Auto characterization of PEDs for digital references

towards iterative process optimisation

Project Nº: 43927229

Deliverable D5.1

Review report of machine learning algorithms for in-depth learning of PED

Author: Mengjie Han

This project is supported by the European Commission and funded under the Horizon 2020 ERA-NET Cofund scheme under grant agreement N° 875022

Document information Table

Project Data			
Project Acronym	PED-ACT		
Project Title	Auto characterization of PEDs for digital references towards iterative process optimisation		
Project n.	43927229		
Type of Project	Applied Research		
Topic identifier	Topic 2: Making PEDs happen: process innovation and business models		
Project duration	36 months		
Coordinator	Dalarna University, Sweden; Xingxing Zhang		
Deliverable Document Sheet			
Deliverable No.	D5.1		
Deliverable title	Review report of machine learning algorithms for in-depth learning of PED		
Description	D5.1 goes through the state-of-the-art machine learning algorithms as sub models for the in-depth characterization of the PED database, and the generation of digital PED reference.		
WP No.	5		
Lead beneficiary	Umeå University		
Author(s) & contributor(s)	Mengjie Han		
Type	X		
Dissemination L.	X		
Due Date	X	Submission Date	X

Version	Date	Author(s)	Organisation	Comments
v1	2023-09	Mengjie Han	Dalarna University	
v2	2023-10	Mengjie Han	Dalarna University	
v3				

Copyright notice

Disclaimer

This project receives funding under the Joint Programming Initiative (JPI) Urban Europe framework under project number 43927229.

It receives funding support from the strategic innovation program 'Viable Cities', which is financed by Vinnova, the Swedish Energy Agency and Formas (P2022-01000), The Scientific and Technological Research Council of Türkiye, and the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the funding agencies involved.

Table of contents

Innehåll

List of Figures

Executive Summary

Positive Energy District (PED) has emerged as a critical component of endeavors to accelerate the shift towards zero carbon emissions and climate-neutral living environments. While numerous innovation projects, programs, and activities have yielded valuable insights into PED implementation and operation, there remains a lack of consensus on defining a PED and assessing its constituent elements. D5.1 aims to establish a systematic process for characterizing PEDs. Initially, nineteen distinct elements of a PED are identified. Subsequently, the potential of two AI techniques, Machine Learning (ML) and Natural Language Processing (NLP), in modeling, extracting, and mapping these PED elements are explored. Finally, a comprehensive review of state-of-the-art research papers is conducted to evaluate their contributions towards assessing the effectiveness of ML and NLP models. D5.1 indicates that both ML and NLP exhibit significant promise in modeling various PED elements across optimization, control, design, and stakeholder mapping domains. Leveraging extensive datasets empowers these models to generate precise and actionable insights crucial for PED planning and implementation. It is imperative to develop an integrated approach that combines existing and innovative techniques for PED characterization.

1. The need for AI techniques to characterize PEDs

While PEDs have shown certain shared characteristics, they are inherently shaped by their unique local contexts. Cities vary significantly due to diverse factors in geography, history, politics, structure, society, law, and economics [1]. Given the inherent complexity of replicating a PED, it becomes important to maximize its potential for replication. The attributes of established PEDs can facilitate the creation of customized solutions tailored to specific local contexts [2]. These characterizations serve as a solid foundation for constructing an effective plan for replicating PEDs. Thus, the characterization of PEDs plays a crucial role in identifying common solutions that enhance their potential for replication, ultimately contributing to the attainment of climate neutrality and energy surplus. To guide the future development of PEDs, a deeper understanding of existing practice becomes imperative. This necessitates the acquisition of more extensive scientific knowledge concerning PEDs and the most effective methodologies for their operation.

PEDs are notable for their substantial data generation, stemming from diverse origins. Data can be drawn from various entities such as design and construction processes, building services, operational and building management systems, energy infrastructure, transportation systems, and maintenance and replacement systems. Furthermore, the growing utilization of digital twins has streamlined the accumulation of extensive datasets, spanning geometric and non-geometric data (pertaining to building characteristics), weather conditions, and energy data. The analysis of patterns within this data assumes paramount importance in comprehending the interplay between various systems and infrastructures. This scrutiny aids in measuring their effectiveness, a critical aspect as PEDs necessitate the harmonious integration of diverse systems and infrastructures for interaction among buildings, users, local energy sources, mobility options, and information and communication technology (ICT) systems [3].

Artificial intelligence (AI) techniques offer a potent means to scrutinize extensive datasets and derive valuable insights regarding the functioning of PEDs. These methodologies have effectively addressed a number of applications including load predictions, energy pattern profiling, regional energy consumption mapping, benchmarking for building stocks, and the evaluation of retrofit strategies. The capacity to model intricate relationships between input and output renders AI models highly efficient tools for managing vast and complex data [4]. Despite their growing utilization in various domains, there remains a notable need to investigate the extraction of comprehensive insights into PED characteristics through AI methods. While AI models are increasingly integrated into diverse spheres, the absence of a standardized framework for AI techniques in PEDs limits the broader applicability of findings. The establishment of a universal framework for AI techniques holds the potential to streamline the analysis of PED performance, thereby enhancing its general applicability and effectiveness.

2. Elements of PED

Through a comprehensive examination of diverse definitions found in existing literature, a shared set of elements defining PEDs emerges. Rather than offering narrative descriptions, these elements serve to extract the essence of PEDs by presenting a tangible collection of factors crucial to their success. These elements are necessary characteristics that define, implement, measure, and evaluate PEDs. For example, energy balance serves as the foundational basis that distinguishes PEDs from traditional urban districts in terms of producing more energy than it consumes. It is intrinsic quality that make a district "positive energy". By examining environmental aspect, energy justice and comfort, if is more feasible to assess the quality of life offered within a district. A well-balanced environment and sustainable practices can contribute to a healthier, more comfortable living experience for residents. By characterizing PEDs through these elements, stakeholders can have a comprehensive understanding and embody the core objectives, thus enabling informed decisions, effective management, and continuous enhancement.

These elements can be individually scrutinized to characterize PEDs effectively, totaling nineteen in number. Each element either specifies a vital aspect or facilitates one of the four core principles life quality, inclusiveness, sustainability, and resilience and security of energy supply [5]. For instance, "comfort" stands as a pivotal metric, impacting both the quality of life and seamless interaction with the grid. The production of renewable energy enhances resilience and security in the energy supply. These elements are logically grouped into two clusters, as depicted in Figure 1: the lower semicircle represents the three functions of the urban energy system, while the upper semicircle encompasses non-energy considerations. Each function forms a sub-cluster; for instance, energy efficiency assumes a higher priority among the three functions. A similar clustering approach extends to the non-energy elements. In the upper left quadrant, the three sustainability pillars are clustered, alongside a grouping of principle and policy-related elements.

The top right quadrant houses elements that spotlight specific facets of local implementation processes.

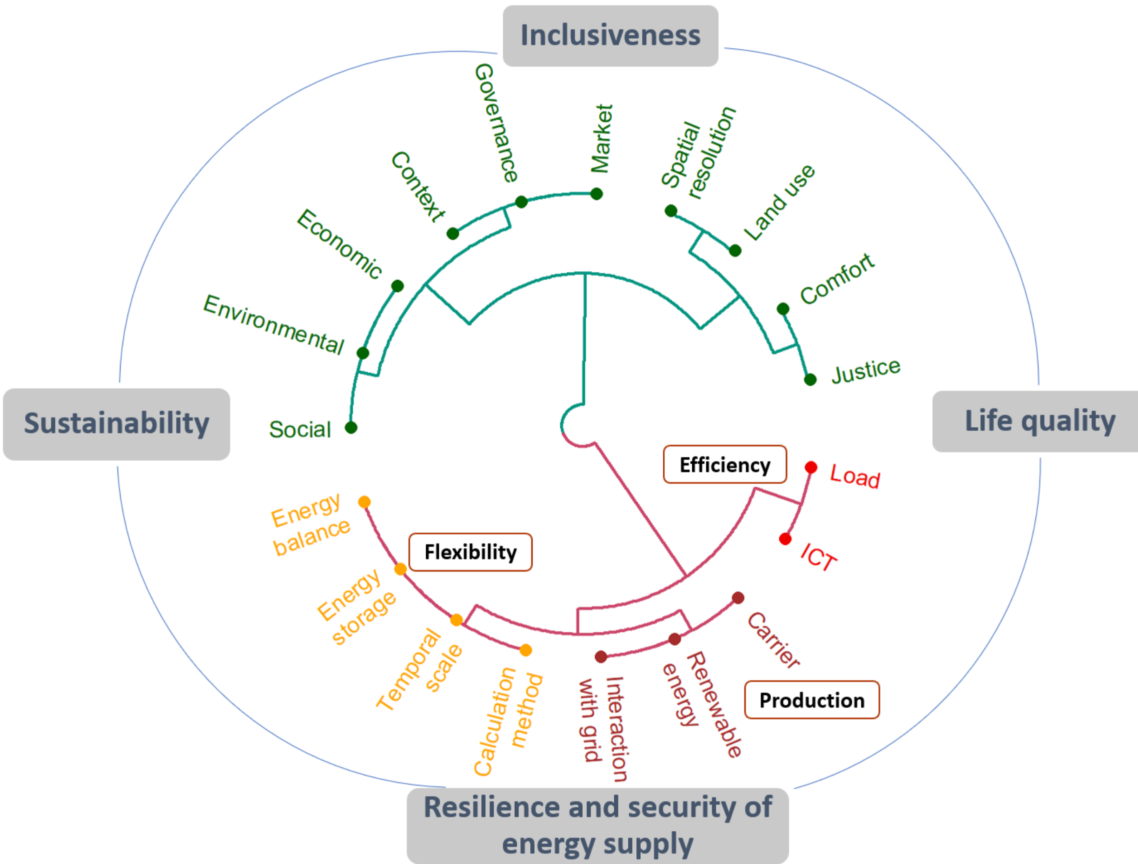

Figure 1: PED elements

3. Machine learning and natural language processing

AI means designing and applying algorithms in a computational environment to simulate human intelligence and solve complex problems. As the application of AI in many domains has assisted people and often worked to improve productivity, the integration of AI techniques into building energy management mainly concerns thermal comfort and energy use prediction, building system control, fault detection, and building information modelling [6–10]. The evidence from these applications, therefore, provides abundant hands-on experience for PED learning and replication. Among the various AI techniques, the versatility and scalability of machine learning (ML) and natural language processing (NLP) make them highly suitable for large datasets and complex problems. These two techniques are also constantly evolving, with new models and algorithms being developed to improve their performance all the time. Training ML and NLP models usually requires feeding them with large amounts of labeled or unlabeled data and utilizing an algorithm to optimize and iteratively refine the parameters to improve their performance. In this section, specific applications of ML and NLP on the PED elements or a combination of them are presented.

The analysis is based on scientific research papers, implying that only those PED elements that have been discussed in literature is presented in the following sub-sections.

3.1. Machine learning

ML learns to automate analytical models from data rather than being taught how to improve its learning abilities. Deep learning, as a subset of ML, is based on multi-layer artificial neural networks that can efficiently model the complex relationships between neurons and recognize complex patterns in the input data. ML represents a data-driven approach designed to complement physical models. Its potency lies in its capacity to handle intricate non-linear relationships within data, accommodating complex interactions and uncertainties. ML models are widely embraced due to their user-friendliness, finding applications in energy demand forecasting, energy pattern profiling, diverse retrofit strategies, and the prediction of renewable energy production [11]. However, the main challenge associated with ML pertains to the acquisition of extensive historical data for model training [12]. Obtaining data with finer time resolutions, for instance, proves more demanding compared to datasets with coarser resolutions, such as quarterly or annual intervals. In instances where real data is unavailable, perhaps due to a lack of monitoring, ML models can be trained using synthetic data generated through simulations.

3.1.1. ML for energy efficiency

The increase in energy consumption and the global energy crisis has amplified the significance of research into energy efficiency. Enhanced energy efficiency not only diminishes energy demand but also lessens reliance on external energy sources for PEDs. Over the past few decades, numerous studies have delved into the application of ML models to enhance building energy efficiency. Furthermore, various investigations have recognized the potential of employing industrial data to advance energy efficiency objectives [13]. A framework has been proposed to serve as a guideline for process industries in selecting appropriate ML tools to enhance energy efficiency goals. In addition, ML has also been used to forecast occupancy behavior and trends, with the aim of augmenting energy efficiency. Evaluations of the suitability of different ML algorithms in energy-efficient applications focusing on occupancy behavior have been also discussed [14].

3.1.1.1. ML for ICT

The term 'Building ICT' encompasses the information technology and communication systems within a building that generate data, which can be collected and analyzed to enhance the operational efficiency of the building. ICT plays a pivotal role in a PED as it establishes a dependable and stable connectivity infrastructure, linking active and passive devices utilized by residents in smart cities [15]. The entire process of designing, making decisions, and implementing a sustainable and intelligent building system relies on an ICT framework that can accommodate the integration of ML methods [16]. To attain the necessary level of ICT performance, ML is indispensable for dynamically and continuously adapting network behavior. Data acquired from IoT sensors strategically placed throughout a smart city can be effectively managed by ML to optimize resource and asset utilization [17]. To enhance energy efficiency, ICT enables the acquisition of extensive data, its processing, and preparation for practical use.

3.1.1.2. ML for building load

ML can play a crucial role in assessing energy loads and balances by forecasting the heating, cooling, ventilation, and electrical energy requirements for buildings, both at the individual building level and at the district level. Among these efforts, 57% have focused on individual buildings, while 43% have addressed multiple buildings [18]. Collectively, research has contributed to the utilization of ML-based energy consumption predictions as a means of evaluating diverse energy-saving techniques. Energy predictions facilitate demand-side management for making intelligent control decisions, analyzing and balancing energy supply and demand, and assessing a building's energy flexibility based on smart grid strategies [19]. It has also been commonly recognized that ML methods demonstrate superior accuracy in the short term and are more adept at forecasting for brief time intervals compared to longer ones, such as a year or more [20]. Although this capability is valuable in the context of PEDs for short-term energy sharing planning, it is imperative for PED development to generate precise long-term predictions for shaping energy supply strategies and making capital investments in energy-efficient applications.

3.1.2. ML for sustainable society

Renewable energy sources play a vital role in the electricity grid, offering advantages in terms of reliability, cost-effectiveness, and environmental sustainability. While the majority of ML applications in renewable energy are focused on predictions for solar and wind energy, there is a broader spectrum of studies utilizing ML models to simulate energy production from renewable sources. ML approaches are instrumental in sustainability assessments, aiding decision-makers in identifying actions to enhance sustainability. This is particularly important as urban areas strive to become more inclusive, safe, resilient, and sustainable [21]. Supervised ML methods have predominantly been applied for prediction tasks, while unsupervised techniques have found utility in developing novel energy sector products and materials. Nevertheless, the availability and refinement of data have played a pivotal role in the adoption of ML within the energy sector.

Indoor comfort significantly influences the well-being and contentment of building occupants. When the indoor environment becomes uncomfortable, occupants may consider altering the building's HVAC system or lighting, potentially leading to detrimental effects on the building's energy efficiency. Consequently, this could disrupt the overall energy equilibrium within the PED. Several ML models can be employed to create personalized comfort systems. These personal comfort models, founded on occupants' heating and cooling preferences, can be integrated into daily comfort management routines, elevating occupant satisfaction levels and optimizing energy utilization [22].

3.2. Natural language processing

As a sub-field of AI, NLP integrates linguistics and computer science to enable a computer to process and understand natural language data. The most common data source for helping a computer to develop rules for decoding information comes in the form of audio and text. Typical NLP tasks comprise named entity recognition, part of speech tagging, topic modeling, machine translation, and text classification. Sometimes, as indicated in Figure 2, NLP tasks need to be executed by using ML or a deep learning method. In these cases, there can be overlaps between solving an NLP task and ML algorithms.

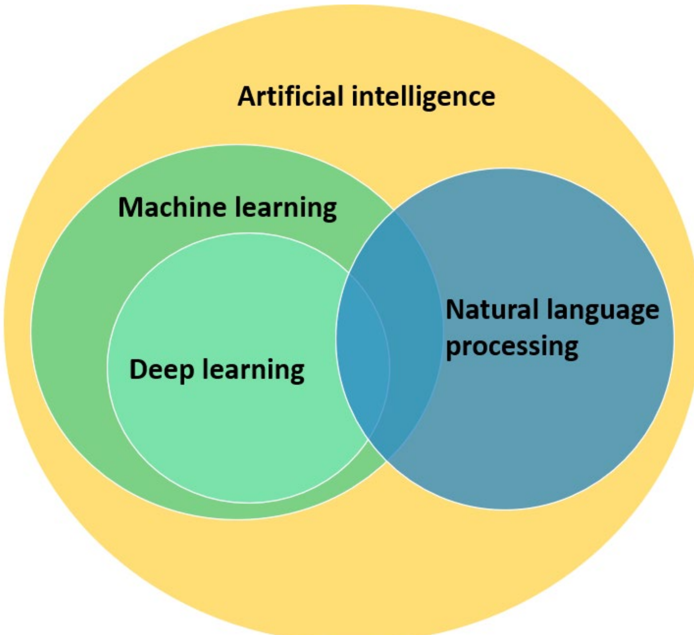

Figure 2: ML and NLP

3.2.1. NLP for energy efficiency

Research papers have been widely assembled to serve as documents for training NLP models [23,24]. One example is to analyze papers with the objective of uncovering the relationships between data science and energy efficiency across four categories: data, data science, energy efficiency, and phase [25]. An NLP method word2vec has been mostly employed to represent each word within these categories as a high-dimensional vector. The resulting usability relation extraction reveals that passive design, demand-controlled ventilation, model predictive controls, fault detection and diagnosis, and retrofit analysis make more frequent use of data. Another method part of speech (POS) tagging can be applied to preprocess energy audit report data containing descriptions of energy conservation measures (ECM). In this method, the frequency of each word was computed to create ECM dictionaries based on the auditors' recommendations [26].

3.2.1.1. NLP for load

Modeling building energy consumption is a valuable tool for enhancing energy efficiency as it provides insights into energy utilization patterns and identifies areas where energy savings can be optimized. By understanding energy usage more comprehensively, it becomes possible to pinpoint inefficient energy consumption areas and suggest strategies for achieving substantial energy savings while maintaining occupant comfort. An innovative approach involved the development of an energy2vec model based on word2vec, which utilized time series building load data with a one-minute window length, incorporating appliance status as well [27]. The embedded vectors captured contextual information about energy profiles, reflecting residents' habits and appliance usage. Varied sliding window lengths for word embeddings can be employed to separate the load attributed to different appliance operating cycles [28]. Addressing the challenge of encoding categorical attributes and identifying the most relevant ones, a solution was found by applying word2vec to these attributes before fitting an attention-based Long Short-Term Memory (LSTM) model [29]. This approach offers distinct advantages, particularly for medium- and long-term load forecasting. Beyond word2vec, other techniques have also contributed to understanding energy-saving trends and outlining the applicability of machine learning methods in assessing energy consumption and intelligent computing [28,30].

3.2.2. NLP for renewable energy

PEDs stand to gain significant advantages from the production of renewable energy, as it grants them access to sustainable, carbon-neutral energy sources. Local renewable energy production ensures energy independence and enhances PEDs' resilience. Topic modeling is a useful method to analyze research papers, aiming to uncover the key factors contributing to the success and growth of renewable energy projects [31]. The identified factors can be prioritized as follows: (1) effective government policies; (2) robust public-private partnerships with risk-sharing mechanisms; (3) community support and engagement; (4) favorable fiscal incentives and terms; and (5) access to skilled talent. POS tagging has also been applied to filter out irrelevant words from abstracts. This process enables the identification of topics, encompassing both trending and less-explored subjects within current research on renewable energy [32]. Hot topics in this research domain may encompass energy storage, photonic materials, nanomaterials, and biofuels, while less-discussed areas could relate to sustainable development and agriculture. A critical research challenge lies in devising methods to establish and optimize renewable energy systems effectively within PEDs.

3.2.3. NLP for context and market

Contextual factors play a pivotal role in optimizing the design of PEDs to align with local resources, policies, and constraints. Topic modeling has emerged as the preferred method for capturing these localized nuances. For instance, data from reports pertaining to low-carbon transitions can be employed to gain insights into how local governments interpret low-carbon transitions and to reveal primary topics. Topic modeling is also instrumental in clarifying the objectives of smart city projects from the perspectives of urban leaders and citizens. This approach enables city officials to effectively involve residents in smart city project development while establishing a communication baseline that took into account diverse cultural, demographic, geographic, and economic factors within the community [33,34].

The establishment of essential infrastructure for managing the supply and demand of renewable energy sources within PEDs also integrates various energy technologies into the energy market. Leveraging a dataset of approximately 100,000 website comments, a hybrid approach combining BERT and bidirectional LSTM has been employed to model public sentiment. This enabled stakeholders to provide precise technical support in the energy market, ensuring effective management of renewable energy resources [35].

3.2.4. NLP for land use

Effective urban planning and land use play pivotal roles in developing PEDs. One notable approach is the adoption of a compact district structure, which can significantly reduce energy transportation and waste. The concept of Points of Interest (POIs) and the utilization of Geographic Information System (GIS) data offer valuable methods for land use identification by analyzing the types and spatial distribution of facilities within a specific geographical area. While quantifying the relationship between the spatial distribution of POIs and land use types has posed challenges, innovative solutions have emerged. For instance, a novel approach was introduced for developing a shortest path connection to represent sequential POIs for word embedding [36]. In this context, each POI is analogous to a word, and each traffic analysis zone is considered as a document. However, it's worth noting that converting spatial data into sequential document data may have limitations due to the 2D distribution of POIs [37]. Instead, word frequency can be used to represent the distribution of POI types, which demonstrated high performance in clustering functional regions. Additionally, mobility patterns have recognized as influential factors and treated as "words" in a topic modeling analysis of regional functions, thus accounting for the impact of metadata within a region [38]. These innovative approaches contribute to our understanding of the intricate relationship between urban planning, land use, and the development of PEDs.

4. Features and algorithms

As illustrated in Figure 3, several key features and algorithms, along with their respective functions or data requirements for specific elements for ML and NLP, have be summarized in D5.1. For instance, the powerful BERT (*BERT* representing variations of BERT) language model can be swiftly deployed to construct a comprehensive representation of a PED by enhancing semantic interoperability among buildings by scrutinizing building metadata. This modeling framework directly facilitates efficient monitoring and optimization of energy consumption within buildings, thereby bolstering overall energy efficiency. Topic modeling is a statistical model used to discover abstract topics within a collection of documents. It is efficient to extract concrete guidelines from policies, regulations and legislations that govern activities to enhance comprehension administrative framework. Another example is the utilization of deep Artificial Neural Networks (ANNs) to enhance indoor comfort through intricate system control. This sophisticated model structure, supported by a wealth of data, effectively captures the nonlinear dynamics of a system, enabling intelligent control actions in complex environments. The choice of these AI paradigms, whether employed in concert or in isolation, should ideally align with the specific phase of a PED's development. During the design phase, stakeholder matching techniques could identify optimal sustainable design strategies. Conversely, evolutionary algorithms could come to the fore during the implementation phase to optimize the placement and sizing of renewable energy systems.

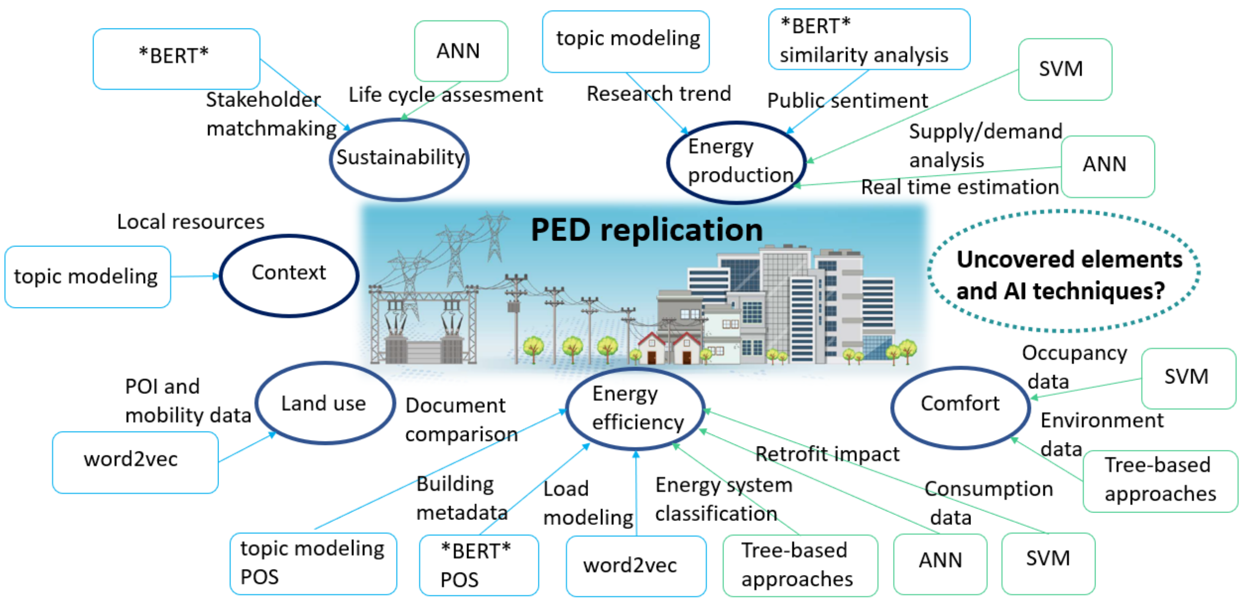

Figure 3: Key features and algorithms of ML and NLP in PED characterization and replication

5. Summary

In D5.1, we have briefly reviewed the state-of-the-art techniques in machine learning and natural language process for characterizing PEDs. Nineteen elements, including both energy and non-energy facets of PED, have been extracted from existing literature. The predominant models in ML focus on ANN, SVM, and tree-based methods, each instrumental for prediction and classification across various stages. For NLP, topic modeling, word embedding, and the training of large language models like BERT are pivotal for tasks for stakeholder matching, sentiment analysis, and metadata examination. However, it's noteworthy that certain elements have not yet been found for demonstrating modeling outcomes through ML or NLP techniques. These include aspects like renewable energy carriers, grid integration, energy balance, and considerations of justice. Given that PEDs encompass a wide range of spatial and temporal scales, along with intricate relationships between buildings, urban infrastructure, and energy systems at the district level, it is imperative to utilize existing urban-scale modeling approaches for PED research. In addition, there is currently no direct link between the modeling outcomes produced by ML or NLP and the actual operation of a PED. PED characterization frequently relies on just one element, neglecting the influence of factors like policy, societal acceptance, and economic feasibility, all of which are pivotal to the success of PEDs. To achieve the intended results, it is imperative to adopt a holistic approach when evaluating PEDs, wherein ML and NLP can be integrated with other dimensions of PED planning and execution.

References

1. Lindholm, O.; Rehman, H. ur; Reda, F. Positioning Positive Energy Districts in European Cities. *Buildings* **2021**, *11*, 19, doi:10.3390/buildings11010019.
2. Uspenskaia, D.; Specht, K.; Kondziella, H.; Bruckner, T. Challenges and Barriers for Net-Zero/Positive Energy Buildings and Districts—Empirical Evidence from the Smart City Project SPARCS. *Buildings* **2021**, *11*, 78, doi:10.3390/buildings11020078.
3. Sareen, S.; Albert-Seifried, V.; Aelenei, L.; Reda, F.; Etminan, G.; Andreucci, M.-B.; Kuzmic, M.; Maas, N.; Seco, O.; Civiero, P.; et al. Ten Questions Concerning Positive Energy Districts. *Building and Environment* **2022**, *216*, 109017, doi:10.1016/j.buildenv.2022.109017.
4. Jordan, M.I.; Mitchell, T.M. Machine Learning: Trends, Perspectives, and Prospects. *Science* **2015**, *349*, 255–260, doi:10.1126/science.aaa8415.
5. Hinterberger, R.; Gollne, C.; Noll, M.; Meyer, S.; Schwarz, H.-G. *White Paper on PED Reference Framework for Positive Energy Districts and Neighbourhoods*; Austrian Research Promotion Agency: Austria, 2020;
6. Ngarambe, J.; Yun, G.Y.; Santamouris, M. The Use of Artificial Intelligence (AI) Methods in the Prediction of Thermal Comfort in Buildings: Energy Implications of AI-Based Thermal Comfort Controls. *Energy and Buildings* **2020**, *211*, 109807, doi:10.1016/j.enbuild.2020.109807.
7. Wang, Z.; Srinivasan, R.S. A Review of Artificial Intelligence Based Building Energy Use Prediction: Contrasting the Capabilities of Single and Ensemble Prediction Models. *Renewable and Sustainable Energy Reviews* **2017**, *75*, 796–808, doi:10.1016/j.rser.2016.10.079.
8. Halhoul Merabet, G.; Essaaidi, M.; Ben Haddou, M.; Qolomany, B.; Qadir, J.; Anan, M.; Al-Fuqaha, A.; Abid, M.R.; Benhaddou, D. Intelligent Building Control Systems for Thermal Comfort and Energy-Efficiency: A Systematic Review of Artificial Intelligence-Assisted Techniques. *Renewable and Sustainable Energy Reviews* **2021**, *144*, 110969, doi:10.1016/j.rser.2021.110969.
9. Zhao, Y.; Li, T.; Zhang, X.; Zhang, C. Artificial Intelligence-Based Fault Detection and Diagnosis Methods for Building Energy Systems: Advantages, Challenges and the Future. *Renewable and Sustainable Energy Reviews* **2019**, *109*, 85–101, doi:10.1016/j.rser.2019.04.021.
10. Sacks, R.; Wang, Z.; Ouyang, B.; Utkucu, D.; Chen, S. Toward Artificially Intelligent Cloud-Based Building Information Modelling for Collaborative Multidisciplinary Design. *Advanced Engineering Informatics* **2022**, *53*, 101711, doi:10.1016/j.aei.2022.101711.
11. Leone, F.; Reda, F.; Hasan, A.; Rehman, H. ur; Nigrelli, F.C.; Nocera, F.; Costanzo, V. Lessons Learned from Positive Energy District (PED) Projects: Cataloguing and Analysing Technology Solutions in Different Geographical Areas in Europe. *Energies* **2022**, *16*, 356, doi:10.3390/en16010356.
12. Österbring, M.; Mata, É.; Thuvander, L.; Mangold, M.; Johnsson, F.; Wallbaum, H. A Differentiated Description of Building-Stocks for a Georeferenced Urban Bottom-up Building-Stock Model. *Energy and Buildings* **2016**, *120*, 78–84, doi:10.1016/j.enbuild.2016.03.060.
13. Narciso, D.A.C.; Martins, F.G. Application of Machine Learning Tools for Energy Efficiency in Industry: A Review. *Energy Reports* **2020**, *6*, 1181–1199, doi:10.1016/j.egyr.2020.04.035.
14. Zhang, W.; Wu, Y.; Calautit, J.K. A Review on Occupancy Prediction through Machine Learning for Enhancing Energy Efficiency, Air Quality and Thermal Comfort in the Built

Environment. *Renewable and Sustainable Energy Reviews* **2022**, *167*, 112704, doi:10.1016/j.rser.2022.112704.
15. Said, O.; Tolba, A. Accurate Performance Prediction of IoT Communication Systems for Smart Cities: An Efficient Deep Learning Based Solution. *Sustainable Cities and Society* **2021**, *69*, 102830, doi:10.1016/j.scs.2021.102830.
16. Ullah, Z.; Al-Turjman, F.; Mostarda, L.; Gagliardi, R. Applications of Artificial Intelligence and Machine Learning in Smart Cities. *Computer Communications* **2020**, *154*, 313–323, doi:10.1016/j.comcom.2020.02.069.
17. Bhattacharya, S.; Somayaji, S.R.K.; Gadekallu, T.R.; Alazab, M.; Maddikunta, P.K.R. A Review on Deep Learning for Future Smart Cities. *Internet Technology Letters* **2022**, *5*, doi:10.1002/itl2.187.
18. Fathi, S.; Srinivasan, R.; Fenner, A.; Fathi, S. Machine Learning Applications in Urban Building Energy Performance Forecasting: A Systematic Review. *Renewable and Sustainable Energy Reviews* **2020**, *133*, 110287, doi:10.1016/j.rser.2020.110287.
19. Khalil, M.; McGough, A.S.; Pourmirza, Z.; Pazhoohesh, M.; Walker, S. Machine Learning, Deep Learning and Statistical Analysis for Forecasting Building Energy Consumption — A Systematic Review. *Engineering Applications of Artificial Intelligence* **2022**, *115*, 105287, doi:10.1016/j.engappai.2022.105287.
20. Liu, Z.; Wu, D.; Liu, Y.; Han, Z.; Lun, L.; Gao, J.; Jin, G.; Cao, G. Accuracy Analyses and Model Comparison of Machine Learning Adopted in Building Energy Consumption Prediction. *Energy Exploration & Exploitation* **2019**, *37*, 1426–1451, doi:10.1177/0144598718822400.
21. De Las Heras, A.; Luque-Sendra, A.; Zamora-Polo, F. Machine Learning Technologies for Sustainability in Smart Cities in the Post-COVID Era. *Sustainability* **2020**, *12*, 9320, doi:10.3390/su12229320.
22. Kim, J.; Zhou, Y.; Schiavon, S.; Raftery, P.; Brager, G. Personal Comfort Models: Predicting Individuals' Thermal Preference Using Occupant Heating and Cooling Behavior and Machine Learning. *Building and Environment* **2018**, *129*, 96–106, doi:10.1016/j.buildenv.2017.12.011.
23. Copiello, S. Economic Parameters in the Evaluation Studies Focusing on Building Energy Efficiency: A Review of the Underlying Rationale, Data Sources, and Assumptions. *Energy Procedia* **2019**, *157*, 180–192, doi:10.1016/j.egypro.2018.11.179.
24. Zeng, R.; Chini, A. A Review of Research on Embodied Energy of Buildings Using Bibliometric Analysis. *Energy and Buildings* **2017**, *155*, 172–184, doi:10.1016/j.enbuild.2017.09.025.
25. Abdelrahman, M.M.; Zhan, S.; Miller, C.; Chong, A. Data Science for Building Energy Efficiency: A Comprehensive Text-Mining Driven Review of Scientific Literature. *Energy and Buildings* **2021**, *242*, 110885, doi:10.1016/j.enbuild.2021.110885.
26. Lai, Y.; Papadopoulos, S.; Fuerst, F.; Pivo, G.; Sagi, J.; Kontokosta, C.E. Building Retrofit Hurdle Rates and Risk Aversion in Energy Efficiency Investments. *Applied Energy* **2022**, *306*, 118048, doi:10.1016/j.apenergy.2021.118048.
27. Nalmpantis, C.; Krystalakos, O.; Vrakas, D. Energy Profile Representation in Vector Space. In Proceedings of the Proceedings of the 10th Hellenic Conference on Artificial Intelligence; ACM: Patras Greece, July 9 2018; pp. 1–5.
28. Nie, Z.; Yang, Y.; Xu, Q. An Ensemble-Policy Non-Intrusive Load Monitoring Technique Based Entirely on Deep Feature-Guided Attention Mechanism. *Energy and Buildings* **2022**, *273*, 112356, doi:10.1016/j.enbuild.2022.112356.
29. Peng, J.; Kimmig, A.; Wang, J.; Liu, X.; Niu, Z.; Ovtcharova, J. Dual-Stage Attention-Based Long-Short-Term Memory Neural Networks for Energy Demand Prediction. *Energy and Buildings* **2021**, *249*, 111211, doi:10.1016/j.enbuild.2021.111211.

30. Abdelaziz, A.; Santos, V.; Dias, M.S. Machine Learning Techniques in the Energy Consumption of Buildings: A Systematic Literature Review Using Text Mining and Bibliometric Analysis. *Energies* **2021**, *14*, 7810, doi:10.3390/en14227810.
31. Kumar, M.; Ng, J. Using Text Mining and Topic Modelling to Understand Success and Growth Factors in Global Renewable Energy Projects. *Renewable Energy Focus* **2022**, *42*, 211–220, doi:10.1016/j.ref.2022.06.010.
32. Bickel, M.W. Reflecting Trends in the Academic Landscape of Sustainable Energy Using Probabilistic Topic Modeling. *Energ Sustain Soc* **2019**, *9*, 49, doi:10.1186/s13705-019-0226-z.
33. Tie, M.; Zhu, M. Interpreting Low-Carbon Transition at the Subnational Level: Evidence from China Using a Natural Language Processing Approach. *Resources, Conservation and Recycling* **2022**, *187*, 106636, doi:10.1016/j.resconrec.2022.106636.
34. Nicolas, C.; Kim, J.; Chi, S. Natural Language Processing-Based Characterization of Top-down Communication in Smart Cities for Enhancing Citizen Alignment. *Sustainable Cities and Society* **2021**, *66*, 102674, doi:10.1016/j.scs.2020.102674.
35. Cai, R.; Qin, B.; Chen, Y.; Zhang, L.; Yang, R.; Chen, S.; Wang, W. Sentiment Analysis About Investors and Consumers in Energy Market Based on BERT-BiLSTM. *IEEE Access* **2020**, *8*, 171408–171415, doi:10.1109/ACCESS.2020.3024750.
36. Yao, Y.; Li, X.; Liu, X.; Liu, P.; Liang, Z.; Zhang, J.; Mai, K. Sensing Spatial Distribution of Urban Land Use by Integrating Points-of-Interest and Google Word2Vec Model. *International Journal of Geographical Information Science* **2017**, *31*, 825–848, doi:10.1080/13658816.2016.1244608.
37. Zhai, W.; Bai, X.; Shi, Y.; Han, Y.; Peng, Z.-R.; Gu, C. Beyond Word2vec: An Approach for Urban Functional Region Extraction and Identification by Combining Place2vec and POIs. *Computers, Environment and Urban Systems* **2019**, *74*, 1–12, doi:10.1016/j.compenvurbsys.2018.11.008.
38. Yuan, J.; Zheng, Y.; Xie, X. Discovering Regions of Different Functions in a City Using Human Mobility and POIs. In Proceedings of the Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining; ACM: Beijing China, August 12 2012; pp. 186–194.