Intent-Based Energy Reduction Using Stream Mining

Al-Amodi Nada[1][QCA5WD], Assendal Fatima[2][XWXRZQ], Farhat Zamam[3][HE3221] Berindza Vilmos Norman[3][NFHJBP], and
Project Supervisor: Nouar Ikrame

Faculty of Informatics, Eötvos Loránd University Budapest, Pázmány Péter stny. 1/C., 1117

Abstract. Rising global energy consumption demands innovative solutions for sustainable management, as traditional systems fail to adapt to real-time fluctuations, leading to inefficiencies. This project addresses this challenge by developing an intent-based energy reduction system using stream mining and machine learning to reduce energy consumption and detect abnormalities. The system integrates Apache Kafka and Spark for real-time data ingestion and processing, with historical data stored in InfluxDB for batch processing and model retraining. Machine learning models, including Gaussian Mixture Model, Isolation Forest, KMeans, and One-Class SVM, were employed to detect abnormal energy patterns, while Grafana dashboards provided real-time visualization of trends and results. The system effectively combines real-time monitoring with periodic retraining, offering a scalable and adaptable solution for multi-building environments. By ensuring efficient resource utilization and providing actionable insights, the system lays the groundwork for integrating renewable energy sources and predictive maintenance, contributing to smarter and more sustainable energy systems. This project highlights the potential of advanced machine learning and stream processing technologies to address critical challenges in modern energy management.

1 Introduction

Global energy consumption is increasing, creating an urgent need for sustainable energy management systems. Traditional systems often fail to adapt to real-time fluctuations in usage and environmental factors, leading to inefficiencies and energy wastage. To address this issue, this project employs stream mining and machine learning techniques to enable real-time optimization of energy consumption. By focusing on sustainable resource usage aligned with user behavior and environmental conditions, the proposed system aims to reduce energy consumption, improve efficiency, and support smarter energy management in modern buildings.

1.1 Literature Review

Recent research on energy management spans multiple areas, including predictive modeling, abnormal energy detection, sustainability, and advanced architectures. Tan et al. [1] utilized Interquartile Range (IQR) and Isolation Forest (IF) to label abnormal power consumption, achieving over 94% accuracy with Random Forest models after dataset balancing via SMOTE. Elkhoukhi et al. [2] developed a real-time occupancy detection system using stream-based machine learning models, where SAM-KNN achieved 87.06% accuracy in dynamic, resource-constrained environments. Ziran et al. [3] focused on simplicity in energy prediction, using AutoReg models to achieve over 75% accuracy with the CU-BEMS dataset. Kong et al. [4] applied K-means clustering to analyze regional energy consumption patterns, facilitating insights for optimized energy allocation. Peña et al. [5] proposed a rule-based system integrating Energy Efficiency Indicators (EEIs) to detect inefficiencies in smart buildings, achieving over 92% sensitivity.

In addition, Himeur et al. [6] presented a taxonomy of methods for abnormal energy consumption detection, highlighting challenges like dataset scarcity and strategies for scalability. Chou et al. [7] combined NNAR and ARIMA models for real-time anomaly detection, achieving up to 97% prediction accuracy in dynamic environments. Sari et al. [8] used k-NN for energy use prediction, reducing energy waste by 34.5%-45.3% and providing a user-friendly dashboard for scalability. Borgato et al. [9] introduced Energy Signature Curves (ESCs) to enable granular energy analysis and dynamic optimization. Abboud et al. [10] proposed a Hybrid Loss Function (HLF) combining RMSLE and MAE, improving prediction accuracy in varying seasonal conditions. Nam et al. [11] demonstrated the capability of GRU models for renewable energy forecasting, addressing erratic energy patterns. Lastly, Sun et al. [12] introduced Re-Stream, an energy-efficient resource scheduling framework using DAGs and heuristic optimization to reduce energy consumption in stream computing environments. Together, these studies inform the design of a data-driven, real-time energy management system capable of adapting to complex energy usage patterns and environmental variables.

2 Dataset Description

The **CU-BEMS Dataset** [13], developed by Chulalongkorn University, provides a comprehensive collection of energy consumption and environmental data from a seven-story academic building in Bangkok, Thailand. Spanning 18 months (mid-2018 to end-2019), this dataset is structured to support research in energy load forecasting, detection of excessive energy usage, and strategies for reducing energy consumption in real-time building management.

2.1 Key Dataset Features

- **Energy Consumption**: The dataset records minute-by-minute energy usage for three primary load types across 33 building zones, providing valuable insights for optimizing energy use. For air conditioning (AC), data from 55 individual units captures cooling load variations throughout the building, enabling targeted cooling adjustments to minimize unnecessary energy consumption. Lighting data tracks power consumption across different floors and zones, offering opportunities to create optimized schedules that reduce energy wastage. Additionally, plug load data reflects general device usage patterns across zones, allowing for the identification of areas where excess consumption can be curtailed. These detailed metrics are essential for understanding energy use and identifying optimization opportunities within the building.
- **Indoor Environmental Conditions**: The dataset also includes: *Temperature (°C)*, *Relative Humidity (%)*, and *Ambient Light (lux)* readings from multiple sensors. These environmental metrics are essential for understanding how indoor conditions affect energy needs, enabling predictive models to balance occupant comfort with energy savings.

2.2 Applications of the CU-BEMS Dataset

1. **Energy Usage Optimization**: By analyzing consumption patterns across different building zones and load types, targeted control measures can be implemented—such as adjusting air conditioning and lighting in response to occupancy. This supports our project's aim of reducing energy consumption by aligning energy supply with user intents and real-time data.
2. **Detection of Excessive Energy Consumption**: The high-resolution data allows for identifying unexpected or excessive energy spikes, helping to address instances of unnecessary consumption. This application is directly aligned with our project's goal of detecting and mitigating instances of unjustified energy use.
3. **Predictive Energy Reduction**: By incorporating environmental data (temperature, humidity, ambient light), predictive models can anticipate energy needs and recommend adjustments without impacting comfort. This supports our intent-based approach by forecasting optimal energy usage and dynamically adjusting supply.

3 System Architecture

Our project's system architecture focuses on intent-based energy reduction using stream mining, structured into two main phases: batch processing and stream processing. The architecture operates within a Dockerized environment, ensuring modularity and scalability for real-time data processing and analysis.

3.1 Phase 1: Batch Processing

The batch processing phase establishes the foundation for real-time abnormal energy detection by preparing the data and training models:

Pre-processing Stage : The CU-BEMS dataset is cleaned, normalized, and formatted to ensure consistency. Feature engineering derives key metrics such as energy consumption per building zone and the impact of environmental conditions.

Modeling Stage : Four machine learning models are trained to identify abnormal energy consumption patterns:

- **Gaussian Mixture Model** for probabilistic clustering and abnormal energy detection.
- **Isolation Forest** for identifying abnormalities based on isolation distances.
- **K-Means Clustering** for grouping energy consumption patterns and detecting outliers.
- **One-Class SVM** for distinguishing normal from abnormal energy usage based on high-dimensional feature space.

The trained models are saved for deployment in the stream processing phase.

3.2 Phase 2: Stream Processing

Real-time data ingestion, abnormal energy detection, and visualization are implemented in this phase:

Data Ingestion: Energy consumption and environmental metrics are streamed into Apache Kafka, leveraging its high-throughput and fault-tolerant capabilities.

Real-Time Detection: Pre-trained models are deployed in Apache Spark to process incoming Kafka streams in real time. Each of the four models (Gaussian Mixture Model, Isolation Forest, K-Means, and One-Class SVM) is applied to detect abnormal energy consumption. This multi-model approach ensures robust detection of varied patterns of abnormal usage, enhancing the system's adaptability to dynamic environments. Detected abnormal energy patterns are published to a separate Kafka topic for further analysis.

Data Storage and Visualization: Processed data and detected abnormal energy patterns are stored in InfluxDB, a time-series database tailored for energy management. Grafana visualizes this data through dynamic dashboards, enabling real-time monitoring of energy trends, correlations, and patterns of abnormal energy usage.

Tool	**Description**
Pandas, NumPy	Used for data loading, cleaning, and normalization of energy consumption data.
Scikit-Learn, TensorFlow	Used for building and training anomaly detection models.
Apache Kafka	Streams real-time data from energy meters for processing.
Apache Spark	Processes real-time data from Kafka for anomaly detection.
InfluxDB	Stores detected anomalies and analysis results in a time-series format.
Grafana	Visualizes real-time energy consumption and detected anomalies.

Table 1. List of Tools with Descriptions used in the Project

4 Experiments

4.1 Stream Processing

We implemented real-time abnormal energy detection using streaming data from the **CU-BEMS dataset**. The system leverages **Apache Kafka** for data ingestion and **Apache Spark** for processing. The streaming data is preprocessed in Spark, where features are normalized, and abnormal energy consumption patterns are identified in real-time using machine learning models.

The key steps involve consuming minute-by-minute energy data from Kafka topics, normalizing and transforming features into a standardized format, applying machine learning models to detect deviations

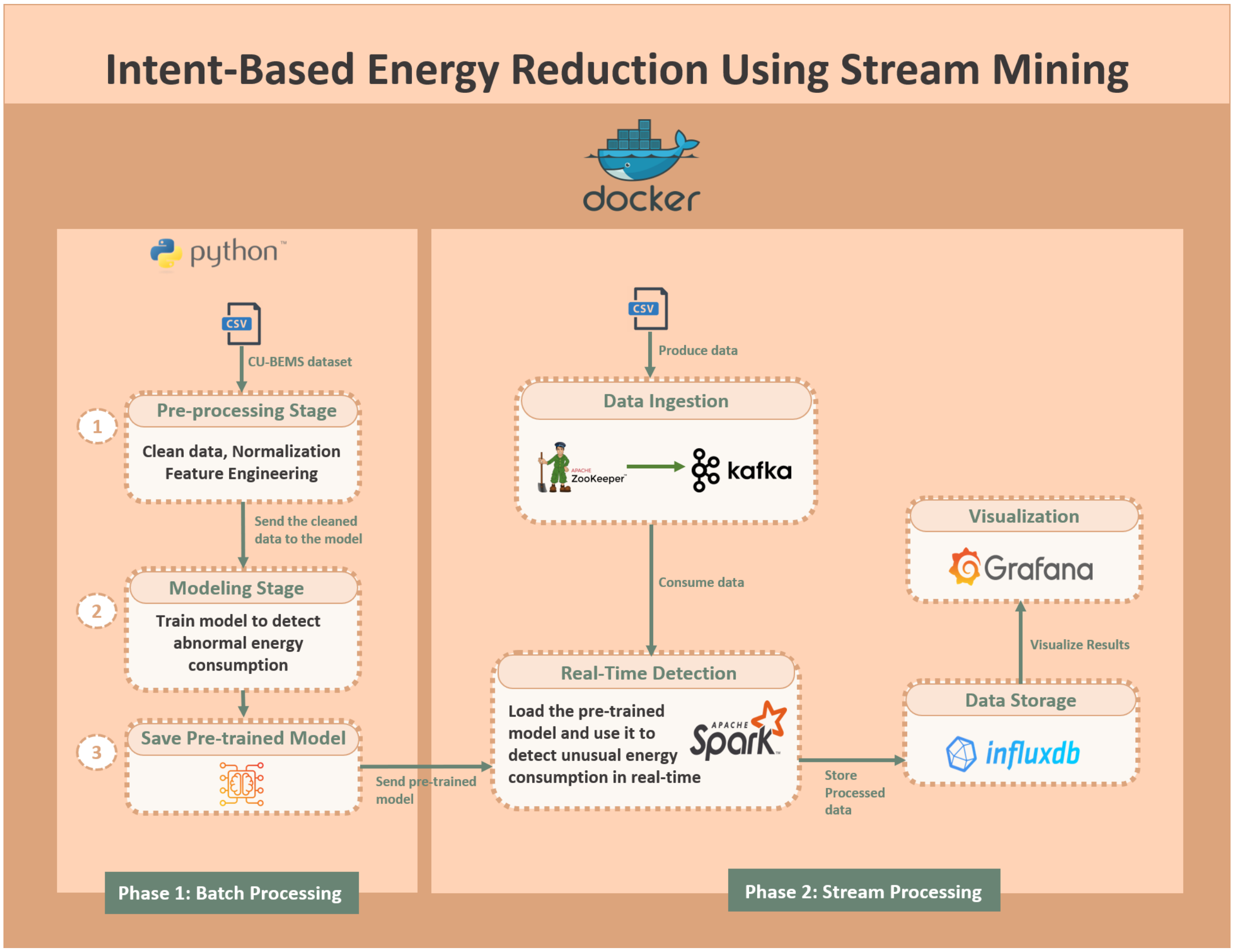

Fig. 1. System Architecture

indicative of abnormal energy consumption, and sending the detected abnormalities to InfluxDB for storage and analysis. This pipeline ensures efficient and timely monitoring of energy consumption trends, enabling the identification of anomalies as they occur.

We implemented and evaluated four different algorithms for abnormal energy detection:

1. **Gaussian Mixture Model (GMM):** A probabilistic model that clusters data points by assuming they are generated from a mixture of Gaussian distributions. It is effective in capturing abnormalities based on deviations from expected cluster distributions but may struggle with high-dimensional data.
2. **Isolation Forest:** Detects sparse, distant outliers by isolating abnormal instances. It is well-suited for detecting isolated spikes in energy consumption but may miss subtle patterns.
3. **K-Means Clustering:** Groups data into clusters based on proximity, with outliers identified as points far from cluster centers. It performs well for distinct, distant abnormalities but struggles with complex patterns and overlapping clusters.
4. **One-Class SVM:** Learns the boundary of normal data and identifies deviations as anomalies. It is capable of detecting both sparse and subtle anomalies but is sensitive to hyperparameter tuning and less effective in high-dimensional datasets.

The results of each model, including detected abnormalities and related statistics (e.g., false positives and detection delays), were stored in **InfluxDB** and visualized in **Grafana**.

4.2 Batch Processing and Training

Batch training was implemented using historical data stored in **InfluxDB**, allowing models to learn from fixed datasets and identify historical energy consumption patterns. Data was retrieved using time-range queries, then split into training (80%) and testing (20%) sets. Models such as Random Forest and XGBoost were trained and evaluated, achieving accuracies of 92% and 90%, respectively. This offline batch learning approach enables periodic retraining and ensures efficient use of historical data while addressing the challenges of detecting complex patterns in energy consumption.

5 Results

5.1 Performance Analysis of Abnormal Energy Detection Models

The figures below illustrate the performance of various models in detecting abnormalities within energy consumption data. Each model demonstrates unique detection capabilities, as described:

In Figure 2, the Isolation Forest model highlights abnormalities as red dots overlaid on the blue line representing energy consumption trends. This model excels at isolating sparse, distant outliers, making it effective for detecting significant deviations. However, it may fail to capture subtle, contextual anomalies due to its focus on sparse points.

In Figure 3, the Gaussian Mixture Model detects abnormalities as red dots in regions of unexpected relationships between features. This model captures complex and non-linear patterns effectively, demonstrating its strength in identifying nuanced anomalies and adapting to intricate energy consumption trends.

In Figure 4, the One-Class SVM model uses boundary learning to differentiate normal instances from anomalies. Red dots systematically highlight deviations, showcasing its ability to detect both sparse and subtle abnormalities. However, its performance depends heavily on hyperparameter tuning and may be less effective in high-dimensional datasets.

In Figure 5, the K-Means model groups data into clusters based on proximity, with red dots marking outliers far from cluster centers. This visualization demonstrates its effectiveness in identifying clear, distant anomalies. However, it struggles to detect nuanced patterns or subtle deviations not aligned with cluster distributions.

Among the models, the Gaussian Mixture Model stands out as the most effective, given its adaptability to complex, non-linear relationships in the data. Its ability to capture both subtle and significant abnormalities makes it well-suited for dynamic energy consumption datasets.

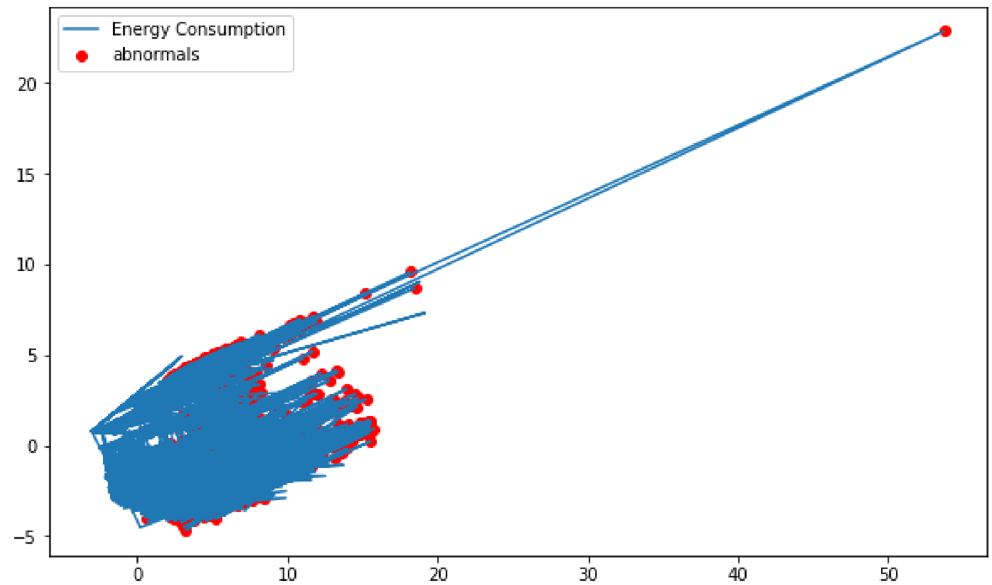

Fig. 2. Isolation Forest Model

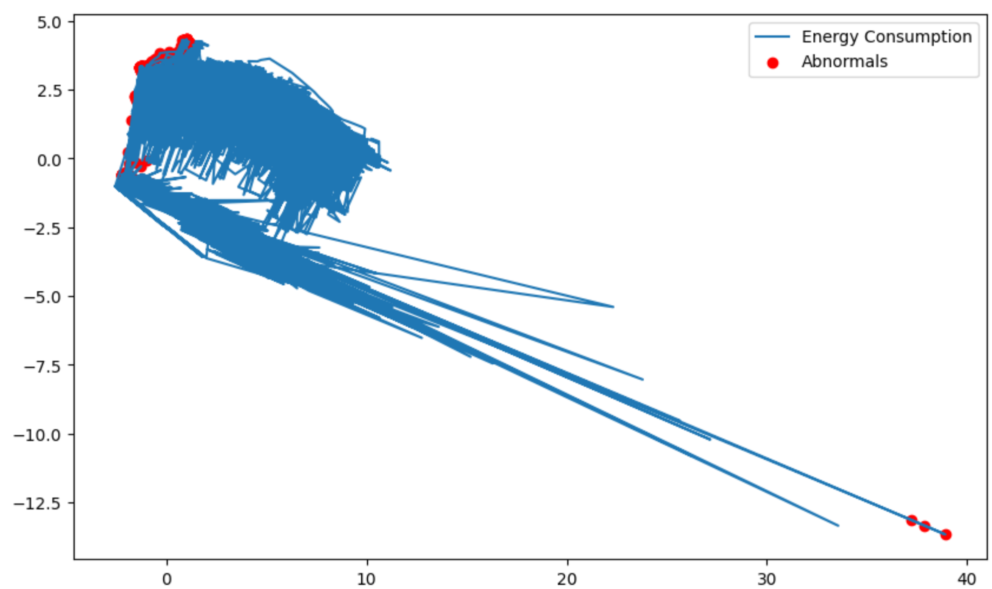

Fig. 3. Gaussian Mixture Model

5.2 Dashboards

The dashboard in Figure 6, created using Grafana, provides a detailed overview of energy consumption and abnormalities across zones. Line graphs display total energy trends, while gauges compare normal and abnormal usage for Zone 1. A stacked bar chart highlights zone-specific abnormal consumption, supported by metrics on average light usage, humidity, and light intensity. Additionally, line plots track air conditioning

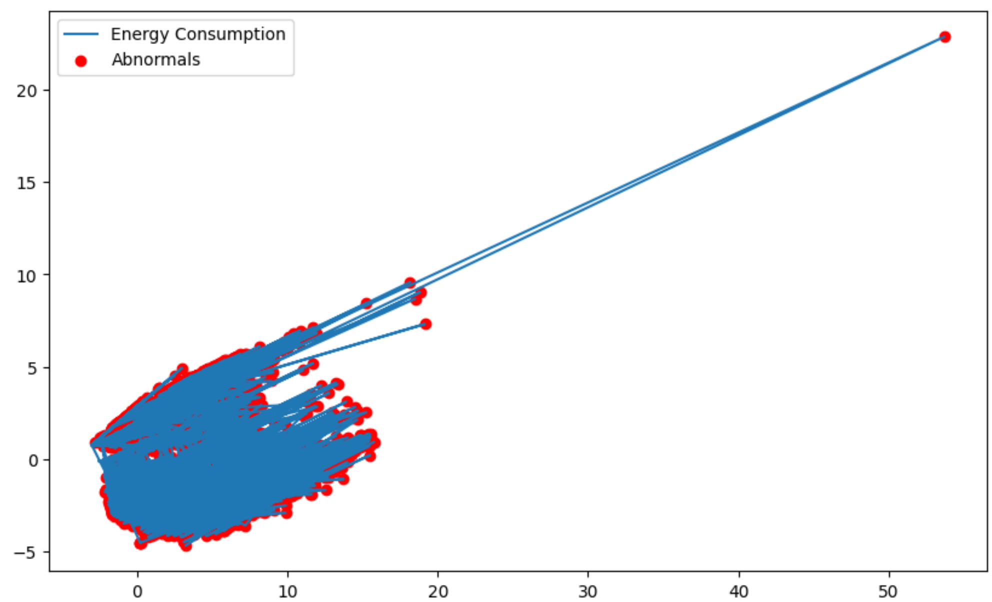

Fig. 4. One Class SVM Model

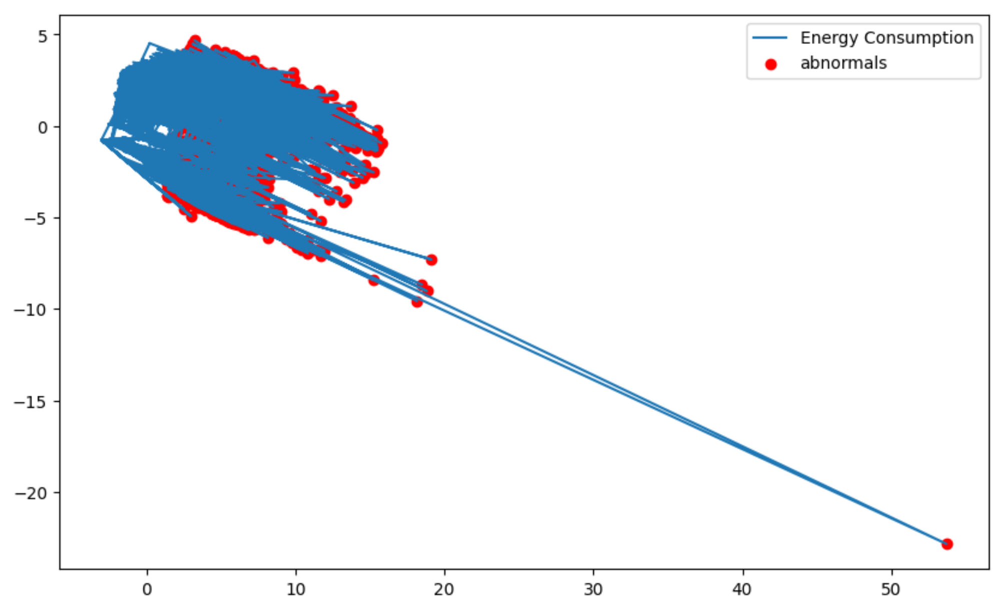

Fig. 5. KMeans Model

energy trends, and seasonal comparisons reveal normal vs. abnormal usage patterns. This Grafana-based visualization offers actionable insights for optimizing energy efficiency and detecting abnormalities.

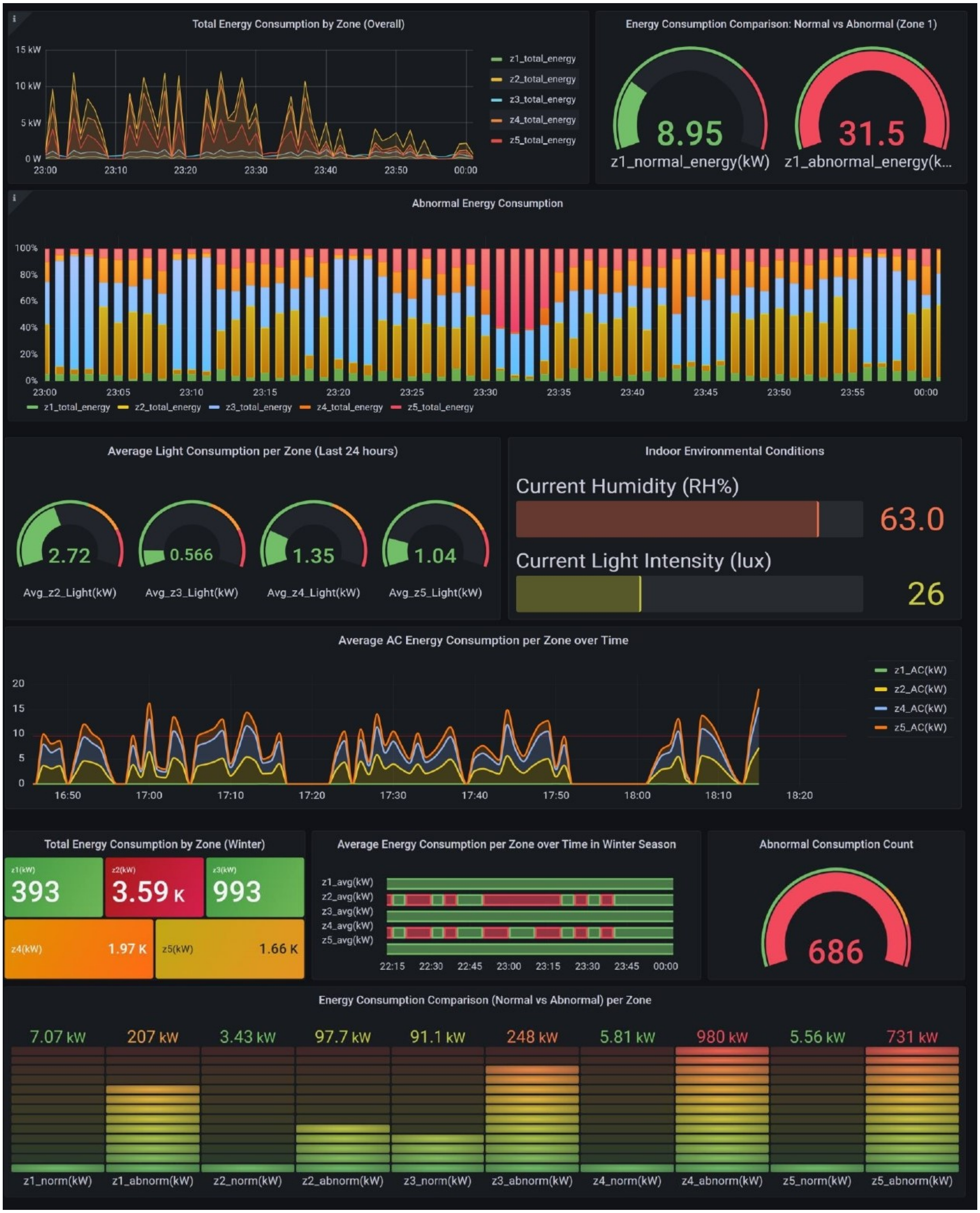

Fig. 6. Grafana Dashboard

6 Future Work

Moving forward, the system's adaptability can be enhanced by incorporating advanced machine learning algorithms, such as deep learning models, to better capture complex patterns in energy consumption. Additionally, integrating predictive maintenance and renewable energy data will improve the system's robustness. Efforts will also be made to scale the solution for deployment in multi-building environments while improving data visualization and reporting for actionable insights.

7 Conclusion

In conclusion, our project successfully establishes a robust and scalable framework for real-time energy consumption analysis and abnormal detection using the CU-BEMS dataset. Leveraging Apache Kafka, Spark, InfluxDB, and Grafana, we have created an integrated environment capable of processing streaming data, implementing machine learning models such as Gaussian Mixture Model, Isolation Forest, K-Means, and One-Class SVM, and providing actionable insights through visualizations. This work demonstrates the potential of combining stream mining and machine learning for sustainable energy reduction, highlighting the importance of real-time monitoring and adaptive analysis in advancing energy efficiency and smart infrastructure.

References

1. Tan, Y. F., Zhao, G. Z., Ooi, C. P., Tan, W. H.: Leveraging interquartile range and isolation forest for abnormal power consumption prediction. In: 2024 IEEE 6th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), vol. 6, pp. 815–819. IEEE, May 2024.
2. Elkhoukhi, H., Bakhouya, M., El Ouadghiri, D., Hanifi, M.: Using stream data processing for real-time occupancy detection in smart buildings. In: Editor, F., Editor, S. (eds.) CONFERENCE 2022, LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2022). https://doi.org/10.3390/s22062371
3. Ziran, Z., Mecella, M., Leotta, F.: A Simplified and Sustainable Approach for Energy Prediction. In: Editor, F., Editor, S. (eds.) Intelligent Environments 2024, LNCS, vol. 9999, pp. 104–113. Springer, Heidelberg (2024). https://doi.org/10.3233/978-1-61499-927-0-104
4. Kong, W., Wang, Y., Dai, H., Zhao, L., Wang, C.: Analysis of energy consumption structure based on K-means clustering algorithm. In: Editor, F., Editor, S. (eds.) E3S Web Conf., vol. 267, pp. 01054. Springer, Heidelberg (2021). https://doi.org/10.1051/e3sconf/202126701054.
5. Peña, M., Biscarri, F., Guerrero, J.I., Monedero, I., León, C.: Rule-Based System to Detect Energy Efficiency Abnormalities in Smart Buildings: A Data Mining Approach. In: Editor, F., Editor, S. (eds.) Expert Systems with Applications 2016, LNCS, vol. 56, pp. 242–255. Springer, Heidelberg (2016). https://doi.org/10.1016/j.eswa.2016.03.002
6. Himeur, Y., Alsalemi, A., Bensaali, F., Amira, A.: Detection of Energy Consumption in Buildings: A Review, Current Trends, and New Perspectives. In: Editor, F., Editor, S. (eds.) ArXiv 2020, LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2020). https://doi.org/arXiv:2010.04560
7. Chou, J. S., Telaga, A. S.: Real-time Detection of Abnormal Power Consumption. Renewable and Sustainable Energy Reviews, vol. 33, pp. 400–411 (2014).
8. Sari, M., Berawi, M. A., Zagloel, T. Y., Madyaningarum, N., Miraj, P., Pranoto, A. R., Susantono, B., Woodhead, R.: Machine Learning-Based Energy Use Prediction for the Smart Building Energy Management System. Journal of Information Technology in Construction, vol. 28 (2023).
9. Borgato, N., Prataviera, E., Bordignon, S., Garay-Martinez, R., Zarrella, A.: A Data-Driven Model for the Analysis of Energy Consumption in Buildings. In: E3S Web of Conferences, vol. 523, p. 02002. EDP Sciences (2024).
10. Abboud, A., Abouaissa, A., Shahin, A., Mazraani, R.: Empowering Energy Consumption Forecasting in Smart Buildings: Towards a Hybrid Loss Function. In: 2024 International Wireless Communications and Mobile Computing (IWCMC), pp. 1118–1123. IEEE, May 2024.
11. Nam, K., Hwangbo, S., Yoo, C.: A deep learning-based forecasting model for renewable energy scenarios to guide sustainable energy policy: A case study of Korea. In: Editor, F., Editor, S. (eds.) Renewable and Sustainable Energy Reviews, vol. 122, pp. 109725. Springer, Heidelberg (2020). https://doi.org/10.1016/j.rser.2020.109725
12. Sun, D., Zhang, G., Yang, S., Zheng, W., Khan, S.U., Li, K.: Re-Stream: Real-time and energy-efficient resource scheduling in big data stream computing environments. In: Editor, F., Editor, S. (eds.) Information Sciences, vol. 319, pp. 92–112. Springer, Heidelberg (2015). https://doi.org/10.1016/j.ins.2015.03.027
13. Pipattanasomporn, M., Chitalia, G., Songsiri, J., Aswakul, C., Pora, W., Suwankawin, S., Hoonchareon, N.: CU-BEMS, smart building electricity consumption and indoor environmental sensor datasets. In: Editor, F., Editor, S. (eds.) Scientific Data 2020, LNCS, vol. 7(1), pp. 241. Springer, Heidelberg (2020). https://doi.org/10.1038/s41597-020-00604-6