

## **АННОТАЦИЯ**

**диссертационной работы Дауренбаевой Нуркамили Алдангаровны «Интеллектуальная система диагностирования ошибок для управления микроклиматом зданий», представленной на соискание степени доктора философии (PhD) по образовательной программе: 8D06102 - «Компьютерная и Программная Инженерия».**

**Общая характеристика работы.** Диссертационное исследование посвящено изучению и практической реализации методов автоматического обнаружения неисправностей и диагностики в системах микроклимата зданий на основе алгоритмов машинного обучения. Основное внимание уделено алгоритмам кластеризации и статистическим методам, позволяющим выявлять аномалии в условиях отсутствия размеченных данных.

**Актуальность темы исследования.** В последние годы методы автоматической диагностики Отопление, вентиляция и кондиционирование воздуха (далее, ОВКВ) систем развиваются особенно активно. Wang и соавт. (2024) показали, что смещение показаний датчиков снижает эффективность охлаждения в дата-центрах и предложили гибридный подход на основе Случайный лес (Random Forest) и Байесовского вывода. Zhao и соавт. (2023) рассматривали диагностику при неполных данных и применили IFWA-LSTM (Improved Fuzzy Weighted Average – Long Short-Term Memory) для их восстановления. Li и соавт. (2024) разработали отказоустойчивое управление (FTC- Fault-Tolerant Control) с использованием Байесовского вывода, подтвердив его эффективность для энергосбережения и поддержания комфорта. Zhang и соавт. (2024) предложили метод трансферного обучения, основанный на энергетическом и массовом балансе, что позволило повысить качество диагностики в разных условиях эксплуатации.

Эти исследования отражают мировую тенденцию: развитие устойчивых методов диагностики, способных работать с шумными, неполными и изменяющимися данными. Однако большинство таких работ проводилось либо на симулированных данных, либо в условиях специализированных систем (дата-центры, крупные общественные здания), где уровень автоматизации существенно выше.

В отличие от них, данное исследование базируется на реальных эксплуатационных данных из жилых и нежилых зданий, где автоматизация ограничена, а измерения характеризуются высоким уровнем шума и сезонными колебаниями. Это обусловило выбор методов, не требующих предварительной разметки данных: статистической очистки Z-score, снижения размерности PCA (Principal Component Analysis) и кластеризации без учителя DBSCAN (Density-Based Spatial Clustering of Applications with Noise).

Дополнительно использование показателей MTBF (среднее время между отказами) и функции надёжности  $R(t)$  даёт возможность количественно оценивать состояние и долговечность микроклиматических

систем, что особенно важно для автономного мониторинга и предупреждения отказов.

Актуальность данного исследования была обозначена в Послании Президента Республики Казахстан К.К. Токаева народу Казахстана от 8 сентября 2025 года. В документе отмечено сложное состояние жилищно-коммунальной инфраструктуры и необходимость ее цифровой трансформации. Особое внимание уделено внедрению систем мониторинга и интеллектуального управления, а также использованию технологий искусственного интеллекта и «Умных городов» для повышения энергоэффективности и устойчивости инфраструктуры.

**Степень разработанности проблемы.** Предыдущие исследования продемонстрировали научный интерес к таким темам, как интеллектуальные системы управления и обнаружение неисправностей. Работы как отечественных, так и зарубежных авторов способствовали формированию теоретической базы в этой области. Однако многие вопросы остаются открытыми для обсуждения, особенно те, которые связаны с практической реализацией методов обучения без учителя, обработкой многомерных сенсорных данных и адаптацией алгоритмов к местным климатическим и инфраструктурным условиям. До настоящего времени уровень исследований в этой области остается крайне ограниченным. Поэтому тема требует дальнейшего изучения, в первую очередь с акцентом на выявление новых подходов, методов и механизмов для разработки и функционирования соответствующих систем. Вышеперечисленные факторы определили выбор темы диссертации, а также ее цель и задачи исследования.

**Целью диссертационной работы** является разработка подхода, основанного на анализе данных с применением статистических методов и методов машинного обучения, для оценки надёжности и выявления скрытых неисправностей в системах микроклимата жилых и нежилых зданий.

**В соответствии с этой целью были поставлены следующие исследовательские задачи:**

- Провести обзор существующих подходов к мониторингу микроклимата и диагностике неисправностей;
- реализовать экспериментальную установку сбора и зарегистрированных данных;
- осуществить сбор и предварительную подготовку данных с сенсоров микроклиматических систем.
- выполнить очистку данных от выбросов с применением метода Z-score;
- снизить размерность данных с использованием метода главных компонент (PCA) для упрощения визуализации и анализа;
- сравнить алгоритмы кластеризации DBSCAN и K-Means для оценки их эффективности в задачах обнаружения аномалий в зашумлённых данных;
- оценить надёжность систем микроклимата с использованием статистических показателей, таких как MTBF и функция надёжности  $R(t)$ .

**Объектом исследования** являются системы микроклимата (ОВКВ) в зданиях.

**Предметом исследования** являются основанные на данных методы обнаружения неисправностей и анализа надёжности.

**Теоретическая и методологическая основа.** Исследование опирается на теоретические положения машинного обучения и интеллектуального анализа данных. В качестве основных инструментов применены алгоритмы кластеризации (в частности, DBSCAN), статистические методы (Z-score), методы снижения размерности (PCA), а также стандартизированная методология CRISP-DM, обеспечивающая структурированный подход к процессу анализа данных. Программная реализация выполнена с использованием языка Python и специализированных библиотек, включая scikit-learn, pandas, matplotlib и др.

**Информационная база исследования** включает в себя источники данных, которые состоят из экспериментально собранных данных из жилых и нежилых зданий, включая параметры микроклимата (концентрация ( $\text{CO}_2$ ,  $\text{CO}_{2\text{out}}$ ), температура ( $T$ ,  $T_{\text{out}}$ ), влажность ( $H$ ,  $H_{\text{out}}$ ), точка росы ( $DP$ ,  $DP_{\text{out}}$ ), летучие органические соединения (TVOC), давление ( $P$ ,  $P_{\text{out}}$ ), ток ( $C$ ), напряжение ( $V$ ), мощность ( $P_{\text{wtr}}$ ), ультрафиолетовое излучение ( $UVr$ ) и уровень освещенности ( $L$ ) измерялись каждые 10 секунд.) полученных с помощью датчиков на базе устройств IoT.

**Научная новизна** заключается в разработке унифицированного диагностического подхода, сочетающего алгоритмы DBSCAN, PCA и Z-score для обработки неразмеченных данных микроклимата. Адаптирована методология CRISP-DM и многомерных сенсорных данных, что позволило повысить точность и устойчивость анализа. Предложен подход, обеспечивающий выявление отклонений от нормального режима работы посредством идентификации выбросов и формирования кластеров, а также позволяющий оценивать надёжность систем микроклимата с использованием показателей MTBF и  $R(t)$ . Эффективность подхода была подтверждена на основе анализа информативности данных, проведённого с использованием объяснённой дисперсии.

**В ходе диссертационного исследования были получены следующие научные результаты, которые определяют его научную новизну:**

- Проведён комплексный обзор современных подходов к мониторингу микроклимата и диагностике неисправностей HVAC-систем с акцентом на алгоритмы обработки данных и методы машинного обучения.
- Разработана и реализована экспериментальная модель мониторинга микроклимата в жилых и нежилых зданиях с использованием многопараметрических сенсорных модулей.
- Мониторинг проводился по 11 параметрам (температура, влажность,  $\text{CO}_2$ , TVOC, освещённость, потребляемая мощность и др.) в соответствии со стандартами ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers) и IEA (International Energy Agency).

- Выполнен сбор, предварительная обработка и очистка данных от выбросов методом Z-score, что позволило повысить точность последующего анализа.

- Применён метод главных компонент (PCA) для снижения размерности данных с сохранением информативности признаков и упрощения визуализации.

- Проведено сравнение алгоритмов кластеризации DBSCAN и K-Means для выявления аномалий в зашумлённых немаркированных данных; установлено преимущество DBSCAN при работе со сложными многомерными наборами с шумовыми точками.

- Методология CRISP-DM адаптирована для задач обнаружения неисправностей в шумных и неразмеченных данных с минимальной предварительной обработкой, включая возможность применения в режиме реального времени.

- Впервые проведена оценка надёжности систем микроклимата на основе реальных данных: Жилые здания: MTBF  $\approx 103$  ч, R(24)  $\approx 79,2$  %; Нежилые здания: MTBF  $\approx 45$  ч, R(24)  $\approx 58,6$  %. Это указывает на более высокую надёжность систем в жилом секторе.

#### **Основные положения, выносимые на защиту:**

- Разработана и реализована экспериментальная установка для сбора и анализа сенсорных данных микроклимата жилых и нежилых зданий, включающая измерение температуры, влажности, уровня CO<sub>2</sub>, освещённости, атмосферного давления, точки росы, скорости движения воздуха, сейсмических колебаний (Aftershocks), концентрации TVOC, ультрафиолетовое излучение (UVr) и потребления электроэнергии в соответствии со стандартами.

- Адаптирована методология CRISP-DM для обнаружения неисправностей в немаркированных реальных данных микроклимата при минимальной предварительной обработке; проведено сравнение алгоритмов DBSCAN и K-Means, подтверждающее преимущество DBSCAN для выявления аномалий в многопараметрических зашумлённых выборках.

- Проведена оценка надёжности систем микроклимата с использованием показателей MTBF и R(t), выявившая более высокую эксплуатационную стабильность систем в жилых зданиях по сравнению с нежилыми.

#### **Основные результаты работы были представлены на следующих работах:**

Основные результаты и научные достижения данного исследования были представлены и обсуждены на семинарах кафедры компьютерной инженерии Международного университета информационных технологий (2021–2025 гг.), кафедры информационных систем Университета Сулеймана Демиреля (SDU, 2024–2025 гг.), Политехнического института Коимбры, Португалия (2023–2024 гг.), а также кафедры искусственного интеллекта Финансового университета при Правительстве Российской Федерации (2024–2025 гг.). Автор также принимал участие в 17-й Международной

конференции по электронике, вычислительной технике и вычислениям (ICECCO) 2023 года и Международной конференции по управлению физическими активами и науке о данных (PAMDAS) 2025 года.

1. **Daurenbayeva, N.**, Nurlanuly, A., Atymtayeva, L., Mendes, M. Survey of Applications of Machine Learning for Fault Detection, Diagnosis and Prediction in Microclimate Control Systems. *Energies* 2023, 16, 3508. <https://doi.org/10.3390/en16083508>.

2. **Daurenbayeva, N.**, Atymtayeva, L., Nurlanuly, A., Bykov, A., Akhmetov, B., Shuitenov, G., Turusbekova, U. A Machine Learning Approach to Microclimate Monitoring and Fault Detection. *AMIS* 2025, 19, 327-334. <https://doi.org/10.18576/amis/190209>.

3. **Дауренбаева, Н.А.**, Нұрланұлы, А., Атымтаева, Л.Б., Быков, А.А., Ергалиев, Д.С., Әбдірашев, Ө.К. Микроклимат параметрлерін кластеризациялау: әдістер мен математикалық сипаттамалар // *ENU Bulletin (Л.Н. Гумилев ЕНУ Хабаршысы), Technical Sciences And Technology Series, №4 (149)/ 2024 pp. 202-214.* <https://doi.org/10.32523/2616-7263-2024-149-4-202-214>.

4. **Daurenbayeva, N.**, Atymtayeva, L., Nurlanuly, A. 17th International Conference on Electronics Computer and Computation (ICECCO-2023). Choosing the intelligent thermostats for the effective decision making in BEMS. 1-4. 10.1109/ICECCO58239.2023.10147131.

5. **Daurenbayeva, N.**, Atymtayeva, L., Mendes, M., Nurlanuly, A. & Yagalieva B., (2025, July 17–18). Machine learning approach to fault detection in microclimate system at residential and non-residential buildings. Paper presented at the PAMDAS 2025 – International Conference on Physical Asset Management and Data Science, Coimbra Institute of Engineering (ISEC), Polytechnic University of Coimbra, Portugal.

6. **Daurenbayeva, N.A.**, Atymtayeva, L.B., Lutsenko, N.S., Nurlanuly A. Integration of machine learning for microclimate management optimization in buildings: perspectives and opportunities. *International Journal of Information and Communication Technologies*, 2024. Vol. 5. Is. 2. <https://doi.org/10.54309/IJICT.2024.18.2.008>.

7. **Дауренбаева, Н.А.**, Атымтаева, Л.Б., Ыбытаева, Г.С., Нұрланұлы, А. Свидетельство на право охраны программы для ЭВМ № 41781 Республики Казахстан. Аппаратный комплекс для реального мониторинга параметров микроклимата с интегрированным датчиком сейсмического воздействия / заявка 04.01.2024; публикация 05.01.2024.

### **Результаты**

В ходе исследования проведён комплексный обзор современных подходов к мониторингу микроклимата и диагностике неисправностей систем отопления, вентиляции и кондиционирования воздуха (HVAC) с акцентом на алгоритмы обработки данных и методы машинного обучения. Разработана и реализована экспериментальная модель мониторинга микроклимата в жилых и нежилых зданиях на основе многопараметрических сенсорных модулей. Мониторинг осуществлялся по одиннадцати

параметрам, включая температуру, относительную влажность, концентрацию CO<sub>2</sub>, уровень летучих органических соединений (TVOC), освещённость и потребляемую мощность, в соответствии с требованиями стандартов ASHRAE и IEA.

Собранные данные прошли предварительную обработку и очистку от выбросов методом Z-score, что повысило точность последующего анализа. Для снижения размерности и упрощения визуализации применён метод главных компонент (PCA), позволивший сохранить информативность признаков.

PCA показал, что первые четыре компонента полностью объясняют вариацию данных; при этом в нежилых зданиях основная часть дисперсии сосредоточена в первой компоненте (40,08%), а в жилых зданиях распределение вариации более равномерное, что отражает различия в структуре микроклимата этих типов зданий.

Проведено сравнение алгоритмов кластеризации DBSCAN и K-Means для выявления аномалий в зашумлённых немаркированных данных, в результате чего установлено преимущество DBSCAN при анализе сложных многомерных наборов с шумовыми точками. Методология CRISP-DM была адаптирована для решения задач обнаружения неисправностей в шумных и немаркированных данных с минимальной предварительной обработкой, а также с возможностью применения в режиме реального времени.

Впервые проведена количественная оценка надёжности систем микроклимата на основе реальных эксплуатационных данных. Установлено, что среднее время безотказной работы (MTBF) в жилых зданиях составляет около 103 часов, с вероятностью безотказной работы в течение суток  $R(24) \approx 79,2 \%$ , в то время как в нежилых зданиях данные показатели существенно ниже —  $MTBF \approx 45$  ч и  $R(24) \approx 58,6 \%$ . Полученные результаты свидетельствуют о более высокой надёжности систем в жилом секторе и подчёркивают необходимость оптимизации эксплуатации микроклиматических систем в нежилом фонде.

Применение алгоритма DBSCAN как одного из наиболее подходящих решений для выявления аномалий в неразмеченных данных микроклимата. Экспериментально подтверждена эффективность выбранных параметров ( $\epsilon = 1.1$ ,  $\min\_samples = 5$ ) для задач диагностики в условиях сходной размерности и структуры признакового пространства.

Проведённый анализ показал, что предложенный подход обеспечивает визуально интерпретируемую кластеризацию, позволяет выявлять отклонения в работе систем микроклимата, и может быть интегрирован в системы автоматического управления.

Таким образом, цели и задачи, поставленные в диссертационной работе, достигнуты, а полученные научные результаты обладают как теоретической значимостью, так и высоким прикладным потенциалом.

Визуальная оценка применяется в связи с отсутствием стандартных метрик точности для алгоритмов кластеризации. Результаты работы DBSCAN оцениваются посредством графического отображения

сформированных кластеров и выявленных выбросов (аутлайеров), что позволяет наглядно судить о качестве разделения данных и выявлении аномалий в системе микроклимата.

**В первой главе** рассматриваются теоретические и практические аспекты управления микроклиматом в зданиях. Описываются ключевые понятия, значение микроклимата для комфорта, а также основные проблемы, возникающие при его контроле. Приведён обзор современных систем управления зданием (BMS).

**Во второй главе** обоснован выбор методологии CRISP-DM для структурирования процесса анализа данных. Представлен обзор существующих подходов и описан процесс применения данной методологии в рамках исследования.

**В третьей главе** рассмотрены алгоритмы обнаружения и диагностики неисправностей микроклиматических систем, включая методы машинного обучения. Описаны принципы работы алгоритма DBSCAN и применение метода главных компонент (PCA) для снижения размерности данных.

**В четвёртой главе** представлена архитектура экспериментальной установки, включающая описание жилых и нежилых помещений, используемого оборудования и стратегии размещения датчиков для сбора данных.

**Пятая глава** посвящена анализу собранных данных: описаны отслеживаемые параметры микроклимата, представлены методы визуализации и анализ корреляций между переменными.

**В шестой главе** описан процесс подготовки данных для анализа, включая очистку, нормализацию и устранение пропусков, что обеспечило надёжность последующего моделирования.

**Седьмая глава** содержит основные результаты кластеризации и диагностики неисправностей с использованием алгоритмов PCA, K-means и DBSCAN. Приведены интерпретации и сравнение результатов.

**В разделе «Обсуждение»** анализируются полученные результаты, сравниваются с другими исследованиями и обсуждаются перспективы дальнейшего применения.

**В Заключении** подводятся итоги работы, формулируются основные выводы и практические рекомендации по внедрению интеллектуальных систем диагностики микроклимата зданий.

#### **Структура и объем работы.**

Диссертация включает введение, 7 основных глав, обзор литературы, описание методологических основ, представление результатов с последующим обсуждением и заключение. Кроме того, работа содержит приложение, в котором представлены дополнительные материалы, дополняющие основное содержание исследования. Диссертация содержит 30 иллюстраций и 17 таблиц. Общий объем основного текста составляет 90 страницы, не считая приложений.