EURASIAN JOURNAL OF MATHEMATICAL AND COMPUTER APPLICATIONS
ISSN 2306–6172
Volume 12, Issue 2 (2024) 4 –15

OPTIMIZATION OF WATER RESOURCES MANAGEMENT USING BIG DATA AND MACHINE LEARNING IN SMART CITIES

Akhayeva Zh.B., Zakirova A.B., Zhmud V.A.

Abstract In the modern world, the concept of Smart Cities brings new challenges and opportunities for effective resource management and improving the quality of life of citizens. One of the key aspects of Smart Cities is the optimization of water resources management, especially in the context of growing urbanization and a changing climate. Working with big data plays a crucial role in this process, allowing you to analyze and predict water consumption with high accuracy and efficiency. This article examines current methods and technologies for working with big data in the context of predicting water consumption in Smart cities, their advantages and prospects for application in everyday life and urban resource management.

Key words: big data, machine learning, forecasting, water consumption, Smart city.

AMS Mathematics Subject Classification: 68T09, 68T07, 60G25, 62R07, 93C85.

DOI: 10.32523/2306-6172-2024-12-2-4-15

1 Introduction

Fresh water is one of the most valuable resources, determining not only economic development, but also the viability of the whole society, therefore, economical use of water resources is extremely important. This can be achieved most effectively only with the use of modern information technologies for monitoring water consumption and water resources management. In Kazakhstan, due to the vast territory and a wide variety of climatic conditions, the issue of water resources management is especially relevant, and its relevance is only increasing over the years. Here, water is a strategically important resource, both for agriculture and industry, as well as for maintaining ecosystems and ensuring the vital activity of the population, therefore, its conservation is of paramount importance. In different regions, the severity of the problem of water use and water conservation is determined by various factors, such as the lack of fresh water in arid areas (especially with an agricultural bias) or the threat of pollution of water sources in industrial centers. In the context of urbanization and industrialization, water management issues are becoming even more complex and require a systematic approach and innovative solutions. In the context of Kazakhstan, special attention is paid to the management of water resources in Smart cities. The rapid development and modernization of cities create new challenges and opportunities in the field of water resources management, especially taking into account the ongoing program to transform them into a modern Smart city. Big cities, as one of the largest cities in the region, face particularly acute problems related to water resources. On the one hand, it is the need to provide water to a growing population and industrial enterprises. On the other hand, the city faces the task of minimizing the negative impact on the environment and preserving the quality of water systems in conditions of rapid urban growth. This article is devoted to the analysis of the role of water resources in the development of smart cities, an overview of the challenges and solutions related to the management of these resources in order to make an informed choice of innovative approaches to forecasting and optimizing water consumption using machine learning. The article describes the development of a program to optimize the use of water resources in order to prevent excessive consumption and infrastructure costs, as well as to minimize water losses due to leaks and improper management. Effective management of water consumption will reduce the cost of maintaining water supply and sewerage infrastructure, which will bring economic benefits to both the city and its residents. Reducing water consumption and optimizing water use processes will help reduce pressure on natural water resources and minimize negative environmental impacts. Sustainable water supply plays a key role in providing comfortable living conditions for city residents, providing them with access to clean water for drinking, agricultural needs, sanitation and industry. The water consumption forecasting program fits into the city's development strategy as a Smart City, improving infrastructure and increasing resource management efficiency with the help of advanced technologies and data analytics [1]. In general, this water consumption forecasting program will not only help ensure a sustainable water supply for the city, but also make it more competitive, sustainable and attractive to residents, businesses and investors.

2 Materials and methods

Modern methods of data analysis and machine learning are used for data analysis and processing. Time series methods have been applied to analyze the dynamics of water consumption in different time periods. In addition, machine learning algorithms such as linear regression and neural networks have been used to build models for predicting water consumption based on historical data and external factors such as climatic conditions and demographic indicators. To conduct the study, data on water supply and pressure in pumping stations were used, according to which water consumption in the city over the past year was calculated. These data are provided by the organization responsible for the city's water supply, and include indicators of water consumption in various areas of the city. Thus, an effective solution to water supply problems continues to be a key issue for many years, and this situation will only worsen in the future. The water dispatch system is an important part of the water management infrastructure in the city. As a municipal water supply company, it is responsible for providing high-quality and reliable water supply to residents and enterprises of the city [2].

The proposed software part of the system is written in a popular programming language such as Python, it will be implemented on the Raspberry Pi controller. The controller will be connected to sensors for measuring quality and level, such as the HC-SR04 ultrasonic range sensor for measuring water level, a pH sensor for determining the pH of water level, etc. For real-time monitoring integrated into the controller of the IoT platform, in Blynk. This platform allows you to control Internet of Things devices, such as Raspberry Pi, over the Internet. With this integration, the water

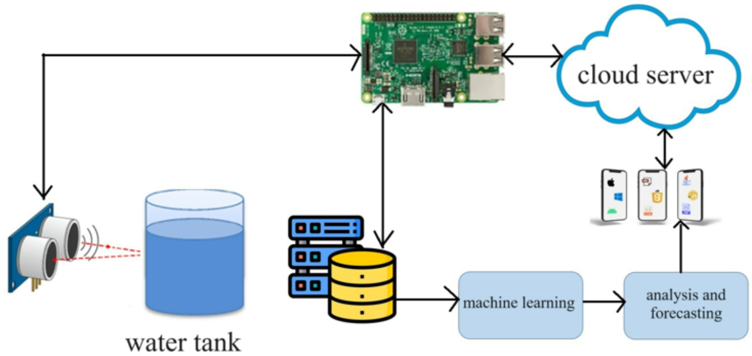

Figure 1: Architecture of the developed system

level can be displayed in real time in a mobile application, as well as transmitted to a database. The Big Data obtained is processed, a forecast is made, and data analysis is performed for water resources management [3, 4]. The data obtained were processed using data analysis and statistical methods to identify trends and patterns in water consumption in the city. This stage included removing outliers, filling in missing values, and aggregating data over time intervals for easier analysis. Various software tools are used for data processing and model building, including the Python and R programming languages, as well as specialized libraries and packages for data analysis such as Pandas, NumPy, scikit-learn and TensorFlow [5, 6]. There are many studies in this field conducted by various authors and research groups. In order to choose the optimal solution, various studies of authors whose work was related to the study of mathematical models for the management of water resources of the city were analyzed.

Table 1 – Researchers in the field of mathematical models for water management

Michael J. Gorelick	Professor of Hydrology and Hydrogeology at Stanford University, USA. His research includes mathematical modeling of hydraulic systems and water management.
Dr. Claudia Maria Zehe	Professor of Hydrology at the Technical University of Munich, Germany. Her research focuses on the development of mathematical models for the assessment and management of surface and groundwater resources.
Dr. David Hanigan	Associate Professor of Hydrology at the University of Texas at Austin, USA. His research includes the development of mathematical models for forecasting and management of surface water resources.
Dr. Claudia D'Ortenzio	Researcher in the field of Hydrology and water resources management at the Pierre-Simon Laplace Institute, France. Her work includes the development of mathematical models for the analysis and forecasting of water systems in various climatic conditions.
Dr. Jianyun Zhang	Professor of Hydrology at Tongji University, China. His research includes the development of mathematical models for the assessment and management of water resources, especially in the context of urban areas.

The mathematical model of the city's water management information system was to use a set of equations describing the main processes and relationships in the system.

The equation of water balance:

$$\frac{dV}{dt} = P - E - R - W \tag{1}$$

where

V - is the total volume of water in the system;

P – water inflow (precipitation, aquifers);

E – evaporation;

R - is runoff (water outlet from the system);

W - is water consumption (for example, consumption by the urban population and industry).

Various machine learning methods such as time series models, regression analysis and neural networks have been used to predict future water consumption. The data is divided into training and test samples to assess the quality of the models. The modeling process consisted of several stages, including data preparation, model selection and configuration, model training and evaluation. Thus, by collecting data, analyzing time series, developing a predictive model and optimizing water consumption, we will build a mathematical model that will predict the water level and optimize water consumption in the water supply system. Let's start with a simple model for predicting the water level in the system. To do this, it is proposed to use the autoregression (AR) method, which is one of the most common methods of time series analysis [7]. Let's assume that there is a time series containing data on the daily water level in the information system over the past few years. Let's denote this time series as Y_t, where t is the ordinal number of the day. The AR(p) model assumes that the current value of the time series depends on the previous p values of the time series. Thus, we can write the model in the form:

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + \varepsilon_t \tag{2}$$

where

Y_t- is the current value of the time series,

c - is a constant,

$\varphi_1, \ \varphi_2, \ \ldots \ \varphi_p$ – autoregression coefficients,

μ_t - is a random error at time t.

The formula for the autoregression model (MAR) was first introduced in works on econometrics, namely in the work of Jan Tinbergen, a Dutch economist and Nobel Prize winner in Economics[8].

He developed autoregression and time series correlation models in his publications in the 1930s.

We visualize the mathematical model of autoregression (MAR) on a graph.

Let's imagine that we have a time series with daily water level data. Let's plot this time series and add lines that will represent the dependence of the current value on the previous p values, as presented in the MAR(p) formula.

Let's assume that we chose *p=2* for this MARK(2) model. Then on the graph we can represent the dependence of the current value Y_t the previous two values of Y_{t-1} and Y_{t-2} as follows:

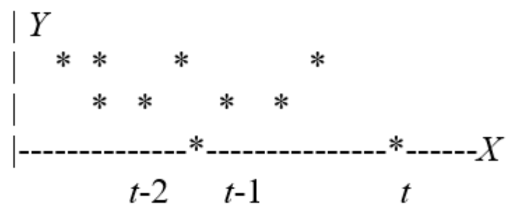

Figure 2: Visualization of the mathematical model of autoregression (MAR) on the graph

In this graph (Figure 2):

The X is represents the time points $t-2, t-1$ and t.

The Y is represents the values of the water level Y_t.

Each point on the graph represents the value of the water level at a certain point in time.

The lines marked with asterisks represent the dependence of the current value on the previous two values of Y_{t-1} and Y_{t-2}

This graphical image helps to visualize how the current value of the water level depends on previous values, which reflects the MAR(2) model.

The quality of forecasts was assessed using various metrics, such as standard error (MSE), coefficient of determination (R^2) and others. This allowed us to determine the effectiveness of various models and choose the most appropriate approach to predicting water consumption in the city.

3 Results and discussion

The implementation of It and monitoring systems includes the installation of sensors and sensors at various points of water and sewer networks for continuous monitoring of the system status. This allows the company to quickly respond to accidents, identify leaks and optimize the use of water resources (Fig.3).

Real-time management using the "Smart Water" system allows the company to manage water supply and sewerage in real time. This includes regulation of water flow, monitoring of network pressure, as well as automatic control of pumping stations (Fig.4).

Analytics and forecasting includes the development of analytical tools and models to predict changes in the water supply system. This allows the company to analyze data on water consumption, identify trends and make informed decisions to optimize the operation of the system.

Thus, the water supply in the reservoirs is an important element of the infrastructure of the water supply system, which ensures reliable and stable operation of

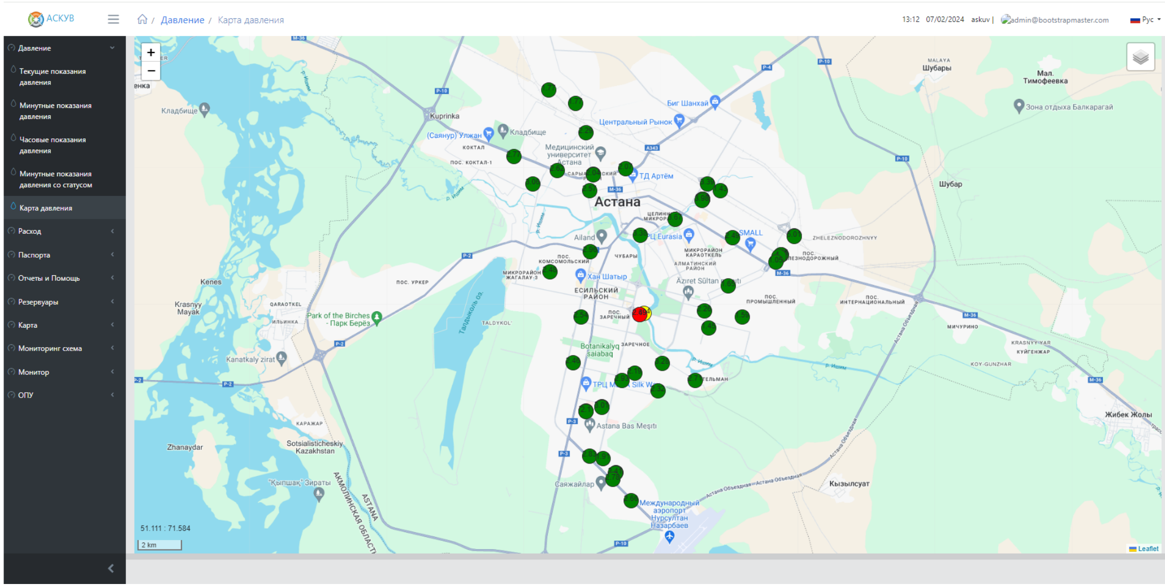

Figure 3: Map of water pressure in the city, data transmission online

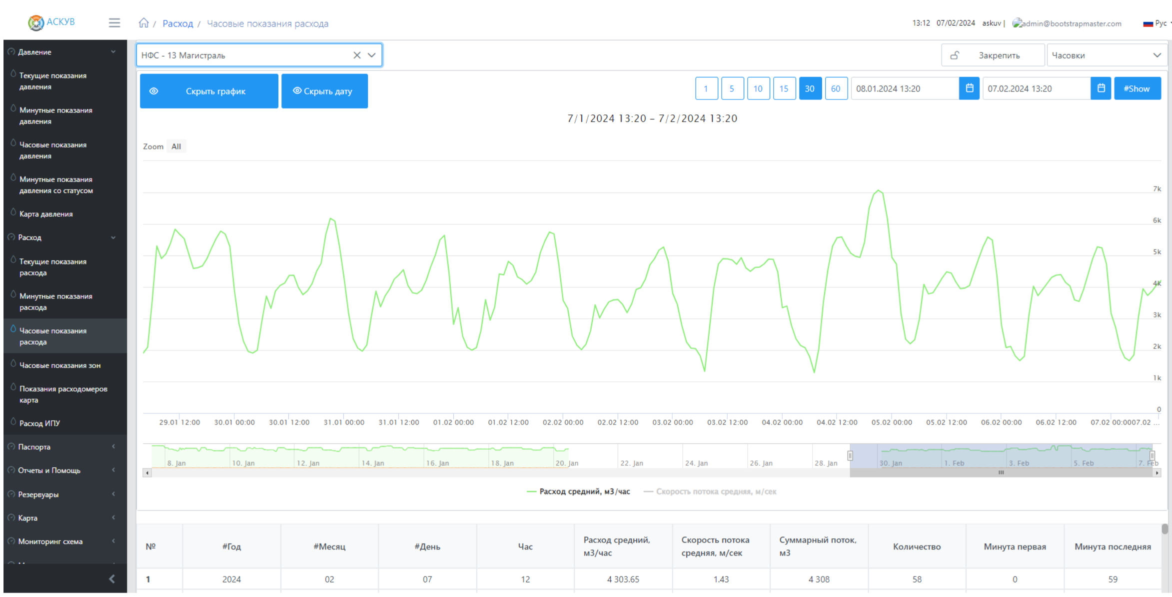

Figure 4: A graph of water consumption in a certain period of time

the system, as well as guarantees access to water in case of unforeseen circumstances (Fig.6).

In general, the water dispatch system plays an important role in ensuring stable and high-quality water supply to the city. It will make it possible to effectively manage water resources, minimize the risks of accidents and provide residents with reliable access to water [2]. A monitoring system based on Internet of Things (IoT) technology, sensors and detectors installed at various points in the water supply and sanitation network collect data online and send it. This allows you to quickly respond to emergencies, detect leaks in a timely manner and optimize the use of water resources. A

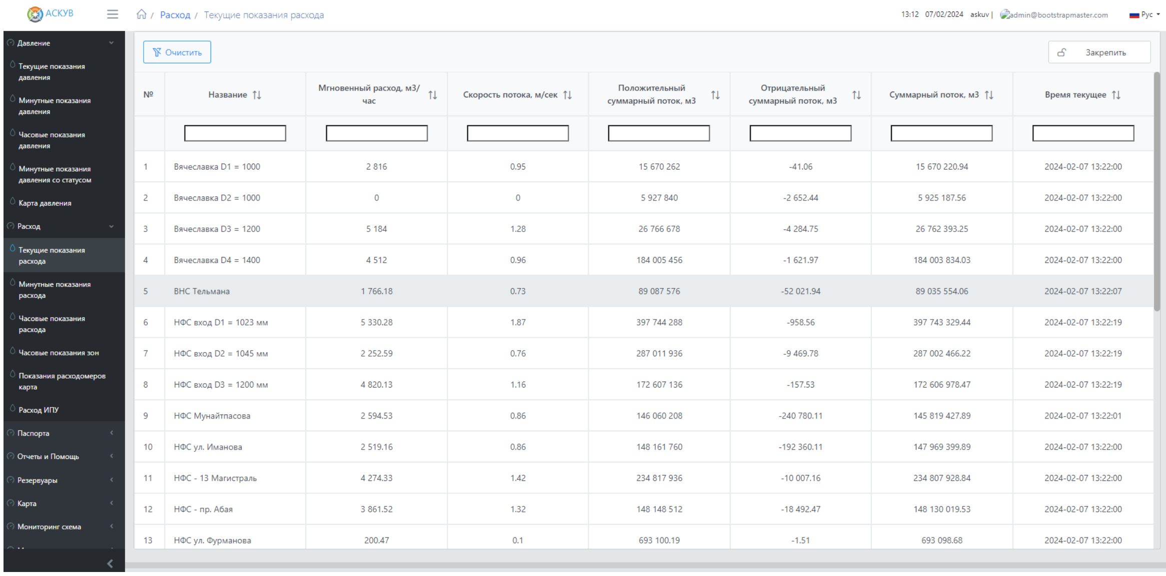

Figure 5: Flow chart

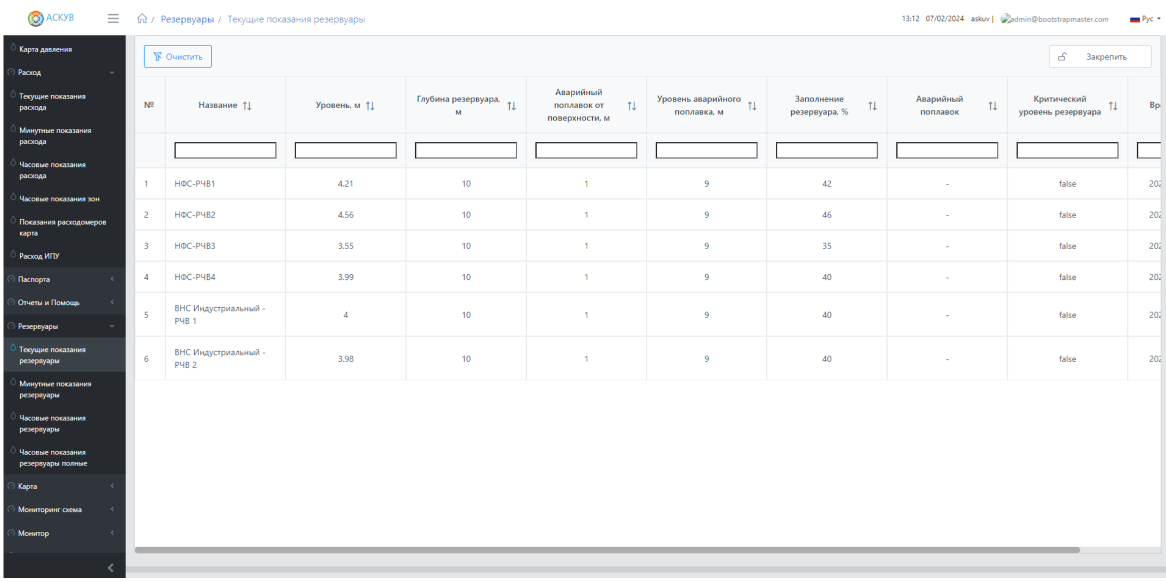

Figure 6: Water supply in tanks

random forest model was chosen to implement the machine learning program. This model is trained by combining several decision trees. This process, known as bootstrap aggregation, is complemented by introducing randomness at each stage of the tree construction. Random forests are a powerful machine learning technique that can be used for both classification and regression. It is based on an ensemble of decision trees, where each tree is built independently based on a subsample of data and a subset of features. The final forecast is formed by averaging the results of all trees. [7]. To solve the problem, marked pairs of objects $(x_i, y_i)^N$ $i = 1$ are used, where xi is a feature description of the object, and yi is a class label belonging to a discrete set whose power corresponds to the number of classes in the problem. It is necessary to build a composition of decision trees with two main properties: The final composition should ensure high accuracy of classification of new zi objects whose class label is unknown. The constructed solution should be sufficiently stable: with small changes in parameters, there should be no significant decrease in the quality of the model.

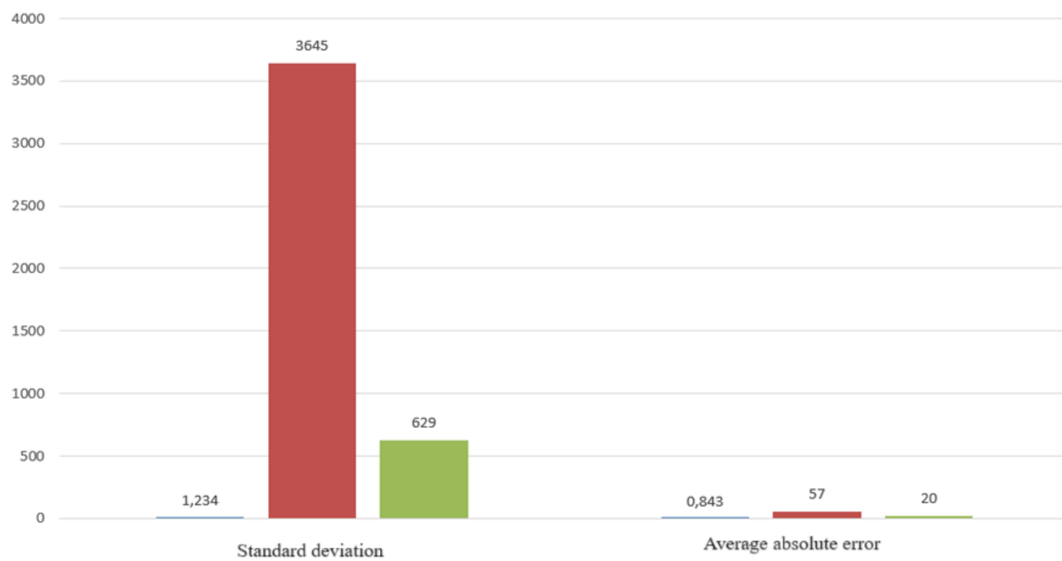

Figure 7: Graph of model deviations and errors

In our example, the standard deviation (MSE) is 629, and the absolute error (MAE) is 20. These values show how much the average error in the forecasts of the linear model is compared to the actual values. The lower these values, the better the quality of the model.

The Logistic regression model gave such a deviation:

Standard deviation: 3645

Average absolute error: 57

This makes it clear to us that these models are not suitable for us.

The Random Forest model

Standard deviation: 1.234.

Average absolute error: 0.843.

The forecasting program was implemented in Python. Using data on water consumption over the past year, machine learning was performed.

The program is divided into modules. The first module is the Database Module (Fig.8).

Also, for more convenient use of this program, a graphical interface has been developed (Figure 8).

Figure 9 shows the graphical interface of the program, where there are 4 windows for data entry and three conditions. In the first window, the required date is entered, in the second window the day is counted, in the third window the daytime temperature and in the fourth window the night temperature. Three conditions: it is raining or snowing, cloudy, sunny. Training a model tied to a specific day in the calendar may have

```python
if __name__ == '__main__':

    # DB name = DB4ML.db
    db_name = 'DB4ML.db'

    create_table_weather(db_name)
    create_table_water_consumption(db_name)

    get_all_data_from_table(db_name, 'weather')
    get_all_data_from_table(db_name, 'water_consumption')

    # drop_table(db_name, 'weather')
    # drop_table(db_name, 'water_consumption')
    print_all_tables_on_DB(db_name)
```

Figure 8: Database Module

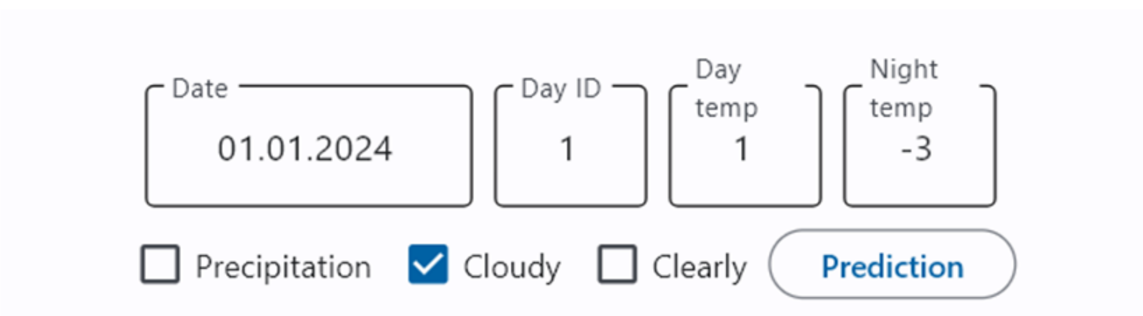

Figure 9: Graphical representation of the program

different reasons, depending on the context and the specific task: Seasonal trends – if the data has a seasonal dependence, for example, sales of goods, weather or economic indicators, it is important to take into account the dates in the calendar for correct modeling of seasonal variations [8]. Holidays and Events – Some days in the calendar may be associated with holidays, events, or other events that affect user behavior or the metrics that we analyze. For example, sales of goods may change significantly during holidays or sales periods. Technical aspects – in some cases, linking the model to a specific day can be useful for updating and retraining the model at a certain time. For example, if data is only available up to a certain point, the user can train the model on data up to that date and use it to predict future events. Thus, we get a forecast of water consumption on a certain day or period, which will help in the future for analysis and making certain decisions. Successful cases of smart water technology implementation:

1. Spain - water supply management in Barcelona. Barcelona has implemented a water supply management system using IoT (Internet of Things) and big data analysis to monitor and optimize water distribution. This made it possible to reduce water losses by 20 % and increase the efficiency of resource use.

2. Singapore is an intelligent water network. Singapore's water management system uses machine learning and big data to predict water demand and manage water reserves. This system helps to significantly reduce water supply costs and improve water quality.

3. USA - the Smart H2O project in California. In California, the Smart H2O system has been implemented, using sensors and analytical platforms to monitor water resources and identify leaks. Thanks to this technology, it was possible to reduce water losses by 15 % and significantly reduce operating costs [9].

4 Conclusions

The study of forecasting water consumption in the city using Big Data and machine learning methods allowed us to obtain accurate and reliable forecasts for the next year. This is important for planning infrastructure investments and optimizing water management. As a result of the study, the main factors affecting water consumption in the city were identified. For example, in summer, the average daily water consumption of the capital is reduced by 20-30 % compared to winter. This is due to the change in the lifestyle of citizens, many of whom leave the city between June and August, as well as the seasonal shutdown of heating and hot water supply. Observations of the nature of hourly water consumption allow us to conclude that such factors as outdoor temperature, state and religious holidays, television programs and others influence it. To improve the efficiency of water supply system management in the city, it is necessary to actively seek and implement innovative solutions. This includes the integrated use of information technologies such as geoinformation systems, modeling of flow distribution in water supply networks and automation of control of water supply and distribution modes. Only this approach will optimize the operation of the water supply system, providing reliable and efficient service to the city. The analysis of the functioning of the water supply system under normal and extreme conditions provides valuable information necessary to identify urgent measures and set tasks for the near future: Ensuring economical modes of operation of equipment at a given pressure in pipelines - this includes optimizing the operation of pumping stations and regulating the pressure in the network to minimize energy consumption and ensure economical water consumption. Increasing the level of automation of the process of controlling the operating modes of the water supply and distribution system is important for rapid response to changes in the system, including automatic control of valves, pumps and other infrastructure elements in accordance with current needs and conditions. Solving these tasks will improve the efficiency of managing the water supply and distribution system in the city. This, in turn, will lead to an improvement in the quality of services provided to residents of the smart city and an increase in the level of comfort and safety of life in the city. In general, the findings emphasize not only the importance of using modern data analysis methods to predict water consumption in the city, but also the need to develop innovative water management strategies to ensure sustainable development of the city in the future. This means that in order to effectively manage water supply, it is necessary to use advanced data analysis methods such as machine learning and geoinformation systems, as well as actively implement innovative resource management technologies. Only such an approach will ensure the sustainable and efficient development of the city in the future, providing its residents with reliable access to water resources and environmental conservation. The introduction of big data and machine learning technologies into water resources management in Kazakhstan has huge potential. For the successful implementation of these technologies, it is recommended to start with pilot projects that will help assess their effectiveness in local conditions. It is also important to organize training and training programs for specialists, as well as attract investments to modernize the water supply infrastructure. The establishment of international cooperation with regions that have already implemented such technologies will allow the exchange of experience and knowledge. Finally, monitoring and evaluation systems need to be implemented to continuously improve and adapt solutions.

References

[1] Akhaeva Zh.B., Merzadinova G.T., Zhmud V.A., Zakirova A.B., Zhukabayeva T.K., *Akyldy kalanyn intellectualdy kolik zhuyesi modelindegi ulken derekterdi ondeu processi*, Scientific journal Bulletin of L.N.Gumilyov ENU, Astana, 2021, P .64-75.

[2] Astana Su Arnasy, *// https://www.astanasu.kz/about /. Astana, 02/10/2024.*

[3] Robles T., Alcarria R., Martin D. et al., *An iot based reference architecture for smart water management processes* , Journal's title, J. Wireless Mobile Networks Ubiquitous Comput. Dependable Appl. – 2015. – Vol. 6. – P. 4-23.

[4] Perumal T., Sulaiman M.N., Leong C.Y., *Internet of Things (IoT) enabled water monitoring system*, Procced. IEEE 4th Global Conference on Consumer Electronics (GCCE 2015). – Osaka, 2016. – P. 86-87.

[5] Malche T., Maheshwary P. , *Internet of Things (IoT) Based Water Level Monitoring System for Smart Village*, Proceed. of internat. conf. on Communication and Networks. – Singapore, 2017. – P. 305-312.

[6] Wang S., Zhang Z., Ye Z. et al., *Application of Environmental Internet of Things on water quality management of urban scenic river* , Int. J. Sustainable Develop. World Ecol. – 2013. – Vol. 20. – P. 216-222.

[7] Ryzhkov A.M. , *Compositions of algorithms based on a random forest* , http://www.machinelearning.ru/wiki/images/d/d8/

[8] *The magazine "Water supply and sanitary engineering"* , https://vstmag.ru/archives-all/2011/2011-1

[9] Verma P. et al. , *Towards an IoT based water management system for a campus* , Procced. 2015 IEEE 1st internat. Smart Cities conf. (ISC2). – Guadalajara, 2015. – P. 1-6.

Zh.B. Akhayeva
Astana International University,
8 Kabanbai batyra st., Astana, Kazakhstan
L.N. Gumilyov Eurasian National University
2 Satbaev st., Astana, Kazakhstan
E-mail: ahaeva07@mail.ru

A.B.Zakirova
L.N. Gumilyov Eurasian National University
2 Satbaev st., Astana, Kazakhstan
E-mail:alma_zakirova@mail.ru

V.A.Zhmud Novosibirsk State Technical University
20 Prospekt K. Marksa, Novosibirsk,630073, Russia
E-mail: zhmud@corp.nstu.ru

Received 05.03.2024, revised 04.06.2024, Accepted 05.06.2024