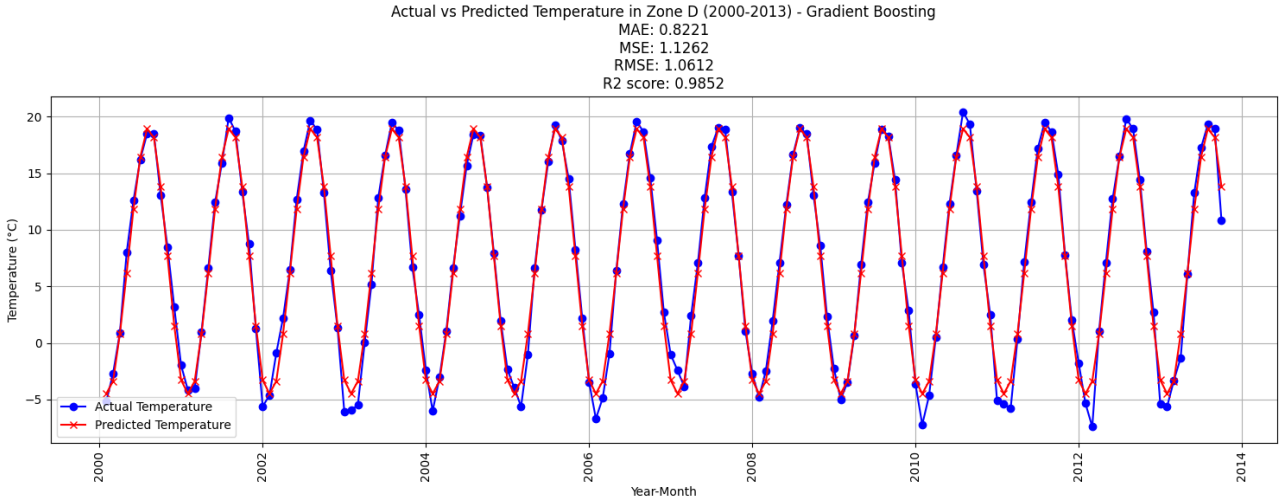


Fig. 5. Gradient Boosting forecast for zone D

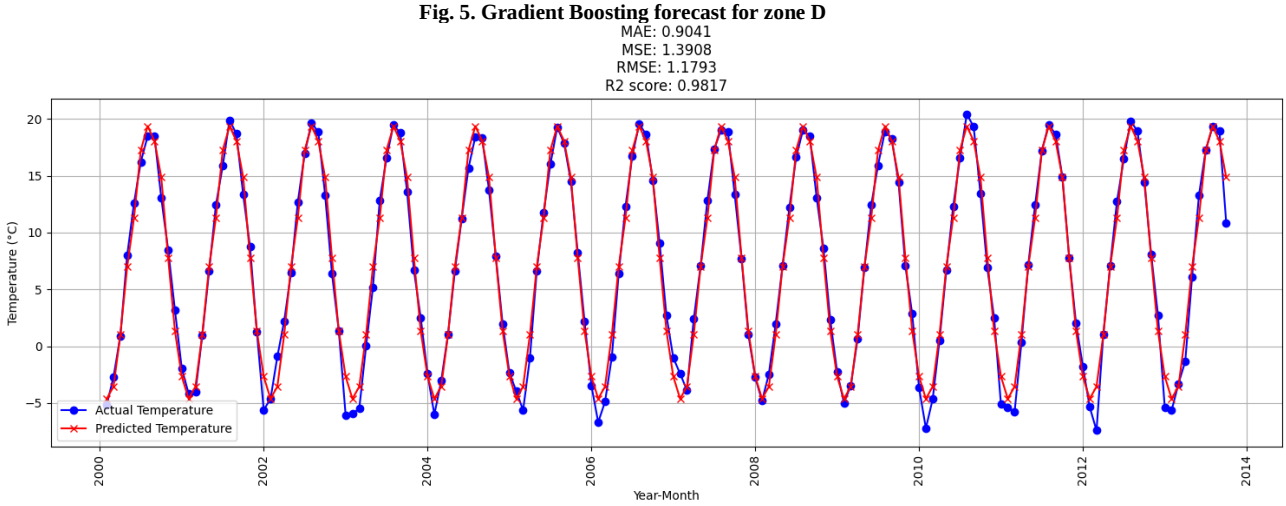


Fig. 6. Decision Tree forecast for zone E

All methods works not very good for forecasting temperature in Zone A. They performs low R2 score (< 0.75). Although for zones B, C and D they works well (R2 score is over 0.96).

According to the specific of countries location on the globe, there are few countries located in the polar climate zone. So, this region remains poorly studied and the dataset has limited information on these territories. Coming out of this, we get the result that the decision tree method performs better than random forest because of its simplicity. In the case of obtaining large amounts of data, we consider it advisable to conduct a separate study for this region in order to select the most optimal forecasting method.

To solve the problem of long-term forecasting, it's advisable to consider more complex models, in particular recurrent neural networks (RNN), which include the LSTM (long-short term memory) method.

Results

An information system for forecasting the temperature of the earth's surface was developed as a web application using the Django framework. This choice is justified by the support of the Python programming language, which is widely used in the field of artificial intelligence and machine learning due to its rich ecosystem of libraries and frameworks, good scalability that allows the development of the project, as well as built-in functions that provide security, which is critical for web applications, which use AI models. Thus, Python's combination of syntax simplicity, flexibility, and compatibility makes Django an excellent choice for developing AI applications. From the user's point of view, web application has a lot of benefits because the access to the system via the Internet can be provided to any user through a browser and does not require software installation, since all calculations are carried out on a remote server.

It was used PyCharmCommunity [15] and Google Colaboratory [16] to test dataset and to develop technical solution and implementation.

The data entered by the user includes country from the proposed list and year for which a forecast should be made. This information is transmitted to the server. After processing the received data, the system selects a model for forecasting and requests the necessary information from the database. The obtained values of expected temperatures are displayed via the User Interface (UI) in numerical and graphical representations.

The architecture of the information system (Figure 7) was presented in previous work [11]. Figure 8 displays sequence diagram of the information system. Figures 9 – 10 demonstrates design of application.

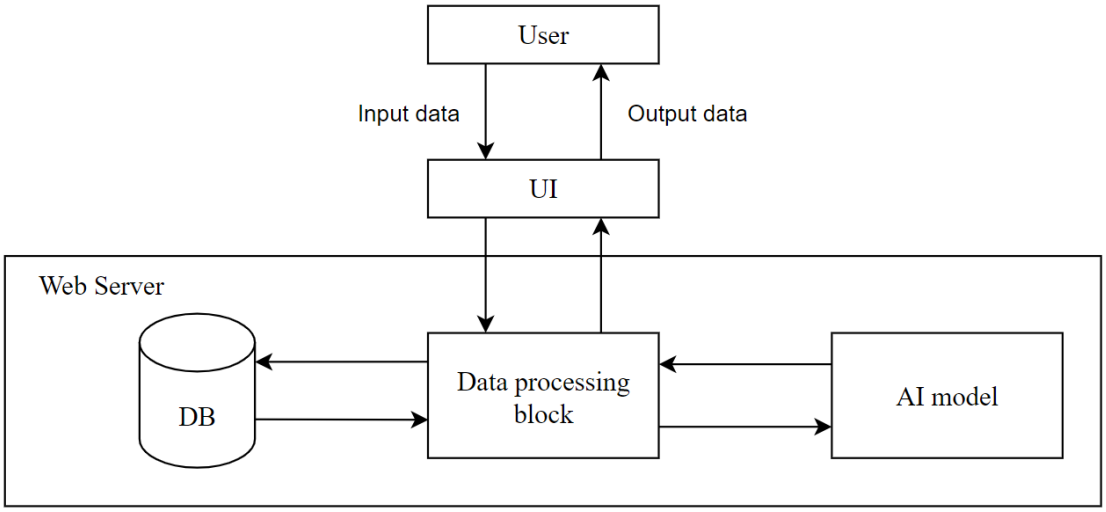


Fig. 7. Architecture of the information system [11]

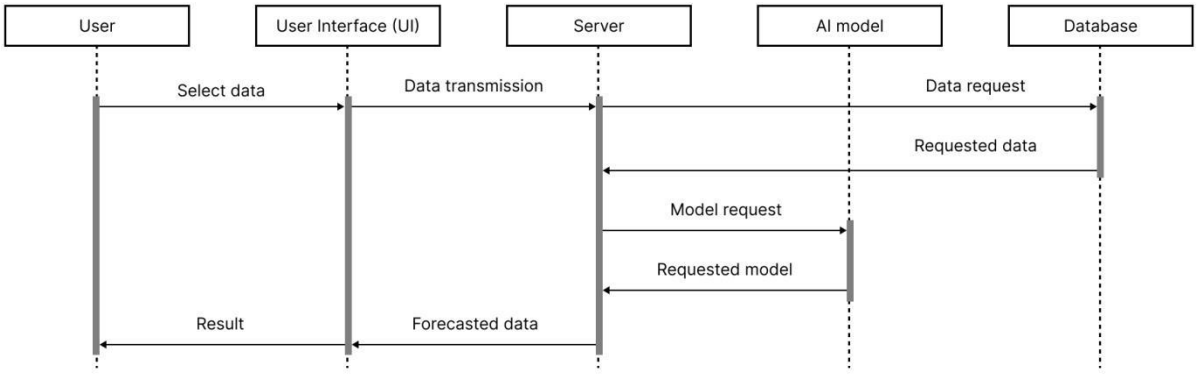


Fig. 8. Sequence diagram of the information system

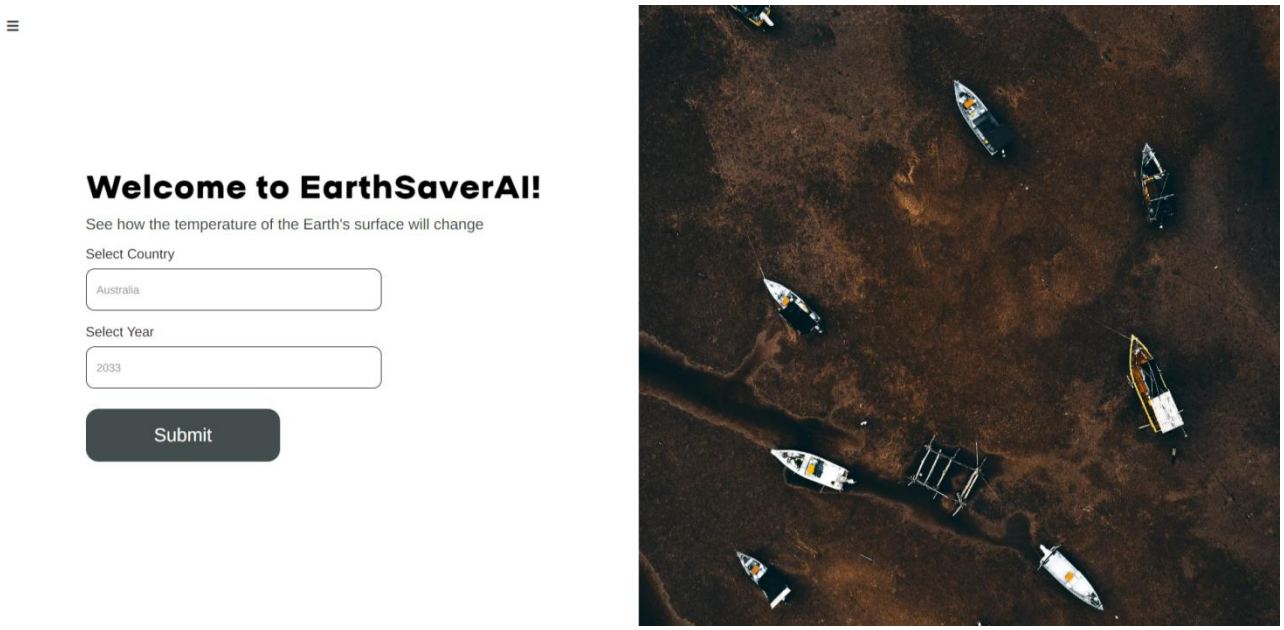


Fig. 9. Design of application

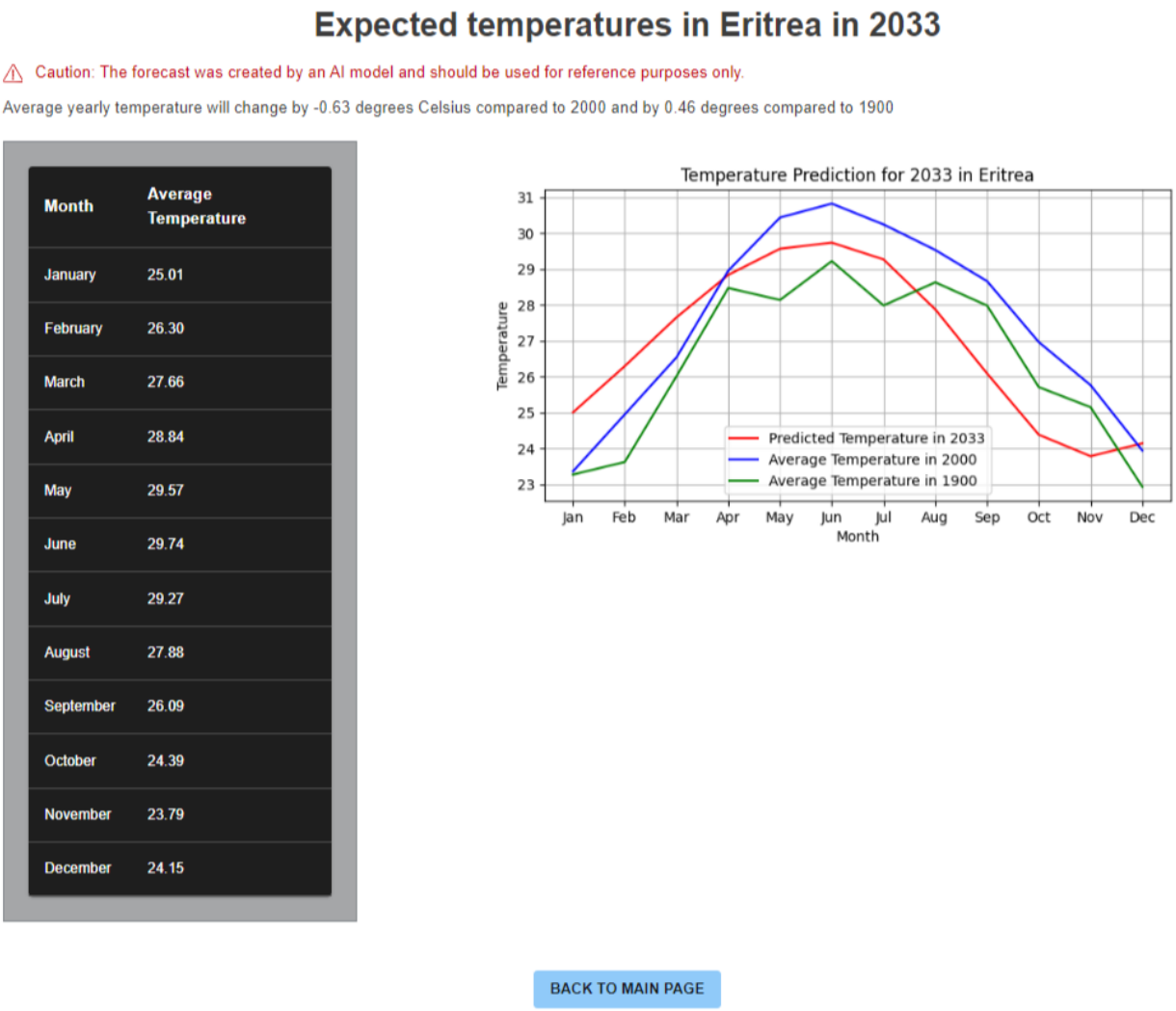


Fig. 10. Design of application

Conclusions

On the basis of the conducted research, the most effective methods for short-term forecasting of the Earth's surface temperature in each climate zone have been determined. An approach to long-term forecasting using recurrent neural networks, namely the long-short-term memory method, is proposed.

The developed web application allows to forecast the temperature of the Earth's surface, using an individual approach to each of the climatic zones.

The software product is ergonomic, easy to use, and has an intuitive interface.

Directing further research to determine regularities in individual climatic subzones will allow to more accurately establish regularities for different regions and promote their development in terms of climatic challenges.

References

1. THE 17 GOALS. Sustainable Development. https://sdgs.un.org/goals
2. P. Hryhoruk, S. Grygoruk, N. Khrushch, T. Hovorushchenko. Using non-metric multidimensional scaling for assessment of regions' economy in the context of their sustainable development. CEUR-WS. (2020). Vol. 2713. 315-333.
3. R. He, L. Zhang, A. W. Z. Chew. Modeling and predicting rainfall time series using seasonal-trend decomposition and machine learning. *Knowledge-Based Systems*, *251*, 109125. 2022. https://doi.org/10.1016/j.knosys.2022.109125
4. R. C. Joshi, D. Ryu, P. N. Lane, G. J. Sheridan. Seasonal forecast of soil moisture over Mediterranean-climate forest catchments using a machine learning approach. *Journal of Hydrology*, *619*, 129307. 2023. https://doi.org/10.1016/j.jhydrol.2023.129307
5. V. Ivanenko, V. Alekseiko, O. Klymenko. Machine forecast evaluation: an algorithm for trends, financial indicators, and business process optimization. *Science and technology today*, *8*(36), 2024. 1018–1033. https://doi.org/10.52058/2786-6025-2024-8(36)-1018-1033
6. S. M. Malakouti. Utilizing time series data from 1961 to 2019 recorded around the world and machine learning to create a Global Temperature Change Prediction Model. *Case Studies in Chemical and Environmental Engineering*, Volume 7, 2023. https://doi.org/10.1016/j.cscee.2023.100312
7. T. Wang, Y. Zhang, X. Zhi, Y. Ji. Multi-Model Ensemble Forecasts of Surface Air Temperatures in Henan Province Based on Machine Learning. *Atmosphere* 14, no. 3: 520. 2023. https://doi.org/10.3390/atmos14030520
8. T. Hovorushchenko, V. Alekseiko. Land surface temperature forecasting in the context of the development of sustainable cities and communities. *Computer Systems and Information Technologies*, *3*, 2024. 6–12. https://doi.org/10.31891/csit-2024-3-1

9. D. Fister, J. Pérez-Aracil, C. Peláez-Rodríguez, J. Del Ser, S. Salcedo-Sanz. Accurate long-term air temperature prediction with Machine Learning models and data reduction techniques. *Applied Soft Computing*. Volume 136, 2023. https://doi.org/10.1016/j.asoc.2023.110118
10. Köppen Climate Classification System. https://education.nationalgeographic.org/resource/koppen-climate-classification-system/
11. O. Pavlova, V. Alekseiko. The concept of an information system for forecasting the temperature regime of the earth's surface based on machine learning. *Computer Systems and Information Technologies, 2*, 2024. 6–13. https://doi.org/10.31891/csit-2024-2-1
12. V. Alekseiko. Web-based information system for land surface temperature forecasting using machine learning methods. *Science and technology today*, *10*(38), 2024. 17–27. https://doi.org/10.52058/2786-6025-2024-10(38)-17-27
13. Kaggle. *Globallandtemperature.* 2018. https://www.kaggle.com/datasets/sambapython/globallandtemperature
14. List of countries by climate zone and average yearly temperatures. https://weatherandclimate.com/countries#google_vignette
15. PyCharm. Jet Brains IDEs. https://www.jetbrains.com/pycharm/download/?section=windows
16. Google Colaboratory. https://colab.research.google.com

Tetiana Hovorushchenko **Тетяна Говорущенко**	DrSc (Engineering), Dean of the Faculty of Information Technologies, Professor of Computer Engineering & Information Systems Department, Khmelnytskyi National University https://orcid.org/0000-0002-7942-1857 e-mail: govorushchenko@gmail.com	Доктор технічних наук, декан факультету інформаційних технологій, професор кафедри комп'ютерної інженерії та інформаційних систем, Хмельницький національний університет
Vitalii Alekseiko **Віталій Алексейко**	PhD Student majoring in Information Systems and Technologies, Assistant Lecturer of Computer Engineering & Information Systems Department, Khmelnytskyi National University https://orcid.org/0000-0003-1562-9154 e-mail: vitalii.alekseiko@gmail.com	Аспірант кафедри комп'ютерної інженерії та інформаційних систем, Хмельницький національний університет
Valeriia Shvaiko **Валерія Швайко**	Master student majoring in Computer Engineering, Khmelnytskyi National University https://orcid.org/0009-0001-4267-3479 e-mail: lera.schvajko@gmail.com	Магістрантка спеціальності «Комп'ютерна інженерія», Хмельницький національний університет
Juliia Ilchyshyna **Юлія Ільчишина**	Master student majoring in Computer Engineering, Khmelnytskyi National University e-mail: juliaillchishina@gmail.com	Магістрантка спеціальності «Комп'ютерна інженерія», Хмельницький національний університет
Andrii Kuzmin **Андрій Кузьмін**	Master student majoring in Computer Engineering, Khmelnytskyi National University https://orcid.org/0009-0005-6489-225X e-mail: andriy1731@gmail.com	Магістрант спеціальності «Комп'ютерна інженерія», Хмельницький національний університет