International Journal of Electrical and Computer Engineering (IJECE)
Vol. 14, No. 6, December 2024, pp. 6656~6674
ISSN: 2088-8708, DOI: 10.11591/ijece.v14i6.pp6656-6674

Development of a digital twin of a network of heating systems for smart cities on the example of the city of Almaty

Gulnur Tyulepberdinova[1], Murat Kunelbayev[1,2], Olga Shiryayeva[3], Meruyert Sakypbekova[1], Nurlan Sarsenbayev[3], Gulmira Bayandina[3]
[1]Department of Big Data and Artificial Intelligence, Faculty of Information Technology, Al-Farabi Kazakh National University, Almaty, Kazakhstan
[2]Institute of Information and Computing Technologies of the Republic of Kazakhstan, Amaty, Kazakhstan
[3]Department of Automation and Control, Institute of Automation and Information Technologies, Satbayev University, Almaty, Kazakhstan

Article Info

Article history:

Received Feb 12, 2024
Revised Aug 3, 2024
Accepted Aug 14, 2024

Keywords:

Digital twin
Heating system
K-neighbor regression
Linear regression
Machine learning
Random forest regression
Smart city

ABSTRACT

In this paper, a digital twin of the network of heating systems for smart cities is developed using the example of the city of Almaty. The study used machine learning algorithms to estimate future thermal energy consumption and develop thermodynamic formulas. This work offers a thorough and in-depth analysis of thermal energy consumption. In addition, the paper identifies the relationship between thermal energy consumption and ambient temperature, and wind uncertainty in certain urban areas using machine learning methods to predict thermal energy consumption. Using both training and regression models, this interdependence is revealed. The obtained forecasts provide useful information for studying the structure of heat consumption in Almaty and reducing heat losses by reducing overheating in the zones of heating networks. In addition, the study analyzes high-resolution spatial data collected from 385 homes and 62 heat transfer circuits located throughout the city during the heating season. The study examines the degree of relationship between the ambient temperature and the amount of heat energy used in the areas of Astana. A minor impact of wind speed is also estimated. These discoveries allow us to use machine learning algorithms to find the location of hot spots and inefficient zones with high losses.

Corresponding Author:

Gulnur Tyulepberdinova
Department of Big Data and Artificial Intelligence, Faculty of Information Technology, Al-Farabi Kazakh National University
Almaty, Kazakhstan
Email: Gulnur.Tiulepberdynova@kaznu.edu.kz

1. INTRODUCTION

In recent years, the idea of "smart cities" has gained considerable momentum, driven by the need for stable and efficient municipal development. Urbanization is projected to grow to 70% by 2050 [1], [2]. The rapid step in the transition to renewable energy and improving energy efficiency is proving to scientists to look for fresh approaches in several academic disciplines, starting with information technology. Significant problems related to the environment, management, and law can be solved by supporting digitalization. A fundamental and ecological challenge for humanity is global warming, which has several causes, starting with a strong shift in energy consumption. To solve this problem, it is boundless to significantly correct the attitude and behavior to the environment and the use of natural resources, especially energy. Several progressive climate protection strategies are currently being implemented in practice [3]. In [4], at the level of municipal planning, the establishment of conclusions is based on an innovative method of analyzing information about the use of heat. Ullah *et al.* [5] investigated the regulation of data analysis, the transfer of information and the effective execution of complex strategies are infinitely necessary to ensure the efficient and safe operation of a megalopolis. One of the main aspects of the concept of smart megacities is the optimization of energy systems, especially heating networks, which play a significant role in providing thermal comfort for residents and businesses. A promising conclusion for achieving rational planning of the formation of heating network data calls for the development of the field of digital twins in combination with advanced mathematical models and big data analysis.

The developers have considered the key characteristics of a solar collector with thermosiphon circulation for a solar heat supply system. A flat solar collector station with thermosiphon circulation was invented, in which the heat transfer coefficient is increased by lengthening additional fences between the panel and thermal insulation. The efficiency of the solar collector is achieved due to the presence of a tank and a heat pump in the installation, where the condenser and evaporator are made in the form of a spiral heat exchanger, and the combinations of heat exchanger pipes are located on top of each other, which allows to increase the area and intensity of heat exchange. The result of the study was the details of the mathematical analysis of the movement of the coolant in the storage tank. They also showed the dependence of the thermophysical parameters of the coolant on temperature [6]. Moreover, the use of such a methodology, as methods of mathematical modeling, and the use of big data analysis capabilities performs a number of functions, including real-time forecast: the digital twin constantly accumulates information from the material heating network. This allows the system operator to track the productivity of the system, uncover irregularities and immediately detect possible problems.

By looking at previous and ongoing data, the digital twin is able to perform punctual monitoring of future demand and heat usage patterns through the use of predictive analysis. This data allows you to pre-design and optimize the heating network, which guarantees the best direction and reduces the amount of energy that is wasted. By considering previous and flowing data, the digital twin is able to make punctual monitoring of future demand and heat usage patterns through the use of predictive analysis. This data allows you to pre-design and optimize the heating network, which guarantees the best direction and reduces the amount of energy that is wasted.

The digital twin uses all kinds of advanced optimization technologies to determine the appropriate configuration of the heating network. When choosing to expand, modify or replace parts of the infrastructure, the digital twin orients to perceive solutions, since it takes into account factors such as energy efficiency, economic efficiency and environmental impact. However, a study of impressive academic papers on this topic leads to the conclusion that difficulties with successful approaches to data, their collection and exchange of information create the greatest difficulties for the effective use of information technologies (ITs). Many technological limitations (usability, disclosure, and accessibility), as well as strategic, economic, legal and conceptual limitations are found [7]. Improving large data sizes can be a difficult topic and can lead to delays, which will lead to losses as firms wait for information to complete their heating projects. Forcing these tasks calls for a win-win data infrastructure, comparability stereotypes, and effective data management strategies. The importance of the topic is determined by the initiative "Development of energy saving policy" within the framework of the Strategic Formation Program of Astana 2050, approved in 2019 [8]. With the help of a complex of digital predictive analytics, it is planned to implement regulation in the field of energy saving in practice. Technical networks will be systematically tested for costs and unauthorized use. It is expected that as a result of the implementation of this program, energy-efficient homes will reduce utility costs, lower household utility bills and improve the well-being of low-income populations.

During the colder months, the average internal temperature will rise. It is possible to reduce the basic heating requirement of the tested premises to a quarter of the available value by modernizing systems using the principles of inert design. The insulation of external walls should always be used in combination with the concept of natural ventilation and shading of openings. In [9], effective management of energy flows to buildings and other structures requires the development and implementation in practice of engineering methodologies for forecasting and evaluating energy consumption processes. The actual basis of the study was the use of machine learning methods to predict the consumption of thermal energy in urban areas and its relationship to variables associated with weather and wind. Three different regression models: linear regression, K-neighbor regression, and random forest regression, were used to establish correlation and compare learning and testing outcomes. Advanced information technologies, intelligent decision-making assistance systems and energy management of homes are important for analyzing a decent amount of information in real time. A detailed and objective examination of how much heat and energy buildings and structures use, taking into account weather and climatic elements, is possible thanks to an automated power supply management system. In [10], [11], systematic comparisons of heat consumption information with weather data and other information, such as the number of people present, and the state of isolation of pipelines, are part of the process of forecasting heat consumption of a certain object. The present method evaluates the effectiveness of heat supply systems to identify possible savings in thermal energy costs.

2011 was the year of the formation of an intelligent energy system. Due to the synergy between energy storage and available heat sources, this provision plays a necessary role in systems that absolutely operate on renewable energy sources. In [12], a variety of possible scenarios were investigated, with particular emphasis on the use of renewable energy sources in transport, taking into account the small access to bioenergy. Modeling of energy systems allows a formalized description of systems that can be used in the analysis of energy and environmental policy objectives. Omirgaliyev *et al.* [13] conducted a study using forecasting of energy systems. When modeling electric power systems, it is important to understand the importance of collecting and analyzing all publicly available information before including them in the model. Thanks to the digital twin concept, users can explore data, monitor systems, and run simulations as if they were dealing with a real asset. This approach was first developed in the manufacturing industry; however, it has subsequently found application in many other areas. This allows virtual testing of new assets, such as aircraft engines, which eliminates the possibility that a real asset will fail during testing. In addition, digital twins are used in testing new tool designs or locations, assessing the condition of organizations, providing digital twin identifiers for public access, and promoting more efficient municipal governance. In the context of a smart city, a digital twin is a system that continuously collects data from an artificial environment using sensors, drones, mobile devices and other technologies to create a real—time representation of the city [14]. Data is collected from a wide variety of sources, including vehicles, structures and infrastructure, and even individuals, before being transmitted to the digital counterpart of the urban area. The digital twin's ability to process and synthesize static, historical and real-time data is improved by integrating data with the internet of things (IoT) device, smart city technologies, as well as the installation of artificial intelligence (AI) and sophisticated analytics. As a result, it provides invaluable information about the city's work, which turns the digital twin into a "strategy accelerator" for public sector organizations, allowing them to recognize connections, develop more effective solutions and improve their solutions manufacturing procedures [15].

Due to the rapid population growth and increasingly complex urban problems faced by cities around the world, municipal administrations in Kazakhstan, in particular, are under increasing pressure to properly allocate urban resources and manage urban areas. Real-time digital data is increasingly recognized as a vital resource for improving urban planning, evaluating the effectiveness of service delivery and optimizing decision-making processes. Even before physical implementation, digital twins allow you to test concepts and simulate "what if" scenarios for objects, processes and urban landscapes. Such modeling provides insight into the potential outcomes and consequences of urban development. The use of digital twins opens up new opportunities for optimizing planning, operations, financing and decision-making related to large-scale construction projects and assets. This, in turn, leads to significant cost savings, increased productivity and reduced carbon dioxide emissions in the city as a whole. Optimization of urban planning, as well as the operation and maintenance of tangible assets, can be achieved by creating a digital twin of a building, district, or even the entire city from the point of view of municipal administration. The architectural environment can be made more sustainable and environmentally friendly by modeling, for example, potential hazards such as high temperatures or dust storms. In addition, preventive maintenance of physical assets or urban networks can be carried out to reduce the number of outages caused by operational problems. The concept of a digital analogue opens up a number of opportunities to improve the planning and management of smart cities. Using real-time data, more advanced analytics and artificial intelligence, digital twins enable public sector organizations to make better informed decisions, improve resource allocation efficiency, and enhance fault tolerance and sustainability of the urban environment.

2. METHOD

2.1. Linear regression

A method used in statistics, known as linear regression, can be used to describe the relationship between a long-running decision and one or more items, which are referred to as the dependent variable and the independent variable, respectively. It can be used for an infinitely large set of response information. Linear modeling is used to describe relationships. In situations where there is only one explanatory variable, the statistical method is called simple linear regression. On the other hand, when there are many explanatory variables, this method is called multiple linear regression. There may be confusion between multivariate linear regression and multiple linear regression that predicts a single scalar-dependent variable. However, multivariate linear regression involves many interrelated dependent variables. To accurately use linear regression, you need to create a linear dependency function. It is possible to restore the interdependence of the variable x from one or some other variables by using this factor function. To properly implement a linear regression model, you need a linear dependency function. It is possible to create a relationship between the variable x and one or several other variables (factors, regressors, and independent variables) with the support of this function, using the function [16], [17]. You need to set a dependency to extract a preconditioned set of m pairs of values for a free variable and a dependent variable, from $i = 1$ to m. This is because values can only be used when the relationship is initially defined. The equation was presented to a linear model using an additive random variable as a representation.

$$y_i = f(w, x_i) + \varepsilon_i$$

It is assumed that the random variable obeys the usual distribution with a long-term variance of 2, which does not depend on the variables x and y. The method of minimizing squared errors is used to find the values of the parameters w of the regression model using these assumptions.

The definition provided by the dependency model is:

$$y_i = w_1 + w_2 x_i + \varepsilon_i$$

The least squares method specifies that the required parameter vector is written as $w = (w_1, w_{w1}, w_2)$ to be valid. T is the solution of a normal equation, which looks like this:

$$w = (A^T * A)^{-1} A^T y$$

where y is a vector consisting of the values of the dependent variable y, which are equal to $(y1y1, \dots, ymym)$. The values of the free variables xxi, $x00$, $aiai2$ and $aiai2$ for $i\ i = 1, \dots, m$ represent substitutions that can be found in the columns of matrix $A.A$.

Using elementary linear regression models, the relationship between two variables is analyzed. Based on X, the model predicts Y in this scenario. In Figure 1 data points are represented by points. Any data point represents the values of X and Y. The linear regression model compares these data points to a straight line. The regression line on the graph is the maximum fit line. Minimize the distance between the line and the data point. This distance is calculated using the least squares method. The equation for the regression line is $Yi = \beta_0 + \beta_1\ X_1 + \varepsilon_1$ where β_0 is the y-intercept (Y when X is zero) and β_1 is the slope (the change in Y for every one-unit increase in X). Regression models estimate these coefficients in order to best suit the data. The regression line is capable of foreshadowing future values of X. The ability to predict the appropriate Y value by substituting a certain X value into the equation. The forecast of the model is an impression of the Y regression line. The elementary linear regression model requires that the independent and dependent variables have a linear relationship. This simple model may not necessarily represent nonlinear time-consuming interactions. There may be a need for innovative regression methodologies. Experts prefer the interpretability of linear regression. The simplicity and certainty of the linear model orient specialists to understand the relationships of variables. The regression coefficients of the model illustrate the effect of any variable on the result. Regression coefficients are directed to correct the dependent variable when the independent variable changes by one unit, provided that all remaining variables remain constant. Specialists are able to numerically assess the impact of certain factors. The signs and values of the coefficients are informative in interpretation. Significant predictors appear using linear regression. Using hypothesis testing or the significance of coefficients, experts can detect statistically important variables. This information focuses on identifying predominantly significant components of the system and understanding its dynamics.

Linear regression allows you to extrapolate solutions and make a direct impact on independent variables. The circumstances of the interaction in the model allow specialists to analyze the interaction variables. This allows you to check whether the relationships of variables change under all possible conditions. These ideas can make it easier to establish solutions and understand complex interactions. The validation of the linear regression model is transparent and interpretable. The R-square model and the adjusted R-square measure are suitable. These indicators demonstrate how much the variance explains the model, and how well independent conditions reproduce the instability of the dependent variable. The accessibility and interpretability of linear regression orient specialists to understand the correlations of variables and their impact. Evidence makes it easier to establish reasoned solutions, test hypotheses, and provide meaningful information about regressions [18]. The advantage of linear regression models is that they are applicable on a large scale and provide the necessary confidence to characterize a wide range of true processes in the business sector and the economy. They are a universal tool that can be used in various industries, recording and approximating difficult phenomena with high accuracy. The interpretability of the results makes it possible to obtain reasoned decisions, and the efficiency of calculations simplifies the study of large data sets. The established methodology and the linear regression function as a reference model guarantee additional support for its applicability and punctuality on a larger scale [19]. Common problems such as multicollinearity, heteroscedasticity, outliers, influential observations, and model classification have well-known solutions for linear regression. Scientists and practitioners have created methods to eliminate these obstacles and guarantee the punctuality of models. Static significance assessment operations for models, such as hypothesis testing and statistical measurements, have been developed and widely implemented. Using these conclusions and adhering to the established procedures, specialists can create reliable linear regression models and make perfect conclusions based on the results [20].

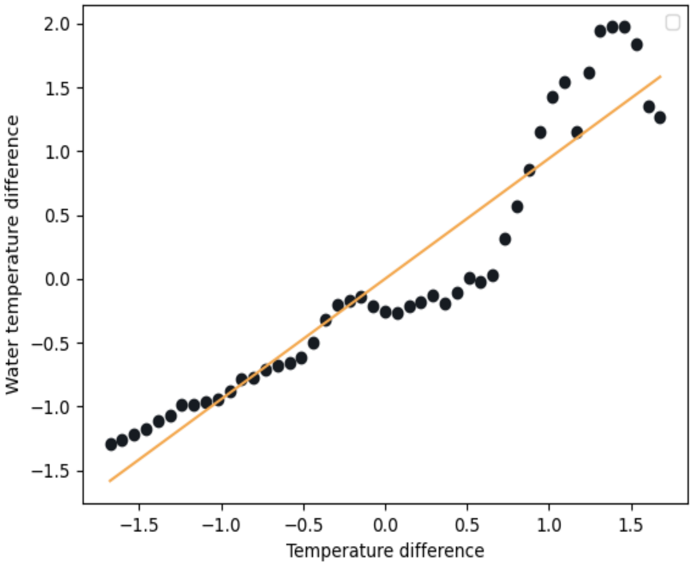


Figure 1. A simple linear regression model to demonstrate the relationship between two

Figure 2 shows a linear regression scheme that examines in more detail the relationship between heat consumption in Almaty and an independent variable such as ambient temperature. By plotting data points on a scattering diagram to compare and contrast the results, you can mentally study how heat consumption rates vary across the temperature spectrum. Any data point on the scattering diagram represents a unique observation: the X coordinate represents the temperature of the outside air, and the Y coordinate represents the amount of heat used. Scrupulously considering the direction of the data points, identifying all possible universal trends or patterns that may exist. A regression line is a line that best fits the data points in order to minimize the gap between the line and the observed data. This line reflects the calculated inconsistency between the amount of heat consumed and the ambient temperature. The slope of the line indicates the rate of change in heat consumption with each increase in temperature by one degree. For example, when the slope is positive, it means that as the outdoor temperature increases, there is a tendency to increase the amount of heat used. On the other hand, a critical slope will indicate that the amount of heat consumed decreases with increasing temperature. The value represented by the intersection point of the regression line represents the estimated amount of heat needed if the outdoor temperature is zero. It is very important to be extremely careful when interpreting an intercepted message. In the vast majority of cases, setting or maintaining zero temperature is impossible or illogical. Therefore, it is quite possible that the interception itself is not of the greatest importance. In order to understand a more meaningful interpretation, it is essential to consider a certain connection and research in the present field. One of the most significant steps in finding whether a linear regression model is suitable for data is to perform a residue analysis. The residuals are attributed to the vertical distances between the observed data components and the regression line. They represent the difference between the true values of heat consumption and the values predicted using the regression line. By examining the distribution of residues, which is a statistical indicator, by revealing different trends or systematic deviations. In an ideal world, the residuals would be seamlessly distributed around zero, indicating that the linear regression model has successfully captured key data patterns. In order to draw the necessary conclusions about the relationship between heat consumption in Almaty and the independent variable of interest, it is infinitely important to consider the quality and specificity of the data, as well as knowledge of the subject area and other relevant factors. Only then will it be possible to draw the necessary conclusions about the nature of the relationship.

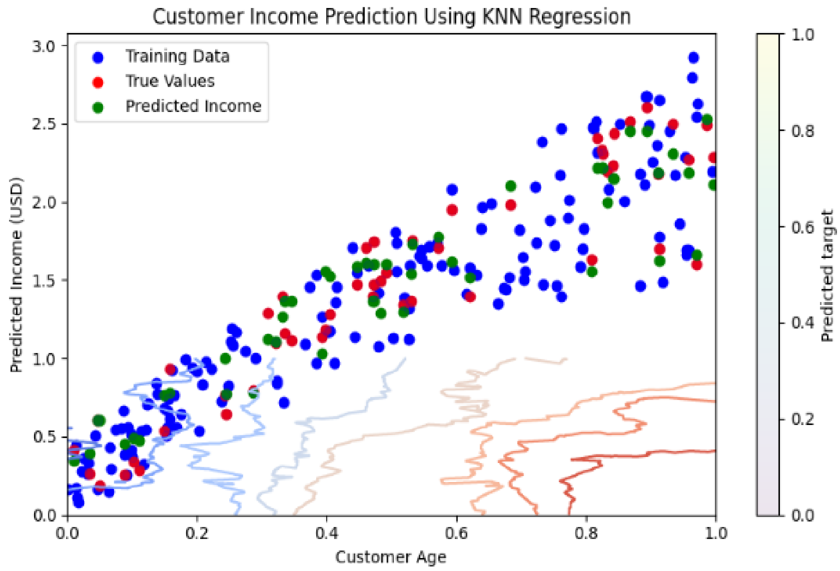


Figure 2. Analysis of a linear regression model based on a data set on heat consumption in Almaty

2.2. K-neighbor regression

The k-nearest neighbor algorithm, commonly known as KNN, is a nonparametric path to supervised learning that was first presented in the field of statistics in 1951. Problems associated with classification and regression are good candidates for using this method. Figure 3 shows an example of comparing KNN regression and linear regression with varying values of k. The output data obtained by the KNN classification method is suitable for the preconditioned membership of an object in its assigned class. The classification process is based on the results of numerous voting conducted by the close environment of the facility. To be clearer, an element is assigned to a class that dominates among the classes of its k-nearest neighbors, where k is a positive integer that is often chosen so that its value is at the lower end of the spectrum. In a specific situation, when k is equal to 1, an element is assigned a class label exclusively of such a unique neighbor to which it has the greatest real proximity. As a consequence, the property of an element to a conditioned class is determined solely by the class of its shortest neighbor. This method operates on the assumption that the information provided by the shortest neighbor of an object is information that is both predominantly win-win and reproduces the object [21], [22]. When features have different physical units or contain significantly contrasting scales, the algorithm should take into account the normalization of the training data to achieve optimal likely results. This normalization method focuses on standardizing data ranges in order to reduce other disproportionate effects that certain characteristics may detect on distance calculations. The k-nearest neighbor method can provide more punctual classification by standardizing data and assigning similar weights to objects that have different scales or units of measurement.

Figure 4 shows the effectiveness of KNN regression models with different values of k, as well as the corresponding learning outcomes. The k-nearest neighbor approach preferably handles a wide range of feature characteristics because it normalizes the data to fit it to a uniform scale. This step helps to reduce any possible regular error that could occur due to the fact that objects have radically different scales. The immediate consequence of this is an impressive increase in the accuracy of the algorithm, which ultimately leads to more reliable and stable classifications. The KNN method differs from other approaches because it attaches more importance to local structure and relationships than other methods that may exist based on global patterns or assumptions. This allows the algorithm to be multi-purpose and adaptable to a variety of data varieties, as it is able to intercept the subtleties and differences present at the local level. KNN regression offers several advantages, starting with reliability to outliers, interpretability of totals and ease of implementation. Although it also has inherent limitations, such as the lack of generalization of the model, high computational complexity and computational costs. The software implementation of the KNN regressor stands out for its inherent simplicity and intuitiveness. The simplicity of the algorithm dramatically simplifies the implementation process and reduces the possibility of coding errors. Using the fundamental strategy of searching for neighboring neighbors and averaging their values, the implementation becomes more accessible, which makes it easier for programmers to convert the algorithm into reproducible code. This simplicity also facilitates efficient debugging and maintenance, since the logic of the KNN regressor is transparent and understandable. In addition, the simplicity of the KNN regressor implementation contributes to the secondary application and expansion of the code. Due to a thorough understanding of the algorithm concept, developers are free to adapt and customize its implementation to meet certain needs or integrate it into larger software systems. This adaptability guarantees smooth integration of the KNN regressor into a variety of applications covering many difficult places and areas [23].

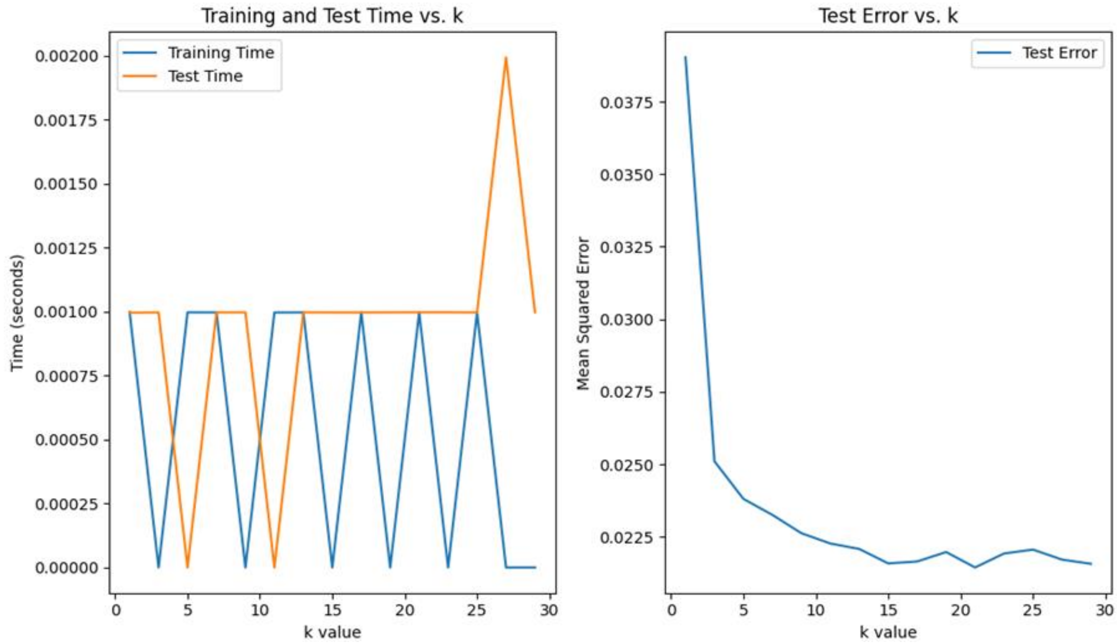


Figure 3. An example comparing KNN regression and linear regression with varying k values illustrate the flexible, data-adaptive nature of KNN regression as opposed to fixed-line linear regression

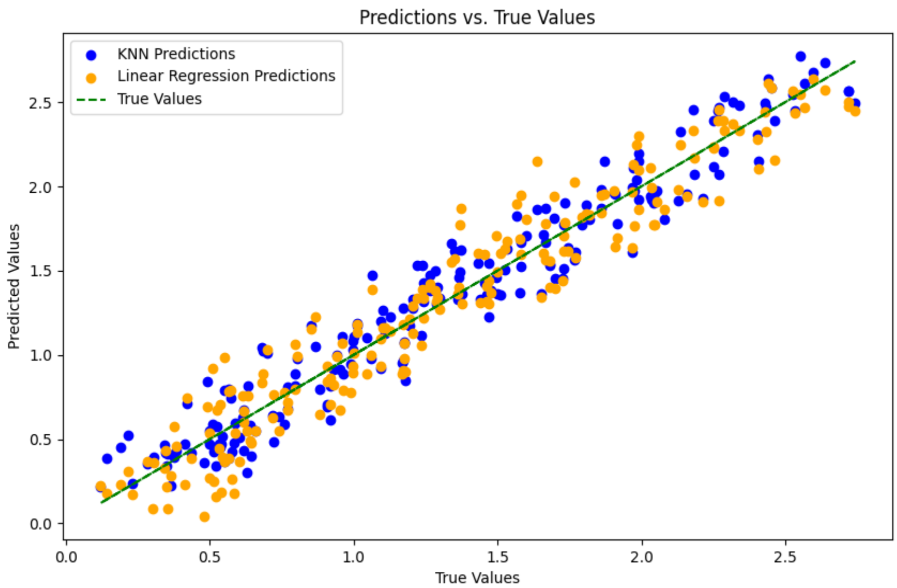


Figure 4. Efficiency of KNN regression models with different k values, as well as their corresponding training results

The interpretability of the results of the KNN regressor seems to be an important advantage, predetermined by its simplicity and openness. The algorithm makes predictions, the logic of which is simply clear and explicable, which simplifies effective communication between experts on the provided data and specialists in the subject area. The KNN regressor improves the interpretability of results, which is essential for decision-making processes by presenting monitoring in a simple and clear form. The interpretability of the results of the KNN regressor has actual significance in areas where it is infinitely important to understand the key factors acting on predictions. For example, in the field of healthcare, the ability to explain and defend forecasts is crucial for gaining trust and recognition from medical professionals. By allowing domain specialists to comprehend the logic of the KNN regressor predictions, the algorithm becomes a significant corporate analysis tool that promotes synergy between data scientists and domain experts. In addition, the interpretability of the KNN regressor orients testing the results of the algorithm. The predictability and quality of the results are supported by the predictability and clarity of the various possible deviations, errors or unusual patterns that may be present. This aspect of interpretability contributes to the joint veracity of the consequences of the KNN regressor, contributing to transparency and accountability in the process of making conclusions [24]. The KNN regressor shows remarkable reliability when it comes to handling messy and incorrect data points. Using the theory of proximity between neighboring instances, the algorithm detects a reduced location for individual outliers, which makes it infinitely constant to atypical observations. Unlike other regression methods, which may be unduly influenced by separated extreme values, the KNN regressor provides for joint characteristics of neighboring points, thereby providing clear forecasts that are less prone to distortion due to individual outliers. This exceptional resistance to anomalies and incorrect values is a consequence of the fundamental principle of the KNN regressor, which describes the forecast for a given data point by calculating a mediocre value or a weighted average of its close neighbors. By counting on a local neighborhood rather than a single data point, the algorithm reveals a remarkable ability to neglect the effects of outliers or unnatural values that may have thoroughly deviated from the overall data structure. Such reliability allows the KNN regressor to protect its prediction reliability and stability in the presence of erratic or incorrect observations. The reliability of the KNN regressor to outliers and incorrect values has impressive actual consequences in various fields. In areas where data quality and integrity are of paramount importance, such as financial analysis or anomaly detection, the KNN regressor's ability to manage low-profile data points without degrading its predictive data is of great importance. Such fault tolerance ensures that the algorithm can provide punctual information and forecasts even in situations where the presence of outliers or deviating values would otherwise make other regression methods incorrect [25].

Figure 5 shows a comparison of energy consumption at temperature in January and April 2023 with actual energy consumption (points) and projected values (line). At the classification stage, when a fresh object is presented that does not have a class attribute, the algorithm wants to assign it a class label based on publicly available training data. To achieve this goal, the algorithm measures the correspondence or similarity between the new object and previously systematized observations using a suitable metric. The metric can be, for example, the Euclidean distance for continuous variables, or an overlap metric such as the Hamming distance for discrete variables. The next step involves identifying the k-nearest neighbors to the new object from the training data based on their proximity. These neighbors are represented by examples in the training set that have mostly identical characteristics with the new object. The choice k describes the number of neighbors considered, and it is usually a positive integer, often small. The development of a real algorithmic process further increases its clarity and clarity. Due to the clear identification of the steps involved in the training and classification stages, a long-term basis is formed for understanding the inner workings of the KNN algorithm. Such a representation is useful for a variety of fields and facilitates its implementation and interpretation by specialists in the provided field. The relative weight given to various da ta by various classification orders is, what distinguishes them from each other when finding the category to which the element relates. Perhaps the reliability of the classification could be improved by understanding the prestige of the algorithm's data. Taking into account the significance of the features used in the algorithmic process, it is possible to achieve improved classification results. In traditional approaches to categorization, all properties are viewed uniformly, assuming that they all contribute to the classification of an object at the same time. However, it is possible that this assumption will not be clear in the real world. The rest may be less relevant or may not be related to a specific task.

When considering the meaning of characteristics, it becomes obvious that different characteristics show a diverse degree of impact on the course of classification. Because of this understanding appropriate weights or significance can be assigned to different conditions, allowing more attention to be paid to characteristics relevant to classification. As a result, the algorithm becomes more efficient in recognizing the underlying patterns and nuances, which leads to an increase in the accuracy of categorization. To include the significance of features in the classification process, you need methods for finding the significance of features. Countless methods, including statistical estimates, algorithmic feature analysis, and domain expertise, are viable alternatives to achieve this goal. The purpose of these methods is to distribute mostly informative and important characteristics for the categorization process. By highlighting the authority of attributes, you can discover previously unknown relationships and get a better view of the data. This information allows us to organize classification models that appear to be better and use mostly characteristic features, which leads to improved correctness and performance.

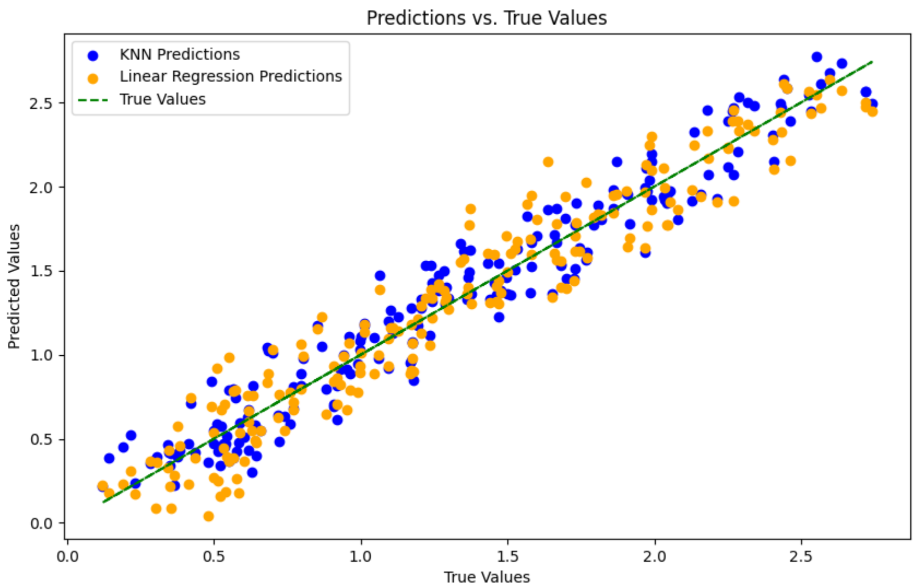


Figure 5. Comparison of the dependence of energy consumption at temperatures in January and April 2023 with actual energy consumption (dots) and projected values (line)

2.3. Random forest regression

Random forest regression is a powerful general machine learning algorithm that performs regression tasks by combining the principles of decision trees and collective learning. It is widely used in many fields, including finance, healthcare, and environmental sciences, where punctual assumption of continuous numerical values is important [26], [27]. It stands out among other systematization algorithms due to its versatility. The ability to compute classification, regression, clustering, anomaly detection, and feature selection tasks contributes to adaptability. In the regression of a random forest, the compilation of decision trees is based. Each tree is trained on a random subset of the training data, and a random subset of features is considered for partitioning at each node. This randomness increases the set among trees, thereby reducing overfitting and improving generalization efficiency. Aggregate estimates are obtained by combining forecasts of individual trees. This aggregation is usually performed by averaging the predictions for regression assignments. The final estimate of a random forest is a boundless estimate that is based on the collective knowledge of many trees.

Figure 6 illustrates the regression of a random forest. Each tree should be constructed using the following method: first, it should take a sub-sample with a sample size from the training sample and build trees with separate sub-samples based on this; second, the largest feature from random objects is considered for modeling each partition; and third, the clearest partition and a feature can be selected based on a previously specified criterion. Items are organized by voting: each member of the committee assigns an item to one of the classes. In addition, the class that received the maximum number of votes is a better-known option. The average number of trees is selected in such a way as to minimize the classifier's error in the test sample. If it is omitted, the estimation error for samples that are not included in the set is minimized. Two factors are needed for more accurate forecasts. First, there are significant explanatory features. Second, forecasts of trees and the forest as a whole should not be correlated.

Compared to other regression methods, random forest regression is less susceptible to outliers and noisy data. Averaging forecasts over multiple trees serves to reduce the impact of outliers and smooth out separate errors [28]. Random forest regression can model non-linear relationships between input features and the target variable. Even in the presence of nonlinear interactions, it can manage complex data patterns and make accurate predictions [28]. Random forest regression guarantees a measure of the feature value. By evaluating the impact of each feature on model performance, you can determine the most significant properties for assigning a forecast. This information can help you choose a feature and interpret the results [29]. Random forest regression is effective when working with multidimensional data sets where the number of features is greater than the number of samples. It can manage both continuous and categorical characteristics without extensive data preprocessing [30]. The ensemble nature of random forest regression reduces the likelihood of overfitting. The randomness inherent in tree construction and feature selection makes the model more win-win and generalizable by decorrelating forecasts [31]. Random forest regression includes hyperparameters such as the number of trees in the ensemble, the maximum depth of any tree, and the abundance of features considered for partitioning. It is absolutely necessary to set up full-fledged hyperparameters to optimize the performance of the model. A suitable combination can be found using methods such as grid exploration and independent search. Random forest regression uses out-of-bag (OOB) samples that are not used during training to evaluate the performance of the model. The OOB error seems to be a useful feature for evaluating the rigor of a model without a single set of checks. Due to the ensemble nature of the random forest regression model, it may not be easy to interpret the outstanding predictive characteristics of the model. Random forest regression allows productive management of large data sets. However, as the number of features or examples increases, training time and memory requirements may increase. In such circumstances, consider dimensionality reduction methods or distributed computing. Presenting the data of genuine judgments and effectively adapting the hyperparameters of the model can help in maximizing the correctness and reliability of random forest regression for regression tasks.

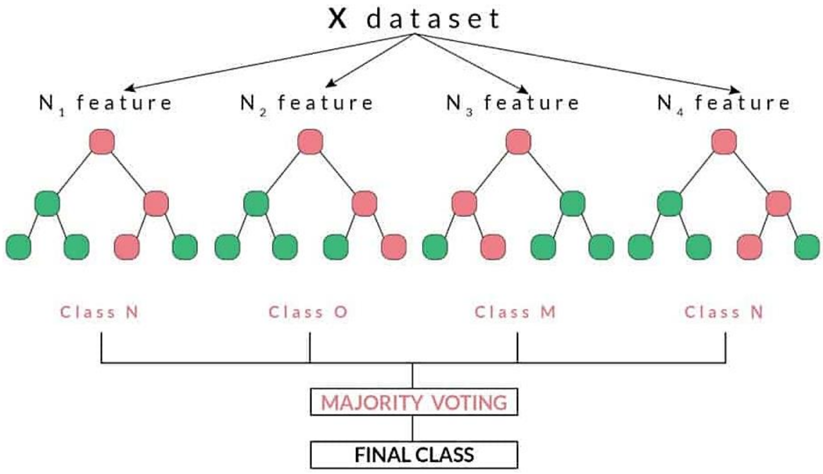


Figure 6. Random forest regression

3. RESULTS AND DISCUSION

3.1. Analysis of digital heating network creation

The study of the growth of digital heating networks that has just been presented includes data collection, mathematical modeling, big data analysis, optimization and planning, integration of the internet of things and management systems, visualization, maintenance and scalability. These are all the nuances of forming digital heating networks. In the context of smart cities, organized information can be analyzed to create a punctual and usable digital twin that allows heating networks to operate in an efficient and environmentally friendly manner.

3.1.1. Correlation analysis

The selection covers 8,754 individual lines of information. When a node is not used, its minimum values are always zero. When the sensors themselves fail or when environmental conditions are undesirable (for example, when it is very cold or very hot), the sensors display values of 0. Due to the fact that there are separate contours, the average value for each of them is different. However, in the current data set, there are rows that have the value 0, and there are no instances of null values. As a consequence, the correct conclusion is to replace them with null values. Figure 7 shows additional data about this data set.

Figure 8 shows how the enthalpy changes over five months in response to fluctuations in ambient temperature. Two different polygons are very clearly represented on the graph, and they are marked with the names kontur-1 and kontur-2. Contour-1, as its name suggests, is a contour located in the immediate vicinity of a thermal power plant (TPP), while contour-2 is located further away. Looking at the graph, it is easy to see how the external temperature affects the enthalpy levels, as well as how this influence is modified on countless objects located at different distances from the TPP.

	Contour-1 (T1)	Contour-1 (T2)	Contour-1 (P1)	Contour-1 (P2)	Contour-62 (T1)	Contour-62 (T2)	Contour-62 (P1)	Contour-62 (P2)
count	8754.00	8754.00	8754.00	8754.00	8754.00	8754.00	8754.00	8754.00
mean	71.88	43.32	7.54	4.37	72.80	45.67	7.34	4.83
std	15.63	10.49	1.79	0.84	16.12	10.48	1.41	0.82
min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25%	66.03	38.41	7.32	4.39	66.89	44.42	7.37	4.90
50%	68.15	41.06	8.01	4.51	69.84	46.76	7.66	4.97
75%	82.43	52.98	8.60	4.71	84.44	50.18	8.03	5.04
max	106.19	59.03	9.64	6.26	111.40	75.11	8.84	6.20

Figure 7. Descriptive statistics on 4 parameters for each of the 62 heating circuits

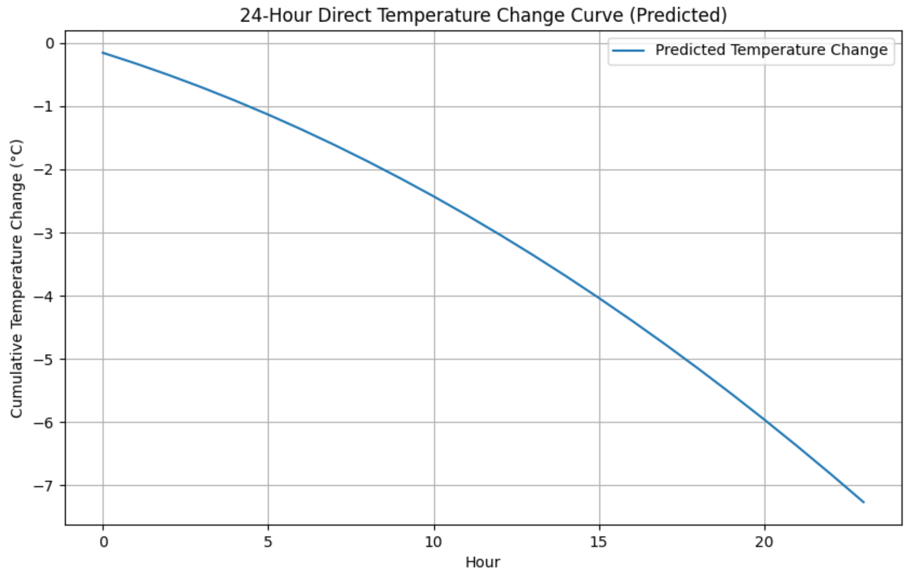


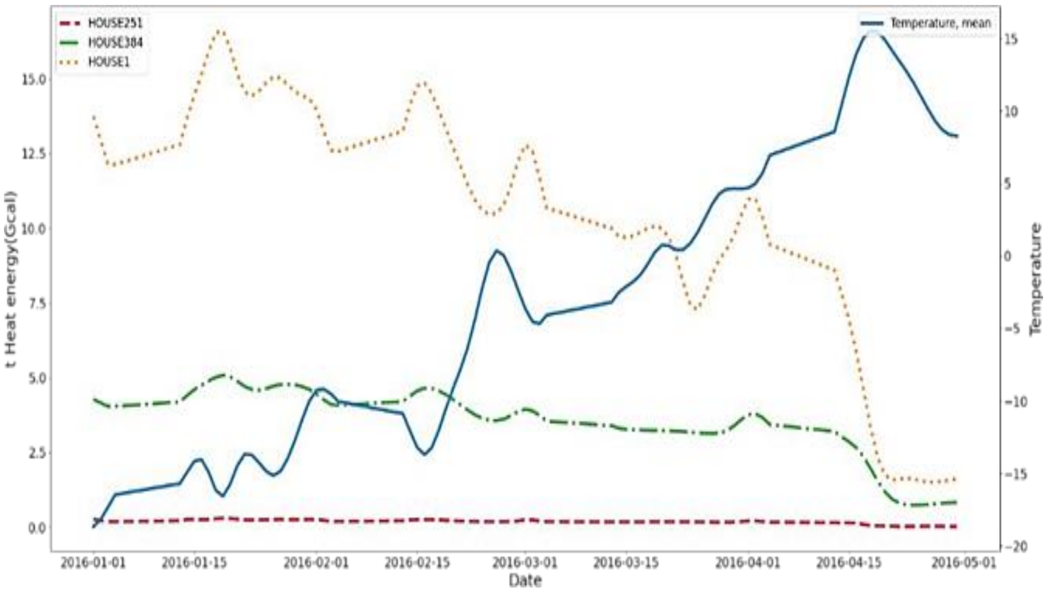
Figure 8. 24-hour direct temperature change curve

When the same process is applied to houses, the result is the graph shown in Figure 9(a), which shows the unequal distances separating the houses. It was found that dom-1 and kontur-1 are mostly located far from each other, but dom-251 and dom-384 are, respectively, a close house and a house placed in the middle. It is noteworthy to note that while in Figure 9(a) the heat consumption of House 251 is visibly zero, in Figure 9(b) it is represented as a multiplied correlation, which allows us to study the consumption structure more seriously. This is something to keep in mind. As a result, this data provides meaningful information about the structure of consumers ' energy consumption, which, in turn, can be used to develop plans for better energy savings.

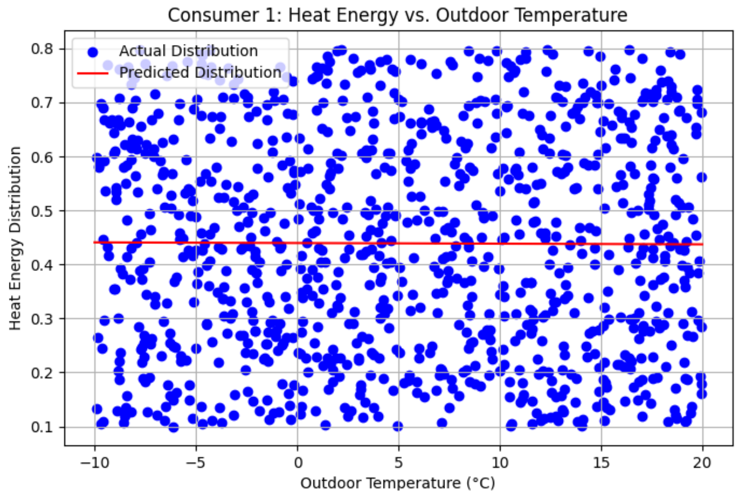
Figure 9(a) show the distribution of thermal energy by day, Figure 9(b) show the distribution of outdoor air temperature. By tracking the changes that occur in the enthalpy readings, you can see that an increase in temperature initiates a comparable drop in enthalpy-1. These shifts in enthalpy values are most easily detected in February and March, when there is an impressive difference in the average ambient temperature. It is important to keep in mind that the ambient temperature tends to increase with decreasing enthalpy values and decrease with increasing enthalpy values. It is important to have an understanding of the relationship between the signs of temperature and enthalpy in order to be able to understand the patterns of heat consumption in homes and develop the best energy saving methods.

$$q = (M_{cw}(t_{out} - t_{in}))/Mhw \tag{4}$$

where q is the specific heat capacity of water, 4.19 kJ/kg; M_{cw} is the mass of indirect water, kg; t_{out} is the water temperature at a contour's output, °C; t_{in} is the water temperature at a contour's input, °C; and Mhw is the mass of condensate, kg.



(a)



(b)

Figure 9. Distribution of (a) thermal energy by day and (b) outdoor air temperature

Figure 10 shows a graph that illustrates the relationship between hourly average temperature and time of day in a simple and explicit way, showing how temperature patterns change throughout the day. It is observed that the rate of temperature increase reaches its maximum between 5 am and 10 am and that it reaches its highest point between 10 pm and 15 pm. It starts at a high level, but gradually falls over the course of the day, until it reaches its lowest point. In addition, it is interesting to see that the temperature at the door and at the exit increases simultaneously with the air temperature; this demonstrates that there is a significant relationship between these two variables. In addition, Figure 10 shows a complete study of the location of residential buildings in Astana in accordance with the data set, which makes a solid contribution to our knowledge of regional energy consumption models.

Figure 11(a) shows a map of the city of Almaty and Figure 11(b) a map of the city of Astana of the Republic of Kazakhstan. Energy consumption models are monitored using enthalpy values that relate to thermal energy. "Thermal energy in Gcal" is the target value of the monitored variable. This information can be determined in Figure 10, which shows the distribution of the values of the highest and lowest temperatures for the day. In addition, the ambient temperature in Astana seems to be a necessary component in the analysis of energy consumption trends. Figure 12 also shows wind speed, which is another component that affects the energy consumption model. These variables can be studied in order to learn more about energy consumption patterns and develop an effective energy-saving strategy.

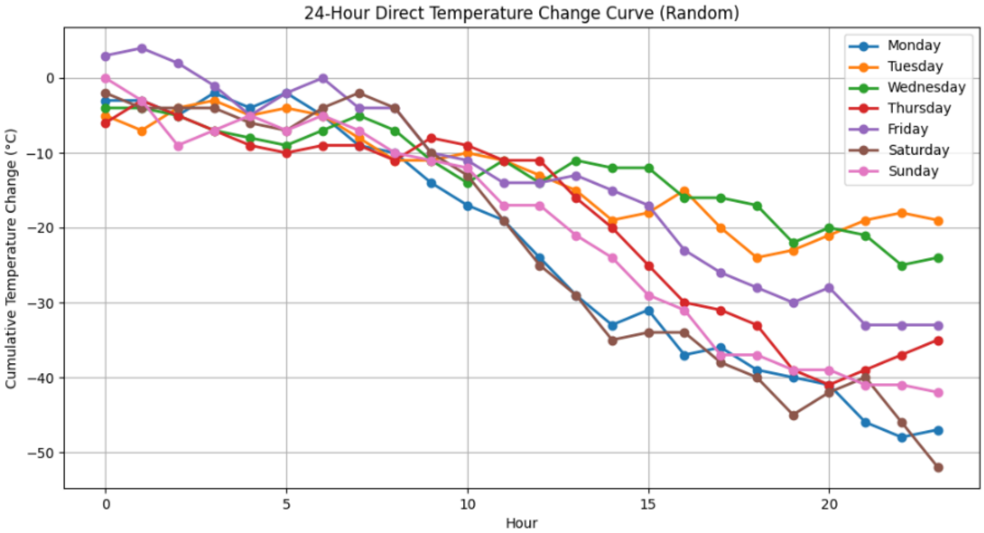
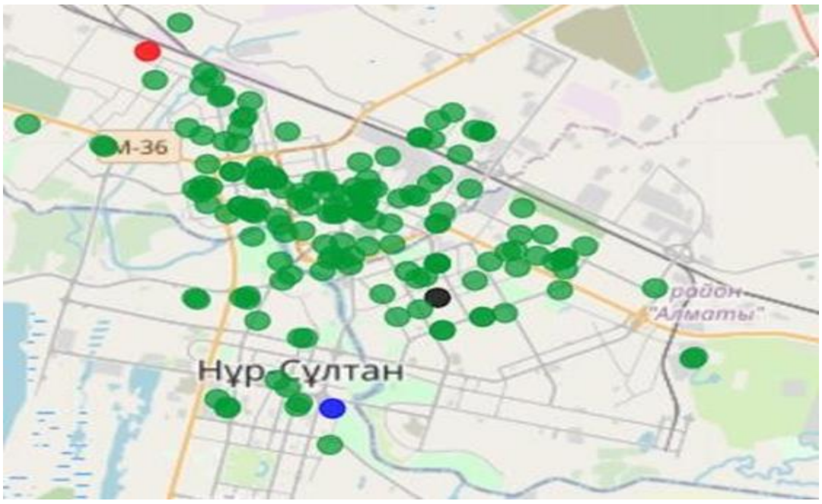


Figure 10. A 24-hour curve of direct temperature change

(a)



(b)

Figure 11. Map of (a) Almaty and (b) Astana (the position of 385 metering points on a map of the city of Astana), Republic of Kazakhstan

Figure 12 shows that the wind speed of 5 meters per second was measured and recorded four times. This pattern has attracted the attention of both meteorologists and environmentalists. This repeated pattern may indicate that there are some synoptic conditions that support a continuous wind speed of 5 meters per second. The significance of this pattern, as well as its likely connections with a number of other components, deserve further study. Figure 13 illustrates the association between these components, representing the coordination between wind speed and other variables that may possibly affect it. When we examine these relationships between variables in more detail, we may notice that they have certain types of positive correlation. When we want to create effective approaches to energy conservation that can help optimize energy use and reduce the human carbon footprint, we must thoroughly understand these patterns and relationships. The global community is increasingly aware of the urgent need to reduce the impact of humanity on the environment and find more environmentally decisive opportunities for life. Knowing the many factors that can contribute to a continuous wind speed of 5 meters per second has significant implications for answering this problem.

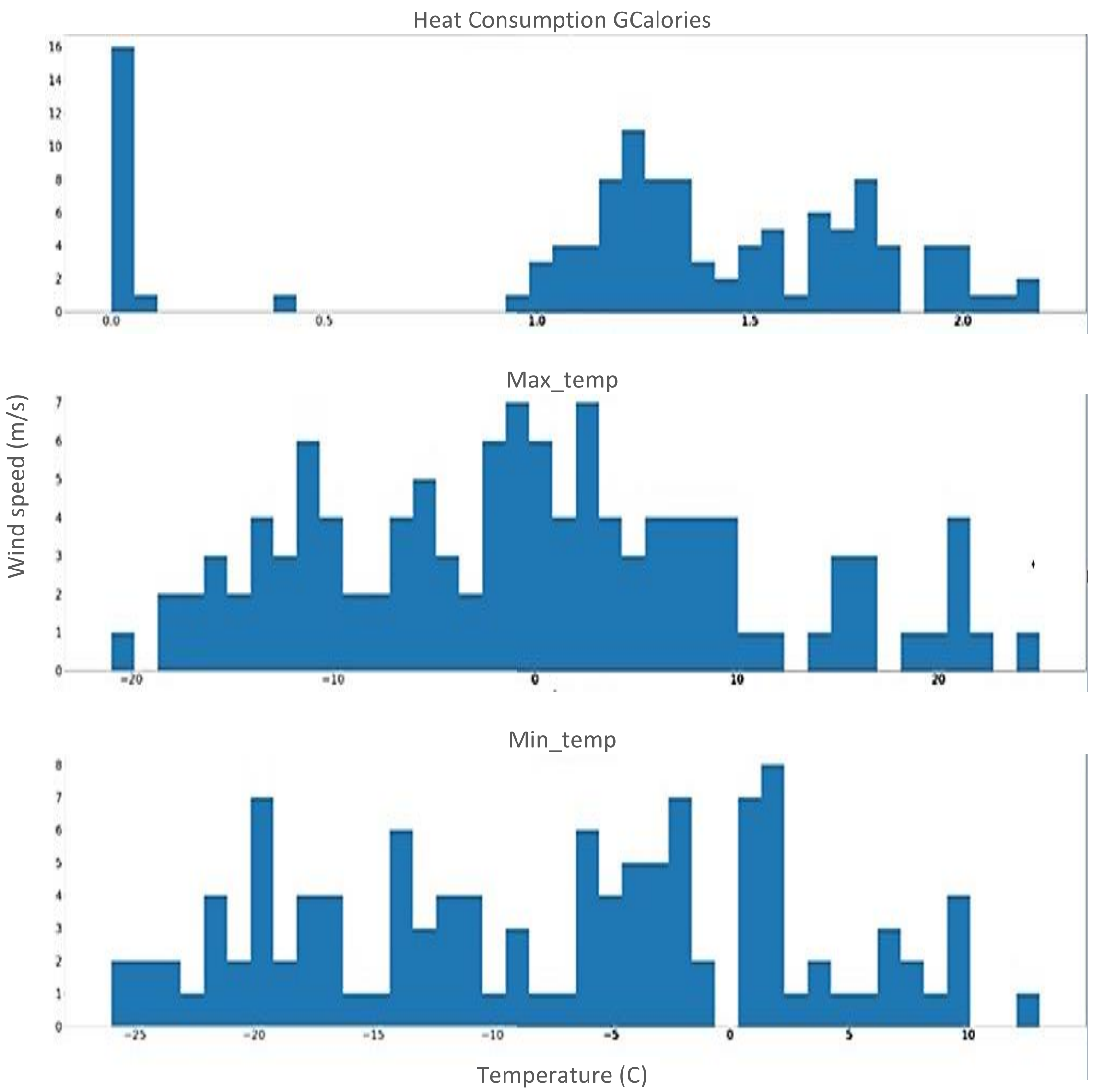


Figure 12. Temperature distribution: maximum and minimum outdoor temperatures in Celsius, heat usage in GCal

3.1.2. Comparative analysis

To begin with, the daytime temperature, which is the highest, and the wind speed, which is the highest, should be used as input characteristics to predict the variable that is at the center of the prediction, using one of the models that were specified, such as linear regression, k-neighbor regression, or random forest regression.

After that, we compared different models to find the one that was most accurate. Figure 14 shows a visual representation of this contrast between the two.

Second, for a better understanding, thermal energy was predicted using the lowest temperature of each day and wind speed, as shown in Figure 15. This allowed us to understand how these elements affect directionality in energy consumption. Third, as shown in Figure 16, the wind speed and the highest and lowest temperatures were used as input data. To further test the results, these three models were used to average the largest dataset that was trained and validated.

A comparison of the data from the three models for the training and test sets is presented in Table 1. For the purpose of modeling the target variable, the data set was divided into a training set (80%) and a test set (20%). With or without the wind component, the highest value of the prediction coefficient for the linear regression method was below 83% accuracy. However, the expected available values improved to an accuracy of 89%-90% (green boxes) when K-neighbor regressor or random forest regressor models were used, demonstrating that these ML model methods are effective in predicting energy consumption trends.

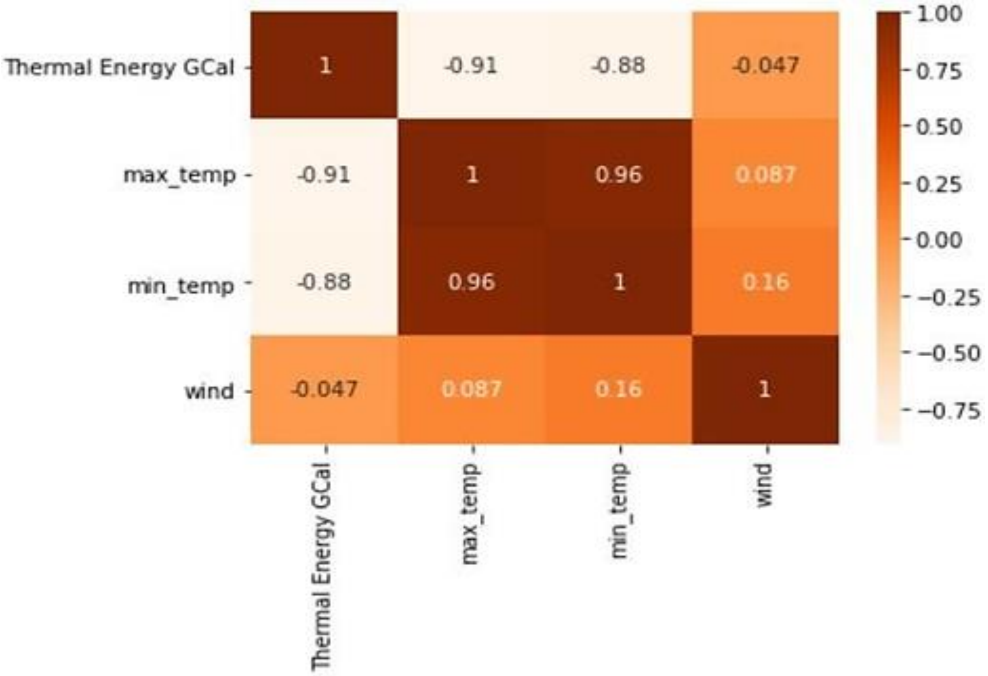


Figure 13. Correlations among the input parameters (heat consumption, temperatures, and wind speed)

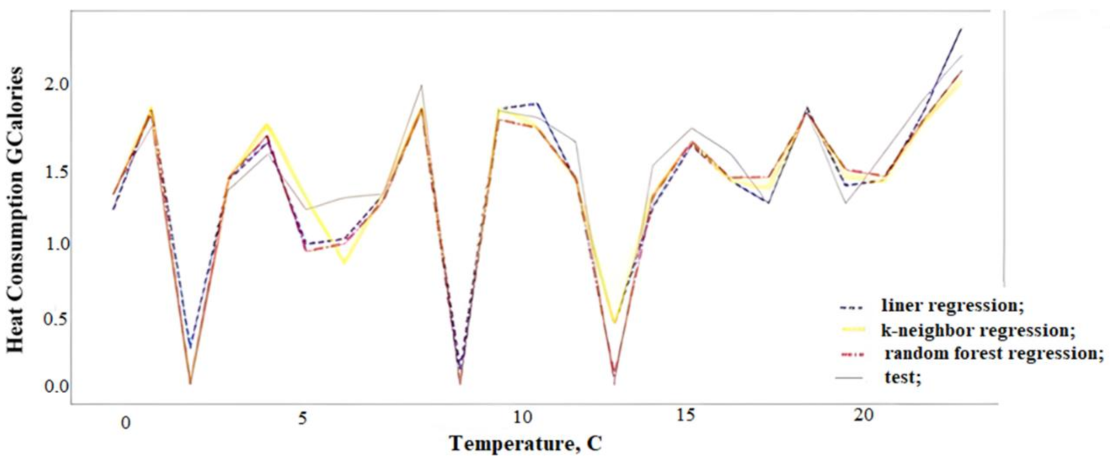


Figure 14. Models with maximum temperature prediction

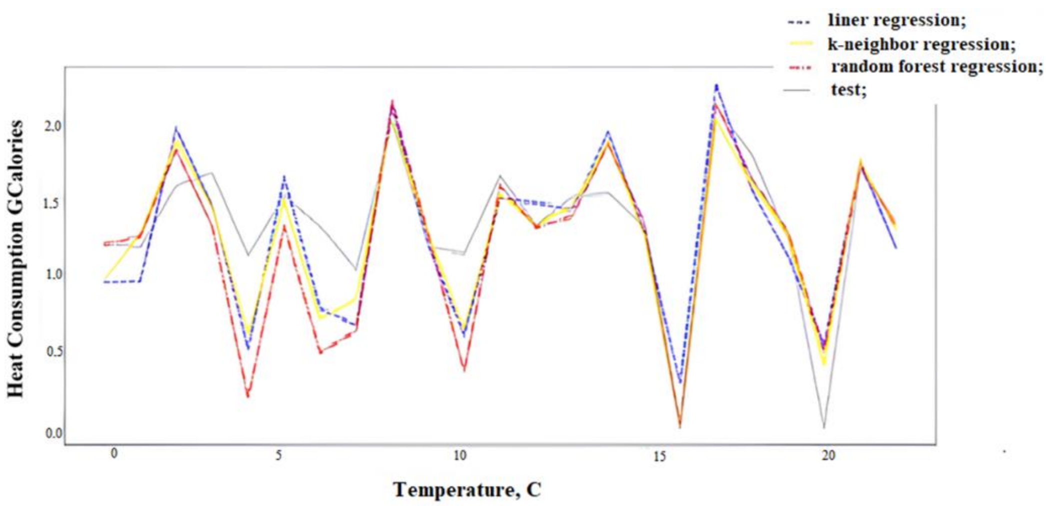


Figure 15. Models with minimum temperature prediction

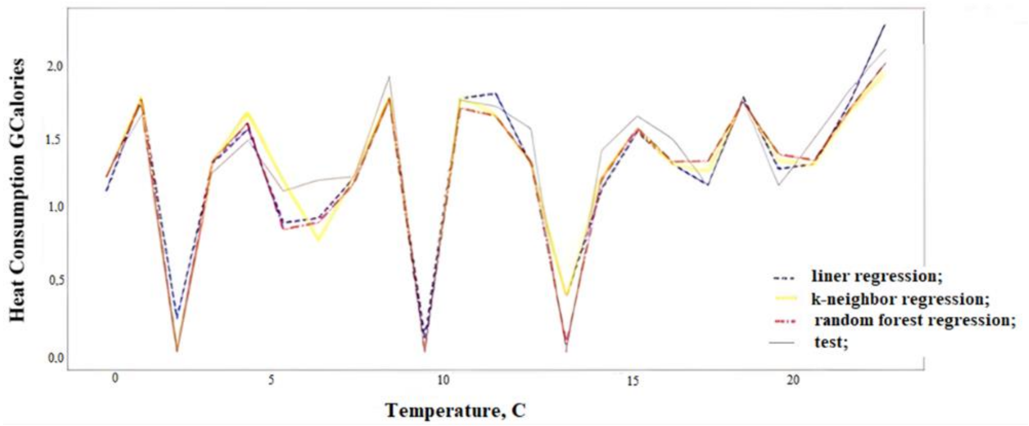


Figure 16. Models with both maximum and minimum temperature prediction

Table 1. Comparison of the three models' prediction accuracy

Model	With wind		Without wind		Input features
	Train	Test	Train	Test	
Linear Regression	0.81	0.54	0.79	0.58	Min temperature
	0.84	0.74	0.84	0.77	Max temperature
	0.83	0.82	0.82	0.83	Min and Max temperature
K-Neighbor Regression	0.89	0.53	0.83	0.90	Min temperature
	0.84	0.71	0.87	0.72	Max temperature
	0.87	0.77	0.87	0.81	Min and Max temperature
Random Forest Regression	0.98	0.42	0.85	0.89	Min temperature
	0.96	0.70	0.92	0.62	Max temperature
	0.96	0.65	0.93	0.78	Min and Max temperature

4. CONCLUSION

In conclusion, the creation of a digital twin for the heating network of smart cities, along with a mathematical model and big data analysis, has great potential for improving development planning and improving the efficiency and stability of urban heating systems. To overcome the difficulties that smart cities face in regulating their heating networks, the initial study was aimed at exploring a combination of digital twin technologies, mathematical modeling, and big data analysis. A review of the literature revealed the importance of heating systems in smart cities, as well as the function of digital twin technology in creating a virtual version of the real infrastructure. Thanks to real-time monitoring, analysis and forecasting, which are made possible by the digital twin, it is possible to better distribute and optimize the operation of heating networks.

To create an accurate digital double, promote a mathematical model, and perform analysis of large amounts of data, a data collection method was needed. The concept and its dynamics were fully understood through the integration of heating network data, real-time data from sensors and smart meters, historical data and certain external data sources. These results have implications for existing Smart City concepts, such as predicting the optimal pumping pressure for each distribution network using machine learning techniques. It is noteworthy that this area of research has not received the most attention in Kazakhstan, but it is expected that its importance will grow in the coming years, as the problems of energy consumption around the world become more urgent, and more states are trying to coordinate energy-intensive industries, especially those associated with buildings and construction. In addition, the machine learning algorithms built in this paper are ready to calculate many problems related to heat consumption in many megacities around the world.

REFERENCES

[1] E. O'Dwyer, I. Pan, S. Acha, and N. Shah, "Smart energy systems for sustainable smart cities: Current developments, trends and future directions," *Applied Energy*, vol. 237, pp. 581–597, Mar. 2019, doi: 10.1016/j.apenergy.2019.01.024.

[2] Y. Liu, C. Yang, L. Jiang, S. Xie, and Y. Zhang, "Intelligent edge computing for IoT-based energy management in smart cities," *IEEE Network*, vol. 33, no. 2, pp. 111–117, Mar. 2019, doi: 10.1109/MNET.2019.1800254.

[3] O. K. Bishoge, G. G. Kombe, and B. N. Mvile, "Energy consumption efficiency behaviours and attitudes among the community," *International Journal of Sustainable Energy Planning and Management*, vol. 31, pp. 175–188, 2021.

[4] J. Knies, "A spatial approach for future-oriented heat planning in urban areas," *International Journal of Sustainable Energy Planning and Management*, vol. 16, pp. 3–30, 2018.

[5] Z. Ullah, F. Al-Turjman, L. Mostarda, and R. Gagliardi, "Applications of artificial intelligence and machine learning in smart cities," *Computer Communications*, vol. 154, pp. 313–323, Mar. 2020, doi: 10.1016/j.comcom.2020.02.069.

[6] A. Yedilkhan *et al.*, "Mathematical justification of thermosyphon effect main parameters for solar heating system," *Cogent Engineering*, vol. 7, no. 1, Jan. 2020, doi: 10.1080/23311916.2020.1851629.

[7] T. Jetzek, "Managing complexity across multiple dimensions of liquid open data: The case of the danish basic data program," *Government Information Quarterly*, vol. 33, no. 1, pp. 89–104, Jan. 2016, doi: 10.1016/j.giq.2015.11.003.

[8] Maslikhat of Astana, "The strategic development program of Astana 2050," *Maslikhat of Astana*. 2019, Accessed: Jun. 01, 2022. [Online]. Available: https://maslihat01.kz/ru/news/strategiya-razvitiya-2050.

[9] S. Schelbach, "Developing a method to improve the energy efficiency of modern buildings by using traditional passive concepts of resource efficiency and climate adaptation," *International Journal of Sustainable Development and Planning*, vol. 11, no. 1, pp. 23–38, Feb. 2016, doi: 10.2495/SDP-V11-N1-23-38.

[10] K. Bogner, F. Pappenberger, and M. Zappa, "Machine learning techniques for predicting the energy consumption/production and its uncertainties driven by meteorological observations and forecasts," *Sustainability*, vol. 11, no. 12, Jun. 2019, doi: 10.3390/su11123328.

[11] M. Bujalski and P. Madejski, "Forecasting of heat production in combined heat and power plants using generalized additive models," *Energies*, vol. 14, no. 8, Apr. 2021, doi: 10.3390/en14082331.

[12] P. A. Østergaard, R. M. Johannsen, H. Lund, and B. V. Mathiesen, "Latest developments in 4th generation district heating and smart energy systems," *International Journal of Sustainable Energy Planning and Management*, vol. 31, pp. 1–4, 2021.

[13] R. Omirgaliyev, A. Salkenov, I. Bapiyev, and N. Zhakiyev, "Industrial application of machine learning clustering for a combined heat and power plant: A Pavlodar case study," in *2021 IEEE International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo)*, Nov. 2021, pp. 51–56, doi: 10.1109/UkrMiCo52950.2021.9716605.

[14] B. Chun and J.-M. Guldmann, "Spatial statistical analysis and simulation of the urban heat island in high-density central cities," *Landscape and Urban Planning*, vol. 125, pp. 76–88, May 2014, doi: 10.1016/j.landurbplan.2014.01.016.

[15] N. Fumo and M. A. Rafe Biswas, "Regression analysis for prediction of residential energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 47, pp. 332–343, Jul. 2015, doi: 10.1016/j.rser.2015.03.035.

[16] F. Lei and P. Hu, "A baseline model for office building energy consumption in hot summer and cold winter region," in *2009 International Conference on Management and Service Science*, Sep. 2009, pp. 1–4, doi: 10.1109/ICMSS.2009.5301031.

[17] R. Wang, S. Lu, and W. Feng, "A novel improved model for building energy consumption prediction based on model integration," *Applied Energy*, vol. 262, Mar. 2020, doi: 10.1016/j.apenergy.2020.114561.

[18] A. Al-Imam, "A novel method for computationally efficacious linear and polynomial regression analytics of big data in medicine," *Modern Applied Science*, vol. 14, no. 5, Apr. 2020, doi: 10.5539/mas.v14n5p1.

[19] S.-J. Kwon, J. Park, J. H. Choi, J.-H. Lim, S.-E. Lee, and J. Kim, "Polynomial regression method-based remaining useful life prediction and comparative analysis of two lithium nickel cobalt manganese oxide batteries," in *2019 IEEE Energy Conversion Congress and Exposition (ECCE)*, Sep. 2019, pp. 2510–2515, doi: 10.1109/ECCE.2019.8912625.

[20] H.-I. Lim, "A linear regression approach to modeling software characteristics for classifying similar software," in *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, Jul. 2019, pp. 942–943, doi: 10.1109/COMPSAC.2019.00152.

[21] T. Parhizkar, E. Rafieipour, and A. Parhizkar, "Evaluation and improvement of energy consumption prediction models using principal component analysis based feature reduction," *Journal of Cleaner Production*, vol. 279, Jan. 2021, doi: 10.1016/j.jclepro.2020.123866.

[22] T. Ahmad and H. Chen, "Nonlinear autoregressive and random forest approaches to forecasting electricity load for utility energy management systems," *Sustainable Cities and Society*, vol. 45, pp. 460–473, Feb. 2019, doi: 10.1016/j.scs.2018.12.013.

[23] S. Li, Z. Shen, and G. Xiong, "A k-nearest neighbor locally weighted regression method for short-term traffic flow forecasting," in *2012 15th International IEEE Conference on Intelligent Transportation Systems*, Sep. 2012, pp. 1596–1601, doi: 10.1109/ITSC.2012.6338648.

[24] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967, doi: 10.1109/TIT.1967.1053964.

[25] S. Zhang, "Challenges in KNN classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 10, pp. 4663–4675, Oct. 2022, doi: 10.1109/TKDE.2021.3049250.

[26] Z. Guo, K. Zhou, X. Zhang, and S. Yang, "A deep learning model for short-term power load and probability density forecasting,"

Energy, vol. 160, pp. 1186–1200, Oct. 2018, doi: 10.1016/j.energy.2018.07.090.
[27] I. El Kafazi, R. Bannari, A. Abouabdellah, M. O. Aboutafail, and J. M. Guerrero, "Energy production: A comparison of forecasting methods using the polynomial curve fitting and linear regression," in *2017 International Renewable and Sustainable Energy Conference (IRSEC)*, Dec. 2017, pp. 1–5, doi: 10.1109/IRSEC.2017.8477278.
[28] J. K. Jaiswal and R. Samikannu, "Application of random forest algorithm on feature subset selection and classification and regression," in *2017 World Congress on Computing and Communication Technologies (WCCCT)*, Feb. 2017, pp. 65–68, doi: 10.1109/WCCCT.2016.25.
[29] Y. Xiao, W. Huang, and J. Wang, "A random forest classification algorithm based on dichotomy rule fusion," in *2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, Jul. 2020, pp. 182–185, doi: 10.1109/ICEIEC49280.2020.9152236.
[30] D. Yao, J. Yang, and X. Zhan, "A novel method for disease prediction: Hybrid of random forest and multivariate adaptive regression splines," *Journal of Computers*, vol. 8, no. 1, Jan. 2013, doi: 10.4304/jcp.8.1.170-177.
[31] S. Wang, Y. Wang, D. Wang, Y. Yin, Y. Wang, and Y. Jin, "An improved random forest-based rule extraction method for breast cancer diagnosis," *Applied Soft Computing*, vol. 86, Jan. 2020, doi: 10.1016/j.asoc.2019.105941.

BIOGRAPHIES OF AUTHORS

Gulnur Tyulepberdinova graduated from Al-Farabi Kazakh National University in 2002 with a bachelor degree in applied mathematics. In 2004, she graduated from the Master's degree program of Al-Farabi Kazakh National University. Since June 30, 2011, the academic degree of Candidate of Physical and Mathematical Sciences has been awarded. Associate professor, specialty 05.13.00 – Computer science, computer engineering, and management. She has more than 60 scientific papers, including scientific publications with a non-zero impact factor, indexed in the international information resources of Scopus, as well as in publications recommended by the of the Ministry of Education and Science of the Republic of Kazakhstan. Main research areas: system analysis and mathematical programming problems, artificial intelligence and bigdata. She can be contacted at email: Gulnur.Tiulepberdynova@kaznu.edu.kz.

Murat Kunelbayev received a bachelor's degree in radio physics and electronics from Al-Farabi Kazakh National University, Almaty, Kazakhstan in 1997, as well as a master's degree in engineering from Al-Farabi Kazakh National University, Almaty, Kazakhstan in 2003. He is currently a senior researcher at the Institute of Information and Computational Technologies CS MES RK. He is also an editor and reviewer of many scientific journals. He has authored or co-authored more than 70 peer-reviewed journal and conference papers, 7 book chapters, and three co-edited books with Springer. His research interests include power system management, electronics, artificial intelligence, sensor technologies, and renewable energy sources. He can be contacted at email: murat7508@yandex.kz.

Olga Shiryayeva received the Ph.D. degree in electrical engineering from the Satbayev University, Almaty, Kazakhstan, in 2001. Since 2002, she has been an associate professor with the Automation and Control Department, the Satbayev University, Almaty, Kazakhstan. Her current research interests include artificial intelligent, smart-technology, digital twins, robotics control systems and MIMO-systems. She can be contacted at email: o.shiryayeva@satbayev.university.

Meruyert Sakypbekova received her bachelor's degree in applied mathematics from Al-Farabi Kazakh National University, Almaty, Kazakhstan, in 1997, and her Master's degree in information system from Al-Farabi Kazakh National University, Almaty, Kazakhstan, in 2003. He is currently a senior lecturer at Al-Farabi Kazakh National University. She can be contacted by e-mail: sakypbekova.meruyert@gmail.com.

Nurlan Sarsenbayev In 2008, he successfully defended his dissertation for the degree of Candidate of Technical Sciences. Currently working at the Kazakh National Research Technical University named after K. Satpayev, as an associate professor in the Department of Automation and Control. N. Sarsenbaev's research work is related to the research of electrical systems, the development of energy-efficient electromechanical coordinated motion systems, as well as simulation modeling of energy systems in the field of renewable energy sources. Currently he is conducting scientific research within the framework of the initiative topics: "Development of a generator based on a dual-power machine for wind turbines and small hydroelectric power plants" and "Mobile biogas plant with a control system". He can be contacted at email: n.sarsenbayev@satbayev.university.

Gulmira Bayandina received a bachelor's degree of Automation of Technology Processes and Manufactory in named Satpayev K.I. Kazakh National Technical University, Almaty, Kazakhstan in 2000, as well as a Master's degree in of automation and control from the named Satpayev K.I. Kazakh National Technical University, Almaty, Kazakhstan in 2001. She is currently a senior lecturer in a Department of Automation and Control in the named Satpayev K.I. Kazakh National Technical University, Almaty. She works with the SCADA systems, automation and control of technological processes. She has a lot of many scientific articles of Republican and International level. She can be contacted at email: g.bayandina@satbayev.university and gul_bayan@mail.ru.