Efficient Energy Disaggregation using DBSCAN: A Novel Approach for Enhanced Energy Management

Emmanouela-Electra Economopoulou[1], Gerasimos Vonitsanos[1], Phivos Mylonas[2] and Andreas Kanavos[3]

1. Computer Engineering and Informatics Department
University of Patras, Patras, Greece
st1057466@ceid.upatras.gr, mvonitsanos@ceid.upatras.gr
2. Department of Informatics and Computer Engineering
University of West Attica, Athens, Greece
mylonasf@uniwa.gr
3. Department of Informatics
Ionian University, Corfu, Greece
akanavos@ionio.gr

Abstract. In the rapidly evolving technology landscape, smart homes are becoming increasingly common, driven by the demand for integrated management of information and services. However, despite advancements, managing electricity consumption efficiently remains a significant challenge, primarily due to the lack of detailed usage data and the complexity of predicting device behavior. This study addresses these challenges by utilizing the DinRail Cerberus meter for granular data collection on household electricity use and applying the DBSCAN clustering algorithm for unsupervised learning. Our research aims to develop a forecasting system that accurately discerns the operational status of household devices—active or inactive—based on energy consumption patterns. This innovative approach promises to revolutionize energy management in smart homes, offering detailed insights into device usage that facilitate more informed decisions for efficient electricity consumption.

Keywords: Apache Spark, Clustering, Consumption, DBSCAN, DinRail Cerberus, Energy, Machine Learning

1 Introduction

The escalating concern over electricity overconsumption and the energy footprint of individuals has underscored the urgency for innovative solutions. Until recently, the granularity required to assess electricity usage, detect underperforming household appliances, or understand a household's energy distribution was unattainable. However, advancements in technology and the pressing need for sustainable energy management have catalyzed the development of specialized meters, such as the DinRail Cerberus meter. These devices not only measure

a household's total energy consumption but also equip consumers with detailed insights into their energy load [22].

In the rapidly evolving technology landscape, efficient management of electricity consumption remains a significant challenge. This research aims to bridge the gap in energy management by leveraging the DinRail Cerberus meter, a cutting-edge tool for collecting granular electricity usage data. By applying machine learning techniques, specifically the DBSCAN clustering algorithm, we introduce an innovative approach to analyze energy data for identifying the operational status of household appliances, employing Apache Spark to handle large datasets effectively.

The utility of these hardware solutions hinges on the processing and interpretation of the data they collect. Machine learning algorithms emerge as pivotal tools, capable of extracting meaningful patterns and predictions from energy consumption data, irrespective of its dimensionality or labeling [28,34]. This approach offers the potential for broad implementation across various household contexts, enhancing energy efficiency at a granular level [17,19,40].

Machine learning enables devices to emulate human behavior without explicit programming, with algorithms and statistical methods that forecast future outcomes [1,36]. It covers supervised learning, using labeled data for precise predictions, and unsupervised learning, revealing patterns in unlabeled data [29,32,37].

Apache Spark, a potent framework for data processing and analysis, supports the execution of these complex algorithms [33]. Since its inception at UC Berkeley in 2009, Spark has evolved into a comprehensive analytics engine, ideal for managing 'Big Data' and executing machine learning tasks. Its suite of libraries, including Spark SQL, Spark Streaming, MLlib, and GraphX, underscores its versatility and power in data analytics [15,21,35].

This research integrates the DinRail Cerberus meter's hardware capabilities [20] with advanced machine learning algorithms [23] to mine and interpret energy consumption data effectively. This research makes several notable contributions to the field of energy management and machine learning. First, it demonstrates the effective integration of the DinRail Cerberus meter with machine learning algorithms for detailed energy consumption analysis. Second, it employs the DBSCAN clustering algorithm within an unsupervised learning framework to distinguish between active and inactive states of household appliances, a novel approach in energy disaggregation. Lastly, it showcases the application of Apache Spark for handling and analyzing large datasets.

The paper is structured as follows: Section 2 reviews related work and delves into the theoretical background of Unsupervised Learning. Section 3 describes the System Architecture, detailing the deployment of the DinRail Cerberus meter for granular data collection and the configuration of our analysis framework, integrating Apache Spark. Section 4 discusses the implementation of the DBSCAN algorithm and the parameter optimization. The methodology is presented in Section 5, where we unveil the research findings, demonstrating the efficacy of DBSCAN in energy disaggregation and evaluating the algorithm's accuracy and scalability. Finally, Section 6 offers conclusions and discussions.

2 Related Work

Effective electricity management, particularly during peak demand periods, has become a critical challenge. The adoption of time-of-use rate plans by power suppliers highlights the need for real-time power consumption data access for consumers. In this context, Internet of Things (IoT) technologies and smart sockets have been identified as potential solutions, offering aggregated electricity consumption data [25]. However, these approaches often lack the granularity needed for individual appliance monitoring. A notable study addressed this gap by enhancing smart socket functionalities to include appliance recognition capabilities, utilizing a recursive DBSCAN approach for the identification of power consumption patterns of individual appliances without prior knowledge of their characteristics [5]. This advancement underscores the potential of integrating more sophisticated machine learning techniques for detailed energy usage analysis [39].

Another significant application of the DBSCAN algorithm is in Building Energy Consumption Anomaly Detection (BECAD), crucial for green building assessments [18]. The challenge of parameter setting in DBSCAN was addressed by introducing an adaptive parameter adjustment method, enabling the algorithm to determine the MinPts and ε parameters based on data distribution characteristics [31]. This approach facilitated improved clustering performance and was effectively applied in BECAD, aiding in the identification of energy utilization patterns and anomalies in buildings, thus offering valuable insights for energy management and conservation strategies [38].

The Swiss residential sector presents a case study where building-related heating represents a major portion of final energy consumption. The challenge of disaggregating heating energy from overall electrical loads in residences equipped with electrical resistance heating systems was tackled using a data mining approach employing DBSCAN. This method proved effective in isolating space and domestic hot water heating consumption, demonstrating the versatility of clustering algorithms in addressing specific energy management challenges [6].

Beyond traditional machine learning algorithms, deep learning offers a promising avenue for energy disaggregation by eliminating the need for extensive preprocessing [30,12]. This is particularly advantageous in handling unstructured data, streamlining the analysis process without compromising accuracy. Neural networks, especially Recurrent Neural Networks (RNN) for time series data and Convolutional Neural Networks (CNN) for image data, exemplify the application of deep learning in simulating complex decision-making processes akin to the human brain [11,14].

Furthermore, the application of basic machine learning techniques, such as k-Nearest Neighbors [13], Naive Bayes [2], Decision Trees [7], Random Forest [9,26], and Support Vector Machines [9], has been explored in the context of energy management and disaggregation. These studies highlight the broad spectrum of methodologies that can be employed to tackle various aspects of energy monitoring, efficiency improvement, and predictive analysis, underscoring the dynamic nature of research in this field.

2.1 Unsupervised Learning

The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) stands out as a premier algorithm within the domain of clustering algorithms [8], itself a subset of the broader field of Unsupervised Learning. Unlike Partition-Based and Hierarchical Clustering algorithms, which excel with spherical or convex cluster shapes, DBSCAN thrives in environments where cluster shapes are irregular or datasets contain outliers [16]. Introduced by Martin Ester in 1996 [10], DBSCAN revolutionized clustering methodologies by employing a density-based approach, effectively distinguishing clusters based on varying density regions. This algorithm's resilience to outliers and its independence from prior cluster quantity knowledge render it particularly adept for analyzing large spatial datasets, assessing local data point density.

Central to DBSCAN's methodology are two parameters: eps (Epsilon) and MinPoints, both critical to the algorithm's operational efficiency and outcome precision. The Epsilon parameter delineates the radius surrounding each data point, serving as a metric for neighborhood density [4,27]. Its optimal value, delicate to determine, directly influences the algorithm's sensitivity to outliers and its ability to discern distinct clusters. Conversely, MinPoints establishes the minimum requisite data points within an Epsilon-defined radius, a value that inherently scales with dataset size [27]. The formula for determining MinPoints, considering the dataset's dimensionality (D), ensures a baseline for cluster density and integrity:

$$minPoints \geq D + 1 \tag{1}$$

DBSCAN classifies data points into three categories based on these parameters: Core Points, Border Points, and Noise (or Outliers). Core Points have at least MinPoints within their Epsilon radius, Border Points are adjacent to Core Points but lack sufficient local density, and Noise points are isolated from dense areas, underscoring the algorithm's nuanced approach to cluster analysis.

2.2 Application in Energy Management

The adaptability of DBSCAN to detect varied cluster shapes makes it particularly suitable for energy consumption data analysis. In the context of energy disaggregation and anomaly detection, DBSCAN's ability to identify outliers and pattern irregularities offers valuable insights into abnormal energy usage or inefficient appliance performance. By applying DBSCAN, this research aims to elucidate distinct energy consumption patterns within households, enabling the identification of opportunities for energy optimization and contributing to the broader goal of sustainable energy management.

Through this exploration of DBSCAN within unsupervised learning, we underscore its instrumental role in our research methodology. Its precision in clustering irregular data shapes and discerning outliers aligns with our objectives of enhancing energy disaggregation and management practices, promising to bring forth novel insights and improvements in the domain of smart energy solutions.

3 System Architecture

The foundation of our research is a comprehensive dataset collected from the DinRail Cerberus meter by Meazon S.A., which has also supported a thesis contributing to this study's development. The DinRail Cerberus meter [20], capable of measuring electrical characteristics in both single-phase and three-phase power supplies, plays a pivotal role in our data acquisition process. Equipped with two Solid-Core current transformers rated up to 600 Amperes per phase, it offers precision in recording electrical current characteristics. For our research, this meter was specifically utilized within a three-phase power supply context, operating with a 20 ms sampling period to ensure high accuracy. Data packets gathered are then systematically stored within our database.

Installation of the DinRail Cerberus electric meter was carried out in the main panel of a residential house, capturing data from five household electrical appliances: vacuum cleaner, electric oven, air conditioning unit, electric stove, and iron press. The dataset encompasses 15-minute intervals of both operational and non-operational periods for each appliance, facilitating a detailed analysis of energy consumption patterns.

The dataset features, as outlined in the following Table 1, encompass a range of electrical measurements:

Table 1. Features of the Dataset

Active Power	Angle	Apparent Power
Crest Factor	Current	Old Active Power
Old Apparent Power	Old Reactive Power	Output
Reactive Power	Timestamp	

For the purposes of our study, we concentrated on features directly influencing energy consumption insights: Active Power, Apparent Power, Reactive Power, Current, Angle, Crest Factor, and the Output. The 'Output' feature is particularly instrumental, indicating appliance activity with a value of 1 and inactivity with a value of 0. The measurements of Old Active Power, Old Apparent Power, and Old Reactive Power, representing readings taken 0.5 seconds prior, were excluded from analysis due to their minimal impact on the model's efficiency. Each appliance's data constitutes a separate dataset, aggregating to approximately 285,000 records, enabling a granular investigation into energy consumption patterns.

This meticulously curated dataset not only facilitates the detailed examination of individual appliance energy usage but also underscores the significance of leveraging high-resolution electrical data to enhance the accuracy of machine learning models. By strategically selecting features that contribute meaningfully to our analysis, we aim to advance the understanding and management of household energy consumption, aligning with our research goals of developing efficient energy disaggregation techniques through unsupervised learning algorithms.

4 Implementation

4.1 Hardware Specifications

The computational framework for this research was established on a high-performance computer, ensuring efficient processing of complex machine learning algorithms and large datasets. The system specifications include an Intel® Core™ i7-7700HQ CPU @ 2.8GHz, NVIDIA GeForce GTZ 1060 6 GB GPU, 32 GB of RAM, and dual 1TB hard drives. This setup provided the necessary computational power for data processing, model training, and analysis.

4.2 Software Tools

The research utilized Python 3.9.72 within the Jupyter Notebook 6.4.5 environment, selected for its support of an interactive computing and development process. Key Python libraries employed in this study include:

- **NumPy 1.23.3** and **Pandas 1.5.0** for data manipulation and numerical computations. These libraries facilitated efficient handling and preprocessing of the dataset.
- **Scikit-learn 1.1.2** for applying machine learning algorithms, particularly the DBSCAN clustering algorithm and utilities for model evaluation [24].
- **Matplotlib 3.6.0** and **Seaborn 0.12.0** for data visualization, enabling the graphical representation of the data analysis and clustering results [3].

These software tools were integral to the research, supporting various stages of the implementation from data processing to analysis and visualization.

4.3 DBSCAN Implementation

The core aim of employing the DBSCAN algorithm was to effectively categorize appliance usage data into two principal clusters based on operational status, utilizing the hyperparameters Epsilon and minPoints for this purpose. These clusters distinguish between periods when the device is active versus inactive, providing a nuanced understanding of energy consumption patterns.

4.3.1 Data Preparation and Feature Selection For each of the five appliances—air conditioner, electric oven, iron press, electric stove, and vacuum cleaner—datasets comprised approximately 300,000 records, reflecting varied energy usage profiles. To manage this extensive data, a sampling strategy was implemented, selecting 90,000 records per appliance. This process ensured that the sample maintained statistical parity with the original dataset, particularly in terms of mean and standard deviation, thus allowing for extrapolation of findings.

Subsequent to sampling, non-contributory features such as timestamps and historical power readings were excluded to streamline the analysis. This filtration resulted in a focused dataset, aptly named "dataset", prepared specifically for DBSCAN clustering. This dataset retained only relevant features while incorporating a "new_output" column to document the clustering outcome.

4.3.2 Optimization of Clustering Parameters The optimization process commenced with the application of the Elbow Plot technique to determine the ideal Epsilon value for each dataset. By assessing the average distance between each data point and its k-nearest neighbors, this method facilitated the identification of a significant bend or "elbow" in the plot, indicative of the optimal Epsilon value for clustering.

An iterative approach was then applied to fine-tune both Epsilon and minPoints values. The range of exploration spanned from 350 to 1,000 for Epsilon and from 1,000 to 30,000 for minPoints. This exhaustive search aimed to strike a balance between cluster integrity and the inclusiveness of data points, ensuring the clusters formed were both meaningful and representative of the underlying data.

4.3.3 Evaluation and Parameter Selection The meticulous parameter tuning yielded distinct optimal values for Epsilon and minPoints, as summarized in the following Table 2:

Table 2. Optimal DBSCAN Parameters for Each Appliance

Appliance	Epsilon	minPoints
Air Conditioner	720	20,500
Electric Oven	720	20,500
Electric Stove	740	20,500
Iron Press	746	20,500
Vacuum Cleaner	746	20,500

Notably, these optimal parameters reflect a deliberate calibration to accommodate the unique characteristics of each appliance dataset. The consistent value across appliances for minPoints underscores a shared threshold for cluster density, while the slight variations in Epsilon values cater to the distinct spatial distributions inherent to each dataset.

In complement to hyperparameter optimization, the Manhattan distance metric was chosen over the Euclidean, especially given the increased dimensionality of the data. This metric, calculating the sum of the absolute differences between coordinates of a pair of points, proved more suitable for the datasets at hand, aligning with established best practices for high-dimensional data clustering.

The culmination of this implementation phase involved the application of the DBSCAN algorithm with the optimized parameters, resulting in the classification of data points into operational clusters. These clusters were then meticulously compared against the original dataset ("test") to assess the model's accuracy, taking into account both the presence and absence of outliers. This dual-faceted evaluation strategy provided a comprehensive understanding of the algorithm's effectiveness in discerning between operational states of the appliances, setting the stage for in-depth results analysis.

5 Experimental Evaluation

This section presents the accuracy measures of the DBSCAN algorithm for classifying operational states of various electrical appliances. The accuracy is assessed both with and without considering outlier records. Detailed results for each appliance are presented in the following Table 3:

Table 3. Accuracy Measures and Outlier Analysis for Each Appliance

Appliance	Accuracy with Outliers	Accuracy without Outliers	Outliers	OFF Cluster	ON Cluster
Air Condition	44.88%	78.62%	38,632	29,141	22,227
Electric Oven	77.56%	77.60%	47	65,872	24,081
Electric Stove	36.14%	63.07%	38,438	29,351	22,211
Iron Press	42.16%	74.12%	38,805	28,547	22,648
Vacuum Cleaner	48.28%	66.08%	24,240	29,172	36,588

The application of the DBSCAN algorithm to classify appliance usage into operational states highlighted the profound impact of outliers on the clustering accuracy, as evidenced by the varied results across different appliances. For the air conditioner, a notable discrepancy in accuracy was observed when outliers were accounted for (44.88%) versus when they were excluded (78.62%). This significant difference is attributed to the substantial portion of the dataset deemed as outliers, approximately 43%, underscoring the critical influence of outlier management on clustering outcomes. Conversely, the electric oven presented a contrasting scenario where the presence of outliers had a negligible impact on accuracy, with figures standing at 77.56% when including outliers and marginally higher at 77.60% when excluding them. This minimal difference is due to the exceptionally low number of outliers within the dataset, suggesting an ideal condition where outlier management does not substantially alter the accuracy of clustering results.

Similarly, the iron press dataset mirrored the pattern observed in the air conditioner, with a considerable improvement in accuracy from 42.16% to 74.12% upon excluding outliers, again highlighting the detrimental effect of a large volume of outliers on clustering efficacy. The electric stove, however, presented a unique challenge where despite attempts to optimize clustering parameters, the achieved accuracy fell short of expectations, particularly when outliers were considered (36.14%). Even after excluding outliers, the accuracy obtained (63.07%) did not meet the desired threshold, indicating potential mismatches between the dataset characteristics and the algorithm's capabilities. The vacuum cleaner dataset provided an interesting insight where the number of records classified as ON exceeded those classified as OFF, a unique occurrence among the appliances analyzed. Despite this, the improvement in accuracy from 48.28% to 66.07% upon excluding outliers reaffirms the sensitivity of the DBSCAN algorithm to data quality and outliers.

These observations across different appliances emphasize the critical role of outlier management in unsupervised learning tasks like clustering. The variance in accuracy due to outliers underscores the necessity of robust preprocessing and parameter optimization tailored to the specificities of each dataset to enhance the effectiveness of clustering algorithms. Moreover, these results shed light on the potential of DBSCAN to offer valuable operational insights into appliance usage patterns, suggesting avenues for future research to explore adaptive strategies for parameter selection and outlier handling to improve clustering accuracy in energy consumption analysis.

5.1 System Optimization

The initial accuracy results from the DBSCAN algorithm, while satisfactory in some cases, presented an opportunity for enhancement. A thorough examination of the algorithm's performance revealed a common issue across datasets: a significant imbalance between the active (ON) and inactive (OFF) states. This imbalance was particularly pronounced in the datasets for air conditioners and vacuum cleaners, where approximately 84% of records indicated the appliances were turned off. The electric oven and iron press datasets exhibited a similar trend, with around 85% of records in the OFF state. The electric cooker dataset displayed a slightly more balanced scenario, with about 63% OFF records.

Recognizing the critical role that this imbalance plays in affecting the algorithm's accuracy led to the hypothesis that adjusting the dataset composition could yield better results. The proposed solution involved generating new, balanced datasets for each appliance after the initial application of the DBSCAN algorithm. This approach aimed to mitigate the skew by equally representing active and inactive states, thus addressing both the excessive presence of outliers and the disproportionate number of inactive appliance records.

Table 4 illustrates the effectiveness of this optimization strategy by comparing the accuracy metrics before and after the optimization process:

Table 4. Accuracy Improvement Through System Optimization

Appliance	Accuracy with Outliers	Accuracy without Outliers	Accuracy after Optimization
Air Condition	44.88%	78.62%	87.79%
Electric Oven	77.56%	77.60%	77.60%
Electric Stove	36.14%	63.07%	66.48%
Iron Press	42.16%	74.12%	80.64%
Vacuum Cleaner	48.28%	66.08%	75.20%

The data post-optimization reflect a clear improvement in clustering accuracy for all appliances, with increases ranging from 3.5% to 9%. This substantial enhancement validates the effectiveness of addressing dataset imbalances and

underscores the potential of targeted dataset adjustments in optimizing unsupervised learning tasks. The optimized results not only demonstrate the DBSCAN algorithm's adaptability but also highlight the importance of preparatory data handling in achieving precise clustering outcomes.

5.2 Discussion

The comprehensive evaluation of the DBSCAN algorithm's application to various electrical appliances has unveiled critical insights into the algorithm's efficacy and the pivotal role of data quality in clustering accuracy.

The profound influence of outliers on the clustering process is unmistakable. Appliances like the air conditioner and iron press exhibited a marked improvement in accuracy when outliers were excluded, underscoring the detrimental impact of outlier records on the algorithm's ability to correctly identify operational states. The disparity in accuracy—such as the air conditioner's increase from 44.88% to 78.62% upon excluding outliers—highlights the necessity of robust outlier management strategies in preprocessing stages to enhance clustering outcomes.

Conversely, the minimal impact of outliers on the electric oven's accuracy metrics suggests that certain datasets, inherently balanced or with negligible outlier presence, may not require as intensive outlier mitigation efforts. This variance across appliances underscores the need for a tailored approach to data preparation, emphasizing the uniqueness of each dataset's characteristics.

The System Optimization phase further elucidated the critical role of dataset balance in achieving high clustering accuracy. The imbalance in operational states, particularly pronounced in datasets with a predominant number of OFF records, posed a significant challenge. The strategic creation of balanced datasets post-DBSCAN application effectively addressed this challenge, leading to notable accuracy improvements across all appliances. For instance, the air conditioner's accuracy improved from 78.62% to 87.79%, showcasing the optimization strategy's efficacy.

These findings collectively underscore the complexity of applying unsupervised learning algorithms like DBSCAN to real-world datasets. They highlight the interplay between data quality, including outlier prevalence and dataset balance, and the algorithm's parameterization, including the choice of Epsilon and minPoints, in determining clustering accuracy.

6 Conclusions and Future Work

This research analyzed energy consumption data from five household appliances with significant energy demands, utilizing the DinRail Cerberus meter. The DBSCAN algorithm, a clustering tool within unsupervised machine learning, was applied to categorize operational states of these appliances based solely on their energy usage patterns. The approach proved to classify the appliances' states into active, inactive, or outlier with satisfactory accuracy for most datasets. An

optimization process was introduced to enhance this accuracy further, demonstrating the method's effectiveness across various appliances. Nonetheless, the study faced challenges with the electric cooker dataset, highlighting the algorithm's sensitivity to different appliance behaviors and the complexity of their energy consumption patterns.

Future enhancements to this work are envisioned to focus on the integration of deep learning to refine the analysis of electrical energy data from the examined appliances. A neural network could be designed to process this data, aiming for a more accurate prediction of the appliances' operational states. This approach would involve training the neural network with a substantial portion of the datasets to improve the prediction accuracy beyond the current methodology. Additionally, exploring a broader spectrum of machine learning models, both supervised and unsupervised, will be crucial in identifying the most effective model for energy consumption analysis. A novel strategy proposes merging deep learning techniques with the DBSCAN algorithm, utilizing a neural network for feature correlation and dimensionality reduction. This preparatory step would ideally enhance DBSCAN's clustering accuracy by providing it with optimized, lower-dimensional data, thus refining the state predictions for each appliance.

This research underlines the importance of leveraging advanced machine learning techniques for the analysis of household energy consumption. The findings suggest significant potential for smart energy management solutions that could lead to more sustainable and efficient energy use. Further investigations along the suggested future work directions have the potential to build upon the foundational achievements of this study, opening new avenues for advancements in smart home technologies and energy management systems. Through these efforts, the study contributes to the broader goals of enhancing energy efficiency and promoting the development of intelligent energy solutions within the context of smart homes.

References

1. Alexopoulos, A., Drakopoulos, G., Kanavos, A., Mylonas, P., Vonitsanos, G.: Two-step classification with SVD preprocessing of distributed massive datasets in apache spark. Algorithms 13(3), 71 (2020)
2. Altrabalsi, H., Stankovic, V., Liao, J., Stankovic, L.: Low-complexity energy disaggregation using appliance load modelling. Aims Energy 4(1), 884–905 (2016)
3. Bisong, E.: Matplotlib and seaborn. In: Building Machine Learning and Deep Learning Models on Google Cloud Platform. pp. 151–165. Springer (2019)
4. Campello, R.J.G.B., Moulavi, D., Sander, J.: Density-based clustering based on hierarchical density estimates. In: 17th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD). Lecture Notes in Computer Science, vol. 7819, pp. 160–172. Springer (2013)
5. Cretulescu, R.G., Morariu, D.I., Breazu, M., Volovici, D.: Dbscan algorithm for document clustering. International Journal of Advanced Statistics and IT&C for Economics and Life Sciences 9(1), 58–66 (2019)
6. Deb, C., Frei, M., Hofer, J., Schlueter, A.: Automated load disaggregation for residences with electrical resistance heating. Energy and Buildings 182, 61–74 (2019)

7. Delcroix, B., Sansregret, S., Martin, G.L., Daoud, A.: Quantile regression using gradient boosted decision trees for daily residential energy load disaggregation. In: Journal of Physics: Conference Series. vol. 2069, p. 012107 (2021)
8. Deng, D.: Dbscan clustering algorithm based on density. In: 7th International Forum on Electrical Engineering and Automation (IFEEA). pp. 949–953. IEEE (2020)
9. Elafoudi, G., Stankovic, L., Stankovic, V.: Power disaggregation of domestic smart meter readings using dynamic time warping. In: 6th International Symposium on Communications, Control and Signal Processing (ISCCSP). pp. 36–39. IEEE (2014)
10. Ester, M., Kriegel, H., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: 2nd International Conference on Knowledge Discovery and Data Mining (KDD). pp. 226–231 (1996)
11. Féraud, R., Clérot, F.: A methodology to explain neural network classification. Neural Networks 15(1), 237–246 (2002)
12. He, W., Chai, Y.: An empirical study on energy disaggregation via deep learning. In: 2nd International Conference on Artificial Intelligence and Industrial Engineering (AIIE). pp. 338–342 (2016)
13. Hidiyanto, F., Halim, A.: KNN methods with varied k, distance and training data to disaggregate NILM with similar load characteristic. In: 3rd Asia Pacific Conference on Research in Industrial and Systems Engineering (APCoRISE). pp. 93–99. ACM (2020)
14. Jiang, J., Kong, Q., Plumbley, M.D., Gilbert, N., Hoogendoorn, M., Roijers, D.M.: Deep learning-based energy disaggregation and on/off detection of household appliances. ACM Transactions on Knowledge Discovery from Data 15(3), 50:1–50:21 (2021)
15. Kanavos, A., Panagiotakopoulos, T., Vonitsanos, G., Maragoudakis, M., Kiouvrekis, Y.: Forecasting winter precipitation based on weather sensors data in apache spark. In: 12th International Conference on Information, Intelligence, Systems & Applications (IISA). pp. 1–6. IEEE (2021)
16. Kanavos, A., Vonitsanos, G., Mylonas, P.: Clustering high-dimensional social media datasets utilizing graph mining. In: IEEE International Conference on Big Data. pp. 3871–3880. IEEE (2022)
17. Kelly, J., Knottenbelt, W.J.: Neural NILM: deep neural networks applied to energy disaggregation. In: 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments (BuildSys). pp. 55–64. ACM (2015)
18. Khan, I., Capozzoli, A., Corgnati, S.P., Cerquitelli, T.: Fault detection analysis of building energy consumption using data mining techniques. Energy Procedia 42, 557–566 (2013)
19. Kolter, J.Z., Johnson, M.J.: Redd: A public data set for energy disaggregation research. In: Workshop on Data Mining Applications in Sustainability (SIGKDD). vol. 25, pp. 59–62 (2011)
20. Lykothanasi, K.K., Sioutas, S., Tsichlas, K.: Efficient large-scale machine learning techniques for rapid motif discovery in energy data streams. In: International Conference on Artificial Intelligence Applications and Innovations (AIAI). IFIP Advances in Information and Communication Technology, vol. 646, pp. 331–342. Springer (2022)
21. Meng, X., Bradley, J.K., Yavuz, B., Sparks, E.R., et al, S.V.: Mllib: Machine learning in apache spark. Journal of Machine Learning Research 17, 34:1–34:7 (2016)
22. Mosavi, A., Bahmani, A.: Energy consumption prediction using machine learning; a review. Preprints (2019)

23. Naqa, I.E., Murphy, M.J.: What is machine learning? In: Machine Learning in Radiation Oncology. pp. 3–11. Springer (2015)
24. Pedregosa, F., Varoquaux, G., Gramfort, A., et al, V.M.: Scikit-learn: Machine learning in python. Journal of Machine Learning Research 12, 2825–2830 (2011)
25. Prasad, V.K., Dansana, D., Bhavsar, M.D., Acharya, B., Gerogiannis, V.C., Kanavos, A.: Efficient resource utilization in iot and cloud computing. Information 14(11), 619 (2023)
26. Schirmer, P.A., Mporas, I.: Statistical and electrical features evaluation for electrical appliances energy disaggregation. Sustainability 11(11), 3222 (2019)
27. Schubert, E., Sander, J., Ester, M., Kriegel, H., Xu, X.: DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. ACM Transactions on Database Systems 42(3), 19:1–19:21 (2017)
28. Shin, C., Rho, S., Lee, H., Rhee, W.: Data requirements for applying machine learning to energy disaggregation. Energies 12(9), 1696 (2019)
29. Sinaga, K.P., Yang, M.: Unsupervised k-means clustering algorithm. IEEE Access 8, 80716–80727 (2020)
30. Sirojan, T., Phung, B.T., Ambikairajah, E.: Deep neural network based energy disaggregation. In: International Conference on Smart Energy Grid Engineering (SEGE). pp. 73–77. IEEE (2018)
31. Starczewski, A., Goetzen, P., Er, M.J.: A new method for automatic determining of the DBSCAN parameters. Journal of Artificial Intelligence and Soft Computing Research 10(3), 209–221 (2020)
32. Tiwari, A.: Supervised learning: From theory to applications. In: Artificial Intelligence and Machine Learning for EDGE Computing, pp. 23–32 (2022)
33. Vonitsanos, G., Kanavos, A., Mohasseb, A., Tsolis, D.: A nosql approach for aspect mining of cultural heritage streaming data. In: 10th International Conference on Information, Intelligence, Systems and Applications (IISA). pp. 1–4. IEEE (2019)
34. Vonitsanos, G., Kanavos, A., Mylonas, P.: Decoding gender on social networks: An in-depth analysis of language in online discussions using natural language processing and machine learning. In: IEEE International Conference on Big Data. pp. 4618–4625. IEEE (2023)
35. Vonitsanos, G., Panagiotakopoulos, T., Kanavos, A., Kameas, A.: An apache spark framework for iot-enabled waste management in smart cities. In: 12th Hellenic Conference on Artificial Intelligence (SETN). pp. 23:1–23:7. ACM (2022)
36. Vonitsanos, G., Panagiotakopoulos, T., Kanavos, A., Tsakalidis, A.K.: Forecasting air flight delays and enabling smart airport services in apache spark. In: International Conference on Artificial Intelligence Applications and Innovations (AIAI). IFIP Advances in Information and Communication Technology, vol. 628, pp. 407–417. Springer (2021)
37. Witten, I.H., Frank, E., Hall, M.A.: Data Mining: Practical Machine Learning Tools and Techniques, 3rd Edition. Morgan Kaufmann, Elsevier (2011)
38. Yao, G., Guo, C., Ge, Q., Ait-Ahmed, M.: A practical building energy consumption anomaly detection method based on parameter adaptive setting dbscan. Cognitive Computation and Systems 3(2), 154–168 (2021)
39. Yen, C.W., Ke, Y.L., Chen, S.T., Pai, Y.C., Wei, H.C., Teng, W.G.: Appliance recognition using a density- based clustering approach with multiple granularities. In: CS & IT Conference Proceedings. vol. 9 (2019)
40. Zeifman, M., Roth, K.: Nonintrusive appliance load monitoring: Review and outlook. IEEE Transactions on Consumer Electronics 57(1), 76–84 (2011)