

PORTFOLIO-STRUCTURED SEGMENTATION AND FORECAST-GUIDED TARGETING OF DAYTIME-COMMERCIAL SMES ELECTRIC ENERGY USAGE PATTERNS

Aigars Laizans¹, Andris Backurs¹, Martins Tisenkopfs¹, Nikita Roldugins²

¹Latvia University of Life Sciences and Technologies, Latvia; ²SIA "TET", Latvia
aigars.laizans@lbtu.lv, andris.backurs@gmail.com, martins.tisenkopfs@gmail.com,
nikita.roldugins@tet.lv

Abstract. The study develops a portfolio-structured interpretation of a Latvian daytime-commercial SME electricity cohort based on annual hourly smart-meter data for 610 clients. It addresses a practical gap between descriptive load clustering and actionable energy-management use: in many SME portfolios, high-resolution meter data are available, whereas verified business-register labels are limited or absent. The workflow combines alignment of client profiles on a common hourly grid, comparison of four normalization approaches, feature extraction from 24 hourly shape indicators plus a scale anchor, cohort-level clustering, subclustering of the dominant segment, and cluster-specific LightGBM forecasting benchmarked against a persistence baseline. Three behavioural portfolio groups were identified. Cluster, containing 382 clients (62.6% of the cohort), represents the dominant mainstream daytime-commercial segment. Forecasting results show that operational value is concentrated in the dominant portfolio segment: this cluster achieved a mean WAPE of 15.7% across four 168-hour windows, compared with 33.7% for the persistence baseline. The main contribution of the paper is at some extent the energy user legal-sector recognition, and the conversion of a raw data portfolio into reproducible behavioural segments that can be used by distribution system operators, electricity retailers, and energy-efficiency consultants. The findings show that behaviour-based portfolio structuring can already support scalable targeting, while future linkage with business, tariff, and premise metadata would further improve recognition precision and intervention design.

Keywords: SME electricity consumption, customer segmentation, clustering, forecasting benchmark, energy efficiency, trend use.

Introduction

Small and medium-sized enterprises (SMEs) are important energy users, yet they are often underserved by structured energy-management, flexibility, and benchmarking programmes. In contrast to large industrial sites, SMEs usually operate with limited technical staffing, smaller project budgets, and a wider diversity of premises, equipment, and occupancy patterns. This makes behaviour-based use of smart-meter data especially valuable, because the data can reveal operating rhythms even when detailed process metadata are unavailable [1-3].

The Latvian context further strengthens the relevance of this problem. According to the 2025 SME Country Fact Sheet, Latvia had 106,886 SMEs in the non-financial business economy, while official statistical sources provide enterprise counts by size group and activity class rather than smart-meter-based behavioural categories [4-5]. This means that electricity portfolios may be operationally rich but administratively under-described: utilities and market actors can observe hourly load patterns, yet cannot automatically translate them into intervention-ready customer families.

This challenge sits at the intersection of three active research streams. First, recent SME energy-efficiency research shows that smaller firms are important for decarbonisation and cost reduction, but frequently lack tailored energy-management support [1-3]. Second, smart-meter-based benchmarking and customer segmentation have already demonstrated that repeated load-shape families can be identified for commercial customers and used for demand alerts, operating-time estimation, and customer grouping [6-12]. Third, forecasting and clustering methods are sufficiently mature to support scalable portfolio analysis, but many studies remain either method-centric or focused on building-type recognition under richer metadata conditions [13-18].

The research gap addressed here is therefore practical as well as scientific. Existing studies show that clustering and forecasting are feasible, yet fewer studies convert an SME smart-meter cohort into an explicit portfolio structure that distinguishes scalable mainstream segments, secondary behavioural groups, and rare anchor-load cases under limited label availability. This translation step is important because utilities, retailers, and consultants do not act on clusters in the abstract; they act on portfolios that must be prioritized, differentiated, and costed.

The portfolio-structured contribution of the study is fourfold.

- It converts a raw energy consumption data into a clearly prioritized portfolio structure.
- It evaluates operational usefulness by combining segmentation with a forecasting benchmark against persistence.
- It treats behavioural interpretation as probabilistic and explicitly separates behaviour-family recognition from legal-sector identification.
- It shows how portfolio segments can already support differentiated actions by DSOs, electricity retailers, and energy-efficiency consultants, while also defining the metadata needed for the next step toward higher recognition precision.

Materials and methods

The empirical object of the study is a portfolio of 610 Latvian daytime-commercial SME clients represented by annual hourly electricity-consumption time series. The clients were aligned on a common UTC-anchored annual grid of 8759 hourly slots. In contextual terms, this sample corresponds to 0.571% of the 106,886 Latvian SMEs reported in the 2025 national fact sheet [4]. This proportion is used only as a scale reference. It should not be interpreted as an official national count of a behavioural SME category, because public statistical sources do not publish smart-meter-defined enterprise classes [4-5].

The study is portfolio structured. The primary analytical question is not whether a single cluster can be described in isolation, but how the full cohort can be transformed into differentiated operational segments. Consequently, the first stage identifies the main behavioural families within the 610-client portfolio, while the second stage examines the internal structure of the dominant segment. This design keeps the article focused on the portfolio as a management object rather than on one cluster taken out of context.

The client profiles were first aligned to the shared annual hourly grid so that cluster comparisons could be performed on a common temporal basis. The cohort was strongly right-skewed in scale. The average daily mean consumption had a mean of 0.819, a median of 0.137, and a maximum of 126.113. This asymmetry is analytically important: clustering based only on absolute load would overemphasize a few large clients, whereas clustering based only on normalized shape could suppress commercially relevant differences in scale.

To balance these effects, the workflow compared four alternative normalization approaches using daily and weekly scoring heuristics designed to assess residual seasonality, stability of weekday-weekend contrasts, and robustness under the observed missingness structure. The normalization stage was therefore treated as a model-selection step rather than as a routine preprocessing choice. Among the tested alternatives, Z-score normalization achieved the strongest combined performance and was retained for the clustering workflow.

The retained representation was intentionally complemented by a scale anchor. In practical terms, the normalized shape features support behaviour-family recognition, while the average daily mean preserves information about commercial weight inside the portfolio. This combination is necessary when the intended output is not only statistical grouping, but differentiated portfolio action.

The final feature space contained 25 variables: h00-h23 hourly shape indicators plus average daily mean as a scale anchor. Meteorological variables were deliberately excluded from the clustering stage. The purpose of clustering was to identify persistent behaviour families rather than weather sensitivity. Weather-responsive effects were instead reserved for the downstream forecasting layer, where they are methodologically more appropriate.

For the cohort-level segmentation, a silhouette-based sweep over candidate k values selected $k = 3$, with an approximately 0.204 silhouette score. The resulting clusters were interpreted jointly through client count, aggregate load scale, peak hour, weekend-to-weekday load ratio, and subsequent forecast behaviour. This multi-criterion reading was important because the target of interest was not the most compact cluster in a purely geometric sense, but the most operationally meaningful segment in portfolio terms.

A second-level subclustering step was then applied inside the dominant Cluster 2. The purpose of this step was to test whether the large mainstream daytime segment was internally homogeneous or whether it contained subgroups with materially different operational meaning. The resulting structure separated a large benchmarkable daytime subgroup, a smaller lighter-load daytime subgroup, and one

single exceptional anchor client that would otherwise remain hidden inside the largest actionable portfolio segment.

To test whether the behavioural segmentation had operational value beyond descriptive grouping, cluster-specific LightGBM models were trained for 168-hour forecasting horizons in four rolling windows. LightGBM was selected because gradient-boosted decision trees can efficiently capture nonlinear relationships, autoregressive dependence, and calendar effects in electricity time series [16-18].

The predictor set included short-lag and daily-recurring memory terms together with calendar structure. The most important variables in the fitted models were lag 1 h, lag 24 h, a 24-hour rolling feature, hour of day, and day of week. Model performance was evaluated using MAE, RMSE, R^2 , and WAPE. A persistence benchmark was added as a demanding but interpretable baseline. This comparison was methodologically essential: if the machine-learning models did not improve on persistence, the clustering would remain descriptive and would not justify operational deployment.

Behavioural interpretation was treated as probabilistic throughout the study. The available data support inference about operating schedules, relative scale, daytime regularity, and atypicality, but they do not support verified legal-sector assignment for each enterprise. Accordingly, the article uses cautious formulations such as behaviourally compatible examples or illustrative enterprise analogues rather than categorical sector identification.

The same logic also underpins the hypothetical-client concept. The empirical cluster and subcluster profiles can be used as privacy-preserving behavioural templates for scenario analysis, product design, and portfolio simulation. A hypothetical client is therefore not a claimed real enterprise type, but a synthetic profile assembled from empirically observed shape, scale, and forecastability characteristics [19-21].

Results and discussion

The full energy consumption data pool was separated into three behavioural families. Cluster 1 contained 223 clients (36.6%), Cluster 2 contained 382 clients (62.6%), and Cluster 3 contained 5 clients (0.8%).

The cluster indicators in Table 1 show that the portfolio is differentiated not only by size but also by operating rhythm and forecastability. Cluster 2 peaks at 14:00 and exhibits a weekend-to-weekday ratio of 0.63, which is consistent with a regular business-hours profile. Cluster 1 peaks later, at 16:00, and therefore suggests an afternoon-extended service or workshop-like rhythm. Cluster 3 achieves the lowest forecast error numerically, but because it contains only 5 clients, it is best treated as a niche or anchor-load segment rather than as a mass-market target.

Table 1

Portfolio-level summary of internal behavioural clusters

Group	Clients	Peak hour	Weekend/weekday	Agg. mean load	WAPE, %	Behavioural family (illustrative)
Cluster 1	223 (36.6%)	16:00	0.64	130.8	30.6	Afternoon-extended service/workshop-like segment
Cluster 2	382 (62.6%)	14:00	0.63	366.2	15.7	Mainstream daytime-commercial/service segment
Cluster 3	5 (0.8%)	16:00	0.66	10.7	13.0	Niche or anchor-load segment

Fig. 1 shows that Cluster 2 dominates the portfolio not only by the number of clients, but also by aggregate energy volume. Its load rises rapidly after the early morning hours, remains elevated from late morning until early evening, and then declines in a smooth evening transition. This is the profile most suitable for standardized benchmarking, baseline construction, and forecast-guided outreach because the segment is sufficiently large, sufficiently regular, and sufficiently synchronized.

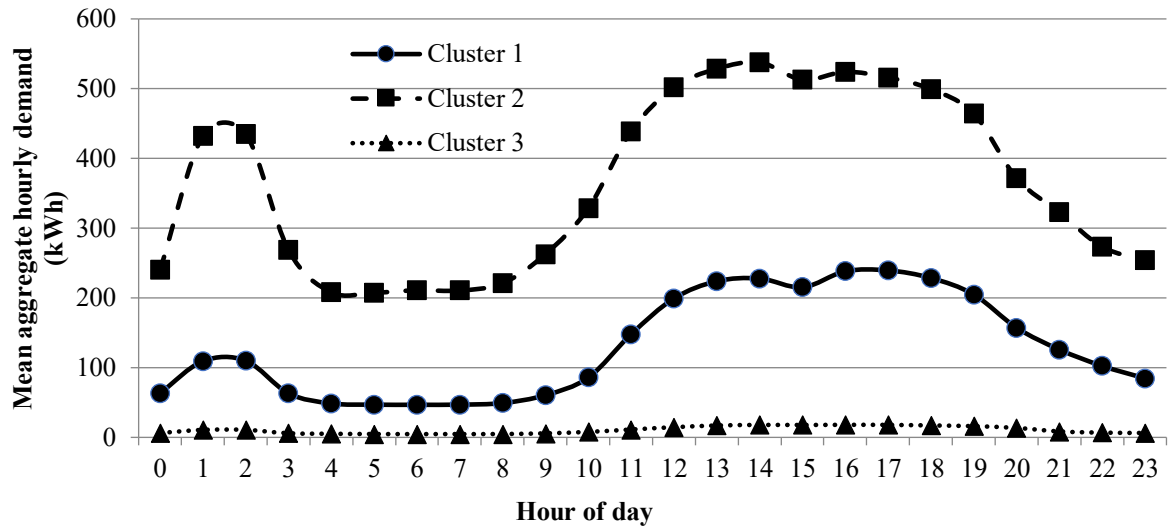


Fig. 1. Mean aggregate hourly electricity demand by internal portfolio cluster

When the same profiles are viewed in normalized form, the behavioural distinction remains clear (Fig. 2). Cluster 2 displays the most conventional weekday business pattern: low night-time demand, a strong daytime plateau, and a relatively smooth decline after the working day. Cluster 1 follows a similar daytime logic but shifts the relative peak later and descends more slowly, which is consistent with later operational activity.

Cluster 3 remains atypical even after normalization. Its profile is therefore not simply a smaller version of the mainstream daytime-commercial segment, but an unusual load family that is better handled as a special-case or key-account category. The portfolio implication is important: the most actionable segment is not merely the largest one, but the largest one which behaviour is also sufficiently regular for scalable intervention.

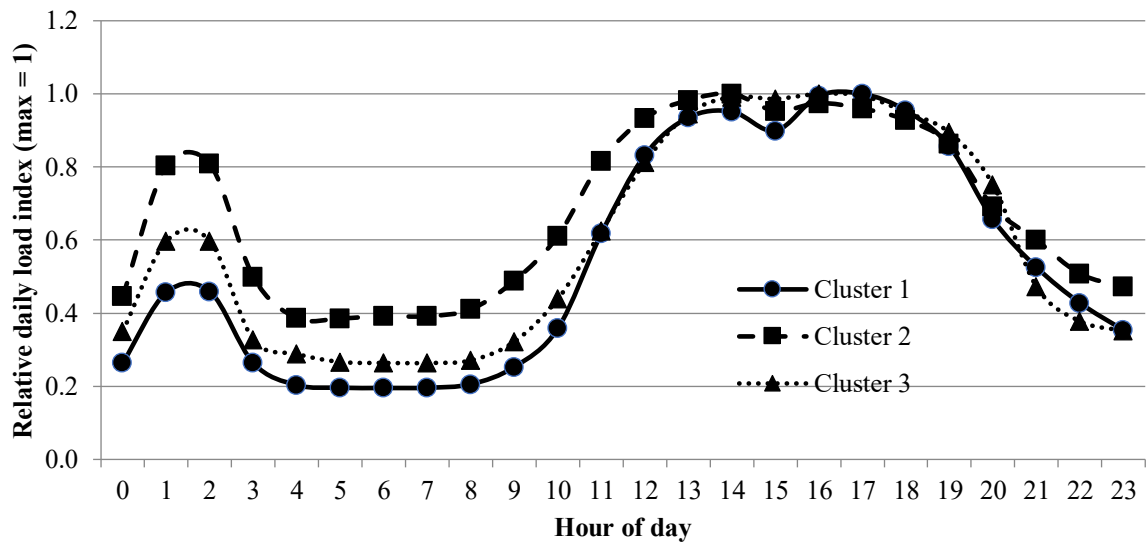


Fig. 2. Relative daily load index by internal portfolio cluster

The current dataset supports behaviour-family recognition more strongly than exact sector recognition. This is a strength for operational targeting, but also a clear boundary condition for scientific interpretation. The clusters reveal operating schedules, relative scale, weekday-weekend contrast, and atypicality; they do not prove verified legal-sector membership for each enterprise.

At portfolio level, the evidence is most consistent with three behavioural families. Cluster 2 can be read as a mainstream daytime-commercial and service segment the operating rhythm of which is compatible with offices, clinics, pharmacies, professional services, showrooms, beauty and personal services, and comparable weekday-oriented premises. Cluster 1 is more consistent with later-running

service, workshop, depot, or mixed service-production activity. Cluster 3 is best interpreted not as a sector, but as a small set of atypical or anchor-load sites that require case-specific handling. These examples are illustrative analytical hypotheses rather than validated firm labels.

This interpretation is aligned with prior load-benchmarking and customer-segmentation studies that identify repeated commercial load-shape families without always having complete business metadata [6-15]. The present contribution is to translate those behavioural families into a portfolio hierarchy that is directly usable for targeting, while retaining explicit caution about what the data can and cannot identify.

The dominant Cluster 2 was not internally homogeneous. The subclustering step separated Subcluster 0 with 118 clients (30.9% of Cluster 2), Subcluster 1 with a single client (0.3%), and Subcluster 2 with 263 clients (68.8%). This result matters operationally because a large behaviourally coherent daytime segment may still contain both scalable mainstream clients and rare cases that should be removed from mass treatment.

Fig. 3 visualizes these internal differences on a relative scale. The Fig. confirms that Subcluster 2 is the benchmarkable mainstream of the daytime-commercial portfolio. It follows a smooth business-hours trajectory and is therefore the strongest candidate for peer benchmarking, standardized advisory bundles, and forecast-guided commercial products.

Subcluster 0 remains a daytime group, but at lower scale and with a flatter profile. It is therefore behaviourally closer to smaller service premises where simple occupancy alignment, equipment shutdown routines, and low-cost controls may already have measurable value. Subcluster 1 is qualitatively different. Its profile indicates a single unusual site which treatment should be individualized rather than absorbed into a mass-market programme.

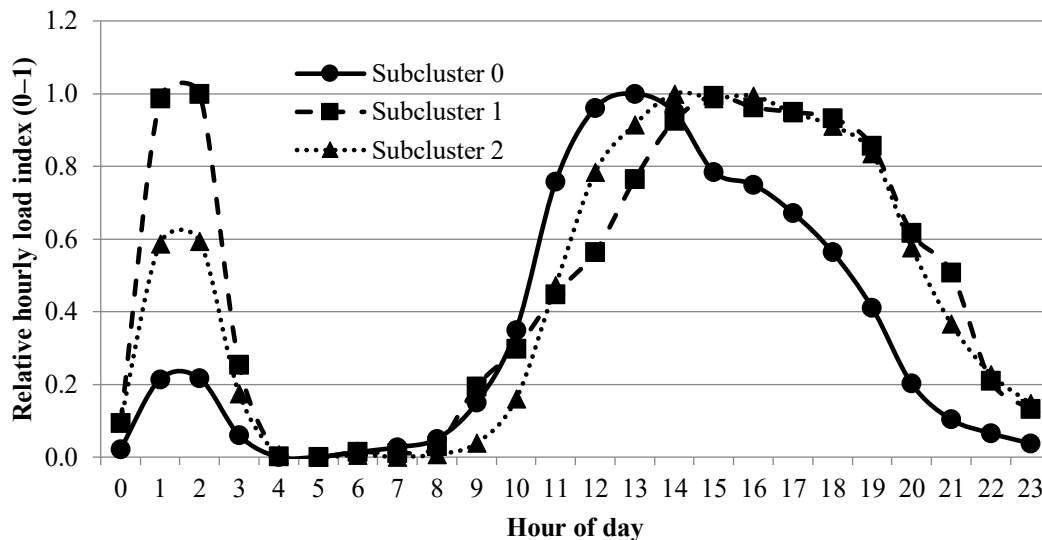


Fig. 3. Relative hourly load index for Cluster 2 subclusters

The subclustering result also clarifies an important portfolio principle: behavioural similarity does not eliminate commercial heterogeneity (Table 2). Even within the dominant daytime segment, a one-client anchor case can materially alter the load structure and should be separated before standardized intervention packages are designed.

Forecasting results further support the portfolio prioritization of Cluster 2. Although Cluster 3 achieved the lowest absolute error numerically, its very small client count prevents it from serving as the main standardized action segment. Cluster 2, by contrast, combines size, interpretability, and forecastability. Across four 168-hour windows, its mean WAPE was 15.7%, compared with 33.7% for the persistence baseline. This margin demonstrates that the dominant daytime-commercial segment is not only behaviourally coherent, but also forecastable enough for operational use.

Cluster 1 also improved relative to persistence, but its mean WAPE remained materially higher at 30.6%. In practical terms, this means that later-running service and workshop-like portfolios may still

benefit from forecasting, but the dominant commercial daytime segment provides the strongest balance between model performance and rollout potential.

Table 2

Internal structure of Cluster 2 with cautious behavioural interpretation

Group	Clients	Load share, %	Load/client ratio	Peak hour	Indicative behavioural reading and illustrative enterprise examples
Sub-cluster 0	118 (30.9%)	21.6	0.70	13:00	Smaller or flatter daytime SMEs; behaviourally compatible examples include micro-offices, salons, local services, and lightly occupied commercial premises where occupancy alignment and standby-load control may matter.
Sub-cluster 1	1 (0.3%)	34.9	134.35	02:00	Single exceptional anchor client; behaviourally compatible examples include refrigerated retail, laboratory or medical premises, mixed-use technical sites, or another unusually large daytime customer. This is a special case, not a scalable subgroup.
Sub-cluster 2	263 (68.8%)	43.4	0.63	14:00	Mainstream benchmarkable daytime SMEs; behaviourally compatible examples include offices, outpatient services, pharmacies, showrooms, and routine commercial-service sites suitable for standard portfolio offers.

The predictor-importance pattern helps explain why Cluster 2 is operationally attractive. Before the feature chart is read, two points should be noted. First, the fitted models rely primarily on short-lag memory and calendar regularity, which is exactly what should be expected in portfolios governed by repeated commercial opening hours. Second, differences between clusters are meaningful: they indicate that the same forecasting engine exploits different combinations of daily rhythm and short-term recurrence in each behavioural family.

Fig. 4 confirms that lag 1 h is the dominant predictor in all three clusters, while hour of day and lag 24 h are also consistently important. The predictive anatomy is strongest in Cluster 2, where short-term memory, daily recurrence, and calendar timing all carry high importance. This is exactly the structure expected for a business-hours portfolio with repeatable operations and limited process volatility.

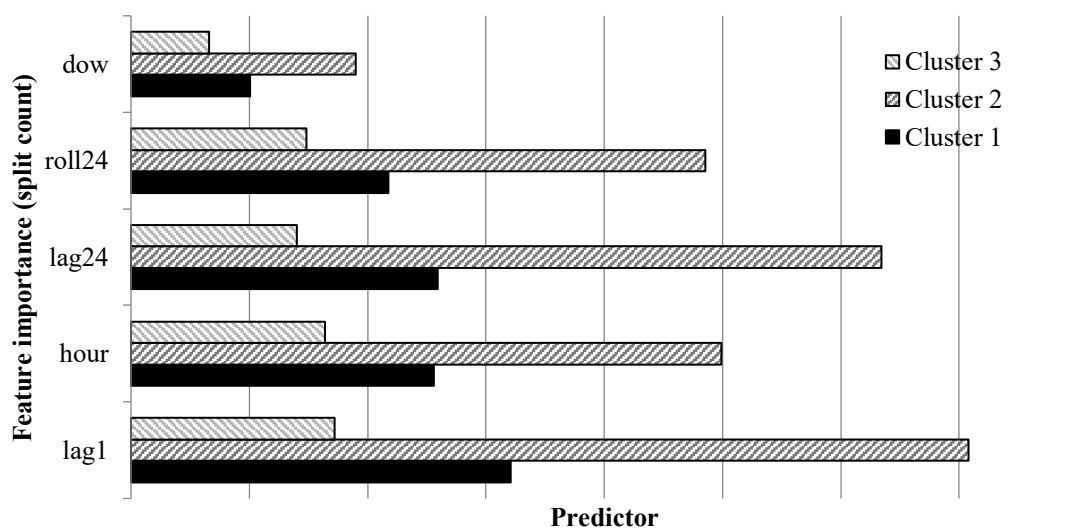


Fig. 4. Top forecasting features by internal cluster

From a scientific perspective, this result links the segmentation to the forecasting literature: the clusters are not arbitrary geometric groups, but load families organized around repeated commercial

routines [16-18]. From a practical perspective, it explains why forecast-guided portfolio actions are already defensible for the dominant daytime-commercial segment.

The value of the portfolio structure lies in differentiated actionability. Because the clusters and subclusters combine different sizes, operating schedules, and forecast profiles, the same electricity-service offer is not equally credible for all groups. Table 3 summarizes the intervention pathways that are already supported by the present evidence.

Several additional important topics for discussion follow from these stakeholder implications.

First, the present smart-meter data are already sufficient for portfolio triage: the cohort can be separated into a mainstream mass segment, a secondary later-running segment, and a small anchor-load segment. Second, Cluster 2 is large enough and regular enough to justify standardized offers. Third, the main scientific caution remains unchanged: the portfolio structure supports behaviour-based targeting, not verified legal-sector classification.

The research provided is the next step towards commercial load-shape benchmarking and smart-meter clustering by making the portfolio itself the central analytical object [6-15]. Earlier studies have shown that commercial load families can be recognized and used for benchmarking or forecast support. Research results add a more explicit operational layer: it shows how a mixed SME cohort can be converted into a ranked portfolio structure that distinguishes scalable mainstream segments from secondary and exceptional groups.

Table 3

Portfolio-level and subcluster-level stakeholder actions

Group	DSO interest	Retailer offer already feasible	Consultant recommendation
Cluster 1	Late-afternoon local peaks, feeder timing, and identification of workshop-like demand pockets.	Late-day time-of-use guidance, demand-cap products, and targeted advisory linked to closing-hour behaviour.	Schedule control, workshop ventilation, motor or compressed-air checks, and late-day HVAC and lighting optimization.
Cluster 2	Mainstream daytime SME portfolio; strongest candidate for scalable flexibility screening and non-wires messaging.	Benchmark-led portfolio offers, dynamic-price guidance, PV self-consumption or storage propositions, and simple flexibility products.	Standardized commercial packages: HVAC schedules, lighting retrofit, standby-load reduction, and occupancy-aligned operation.
Cluster 3	Site-specific capacity planning, exceptional-load review, and bespoke flexibility dialogue.	Key-account pricing, tailored imbalance support, and bespoke contract design rather than mass offers.	Detailed audit, submetering, engineering diagnosis, and bespoke measure design.
Cluster 2 /Subcluster 0	Lower-scale daytime sites suitable for low-cost behavioural measures and operational screening.	Simpler tariff optimization, low-complexity advisory bundles, and targeted off-peak guidance.	Occupancy schedules, standby reduction, equipment shutdown routines, and low-cost controls.
Cluster 2 /Subcluster 1	Single anchor-load client requiring dedicated network and load-risk assessment.	Bespoke commercial negotiation, dedicated forecast monitoring, and account-manager treatment.	Full technical audit, process-level diagnostics, and custom baseline or M&V.
Cluster 2/ Subcluster 2	Largest scalable daytime segment; strongest candidate for standardized portfolio action.	Mass outreach, peer benchmarking, PV self-consumption offers, and simple flexibility or response packages.	Package-type advice: HVAC, lighting, controls, occupancy alignment, and recurring performance reviews.

This portfolio logic matters scientifically because actionability depends on more than cluster existence. A small but statistically clean cluster may have limited deployment value, whereas a large and behaviourally regular cluster can support practical baseline construction, peer benchmarking, and standardized intervention design. In this study, Cluster 2 dominates precisely because it combines all three properties: it contains 382 clients, carries the highest aggregate mean load among the scalable groups, and reaches a WAPE of 15.7% compared with 33.7% for persistence.

Although the empirical data come from Latvia, the methodological message is transferable. Many utilities, aggregators, and market actors in other countries face the same portfolio condition: they possess hourly meter data but lack consistently linked business-register, premise, or process metadata. In such settings, a portfolio-structured workflow can still separate mainstream daytime-commercial groups from late-running service groups and rare exceptional clients, even before full supervised classification becomes feasible.

At the same time, transferability should be understood in behavioural rather than administrative terms. The specific proportions observed here - for example, the dominance of Cluster 2 at 62.6% of the cohort - are sample specific. What is more likely to generalize is the analytical pattern: a mainstream daytime-commercial mass segment, a secondary schedule-shifted segment, and a small number of anchor-load exceptions inside a mixed SME portfolio.

The main limitation is the absence of verified linked metadata at the meter level. Without business-register or NACE labels, contracted capacity, tariff information, premise characteristics, refrigeration indicators, or power-quality variables, the study cannot validate exact enterprise-type recognition. The responsible scientific position is therefore to treat the behavioural readings as plausible but unverified hypotheses. This limitation does not invalidate the portfolio structure, but it defines the boundary of current recognition precision.

Conclusions

1. The study analysed 610 Latvian daytime-commercial SMEs on a common annual grid of 8,759 hourly slots. This corresponds to 0.571% of the 106,886 SMEs, which makes the cohort contextually small but still operationally meaningful.
2. The portfolio is dominated by one mainstream daytime-commercial segment (with 382 clients (62.6%)) plus a secondary later-running group (223 clients (36.6%)) and a small anchor-load group.
3. Daytime-commercial Cluster 2 is the primary rollout target because it peaks at 14:00, has a weekend-to-weekday ratio of 0.63, and carries the highest aggregate mean load among the scalable groups (366.2), which together indicate a large, regular, business-hours portfolio.
4. Subclustering of Cluster 2 revealed Subcluster 0 with 118 clients (30.9% of Cluster 2), Subcluster 1 with 1 client (0.3%), and Subcluster 2 with 263 clients (68.8%), showing that even the dominant daytime segment still requires separation between scalable mainstream clients and a single exceptional anchor case.
5. Forecast-guided use is already feasible: Cluster 2 achieved a mean WAPE of 15.7% across four 168-hour windows, compared with 33.7% for the persistence baseline; Cluster 1 reached 30.6%, while Cluster 3 reached 13.0% but remained too small for mass-market treatment.
6. The current data are strong enough for behaviour-family recognition and stakeholder targeting, but not for verified legal-sector identification; future linkage with business, tariff, and premise metadata is the next required step toward higher recognition precision.

Acknowledgements

The research paper was created as a part of the research project “Development of a new solution providing flexibility services for effective network overload management using the demand response potential of electricity end-users”, No. 2.2.1.3.i.0/1/24/A/CFLA/003, realized within the framework of the European Union Recovery Fund, by AS “Sadales tīkls” and SIA “Tet”. The project is co-financed by the Recovery Fund within the framework of the Operational Program Latvia’s Recovery and Resilience Plan Measure 2.2 Reform and Investment Direction ‘Digital Transformation and Innovation of Enterprises’ Investment 2.2.1.3.i ‘Support for the introduction of new products and services in businesses.

The authors gratefully acknowledge TET Ltd for providing access to the smart-meter dataset used in this study. The research was supported by the Faculty of Engineering and Information Technologies at the Latvia University of Life Sciences and Technologies. The computational framework applied in this work was developed within the scope of collaborative research activities between the authors' institutions.

Author contributions

Conceptualization, A.L.; methodology, A.L. and A.B.; software, N.R.; validation, A.L. and M.T.; formal analysis, A.L. and A.B.; investigation, A.L., A.B., M.T., and N.R.; writing - original draft preparation, A.L.; writing - review and editing, all authors; visualization, A.L. and N.R.; supervision, A.L. All authors have read and agreed to the published version of the manuscript.

References

- [1] Backman C. Drivers for the adoption of eco-efficiency in SMEs. *Energies*, 15(3), 2022, 1072. DOI: 10.3390/en15031072.
- [2] Lisowski B., Lisowska R., Szymanska K. Facing Environmental Goals for Energy-Efficiency Improvements in Micro and Small Enterprises Operating in the Age of Industry 4.0. *Energies*, 15(18), 2022, 6709. DOI: 10.3390/en15186709.
- [3] Konig W., Lobbe S., Buttner S., Schneider C. Establishing Energy Efficiency - Drivers for Energy Efficiency in German Manufacturing Small- and Medium-Sized Enterprises. *Energies*, 13(19), 2020, 5145. DOI: 10.3390/en13195145.
- [4] European Commission. 2025 SME Country Fact Sheet: Latvia. Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs/Joint Research Centre, 2025.
- [5] Central Statistical Bureau of Latvia. Economically active enterprises of market sector in regions, State cities and municipalities by size group according to the number of employees and main economic activity (NACE Rev. 2) units, CSB database.
- [6] Luo X., Hong T., Chen Y., Piette M.A. Electric load shape benchmarking for small- and medium-sized commercial buildings. *Applied Energy*, 204, 2017, pp. 715-725. DOI: 10.1016/j.apenergy.2017.07.108
- [7] Komatsu H., Kimura O. A combination of SOM-based operating time estimation and simplified disaggregation for SME buildings using hourly energy consumption data. *Energy and Buildings*, 201, 2019, pp. 186-197. DOI: 10.1016/j.enbuild.2019.07.037
- [8] Komatsu H., Kimura O. Peak demand alert system based on electricity demand forecasting for smart meter data. *Energy and Buildings*, 225, 2020, 110307. DOI: 10.1016/j.enbuild.2020.110307
- [9] Komatsu H., Kimura O. Customer segmentation based on smart meter data analytics: Behavioral similarities with manual categorization for building types. *Energy and Buildings*, 283, 2023, 112831. DOI: 10.1016/j.enbuild.2022.112831
- [10] Biscarri F., Monedero I., Garcia A., Guerrero J.I., Leon C. Electricity clustering framework for automatic classification of customer loads. *Expert Systems with Applications*, 86, 2017, pp. 54-63. DOI: 10.1016/j.eswa.2017.05.053
- [11] Melzi F.N., Same A., Zayani M.H., Oukhellou L. A Dedicated Mixture Model for Clustering Smart Meter Data: Identification and Analysis of Electricity Consumption Behaviors. *Energies*, 10(8), 2017, 1124. DOI: 10.3390/en10081124
- [12] Banales S., Dormido R., Duro N. Smart Meters Time Series Clustering for Demand Response Applications in the Context of High Penetration of Renewable Energy Resources. *Energies*, 14(12), 2021, 3651. DOI: 10.3390/en14123651
- [13] Bogensperger A., Fabel Y. A practical approach to cluster validation in the energy sector. *Energy Informatics*, 4, 2021, Article 18. DOI: 10.1186/s42162-021-00177-1
- [14] Quintana M., Arjunan P., Miller C. Islands of misfit buildings: Detecting uncharacteristic electricity use behavior using load shape clustering. *Building Simulation*, 14, 2021, pp. 119-130. DOI: 10.1007/s12273-020-0724-0
- [15] Auder B., Cugliari J., Goude Y., Poggi J.-M. Scalable Clustering of Individual Electrical Curves for Profiling and Bottom-Up Forecasting. *Energies*, 11(7), 2018, 1893. DOI: 10.3390/en11071893

-
- [16] Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Advances in Neural Information Processing Systems* 30, 2017.
- [17] Dewangan F., Abdelaziz A.Y., El-Shahat A., Ibrahim A., Elavarasan R.M. Load Forecasting Models in Smart Grid Using Smart Meter Information: A Review. *Energies*, 16(3), 2023, 1404. DOI: 10.3390/en16031404
- [18] Oh J., Min D. Prediction of energy consumption for manufacturing small and medium-sized enterprises (SMEs) considering industry characteristics. *Energy*, 300, 2024, 131621. DOI: 10.1016/j.energy.2024.131621
- [19] Reiter J.P. Releasing multiply imputed, synthetic public use microdata: an illustration and empirical study. *Journal of the Royal Statistical Society: Series A*, 168(1), 2005, pp. 185-205. DOI: 10.1111/j.1467-985X.2004.00343.x
- [20] Hong T., Macumber D., Li H., Fleming K., Wang Z. Generation and Representation of Synthetic Smart Meter Data. *Building Simulation*, 13(6), 2020, pp. 1205-1220. DOI: 10.1007/s12273-020-0661-y
- [21] Fu C., Kazmi H., Quintana M., Miller C. Creating synthetic energy meter data using conditional diffusion and building metadata. *Energy and Buildings*, 312, 2024, 114216. DOI: 10.1016/j.enbuild.2024.114216