FEATURE-BASED IDENTIFICATION OF DISTRIBUTED ENERGY RESOURCES IN BUILDINGS FROM AGGREGATED SMART METER DATA

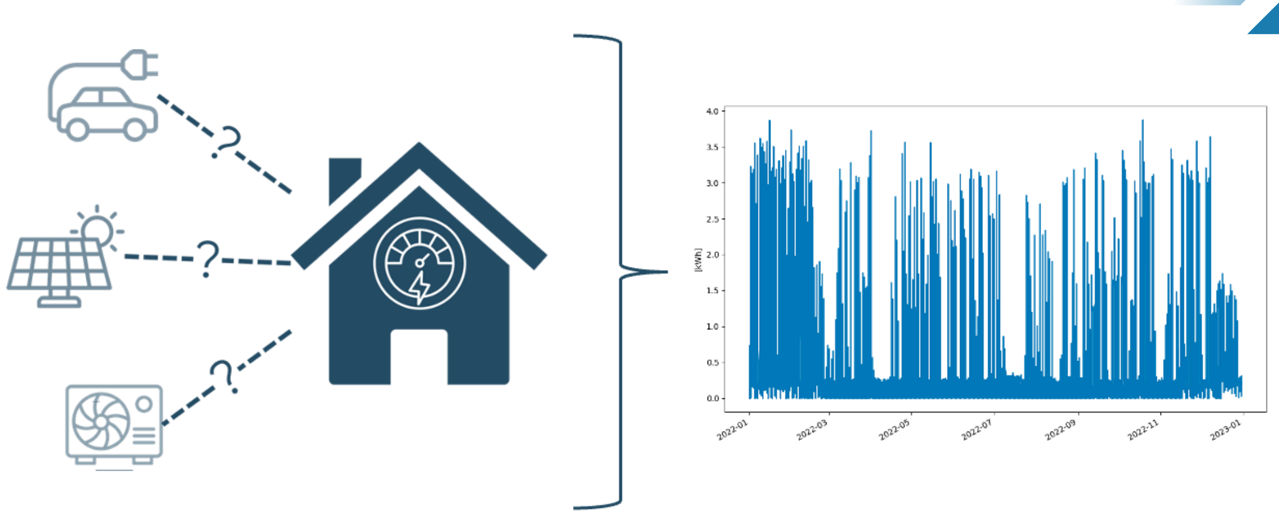
Author

Ada Canaydin

EnergyVille KU LEUVEN

ada.canaydin@kuleuven.be

1. INTRODUCTION



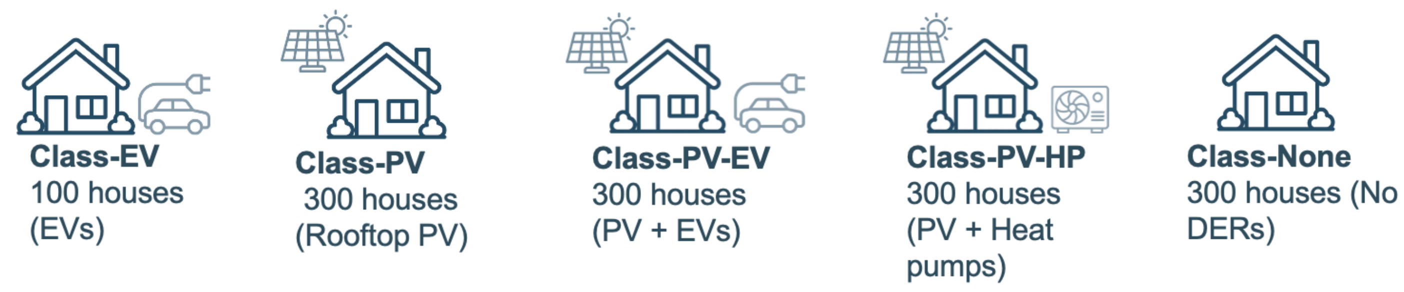
- The rapid rise of DERs (EVs, rooftop PV, and heat pumps) introduces flexibility but also challenges grid stability. DSOs often lack behind-the-meter visibility, as smart meters provide only aggregated load data
- Key challenges:
  1. High dimensional time series
  2. Incomplete, noisy data (e.g., missing values, anomalies)
  3. Need for interpretable, actionable insights from large datasets

2. OBJECTIVE

How can we **extract actionable insights from large volumes of aggregated smart meter data** to **detect DER proliferation without costly high-resolution sub-metering**?

3. METHODOLOGY

- Dataset [1]: Quarter-hourly smart meter data from residential buildings in Flanders, Belgium (Jan 1 – Dec 31, 2022)



- Feature-based analysis:

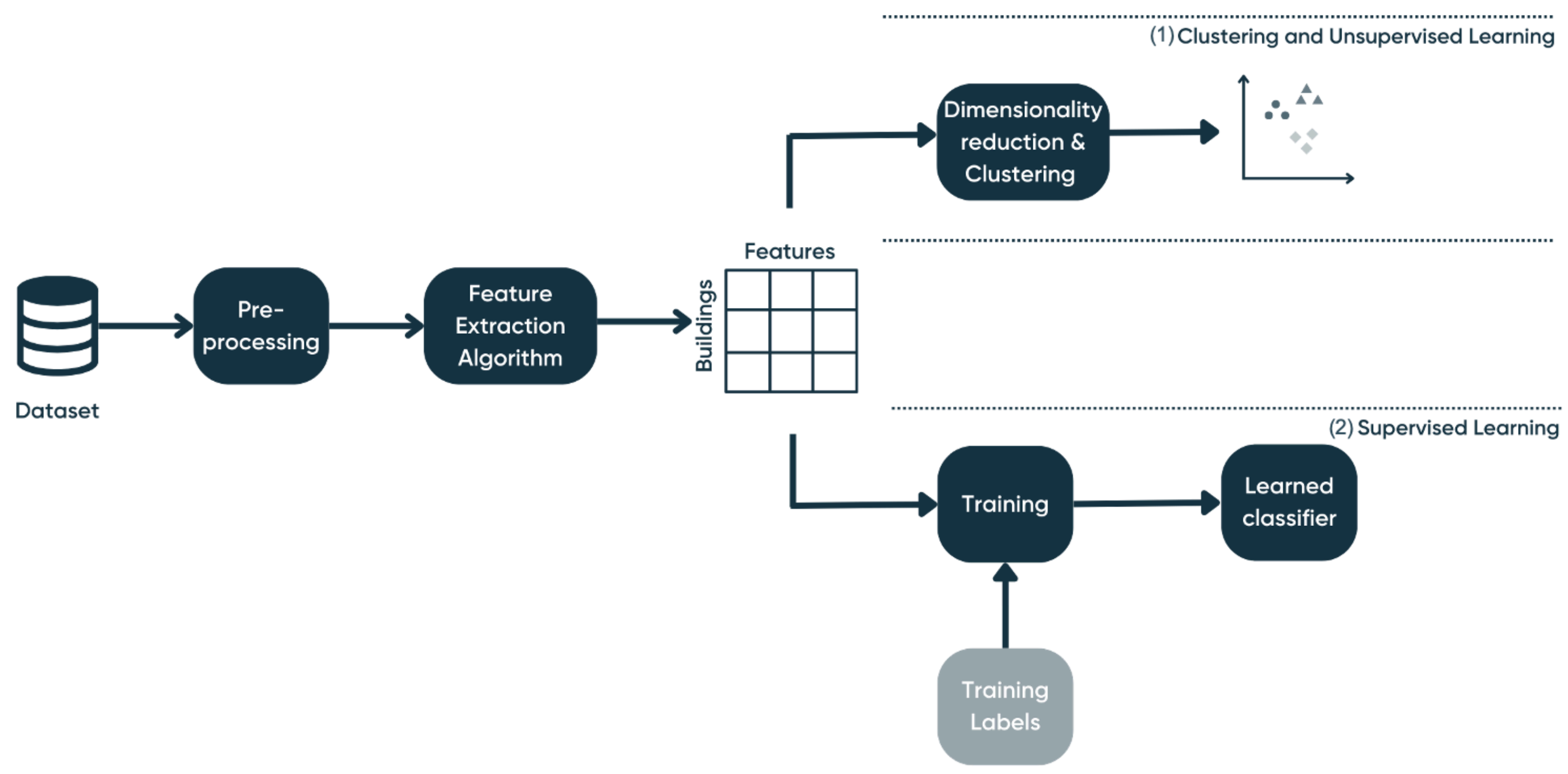


Figure 1: An overview of the end-to-end pipeline

- Two interpretable feature extraction methodologies

Feature Type	Applicability	Focus	Key Indicators
Domain-agnostic features [2]	General time series data	Full time series analysis	Long-term trends, seasonality, spectral patterns, autocorrelation, etc.
Domain-informed features [3]	Electricity demand data	Daily load behaviors (SAX transformation)	Short-term characteristics, peak start times, peak durations, slopes, non-business hour variability, etc.

- Two downstream applications

(1) Clustering and Unsupervised Learning

Apply dimensionality reduction (t-SNE) + K-Means clustering to **group buildings by DER profile**

(2) Supervised Learning

Train a random forest classifier on known labels to **predict DER class directly** from extracted features

4. RESULTS

(1) Clustering and Unsupervised Learning

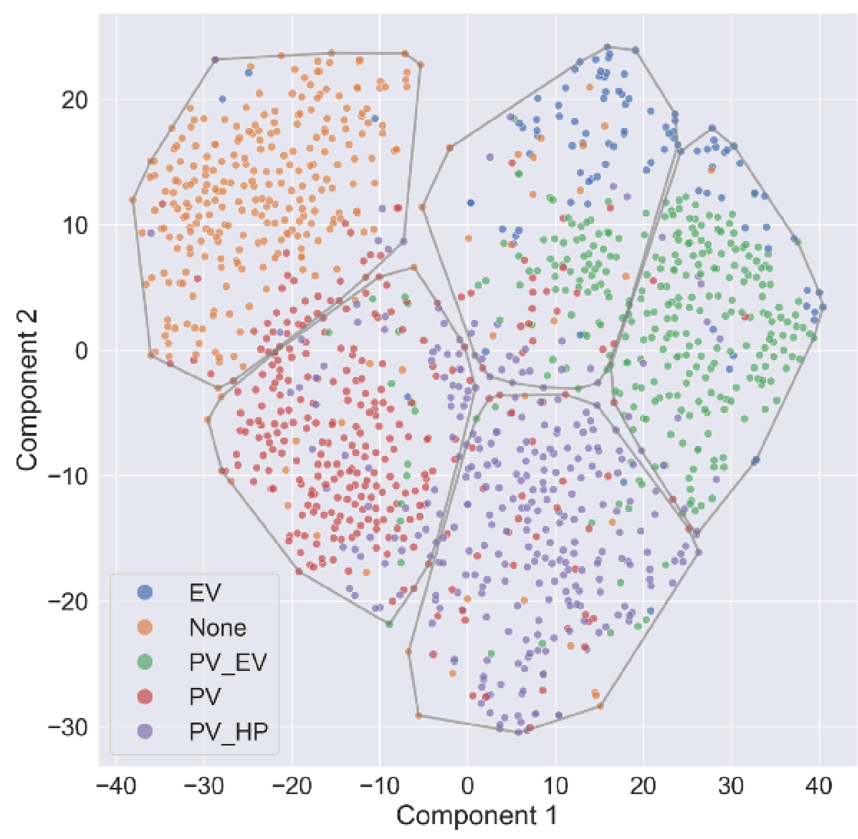


Figure 2: K-Means clusters (K=5) on 2D t-SNE feature space

(2) Supervised Learning

- 85% accuracy on a test set of 390 buildings
- Key features: daily minimum, variability (daily & non-business), past correlations, spikes, consistency

5. CONCLUSION

Beyond classification, feature-based analysis enables metadata completion [4], forecastability assessment [4], synthetic data generation, global forecast improvements, scenario simulation, and data quality control

[1] https://opendata.fluvius.be/explore/dataset/1_50-verbruiksprofielen-dm-elek-kwartierwaarden-voor-een-volledig-jaar/information/

[2] https://www.nixtla.io/.

[3] https://maomaohu.net/software/ifeel/

[4] Canaydin, Ada, et al. "Interpretable domain-informed and domain-agnostic features for supervised and unsupervised learning on building energy demand data." Applied Energy 360 (2024): 122741.