Exploring Data-Driven Approaches for Short-Term Load Forecasting of an Energy Utility's Customer Portfolio

António Luís Lobo Pimentel

Instituto Superior Técnico, Universidade de Lisboa, Portugal
June 2024

Abstract

Energy utilities play a crucial role in the industry by balancing load demand and supply, integrating renewable energy sources, and maintaining grid reliability. From a utility's perspective, accurate Short-Term Load Forecasting (STLF) is essential for optimizing decisions such as the scheduling of generation units and the participation in energy markets, enabling an overall reduction in operational costs. Traditionally, STLF of a utility's customer portfolio has been conducted using statistical regression models, which, although reliable, have limitations in capturing complex consumption patterns and adapting to dynamic market conditions. This dissertation identified an opportunity to improve STLF methodologies by using Machine Learning (ML) techniques to address these gaps. To achieve this, an initial analysis on customer load segmentation was performed using clustering algorithms to categorize consumption patterns. Subsequently, for the forecasting task, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LGBM), and Extra-Trees Regressor (ETR) methods were selected and tuned through Bayesian optimization for each segment. Considered data included measurements of load, temperature, and radiation, and calendar features to model consumption patterns. The developed models were later benchmarked against a statistical approach. Results showed that all ML methods outperformed the statistical regression method in all load segments, with XGBoost and LGBM presenting the best performances. Furthermore, pairing ML models with a segmentation approach using three clusters enabled more accurate forecasts than the traditional injection-focused segmentation method. This highlighted the potential of state-of-the-art supervised and unsupervised ML techniques to enhance forecasting performance within a utility's customer portfolio.

Keywords: Energy Utility; Short-Term Load Forecasting; Load Segmentation; Gradient Boosting; Clustering; Machine Learning.

1. Introduction

The global energy landscape has been subject to significant transformations in recent decades, marked by the liberalization of energy markets and the integration of renewable energy sources. These changes require utilities to adapt to changing conditions while still ensuring a consistent and reliable supply of electricity. Power utilities play a crucial role in balancing electricity supply and demand, integrating diverse energy sources, and maintaining grid stability. As the energy market becomes more competitive and complex, precise and efficient operational strategies become increasingly necessary.

Short-Term Load Forecasting (STLF) is a critical component of a utility's operational framework. Accurate load demand predictions enable utilities to optimize the scheduling of their generation assets and to participate effectively in energy markets. This allows them to minimize operational costs and enhance the overall efficiency of the power system. The increasing penetration of renewable energy sources and the dynamic nature of utilities' portfolios add uncertainty to the STLF task, further emphasizing the importance of adopting robust forecasting methodologies.

Despite the importance of STLF for utility companies, few studies have focused on forecasting load of a customer portfolio. This dissertation aims to provide an effective and standardized framework for this purpose by identifying specific limitations in traditional approaches, exploring methodologies that can potentially improve STLF, developing and applying the most promising approaches, and analysing the obtained results. Traditionally, STLF has made use of statistical regression models, which, although reliable, often fail to capture complex consumption patterns. This article tackles this by exploring the potential of supervised and unsupervised Machine Learning (ML) techniques, such as regression and clustering, respectively, to improve overall forecasting accuracy and reliability within the context of a utility's customer portfolio.

The remainder of this article is organized as follows: Section 2 presents a literature review focusing on STLF methodologies and its intricacies; Section 3 presents the case study to be addressed and specifies the methodologies and data to be used; Section 4 describes the procedures leading to the acquisition of the customer load segmentation results; Section 5 includes a description of the procedures leading to the acquisition of the predictive results; Section 6 presents a discussion of the obtained results; and Section 7 presents the article's conclusions, limitations, and opportunities of future investigation.

2. Literature Review

Energy consumption forecasting, also known as load forecasting, is used to predict electricity usage in specific regions, areas, buildings or power grid connections. It is essential for utilities to manage supply, demand, and to optimize energy production and distribution, preventing power outages and leading to improved operational efficiency [1].

Load forecasts can be categorized by their time scales [2]: long-term (years ahead), medium-term (months or quarters), short-term (daily), and ultra-short-term (hourly or sub-hourly). Most studies focus on forecasts with shorter horizons [3], which are mainly used for operational energy management. In this context, the primary challenges entail correctly addressing the influence of the multiple weather factors, calendar factors, and the growing integration of renewable sources [4]. In turn, for medium to long-term forecasts, economic and demographic growth, access to electricity, and technological progress are the most critical factors [5]. As the load prediction context addressed in this article comprehends day-ahead consumption, the following review will focus on STLF.

Load forecasting aims to estimate future values of a time-dependent function based on known historical data within a specific interval [6]. Forecasting methods can either be based on multi-factor data, examining causal relationships, or on time-series data, relying solely on historical load patterns. Although some researchers tend to prefer this last approach, which comprises a significantly faster and simpler implementation [2], advances in modern science have significantly enhanced load forecasting technology, reducing the impact of these limitations [7]. Load forecasting methods can also be divided into traditional statistical techniques and ML techniques [8], with ML having been gaining a growing popularity due to its effectiveness. Subsections 2.1 to 2.6 present a review of different methods commonly used for STLF. These were selected based on multiple state-of-the-art review articles [1, 3] and on self-conducted research. Furthermore, subsections 2.7, 2.8, and 2.9 address relevant factors for STLF, load segmentation methodologies used, and model tuning approaches, respectively.

2.1. Statistical Regression-Based Methods

Statistical regression methods utilize either linear or non-linear regression analysis to identify causal relationships between variables [1]. Despite being often considered outdated, these models are still used today because their regressed coefficients offer clear implications for validation and explanation [9], making them suitable for benchmarking against more complex models [10]. Moreover, transforming input variables using other functions, such as polynomials, can broaden the applicability of linear regression.

Various studies have explored and compared different regression-based methods for STLF. Mir et al. [11] compared generalized linear regression models with those incorporating polynomials and cross-terms, finding the latter more accurate and interpretable. Kiani and Zeng [6] employed a linear regression approach for utilities in Pennsylvania, New Jersey, and Maryland, demonstrating higher accuracy compared to a basic ARIMA model (see Subsection 2.2). Dudek [12] developed univariate models based on linear regression and daily cycle patterns, and concluded that the linear models demonstrated better extrapolation properties compared to more complex models by avoiding overfitting.

2.2. Time Series Methods

Time series methods are common statistical approaches for STLF. The ARIMA model, a widely used time-series modeling technique [13], is known for its simplicity and ability to capture temporal dependencies. Yet, it has limitations such as assuming linear relationships and stationarity, and not capturing long-term seasonal patterns [1]. To address these limitations, a Seasonal ARIMA (SARIMA) model can be used.

Another frequently used approach is Exponential Smoothing (ES). ES models are computationally efficient, require less historical data, and are effective at capturing non-linear relationships between load and exogenous factors [1].

Various studies have assessed the performance of these time series methods. Dada [13] concluded that ARIMA and SARIMA models outperform ML approaches for shorter learning periods. Taylor and McSharry [14] found that a double seasonal Holt-Winters method outperformed other models, including ARIMA, for electricity demand predictions due to its ability to account for both intra-day and intra-week seasonalities in demand data. Conversely, Dudek [12] reported that univariate linear models outperformed both ARIMA and ES methods for STLF on four different datasets, especially for shorter horizons, with ES models consistently outperforming ARIMA models.

2.3. Artificial Neural Networks

Artificial Neural Networks (ANNs) are ML techniques inspired by the human brain's functioning, widely used in load forecasting. Although these models can easily model non-linear problems, they are computationally expensive when using large networks [3]. Becirovic and Cosovic [8] tested a three-layered feed-forward ANN, showing satisfactory daily and hourly load forecasts. Nwulu and Agboola [15] developed an ANN model for forecasting consumers' load demand in North Cyprus, concluding that the model's effectiveness was dependent on sufficiently-sized training and testing datasets.

2.4. Deep Learning Methods

Deep learning is enabled through Deep Neural Networks (DNNs), which consist of ANNs with multiple hidden layers. These allow models to learn data representations with various levels of abstraction and identify intricate patterns within extensive datasets through back-propagation [16]. However, they require a careful application due to their susceptibility to overfitting and high computational intensity [17]. Most DNNs used for STLF employ two specific architectures: Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs).

CNNs are a specialized category within DNNs renowned for their ability to organize input units using convolutional layers and parameter sharing, thereby mitigating the risk of overfitting [3]. Tudose et al. [18] employed CNN-based models for STLF during the COVID-19 pandemic, achieving lower error values compared to traditional methods. Zhang et al. [19] demonstrated promising results with a novel hybrid methodology for STLF incorporating Temporal Convolutional Networks (TCNs). Additionally, Eskandari et al. [20] utilized CNNs for feature extraction in STLF, enhancing forecasting capability when combined with RNNs.

RNNs are supervised ML models with feedback loops that enable them to learn features and time dependencies from sequential and time-series data [21], making them suitable for energy prediction. Since regular RNNs may face issues with long-term dependencies due to the exploding and vanishing gradient problems, Long Short-Term Memory (LSTM) networks can be used. Eskandari et al. Eskandari et al. [20] combined CNNs and RNN-based methods (LSTM and Gated Recurrent Unit) for STLF, demonstrating high accuracy and robustness even with noisy data. Slowik and Urban [22] developed a single-layered LSTM model for predicting short-term energy demand in a manufacturing plant, finding it more efficient than double-layered LSTM and CNN models. Shabbir et al. [23] proposed an LSTM algorithm for three-day ahead energy forecasting in Estonia, achieving better accuracy compared to the existing methods used by the Estonian energy regulatory authority.

2.5. Support Vector Regression

Support Vector Machines (SVM) is a supervised learning method widely applied in power system applications [1]. For regression purposes, SVM is known as Support Vector Regression (SVR). It is effective for high-dimensional data, at the cost of slow convergence rates and high execution times [24]. Ceperic et al. [25] presented an SVR-based strategy for STLF using datasets from ISO New England and North America, achieving high accuracy and emphasizing the importance of input variables, feature selection, and hyper-parameter optimization. Moradzadeh et al. [26] proposed a hybrid model combining SVR and LSTM for STLF in a rural microgrid in Africa, showing that the hybrid model outperformed conventional SVR and LSTM models in predictive accuracy.

2.6. Ensemble Learning Methods

Ensemble methods combine the output of individual learning algorithms in order to enhance the prediction performance [3]. Common ensemble methods can be classified as bagging, boosting, and stacking techniques.

Bagging, or bootstrap aggregating, involves training multiple instances of the same learning algorithm on different subsets of the training data [3], with the final prediction being the average or majority vote of the individual predictions. Wang et al. [27] applied a bagging ensemble using a Random Forest Regression model, demonstrating its superiority over Regression Trees and SVR. Dada [13] found that ML ensemble methods, including Extra-Trees Regression (ETR), outperformed ARIMA and SARIMA models for day-ahead load forecasting. Zdravevski et al. [28] compared ETR and Random Forest Regression, concluding that ETR was slightly more accurate and faster.

Boosting involves sequentially training models where each corrects the errors of its predecessor by re-weighting the original training data. Gradient Boosting Machine (GBM)-based algorithms, such as Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LGBM), are popular for their flexibility and interpretability, achieving significant advances in energy sector predictions [29]. Nziyumva et al. [30] demonstrated that a hybrid XGBoost and LGBM model outperformed individual models and other approaches such as AdaBoost and Random Forest Regression. Daoud [31] highlighted XGBoost's scalability and robustness due to its regularization technique, while LGBM excelled in training speed and accuracy for large datasets. Miller et al. [32] noted that the most accurate frameworks used in the 2019 ASHRAE load forecasting competition predominantly used GBM models, particularly LGBM.

Stacking, unlike bagging and boosting, uses diverse models and combines their outputs through a meta-model for final predictions. Several STLF studies [29, 33] developed ensemble models combining multiple ML regressors, and all concluded that the ensemble outperformed individual models.

2.7. Feature Selection

Attributes used in STLF contextualize events and enhance model performance by establishing dependencies with the predicted variable. Electricity demand is primarily influenced by weather conditions and calendar types [4]. Key meteorological variables impacting electricity consumption comprise temperature and radiation [3]. As for calendar factors, different consumption patterns can typically be observed within each day; between weekdays, weekends, and holidays, varying by consumer type [14]; and within each year, due to seasonal variations [25].

Feature selection is critical in STLF to simplify models, reduce training times, and prevent overfitting [1, 34]. Yet, the choice of the feature selection method depends on the dataset and forecasting model. For instance, Ceperic et al. [25] found stepwise regression most effective for enhancing model accuracy, Porteiro et al. [34] used recursive feature elimination to assess feature importance, and Yang et al. [35] employed auto-correlation for time series feature selection.

2.8. Customer Load Segmentation

Customer load segmentation is essential for utilities to enhance predictive accuracy by defining hierarchies and segments within their diverse customer portfolios. Spiliotis et al. [36] identified three common aggregation approaches: bottom-up, top-down, and middle-out. Biscarri et al. [37] developed a clustering methodology for a Spanish utility, selecting data based on contracted power and location, applying feature selection to reduce dimensionality, and using hierarchical clustering to categorize customers based on their daily behaviors. Grigoraş et al. [38] highlighted the effectiveness of hierarchical and k-means clustering methods, emphasizing the importance of data scaling and validation measures like the Silhouette indicator. Tureczek and Nielsen [39] reviewed classification methodologies using smart meter data, noting prevalent methods such as k-means and hierarchical clustering with multiple distance measures, and highlighting the popularity of validity tests with indices such as Davies-Bouldin and Silhouette for evaluating clustering quality.

2.9. Training, Testing, and Validation

In the context of STLF, data splitting methods must consider the time series nature of load data, favoring chronological splits to prevent leakage of future information into the training set, as performed in [6, 9, 23, 36]. Studies have also highlighted the importance of model validation through hyper-parameter tuning for enhancing model performance and preventing issues such as overfitting or underfitting [10]. This entails refining a model's performance by fine-tuning its parameters based on predefined evaluation criteria. Within the reviewed literature, common hyper-parameter tuning algorithms include grid search [34], random search [40], and Bayesian optimization [9]. Moreover, for performance evaluation, the most used criteria were the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), and the Mean Absolute Percentage Error (MAPE), as shown in [13, 34].

2.10. Research Questions

The reviewed literature suggests that there is a clear opportunity for further exploration of several aspects intricate to STLF for a utility's portfolio, particularly customer load segmentation and forecasting model selection. Hence, this article will focus on answering the following research questions:

1. Can the performance of short-term load forecasting within a utility's customer portfolio be enhanced with the adoption of state-of-the-art predictive models?

2. What customer load segmentation approaches within a utility's portfolio may improve forecasting performance?

3. Methodology and Case Study Definition

This chapter describes the article's case study and particularizes the methodology to be followed.

3.1. PZEM Case Study

PZEM, an energy utility founded in 1919, operates as a producer, trader, and supplier of electricity and gas in The Netherlands. PZEM manages a varied energy generation mix, including wind turbines, solar panels, and gas-fired units, along with a diverse customer portfolio. Within its operational landscape, a focus is placed on optimizing the company's assets while still fulfilling the customer's demand. To achieve this, accurate day-ahead load consumption forecasts become paramount.

PZEM's forecasting framework provides exemplary opportunities to investigate the STLF task due to the diversity and scale of its operations, and to the additional complexity introduced by the growing integration of renewable energy sources into the Dutch energy landscape.

3.2. Customer Load Segmentation

Within PZEM's STLF framework, it is impractical to develop a predictive model for each individual power meter on the customer portfolio, due to its size. Instead, meters are grouped based on their main characteristics (load data type, location, and daily consumption patterns). Segmentation procedures are used mainly to distinguish different daily load patterns, as meter's location and data type are typically known. For this purpose, a clustering procedure will be applied as proposed in [37]. This will enable an assessment of the performance impact of using a standardized and state-of-the-art ML approach for customer load segmentation.

3.3. Data Selection

Considered data for PZEM's STLF includes load, temperature, and radiation measurements. The customer load data for PZEM's STLF can be analysed at various levels of abstraction. As this article focuses on improving the load segmentation process, load data will be extracted at the individual grid connection level. Additionally, given the broad customer portfolio of the utility, the scope of the research was narrowed to a group of customers comprehending telemetric grid connections and belonging to a central region of The Netherlands, selected for its representative dataset, high connection count, and significant annual load volume. Currently these grid connections are segmented into two groups (Regular and Injection) based on the impact of radiation in their load profiles. In turn, weather data will include averaged regional measurements for radiation, temperature, and 24-hour prior temperature. All data will be extracted with a 1-hour granularity and restricted to the year of 2023, ensuring consistency and avoiding the portfolio changes that typically take place each calendar year.

3.4. Predictive Method Selection

The literature review on STLF methodologies suggests that models' predictive performance significantly depends on the characteristics of the considered dataset. For this particular study, focusing on hourly load and weather data over one year, the chosen methods should be accurate, reliable, generalize well, and be computationally efficient. Matching these criteria with the reviewed literature, the most promising methods are XGBoost, LGBM, and ETR. XGBoost is known for its versatility and regularization abilities, LGBM offers high accuracy and faster training times, and ETR is robust with simpler hyper-parameter tuning. Despite the popularity of LSTM models, their need for larger datasets makes them less suitable for this study. The selected methods will be developed and assessed following methodologies from recent studies and benchmarked against a traditional statistical regression-based forecasting procedure.

A holistic view on the overall methodology to be adopted for the remaining part of this article is illustrated in Figure 1.

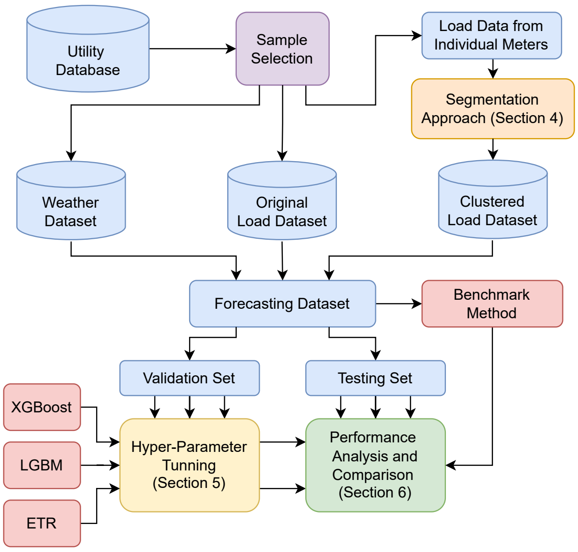

Figure 1: Proposed Research Workflow (Source: Developed by the author).

4. Experimental Segmentation Results

This section presents a detailed description of the procedures leading to the acquisition of the load segmentation results, comprising the first phase of the case study application.

4.1. Data Profiling and Preparation

The dataset used for the clustering comprises telemetric hourly load data from 2023 in a central region of The Netherlands, disaggregated at the single power grid connection level and provided in kWh. The data includes 8760 records across 980 variables (date, time, and load from 978 meters), with a sparsity ratio of 0.26% due to missing values from 55 meters at the year's start. All data was processed using *Python 3.10*. As performed in [37], load data was averaged daily for clustering, eliminating the need for missing value imputation. Since telemetric load data is validated beforehand by grid operators, no outliers were considered. Before clustering, load data was also normalized using z-score standardization to ensure equal contribution from all variables.

4.2. Feature Engineering

Recognizing that customer load behavior varies across different days, a categorical variable was created to specify each observation's day type, distinguishing regular working days, Saturdays, Sundays, and holidays. Subsequently, all observations were grouped by hour and by day type, and the average value for each (day type, hour) group was computed. This enabled a definition of each connection's averaged daily load profile under different calendar circumstances. The obtained dataset was later transposed, resulting in a total of 96 (24 hours * 4 day types) features with averaged load values and 978 meters as records.

4.3. Algorithm Selection and Optimal Number of Clusters

The framework proposed by Biscarri et al. [37] involves selecting the most suitable clustering solution by testing various algorithms with different numbers of clusters and validation measures. The conducted review on load segmentation procedures [36–39] highlighted prevalent clustering methods such as K-means, Hierarchical algorithms (with Single, Complete, Average, Centroid, and Ward distances), PAM, CLARA, and Mixture models. Hence, these methods were considered for a comparative analysis. The validation measures used for assessing clustering quality were the Davies-Bouldin and the Silhouette indices, evaluating cluster compactness and separation, and cluster cohesion and separation, respectively.

The established analysis revealed that Single and Average linkage methods with two clusters initially performed well according to the Davies-Bouldin index, but were ultimately discarded due to the imbalances in volume and number of observations between the resulting clusters. Instead, Ward linkage with two clusters showed better overall performance. In turn, the best performing combination according to the Silhouette index was a K-means approach with three clusters. Since both of these combinations performed well in both indices, both were elected to be further analyzed.

4.4. Results

The two selected clustering approaches aim to divide individual electricity grid connections based on their daily consumption patterns into representative groups. Load values were analyzed hourly, and cubic spline interpolation was applied to create approximated load curves for easier visualization of daily load patterns within each cluster. Load values are displayed from the power grid perspective, with consumption values as negative and injection values as positive. Figure 2 shows the two-cluster segmentation results.

The ward linkage clustering algorithm with two clusters produced very similar segments to the currently implemented segmentation. Cluster 1 aligned with the Regular segment, showing higher consumption during solar hours, particularly on working days, reflecting a business-focused load pattern. Cluster 2 matched the Injection segment, characterized by a pronounced midday load drop due to significant customer solar energy generation, also

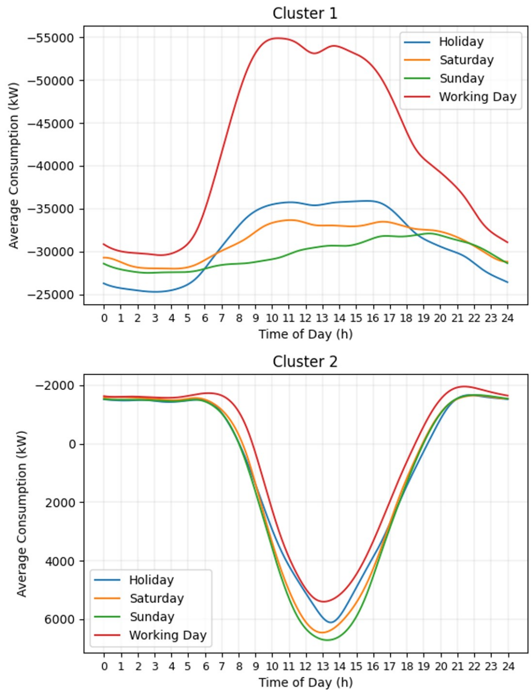

Figure 2: Two-Cluster Segmentation Results (Source: Developed by the author).

known as the duck curve. This cluster included 89% of the meters from the original Injection segment, with the differences having a minimal impact on the averaged load patterns. The similarity in results indicates that the utility's existing segmentation is effective at separating connections into two segments, making the two-cluster approach potentially redundant. Hence, this clustering approach was not further investigated.

The results from the k-means clustering algorithm with three clusters, shown in Figure 3, revealed distinct patterns from the company's current segmentation approach, with the total load volume being more evenly distributed across clusters. Cluster 1 exhibited higher consumption during working days, typical of businesses with reduced activity on weekends and holidays. This cluster showed increased load during early mornings, peaking around 9:00, and maintaining high consumption until the afternoon. Holidays and weekends had lower consumption levels, with a midday dip suggesting some mild solar influence. Cluster 2 had peak consumption in the late afternoon and similar load profiles across different day types, a typical characteristic of commercial entities. Cluster 3's load profiles resembled the original Injection segment, indicating significant solar influence but with higher overall consumption levels, highlighting the different composition of this cluster.

This three-cluster approach provided new insights into customer behavior, redistributing load profiles into more specific categories: Businesses

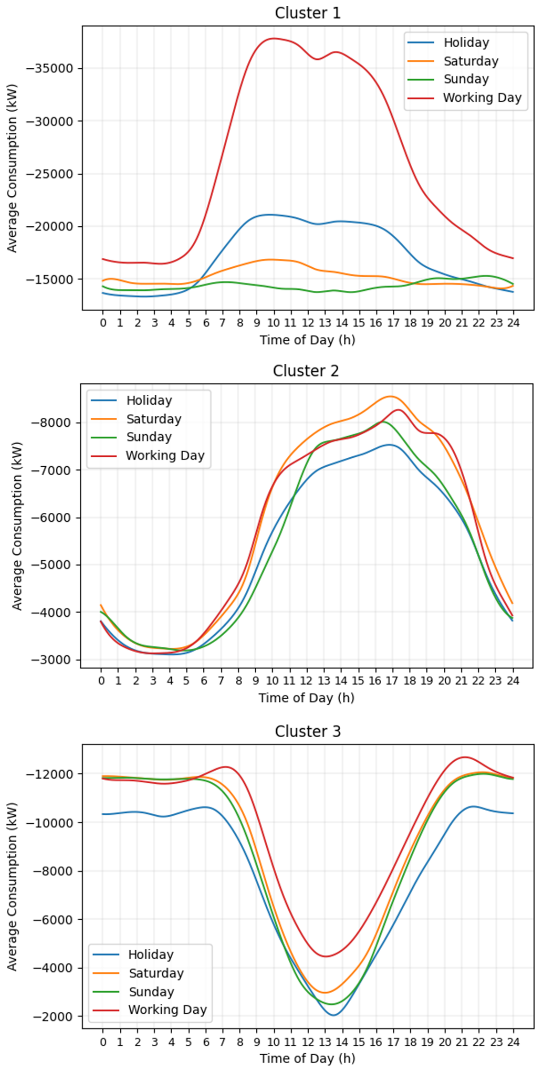

Figure 3: Three-Cluster Segmentation Results (Source: Developed by the author).

(Cluster 1), Commercials (Cluster 2), and Solar (Cluster 3). Thus, the segmentation resulting from this approach was elected to be further analysed, being considered a promising way of enhancing overall load prediction accuracy.

5. Experimental Predictive Results

This section provides a comprehensive overview of the steps involved in the acquisition of the predictive results, corresponding to the second phase of the case study application.

5.1. Data Profiling and Preparation

An initial data analysis provided a detailed overview of the load and weather datasets used for the STLF task. The load data is provided in two separate datasets (original and re-segmented), each containing a different segmentation of the 2023 load measurements from the same grid connections. The original dataset included the Regular and Injection segments, and the re-segmented dataset included the Businesses, Commercials, and Solar segments, as mentioned in 4.4. The weather data is provided as described in 3.3. These datasets were merged into a single dataset, resulting in 8760 records and 9 columns, with each observation including a date-time reference, load values for each of the five segments, and three weather measurements.

A subsequent data distribution analysis further explored the statistical properties of the dataset, revealing the absence of missing values and identifying outliers in the load data from the first 14 days of 2023. These outliers, attributed to the addition of load connections to the portfolio, were removed to ensure data integrity. The distribution analysis also examined the characteristics of each attribute, including skewness and variability, through boxplots and histograms. Moreover, an analysis using Spearman's correlation coefficients was also performed to further assess the relationships between variables. The results obtained revealed distinct correlations between load segments and weather variables, with notable differences in correlation direction and magnitude. Overall, the correlation analysis validated previous assumptions about each load segment's characteristics, and underscored the importance of accounting for temporal patterns in load forecasting models.

5.2. Feature Engineering

In the feature engineering phase, calendar-related variables were incorporated into the dataset to capture the temporal patterns influencing electricity demand. These patterns include daily, weekly, and yearly seasonality. For the daily pattern, the hour of the day was encoded as an integer variable, enabling models to recognize sequential relationships between different hours. To address the weekly pattern, the day of the week was encoded similarly, with additional dummy variables indicating working days and holiday periods. These features allow the models to differentiate between regular and irregular calendar characteristics, aiding in associating specific load behaviors with different calendar contexts. Regarding yearly seasonality, since the available data is restricted to one single year, the historically averaged number of daily solar hours in central Netherlands was included, aiming to help models recognize similar consumption behaviors due to similar positions in the yearly earth cycle. To let models learn the time proximity between different days, the day of the year was also included.

Considering that the selected forecasting methods already include automatic feature selection, and that the provided features were carefully chosen based on domain knowledge, no manual feature selection was performed and all features were provided to the forecasting models as input variables.

5.3. Data Splitting

For data splitting, as in [36], a rolling origin method was chosen, ensuring that the sequential nature of day-ahead load forecasting is preserved. A train-only period comprising observations from the first eight months of 2023 was defined, representing about 67% of the dataset. Testing rounds were then conducted incrementally, with the training set for forecasting load on day D being limited to observations from the first considered day until the end of day D - 3 (to represent the typical two-day delay in telemetric load data). For validation, 16 non-adjacent days were selected, aiming to assess model performance under diverse calendar characteristics and potential pattern disruptions. These included two non-adjacent days of each weekday, the Christmas day, and the day after the second time change of 2023 in The Netherlands. The selected days were used solely for validation purposes.

5.4. Model Application and Validation

The validation of the forecasting models involved a systematic search for hyper-parameters that maximize predictive performance on the selected validation days. For this, a Bayesian optimization algorithm was chosen, employing 50 iterations to minimize the MSE of forecasts. It uses probabilistic methods to estimate error across continuous hyper-parameter search spaces, requiring a reduced amount of evaluations to find good-performing hyper-parameter combinations.

Each forecasting method (XGBoost, LGBM, and ETR) was tuned separately for each load segment, resulting in 15 validated models. For XGBoost, parameters such as *learning_rate*, *max_depth*, *min_child_weight*, *subsample*, *colsample_bytree*, *reg_lambda*, and *reg_alpha* were tuned. For LGBM tuning, similar parameters were selected (*learning_rate*, *max_depth*, *min_data_in_leaf*, *bagging_fraction*, *feature_fraction*, *reg_lambda*, and *reg_alpha*). ETR tuning involved *max_features*, *max_depth*, *min_samples_split*, *min_samples_leaf*, and *bootstrap*. Table 1 describes the validation RMSE results for each of the tuned models.

Table 1: Validation RMSE Results (in kWh).

Load Segment	Method		
	XGBoost	LGBM	ETR
Regular	1890.9	**1644.8**	1836.3
Injection	395.6	388.9	**383.8**
Businesses	**723.6**	821.9	808.2
Commercials	**266.7**	273.0	279.7
Solar	877.9	**846.1**	980.6

The tuning process revealed to be an important step for enhancing model performance, with the biggest improvements occurring in the ETR and LGBM models, particularly the ones addressing the Regular and Businesses segments. As the iterations progressed, the error metrics converged, indicating the sufficiency of the defined number of iterations. All three methods displayed comparable performances after validation. Thus, all 15 tuned models were deemed eligible for further evaluation in the testing set.

6. Discussion

This section presents a discussion of the performance results obtained through the implementation of the segmentation and forecasting approaches. It analyses and benchmarks the proposed ML models against traditional statistical models, and assesses their aggregated performance across the two different customer load segmentation approaches.

6.1. Model Performance Evaluation and Comparison

Table 2 presents the RMSE and the MAE results obtained when using both the benchmarking models and the tuned ML models on the testing set.

Table 2: Test RMSE Results (in kWh).

Load Segment	Method			
	Benchmark	XGBoost	LGBM	ETR
Regular	2084.2	1678.6	**1625.1**	1777.1
Injection	469.6	412.6	**379.5**	386.7
Businesses	1439.4	**1108.9**	1160.5	1272.6
Commercials	240.5	206.3	**183.1**	184.3
Solar	1132.0	**872.3**	888.9	923.7

An initial performance assessment, based on RMSE scores, shows that all ML models consistently outperformed the statistical regression-based (Benchmark) models across all load segments, indicating a clear potential for improvement. Despite the LGBM having shown to be the faster ML method, all models revealed to be efficient at providing day-ahead forecasts. The obtained results highlighted that, at the individual segment level, LGBM was the overall most effective ML method, providing top or near-top performance in all load segments. It was closely followed by XGBoost, which presented similar (occasionally better) performances, while ETR showed less reliability.

Further analysis of the models' historical forecasting errors revealed patterns that affected predictive performance. All models showed increased errors during unusual temperature peaks and changes, and during the Christmas to New Year's period, caused by the inherent uncertainty of holiday consumption patterns. The Injection and Solar segments also experienced error spikes correlated with radiation levels, emphasizing the challenge of solar generation in load forecasting. During these circumstances, ML methods showed better generalization abilities compared to the Benchmark method, which typically requires manual adjustments to account for unusual patterns. No significant bias was observed across models, indicating that none suffered from underfitting.

The obtained results answer the first research question, proposed in 2.1, by evidencing that the adoption of state-of-the-art predictive models can significantly enhance STLF performance within a utility's portfolio, providing more reliable forecasts compared to traditional statistical regressions.

6.2. Aggregated Performance Evaluation and Comparison

In this article, two different segmentation procedures were used for customer load segmentation, as defined in Section 3. To assess the overall performance of each segmentation approach, hourly load forecasts for the Regular and Injection segments were summed, as were the forecasts for the Businesses, Commercials, and Solar segments. These aggregated forecasts were then compared with the total hourly load using the proposed ML and Benchmark methods individually for all segments, and using a model combination based on individual RMSE performance for each load segment.

The obtained results, presented in Table 3, showed that the combined method (of the best-performing models for each load segment) achieved the best results in both segmentation approaches. For the original segmentation, the combined method used LGBM models for Regular and Injection segments, matching the performance of the LGBM method. In the clustered segmentation, the combined method used XGBoost for Businesses and Solar segments and LGBM for Commercials, leading to similar aggregated performances across metrics. Overall, XGBoost and LGBM presented the best performances, while ETR was less consistent in providing competitive forecasts. Finally, it can be noted that the proposed clustered approach outperformed the original approach for all methods except for the Benchmark one, where it achieved similar results. This showed that performance improvements resulting from a finer segmentation are dependent on the forecasting methods' ability to handle complex and non-linear relationships within the data.

The aggregated results respond to the second research question by showing that a validated clustering procedure for a 3-group customer load segmentation is effective in improving STLF performance if paired with ML forecasting methods.

7. Conclusions

STLF is essential for decision-making in energy production and procurement. This article aims to enhance a utility's STLF framework by focusing on customer load segmentation and ML forecasting methods. XGBoost, LGBM, and ETR were firstly identified as promising ML forecasting methods, and it was concluded that all these methods significantly outperformed a traditional statistical method, with XGBoost and LGBM consistently providing the most accurate forecasts. It was additionally observed that abrupt temperature changes, exceptional weather conditions, and holiday seasons resulted in a general reduction in model performance. Finally, the study found that a 3-cluster approach for segmenting load data based on averaged daily profiles uncovered distinct patterns among customers, notably enhancing STLF accuracy when combined with ML forecasting models. The best overall performance was achieved by a combination of LGBM and XGBoost models tuned for each respective cluster, highlighting the potential of advanced supervised and unsupervised ML techniques to improve forecasting performance of a utility's customer portfolio.

This article faced certain limitations, including the absence of historical weather forecasts (with only measurements being available) and the reliance on just one year of data. Future work may explore the impact of weather forecast accuracy on load predictions, and assess model performance with differently-sized datasets. Lengthier data should enable models to better learn yearly patterns, while, with less data available, new performance comparisons can be established between statistical and ML forecasting methods. Additionally, more standardized forecasting frameworks could be developed through further investigation into stacking ensemble models (addressing single model's limitations)

Table 3: Test Results per Segmentation Approach (in kWh for the RMSE and MAE).

Method	Original Segmentation			Clustered Segmentation		
	RMSE	MAE	NMAE	RMSE	MAE	NMAE
Benchmark	2213.8	1460.6	0.0370	2214.0	1465.3	0.0371
XGBoost	1807.7	1214.4	0.0307	1562.9	1091.8	0.0276
LGBM	1741.9	1162.0	0.0294	1636.0	1095.1	0.0277
ETR	1915.3	1291.3	0.0327	1830.9	1233.9	0.0312
Combined	1741.9	1162.0	0.0294	**1557.8**	**1087.5**	**0.0275**

Note: The Normalized MAE (NMAE) corresponds to the MAE divided by the average load during testing (39439 kWh).

and into rare event modelling. Finally, future analyses could focus on different regions and customer groups, and explore the impact of finer clustering methods on the overall forecasting performance.

References

[1] S. Akhtar, S. Shahzad, A. Zaheer, H. S. Ullah, H. Kilic, R. Gono, M. Jasiński, and Z. Leonowicz, "Short-Term Load Forecasting Models: A Review of Challenges, Progress, and the Road Ahead," *Energies*, vol. 16, p. 4060, Jan. 2023.

[2] M. A. Hammad, B. Jereb, B. Rosi, and D. Dragan, "Methods and Models for Electric Load Forecasting: A Comprehensive Review," *Logistics, Supply Chain, Sustainability and Global Challenges*, vol. 11, pp. 51–76, Feb. 2020.

[3] Y. Sun, F. Haghighat, and B. C. M. Fung, "A review of the-state-of-the-art in data-driven approaches for building energy prediction," *Energy and Buildings*, vol. 221, p. 110022, Aug. 2020.

[4] J. Ma and X. Ma, "A review of forecasting algorithms and energy management strategies for microgrids," *Systems Science & Control Engineering*, vol. 6, pp. 237–248, Jan. 2018.

[5] J. Steinbuks, J. de Wit, A. Kochnakyan, and V. Foster, "Forecasting Electricity Demand: An Aid for Practitioners," 2017.

[6] H. M. Kiani and X.-J. Zeng, "A Function-on-Function Linear Regression Approach for Short-Term Electric Load Forecasting," in *2019 IEEE Texas Power and Energy Conference (TPEC)*, pp. 1–5, Feb. 2019.

[7] R. Wang, J. Wang, and Y. Xu, "A novel combined model based on hybrid optimization algorithm for electrical load forecasting," *Applied Soft Computing*, vol. 82, p. 105548, Sept. 2019.

[8] E. Becirovic and M. Cosovic, "Machine learning techniques for short-term load forecasting," pp. 1–4, Sept. 2016.

[9] A. Groß, A. Lenders, F. Schwenker, D. A. Braun, and D. Fischer, "Comparison of short-term electrical load forecasting methods for different building types," *Energy Informatics*, vol. 4, p. 13, Sept. 2021.

[10] Z. Wang, T. Hong, H. Li, and M. Ann Piette, "Predicting city-scale daily electricity consumption using data-driven models," *Advances in Applied Energy*, vol. 2, p. 100025, May 2021.

[11] A. A. Mir, K. Ullah, Z. A. Khan, F. Bashir, T. U. R. Khan, and A. Altamimi, "Short Term Load Forecasting for Electric Power Utilities: A Generalized Regression Approach Using Polynomials and Cross-Terms," *Engineering Proceedings*, vol. 12, no. 1, p. 21, 2021.

[12] G. Dudek, "Pattern-based local linear regression models for short-term load forecasting," *Electric Power Systems Research*, vol. 130, pp. 139–147, Jan. 2016.

[13] G. I. Dada, "Analysis of Electric Load Forecasts Using Machine Learning Techniques," Master's thesis, Dublin, National College of Ireland, 2019.

[14] J. W. Taylor and P. E. McSharry, "Short-Term Load Forecasting Methods: An Evaluation Based on European Data," *IEEE Transactions on Power Systems*, vol. 22, pp. 2213–2219, Nov. 2007.

[15] N. I. Nwulu and O. P. Agboola, "Modelling and predicting electricity consumption using artificial neural networks," in *2012 11th International Conference on Environment and Electrical Engineering*, pp. 1059–1063, May 2012.

[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[17] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Nov. 2016.

[18] A. M. Tudose, I. I. Picioroaga, D. O. Sidea, C. Bulac, and V. A. Boicea, "Short-Term Load Forecasting Using Convolutional Neural Networks in COVID-19 Context: The Romanian Case Study," *Energies*, vol. 14, p. 4046, Jan. 2021.

[19] Z. Zhang, J. Liu, S. Pang, M. Shi, H. H. Goh, Y. Zhang, and D. Zhang, "General short-term load forecasting based on multi-task temporal convolutional network in COVID-19," *International Journal of Electrical Power & Energy Systems*, vol. 147, p. 108811, May 2023.

[20] H. Eskandari, M. Imani, and M. P. Moghaddam, "Convolutional and recurrent neural network based model for short-term load forecasting," *Electric Power Systems Research*, vol. 195, p. 107173, June 2021.

[21] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, "Recent Advances in Recurrent Neural Networks," Dec. 2017.

[22] M. Slowik and W. Urban, "Machine Learning Short-Term Energy Consumption Forecasting for Microgrids in a Manufacturing Plant," *Energies*, vol. 15, p. 3382, May 2022.

[23] N. Shabbir, L. Kutt, M. Jawad, M. N. Iqbal, and P. S. Ghahfaroki, "Forecasting of Energy Consumption and Production Using Recurrent Neural Networks," *Advances in Electrical and Electronic Engineering*, vol. 18, pp. 190 – 197, Oct. 2020.

[24] A. Ahmad, N. Javaid, A. Mateen, M. Awais, and Z. A. Khan, "Short-Term Load Forecasting in Smart Grids: An Intelligent Modular Approach," *Energies*, vol. 12, p. 164, Jan. 2019.

[25] E. Ceperic, V. Ceperic, and A. Baric, "A Strategy for Short-Term Load Forecasting by Support Vector Regression Machines," *IEEE Transactions on Power Systems*, vol. 28, pp. 4356–4364, Nov. 2013.

[26] A. Moradzadeh, S. Zakeri, M. Shoaran, B. Mohammadi-Ivatloo, and F. Mohammadi, "Short-Term Load Forecasting of Microgrid via Hybrid Support Vector Regression and Long Short-Term Memory Algorithms," *Sustainability*, vol. 12, p. 7076, Jan. 2020.

[27] Z. Wang, Y. Wang, R. Zeng, R. S. Srinivasan, and S. Ahrentzen, "Random Forest based hourly building energy prediction," *Energy and Buildings*, vol. 171, pp. 11–25, July 2018.

[28] E. Zdravevski, P. Lameski, A. Kulakov, and V. Trajkovikj, "Performance comparison of random forests and extremely randomized trees," Apr. 2016.

[29] P.-P. Phyo, Y.-C. Byun, and N. Park, "Short-Term Energy Forecasting Using Machine-Learning-Based Ensemble Voting Regression," *Symmetry*, vol. 14, p. 160, Jan. 2022.

[30] E. Nziyumva, R. Hu, C.-Y. Hsu, and J. Niyogisubizo, "Electrical Load Forecasting Using Hybrid of Extreme Gradient Boosting and Light Gradient Boosting Machine," in *The International Conference on Image, Vision and Intelligent Systems (ICIVIS 2021)* (J. Yao, Y. Xiao, P. You, and G. Sun, eds.), Lecture Notes in Electrical Engineering, pp. 1083–1093, 2022.

[31] E. A. Daoud, "Comparison between XGBoost, LightGBM and CatBoost Using a Home Credit Dataset," *International Journal of Computer and Information Engineering*, vol. 13, pp. 6–10, Jan. 2019.

[32] C. Miller, P. Arjunan, A. Kathirgamanathan, C. Fu, J. Roth, J. Y. Park, C. Balbach, K. Gowri, Z. Nagy, A. Fontanini, and J. Haberl, "The ASHRAE Great Energy Predictor III competition: Overview and results," *Science and Technology for the Built Environment*, vol. 26, pp. 1427–1447, Nov. 2020.

[33] Y. Huang, Y. Yuan, H. Chen, J. Wang, Y. Guo, and T. Ahmad, "A novel energy demand prediction strategy for residential buildings based on ensemble learning," *Energy Procedia*, vol. 158, pp. 3411–3416, Feb. 2019.

[34] R. Porteiro, S. Nesmachnow, and L. Hernández-Callejo, "Short Term Load Forecasting of Industrial Electricity Using Machine Learning," in *Smart Cities* (S. Nesmachnow and L. Hernández Callejo, eds.), Communications in Computer and Information Science, pp. 146–161, 2020.

[35] A. Yang, W. Li, and X. Yang, "Short-term electricity load forecasting based on feature selection and Least Squares Support Vector Machines," *Knowledge-Based Systems*, vol. 163, pp. 159–173, Jan. 2019.

[36] E. Spiliotis, F. Petropoulos, N. Kourentzes, and V. Assimakopoulos, "Cross-temporal aggregation: Improving the forecast accuracy of hierarchical electricity consumption," *Applied Energy*, vol. 261, p. 114339, Mar. 2020.

[37] F. Biscarri, I. Monedero, A. García, J. I. Guerrero, and C. León, "Electricity clustering framework for automatic classification of customer loads," *Expert Systems with Applications*, vol. 86, pp. 54–63, Nov. 2017.

[38] G. Grigoraș, M. S. Raboaca, C. Dumitrescu, D. L. Manea, T. C. Mihaltan, V.-C. Niculescu, and B. C. Neagu, "Contributions to Power Grid System Analysis Based on Clustering Techniques," *Sensors (Basel, Switzerland)*, vol. 23, p. 1895, Feb. 2023.

[39] A. Tureczek and P. Nielsen, "Structured Literature Review of Electricity Consumption Classification Using Smart Meter Data," *Energies*, vol. 10, p. 584, Apr. 2017.

[40] E. Elgeldawi, A. Sayed, A. R. Galal, and A. M. Zaki, "Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis," *Informatics*, vol. 8, p. 79, Dec. 2021.