**INTEGRATION OF MACHINE LEARNING FOR MICROCLIMATE MANAGEMENT OPTIMIZATION IN BUILDINGS: PERSPECTIVES AND OPPORTUNITIES**

***N.A. Daurenbayeva1\*, L.B. Atymtayeva2, N.S. Lutsenko3, A. Nurlanuly3***

1International Information Technology University, Almaty, Kazakhstan;

2Suleyman Demirel University, Kaskelen, Kazakhstan;

3Academy of Civil Aviation, Almaty, Kazakhstan.

E-mail: n.daurenbayeva@iitu.edu.kz

**Nurkamilya A. Daurenbayeva** ― PhD student, senior lecturer of the Department «Computer Engineering», International University of Information Technologies

E-mail: n.daurenbayeva@iitu.edu.kz. ORCID: 0000-0003-0341-4017;

**Lyazzat B. Atymtayeva** ― Doctor of physical and Mathematical Sciences, Associate Professor of the Department of «Computer Sciences», University named after Suleiman Demirel

ORCID: 0000-0001-9201-1118;

**Lutsenko Natalya Sergeevna** ― Master, senior lecturer of the Department «Aviation Engineering and technologies», Academy of Civil Aviation

ORCID:0000-0001-7358-5490;

**Almas Nurlanuly** ― PhD student, senior lecturer of the Department «Aviation Engineering and technologies», Academy of Civil Aviation

ORCID: 0000-0002-0364-0455.

© N.A. Daurenbayeva, L.B. Atymtayeva, N.S. Lutsenko, A. Nurlanuly, 2024

**Abstract.** Modern machine learning (ML) technologies offer significant opportunities for optimizing microclimate management systems in buildings. In this article, we explore the potential application of ML methods for forecasting, adaptive control, and optimization of heating, ventilation, and air conditioning (HVAC) systems in buildings. We examine ML methods used for analyzing weather data, working hours, thermal needs, and user preferences to automatically optimize HVAC parameters. Additionally, we discuss the application of ML for detecting faults and preventing failures in microclimate systems, contributing to increased reliability and efficiency of building operations. Finally, we consider prospects for personalizing comfortable microclimates in buildings based on user preferences. Our analysis identifies the potential of ML for creating sustainable, energy-efficient, and comfortable buildings that meet modern requirements for microclimate management.

**Keywords:** machine learning, microclimate management, HVAC Optimization, fault detection, predictive maintenance, user preferences, energy efficiency

**For citation:** N.A. Daurenbayeva, L.B. Atymtayeva, N.S. Lutsenko, A. Nurlanuly. INTEGRATION OF MACHINE LEARNING FOR MICROCLIMATE MANAGEMENT OPTIMIZATION IN BUILDINGS: PERSPECTIVES AND OPPORTUNITIES//INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES. 2024. Vol. 5. No. 18. Pp. 00–00 (In Eng.). https://doi.org/10.54309/IJICT.2024.18.2.008.

**ҒИМАРАТТАРДАҒЫ МИКРОКЛИМАТТЫ БАСҚАРУДЫ ОҢТАЙЛАНДЫРУ ҮШІН МАШИНАЛЫҚ ОҚЫТУДЫ БІРІКТІРУ: ПЕРСПЕКТИВАЛАР МЕН МҮМКІНДІКТЕР**

***Н.А. Дауренбаева1\*, Л.Б. Атымтаева2, Н.С. Луценко3, А. Нұрланұлы3***

1Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан;

2Сүлеймен Демирель атындағы университет;

3Азаматтық авиация академиясы.

E-mail: n.daurenbayeva@iitu.edu.kz

**Дауренбаева Нуркамиля Алдангаровна** ― PhD докторант, «Компьютерлік инженерия» кафедрасының сениор-лекторы Халықаралық ақпараттық технологиялар университеті

E-mail: n.daurenbayeva@iitu.edu.kz. ORCID:0000-0003-0341-4017;

**Атымтаева Ляззат Бахитовна** ― физика-математика ғылымдарының докторы, Сүлейман Демирел атындағы университеттің «Компьютерлік ғылымдар» кафедрасының қауымдастырылған профессоры

ORCID:0000-0001-9201-1118;

**Луценко Наталья Сергеевна** ― Магистр, «Авиациялық техника және технологиялар» кафедрасының аға оқытушысы, Азаматтық авиация академиясы

ORCID:0000-0001-7358-5490;

**Нұрланұлы Алмас** ― PhD докторант, «Авиациялық техника және технологиялар» кафедрасының сениор-лекторы, Азаматтық авиация академиясы

ORCID:0000-0002-0364-0455.

© Н.А. Дауренбаева, Л.Б. Атымтаева, Н.С. Луценко, А. Нұрланұлы, 2024

**Аннотация.** Машиналық оқытудың (ML) заманауи технологиялары ғимараттардағы микроклиматты басқару жүйелерін оңтайландыруға айтарлықтай мүмкіндіктер береді. Бұл мақалада біз ҒИМАРАТТАРДАҒЫ жылыту, желдету және ауаны баптау (HVAC) жүйелерін болжау, адаптивті бақылау және оңтайландыру ҮШІН ML әдістерінің әлеуетті қолданылуын зерттейміз. БІЗ HVAC параметрлерін автоматты түрде оңтайландыру үшін ауа райы деректерін, жұмыс уақытын, жылу қажеттіліктерін және пайдаланушы қалауларын талдау үшін ҚОЛДАНЫЛАТЫН ML әдістерін зерттейміз. Сонымен қатар, біз ҚҰРЫЛЫС жұмыстарының сенімділігі мен тиімділігін арттыруға ықпал ететін микроклиматтық жүйелердегі ақауларды анықтау және ақаулардың алдын алу ҮШІН ML қолдануды талқылаймыз. Соңында, біз пайдаланушылардың қалауы бойынша ғимараттардағы жайлы микроклиматтарды жекелендіру перспективаларын қарастырамыз. Біздің талдауымыз МИКРОКЛИМАТТЫ басқарудың заманауи талаптарына жауап беретін тұрақты, энергияны үнемдейтін және жайлы ғимараттарды құру ҮШІН ML әлеуетін анықтайды.

**Түйін сөздер:** машиналық оқыту, микроклиматты басқару, HVAC Оңтайландыру, ақауларды анықтау, болжамды техникалық қызмет көрсету, пайдаланушының қалауы, энергия тиімділігі

**Дәйексөздер үшін:** Н.А. Дауренбаева, Л.Б. Атымтаева, Н.С. Луценко, А. Нұрланұлы. ҒИМАРАТТАРДАҒЫ МИКРОКЛИМАТТЫ БАСҚАРУДЫ ОҢТАЙЛАНДЫРУ ҮШІН МАШИНАЛЫҚ ОҚЫТУДЫ БІРІКТІРУ: ПЕРСПЕКТИВАЛАР МЕН МҮМКІНДІКТЕР //ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ. 2024. Т. 5. No. 18. 00–00 бет. (ағылшын тілінде). https://doi.org/10.54309/IJICT.2024.18.2.008.

**ИНТЕГРАЦИЯ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОПТИМИЗАЦИИ УПРАВЛЕНИЯ МИКРОКЛИМАТОМ В ЗДАНИЯХ: ПЕРСПЕКТИВЫ И ВОЗМОЖНОСТИ**

***Н.А. Дауренбаева1\*, Л.Б. Атымтаева2, Н.С. Луценко3, А. Нұрланұлы3***

1Международный университет информационных технологий, Алматы, Казахстан;

2Университет Сулеймана Демиреля, Каскелен, Казахстан;

 3Академия Гражданской Авиации, Алматы, Казахстан.

E-mail: n.daurenbayeva@iitu.edu.kz

**Дауренбаева Нуркамиля Алдангаровна** ― докторант PhD, сениор-лектор кафедры «Компьютерная инженерия», Международный университет информационных технологий

E-mail: n.daurenbayeva@iitu.edu.kz. ORCID: 0000-0003-0341-4017;

**Атымтаева Ляззат Бахитовна** ― доктор физико-математических наук, ассоциированный профессор кафедры «Компьютерные науки», Университет им. Сулеймана Демиреля

ORCID: 0000-0001-9201-1118;

**Луценко Наталья Сергеевна** ― магистр, старший преподаватель кафедры «Авиационная техника и технологии» Академии гражданской авиации

ORCID: 0000-0001-7358-5490;

**Алмас Нұрланұлы** ― докторант PhD, сениор-лектор кафедры «Авиационная техника и технологии» Академии гражданской авиации

ORCID:0000-0002-0364-0455.

© Н.А. Дауренбаева, Л.Б. Атымтаева, Н.С. Луценко, А. Нұрланұлы, 2024

**Аннотация.** Современные технологии машинного обучения (ML) предоставляют значительные возможности для оптимизации систем управления микроклиматом в зданиях. В этой статье исследуется потенциальное использование методов ML для прогнозирования, адаптивного мониторинга и оптимизации систем отопления, вентиляции и кондиционирования воздуха (HVAC) в зданиях. Исследованы методы ML, используемые для анализа данных о погоде, времени безотказной работы, тепловых потребностей и предпочтений пользователей для автоматической оптимизации параметров HVAC. Кроме того, обсуждаено использование ML для выявления дефектов и предотвращения дефектов в микроклиматических системах, которые способствуют повышению надежности и эффективности строительных работ. Наконец, рассмотрены перспективы персонализации комфортного микроклимата в зданиях по усмотрению пользователей. Данный анализ определяет потенциал ML для создания устойчивых, энергоэффективных и комфортных зданий, отвечающих современным требованиям управления микроклиматом.

**Ключевые слова**: машинное обучение, управление микроклиматом, оптимизация HVAC, обнаружение неисправностей, прогнозируемое обслуживание, предпочтения пользователей, энергоэффективность

**Для цитирования:** Н.А. Дауренбаева, Л.Б. Атымтаева, Н.С. Луценко, А. Нұрланұлы. ИНТЕГРАЦИЯ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОПТИМИЗАЦИИ УПРАВЛЕНИЯ МИКРОКЛИМАТОМ В ЗДАНИЯХ: ПЕРСПЕКТИВЫ И ВОЗМОЖНОСТИ//МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ. 2024. Т. 5. No. 18. Стр. 00–00. (На анг.). https://doi.org/10.54309/IJICT.2024.18.2.008.

**Introduction**

Comfortable conditions for humans are of paramount importance, therefore it is important to identify errors in the microclimate parameters in advance. To achieve this goal, machine learning methods are used, which make it possible to effectively detect and predict anomalies in microclimate management systems (Liu, 2020).

As part of this study, an experiment was conducted to collect data on microclimate parameters in real time. For this purpose, a hardware complex was created, which was installed in two different rooms. Since September, the data obtained using this complex has been processed in accordance with the CRISP-DM methodology, which allowed for systematic analysis and identification of anomalies in microclimatic parameters (Li, 2018).

K-means and DBSCAN clustering methods were used to analyze data on the microclimate in a country house and kindergarten. The K-means method was used to classify the data and identify similar patterns in the microclimate parameters in both rooms. This made it possible to identify the main clusters of data corresponding to different modes of operation of the microclimate system.

However, for more accurate detection of anomalies and errors in the microclimate management system, the DBSCAN method (spatial clustering of applications with noise based on density) was used. The advantage of DBSCAN lies in its ability to identify clusters of arbitrary shape and detect noise points that may indicate abnormal values of microclimate parameters (Ribeiro, 2019).

In the conditions of a country house and kindergarten, DBSCAN has proven itself as the optimal method, as it effectively identified errors in the microclimate management system. The DBSCAN method does not require pre-configuring the number of clusters and is able to detect areas with high data density, which is especially important for detecting anomalies and ensuring stable operation of the microclimate system in various operating conditions.

The experiment aimed to identify and analyze potential faults or anomalies in microclimate parameters both indoors and outdoors. Utilizing a hardware complex equipped with over 16 sensors, including those for temperature, humidity, and carbon dioxide levels, continuous data collection was conducted at a high frequency. Real-time data was then transmitted to Google Sheets for analysis.

Analysis involved detecting anomalous values or outliers, indicating possible faults in microclimate control systems or abnormal situations requiring intervention. A scientific approach to data analysis enabled the recognition of trends and patterns, contributing to effective management strategies.

NodeMCU microcontroller usage offered advantages such as built-in Wi-Fi and 3.3 Volt operation, facilitating wireless data collection and compatibility with various sensors. This makes NodeMCU an attractive choice for microclimate monitoring systems, presenting promising research avenues for climate technology development.

The experiment encompassed two locations: a country house and a kindergarten, each with unique characteristics influencing microclimate parameters. Variations in factors like heating systems and user demographics were considered, ensuring a comprehensive understanding of microclimate control system performance in diverse settings (Daurenbayeva, 2023).

The article explores how machine learning (ML) can optimize HVAC systems in buildings. It covers forecasting, adaptive control, and fault detection using ML. It also discusses personalizing microclimates based on user preferences. Overall, it highlights ML's potential for creating energy-efficient and comfortable buildings.

**Materials and methods**

The experiment was conducted following the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, a widely recognized framework for data analysis projects. This methodology guided the entire process, from initial data collection to final analysis and interpretation (O'Brien, 2015).

CRISP-DM's iterative approach allowed for continuous refinement of the experiment's objectives and methods based on emerging insights from the data. Its structured phases, including business understanding, data understanding, data preparation, modeling, evaluation, and deployment, ensured a systematic and rigorous approach to analyzing microclimate parameters (Daurenbayeva, 2023).

# *Data understanding and visualization*

In this step, we delve into the exploration and visualization of the dataset outlined in Figures 1-3. This dataset comprises microclimate parameters such as Indoor and Outdoor Temperature, Indoor and Outdoor Humidity, Dew-point, Pressure, TVOC, Power, Current, Voltage, Aftershock, CO2, UV-radiation. Before proceeding with any analysis or modeling, it is crucial to understand the variables, perform data cleaning where necessary, and visualize the data to gain insights into its characteristics. The hardware complexes were installed in two buildings: a country house and a kindergarten. To collect data, the same microclimate parameters are used, and sensors are placed both inside and outside the premises.

 

Fig.1 - Inside and Outside Temperature Over Time Graph (Country house)

 

Fig. 2 - Combined Indoor and Outdoor Temperature Trends for Kindergarten and Country house



Fig. 3 - Inside and Outside Temperature Over Time Graph (Kindergarten)

*Understanding Variance in Principal Component Analysis (PCA)*

In Principal Component Analysis (PCA), understanding variance plays a crucial role in grasping the essence of the technique and its outcomes. Variance serves as a fundamental concept in PCA, delineating how much information each principal component retains from the original dataset. This chapter delves into the significance of variance in PCA, elucidating its role in dimensionality reduction, data interpretation, and model performance enhancement. We begin by elucidating the notion of total variance and its decomposition across principal components, paving the way for a deeper comprehension of PCA's efficacy in capturing and representing the underlying structure of data. Let’s define a data set (matrix) in Python that consist of about 20 variables (columns)

*For country house*

*Explained variance ratio for each principal component:*

PC1: 34.06 %

PC2: 25.77 %

PC3: 20.77 %

PC4: 19.40 %



*For kindergarten*

*Explained variance ratio for each principal component:*

PC1: 40.08 %

PC2: 21.21 %

PC3: 20.94 %

PC4: 17.76 %



PCA, while a powerful tool for dimensionality reduction and data representation, has its limitations, particularly when derived from noisy data. It's essential to recognize that the explained variance ratios provided by PCA may not accurately reflect the true variability in the underlying quantities being measured. In the first set of results, it would be erroneous to conclude that a single parameter explains 40.08 % of the variability in the observed data. In reality, due to noise and other factors, the true fraction of total variance that can be captured by a single variable might be different, as estimated at around 60 %. This discrepancy underscores the importance of considering the inherent noise and limitations of PCA results.

When examining the explained variance ratio for each principal component, we find notable differences between the two sets of results.

**Results and discussion**

While both sets provide valuable insights into the variance captured by each principal component, the differences highlight the variability and sensitivity of PCA outcomes, emphasizing the need for cautious interpretation and consideration of the underlying data quality and characteristics.

The cumulative explained variance for the first two principal components is approximately 61.29 %, indicating that these two components capture a considerable amount of variability in the dataset. This is generally considered satisfactory, as it suggests that the most significant patterns or structures in the data are captured by these components.

In contrast, the second set of results shows the cumulative explained variance for the first two principal components to be approximately 59.83 %. Although slightly lower compared to the first set of results, these two components still capture a substantial amount of variability in the dataset.

Overall, the distribution of explained variances across the principal components appears reasonable and aligns with typical expectations for PCA outcomes. Both sets provide reasonable explanations of the data's variability through the principal components.

 ****

Fig. 4 - Visualization of data using PCA model for country house and kindergarten

 

Fig. 5 - Clustering using K-means for country house and kindergarten

  ****

Fig. 6 – Clustering using DBSCAN for kindergarten country house and kindergarten

Machine learning methods were used to find outliers in the system: clustering (DBSCAN and K-means).

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a widely-used clustering algorithm in machine learning, particularly suitable for identifying outliers or anomalies in datasets with complex structures. Unlike traditional clustering methods like K-means, which require the specification of the number of clusters beforehand, DBSCAN does not require such a parameter and can automatically detect clusters of arbitrary shapes and sizes.

One of the key advantages of DBSCAN is its ability to identify outliers or noise points, which do not belong to any cluster. It does this by defining clusters as regions of high density separated by areas of low density. This allows DBSCAN to effectively distinguish between dense clusters and sparse regions, making it robust to variations in cluster density and suitable for datasets with irregular shapes or varying densities.

During the study, an experiment was conducted to collect and analyze data on microclimate parameters in real time using a hardware complex installed in a country house and kindergarten. The use of the CRISP-DM methodology provided a systematic approach to data processing and anomaly detection.

K-means and DBSCAN clustering methods were used to analyze the data. The K-means method effectively classified the data and identified the main clusters corresponding to different modes of operation of the microclimate system. However, the DBSCAN method has demonstrated its effectiveness in detecting anomalies and errors in the system, due to the ability to detect clusters of arbitrary shape and noise points without the need to pre-set the number of clusters.

Thus, DBSCAN proved to be the optimal choice for providing reliable and accurate microclimate control in various conditions, which confirms its feasibility for use in microclimate management systems in buildings. The results of the study emphasize the importance of using modern machine learning methods to improve the quality of microclimatic control and increase comfort and safety in the premises.

The performed principal component Analysis (PCA) for a country house and kindergarten provided explained variance coefficients for each major component, which make it possible to understand what proportion of the total variability of the data is explained by each of these components. Let's consider their values for a country house and a kindergarten:

For a country house:

PC1: 34.06 %

PC2: 25.77 %

PC3: 20.77 %

PC4: 19.40 %

For kindergarten:

PK1: 40.08 %

PK2: 21.21 %

PC3: 20.94 %

PC4: 17.76 %

What does this give us:

Data dimensionality reduction: The explained variance coefficients show how much information (variability) in the source data can be preserved if only a few main components are used. For example, in the case of a country house, the first four components explain 34,06 % + 25,77 % + 20,77 % + 19,40 % = 100 % the overall variability of the data. In the case of a kindergarten – 40,08 % + 21,21 % + 20,94 % + 17,76 % = 100 %.

Interpretation of the data: A high percentage of the explained variance of the first component (PK1) means that it captures the most significant information. For example, for kindergarten, PK1 explains 40.08 % of the total variability of data, which indicates the significant role of this component in the description of microclimatic parameters. In a country house, PK1 explains 34.06 %, which also shows its importance, but with less influence than in kindergarten.

Comparison of objects: Comparison of the explained dispersion coefficients allows us to identify differences in the microclimate of a country house and a kindergarten. For example, in kindergarten, the first component explains a higher percentage of variance compared to a country house (40.08 % vs. 34.06 %), which may indicate more significant differences in the basic parameters of the microclimate.

Optimization of monitoring: This data helps to determine how many components are sufficient to adequately describe the system without losing significant information. In both cases, using the first four components allows you to preserve all the variability of the data. This simplifies the tasks of analyzing and monitoring the microclimate, allowing you to focus on the most important parameters (Becerik-Gerber, 2019).

Identification of important factors: Analysis of dispersion coefficients allows you to identify the key factors affecting the microclimate. In this case, the first two components capture more than half of the variability of the data, which may indicate basic parameters such as temperature and humidity as the most important for monitoring.

Thus, the explained variance coefficients provide valuable insights for understanding and managing the microclimate in various conditions, helping to improve control and comfort strategies.

**Conclusion**

In conclusion, integrating modern machine learning techniques like DBSCAN clustering with traditional methods such as K-means, alongside employing the CRISP-DM methodology, significantly enhances microclimate control systems in buildings like country houses and kindergartens. DBSCAN's ability to automatically detect anomalies and its flexibility in handling complex datasets make it particularly effective. Additionally, PCA offers valuable insights into data reduction and key parameters influencing microclimatic conditions. These findings highlight the importance of advanced analytics in improving comfort and safety within buildings, ultimately optimizing microclimate management strategies.

**REFERENCES**

Becerik-Gerber B. & Jazizadeh F. (2019). Smart buildings: A review of the concept and its performance. ― Building and Environment, 2019. ― Рp. 39–49.

Daurenbayeva N., Atymtayeva L. & Nurlanuly A. (2023). Choosing the intelligent thermostats for effective decision making in BEMS. ― 17th International Conference on Electronics Computer and Computation (ICECCO), 2023. ― Рp. 1–4.

Daurenbayeva N., Nurlanuly A., Atymtayeva L. & Mendes M. (2023). Survey of Applications of Machine Learning for Fault Detection, Diagnosis and Prediction in Microclimate Control Systems. ― Energies, 2023. ― Р. 3508.

Li H., Hong T. & Lin B. (2018). A review of building energy modeling for control and operation. ― Renewable and Sustainable Energy Reviews, 2018. ― Рp. 1793–1812.

Jin X., Kusiak A. & Zhang Z. (2018). Data-driven building energy management: ― A review. Energy and Buildings, 2018. ― Рp. 1546–1565.

Tahmasebi F., Buswell R.A., Darby S.J. & Oreszczyn T. (2013). Energy performance gap of refurbished buildings: An analysis of UK non-domestic building stock. ― Energy and Buildings, 2013. ― Рp. 433–1451.

Li C., Li M. & Chen B. (2020). A review of machine learning applications in HVAC systems. ― Energy and Buildings, 2020. ― Р. 110100.

Liu H., Guo S., Xu P. & Ren H. (2020). Smart building: An insight into the use of artificial intelligence in smart heating, ventilation, and air conditioning systems. ― Sustainable Cities and Society, 2020. ― Р. 102248.

Ribeiro A., Guedes C., Pereira L., Rodrigues E., Novais P. & Neves J. (2019). Smart Buildings Energy Efficiency: ― A Review. Sustainability, 2019. ― Р. 7072.

Joh J., Sohn J. & Kim G. (2017). An occupancy-based energy management system using an artificial neural network. ― Energy and Buildings, 2017. ― Рp. 168–180.

O'Brien W. & Wright J. (2015). A review of building energy analysis tools. ― HERD: Health Environments Research & Design Journal, 2015. ― Рp. 92–109.

Yan D., O’Brien W. & Hong T. (2018). Occupancy learning-based building energy modeling: A review. ― Energy and Buildings, 2018. ― Рp. 296–313.