Electric Utility Customer Segmentation from Advanced Metering System Data Using K-Shape Clustering — A Norwegian Case Study

Kari Walstad
Lnett AS
Stavanger, Norway.
Kari.Walstad@l-nett.no

Vijay Venu Vadlamudi
Department of Electric Power Engineering
Norwegian University of Science and Technology (NTNU)
Trondheim, Norway.
Vijay.Vadlamudi@ntnu.no

***Abstract*—In this paper, a framework was developed for the segmentation of the customer base of Norwegian Distribution System Operators (DSO), based on Advanced Metering System (AMS) time series data of the electricity consumption of DSO customers. A computer programme for customer segmentation was synthesised in the programming language Python, using shape-based clustering, and a Cluster Validation Index (CVI) algorithm. Additionally, an option to perform a simple outlier analysis based on user input of the AMS input data was included. The assessment of the developed customer segmentation programme and the underlying methodology was first done through tests on a known data set to verify the results. Following this, an assessment was made on the basis of two actual AMS-data sets provided by the Norwegian DSO Lnett AS. AMS-data was more challenging for the algorithm to cluster than the known data set, possibly because the AMS-data set was more homogeneous with more similarly shaped and less discernible time series groups. Outlier analysis was shown to improve the programme performance by removing irregular (i.e. flat) time series. Based on the second AMS-data set, a comparison with the current standard method of customer segmentation utilised by Norwegian DSOs was performed. The developed customer segmentation method, when measured with a CVI, was shown to produce a better partition compactness and distinctness of the AMS-data set than with the standard DSO method.**

***Index Terms*—Advanced Metering Systems, customer segmentation, Distribution System Operator, k-Shape, machine learning, time series.**

I. Introduction

Customer segmentation, according to Electric Power Research Institute (EPRI), is "*a method of identifying homogeneous groups of consumers within a greater population based on common purchasing patterns and behavioral traits*" [1]. When electricity is the commodity in question, the energy consumption data of consumers can be processed using several customer segmentation methods in vogue [2]–[5]. With the advent of smart meters and the accompanying advanced metering infrastructure (AMI), the required data that forms the backbone for customer segmentation methods is easily accessible, and improvements in the methods used for effectively segmenting customers are on the rise. Applications of utility-scale customer segmentation include targeting customers for unique demand response programs, investment in AMI, improving energy profile modeling, design of unique customer incentive programmes, etc. [2]–[5]. Data from AMS could be utilised to more accurately group customers with similar electric consumption behaviour, where "behaviour" could represent time series characteristics such as periodicity, amplitude, trend, etc. Understanding customer behaviour is essential for designing tariff systems where the DSOs target specific customer consumption behaviours [8], [9]. Another significant application of customer segmentation is in obtaining accurate estimates for customer interruption costs (CIC), which are inputs to the cost-benefit analysis needed for power system reliability investments [10].

Historically, AMS-data has been primarily gathered from Norwegian electric utility customers with large electric power consumption, typically above 100 MWh/year [6]. According to changes implemented in 2011 to the Norwegian Regulations on Power Trade and Utility Grid Services, AMS meters would be required for all utility grid consumers from 2019 [7].

Currently, Norwegian DSOs utilize a method for customer segmentation which assigns a customer to a broadly and ambiguously defined segment (customer category) at the time of connection to the grid, based on their *assumed* future consumption behaviour.

In the context of this paper Norwegian DSOs were contacted for conducting research on the extent and purposes of the usage of their collected AMS-data: Elvia AS [11], Lnett AS [12] and BKK Nett AS [13]. The survey and additional direct contact with other DSOs unearthed that none of the large Norwegian DSOs had yet to actively utilise AMS-data in day-to-day operations or for customer research and analysis.

New technologies, e.g., machine learning, have become easily available for processing large volumes of data. Now there exists a variety of computer algorithm libraries for machine learning, in various coding languages (Keras [14], scikit-learn [15], TensorFlow [16] etc.). This opens the possibility for utilisation of freely accessible computer algorithms for machine learning for efficient analysis of AMS-data at relatively insignificant costs to DSOs.

This paper was written in collaboration with Lnett, and therefore it was considered relevant to investigate the possibil-

978-1-6654-8032-1/22/$31.00 ©2022 IEEE

ity for DSO personnel to utilize AMS-data to a larger extent and potentially extract commercial benefits of AMS-data use.

Potential areas of AMS-data use were assessed in an Analytical Hierarchy Process (AHP), the result of which was that customer segmentation was chosen as the priority. A literature review revealed that customer segmentation done with clustering had never been compared with the current customer segmentation methods of DSOs in Norway, although several studies [17] have shown that certain clustering algorithms produce promising results when used on AMS-data.

The main contributions of this work are listed below:

- A computer programme for shape-based (time and offset invariant) DSO customer segmentation using AMS time series data in the programming language Python.
- A comparison of the ability of the identified k-Shape clustering algorithm to segment DSO customers on the basis of the AMS-data, with the current DSO customer segmentation method that is based on segmenting customers from their *assumed*, and not *observed* load profile. K-Shape clustering is not commonly found employed in AMS-data analysis [17]–[22].
- An iterative data refinement method referred to as outlier analysis (not found described in literature).
- A decision to only employ algorithms which use distance metrics particularly suited for time series-type data in clustering algorithm, Cluster Validation Index (CVI) and outlier analysis. This has not been the emphasis of any prior study which involved clustering AMS-data, insofar as a literature review could reveal [17]–[22]

In the first part of Section II, the reasoning behind the composition of the synthesised computer programme is explained, in addition to information on each individual component of it. From Sections II-E to II-F, the different tests performed on the developed programme are explained. Section III describes the results of the tests outlined in Sections II-E to II-F. Section IV contains the conclusions from the conducted research study. The work is based on the Masters thesis of the corresponding author, and the reader is referred to it for a more comprehensive version of the content and findings presented in this paper [30]; the source code used in this paper can be accessed from the appendices in the thesis.

II. Methodology

The development of the customer segmentation programme was approached as follows. Firstly, `tslearn` a machine learning library for Python specifically written for time series, freely accessible to the public, was identified. Secondly, based on the available Python libraries, the development of a customer segmentation computer programme was evaluated. The algorithm type which was a best fit for this task was found to be clustering. A Cluster Validation Index (CVI) was chosen in order establish an appropriate number of clusters for a given time series data set. Lastly, functionality for identifying irregularities in the data set was added to the programme (not explained in this paper). Refining the data set based on removing outliers, on rolling averages, and on dividing the data set into subsets based on mean values was evaluated. Outlier analysis was the functionality implemented in the programme code. The outlier analysis was based on guidance and input from the programme user. Potential implementation of rolling averages and division based on mean values was described, but not implemented.

The customer segmentation method would aspire to have the following properties:

- Be computer-based.
- Be scalable to handle large amounts of data.
- Be freely accessible to anyone who wishes to use it for their own purposes.
- Be modular, which means that a user may easily apply additional functionality to the solution, or change the existing algorithms without necessarily changing the structure of the programme.
- Be able to demonstrate feasibility and promising results.
- Be described in this paper in an accessible manner to the reader.

A. Algorithms

Three types of algorithms represent the building blocks of the programme. These are: distance metric algorithm, clustering algorithm and CVI algorithm. The customer segmentation programme is modular, which means that the selected distance metric, clustering algorithm and CVI may be changed by a future user.

1) Distance metric: The Distance metrics highlighted for use in the programme are Eucilidan distance, Dynamic Time Warping [23], [24], and Shape-based distance.

2) Clustering: In this paper, a clustering approach which is both centroid-based and shape-based was selected. Centroid-based because then each time series cluster can be represented by a single average time series (the centroid). Shape-based because it was considered desirable to be able to capture similarities in customer consumption profiles despite time shifts and amplitude/offset variations. It was also important to select a clustering algorithm which could handle large data sets. To that end the centroid- and shape-based clustering algorithm called "k-Shape" was selected as the algorithm of choice in this paper [25], implemented with the machine learning library for Python `tslearn` [26]. In this paper, the `tslearn KShape` algorithm is considered to be in a "stable zone" when, for a constant `K`, it produces consistent results (i.e. a stable CVI), for the input parameter `n_init` above a certain value.

3) Cluster Validation Index: A Cluster Validation Index (CVI) aims to validate how good a given cluster partition is. In this paper, a CVI was used to evaluate the partition compactness and distinctness for a given number of clusters `K`, with the goal of assisting the identification of an optimal number of partitions, `K_optimal`, for a given time series data set. The Silhouette Index is regarded as a highly accurate CVI, which uses DTW as its distance metric in `tslearn` (`silhouette_score`) [26], and can in many cases achieve satisfactory results on its own [27]–[29].

B. Outlier analysis

The outlier analysis methods are based on DTW [23], and on the user of the programme visually inspecting the data set and the clustering results. The following method is proposed: separate the data set into inliers and outliers and analyse these separately. The inlier data set will contain the majority of data, and the outlier data set will contain the data which is observed to impact the partition quality of the inlier data set negatively.

C. Developed programme

The developed customer segmentation method consists of the following steps:

1) The algorithm starts with a set of AMS-data which is listed row-wise in a .txt file as input.
2) The AMS-data is converted with code in Python such that it is stored in a format that is readable to the algorithms in the Python `tslearn` library.
3) The AMS-data is normalised to facilitate a shape-based customer segmentation approach. Before Step 4), the programme user selects a value for `n_init`, which will decide the number of iterations that the `KShape` algorithm performs for each `K` in Step 4). The value of `n_init` that produces stable Silhouette scores for any given `K` can be found experimentally by varying `n_init`, while keeping `K` constant.
4) `K` is used to denote the number of clusters that the clustering algorithm in the customer segmentation programme is instructed to find in the AMS-data set. `K` is an integer which is incrementally increased in a for-loop from some minimal `K` (`K_min`), to some maximal `K` (`K_max`), defined by the user of the programme. An interval of `K` values is tested in order to generate Silhouette scores for multiple `K` values in Step 5).
5) The partition quality of each `K` used in the for-loop is then tested by a CVI when the for-loop is completed, i.e. the CVI (Silhouette scores) for all `K_min` to `K_max` is calculated. The programme presents the user with the complete set of CVI scores for all `K`. This aids the user in determining a suitable `K`, i.e. `K_optimal`.
6) `K_optimal` is used to denote the optimal number of clusters `K`. For the Silhouette algorithm, the score closest to +1 is the optimal score. The programme user may chose another `K`, for example in the case where `K_optimal` is calculated to be small, e.g. `K_optimal=2` or to be large, e.g. `K_optimal=K_max`.
7) After selecting a `K`, the programme will plot the generated clusters, as well as the highlighted cluster centroids. The user is then able to visually inspect the clustering, and evaluate the programme results.
8) Based on the evaluation of the results in Step 7), the user can choose to refine the data set by performing an outlier analysis.
9) Based on the simple outlier detection methods, the AMS-data is sorted by the user of the programme into an outlier data set and an inlier data set. After one of the outlier detection methods has been applied, the steps 4) to 8) are repeated with the refined (inlier) data set. This is because after removing outliers the Silhouette scores for different `K` values change, and partition quality may improve.
10) When the programme user is of the opinion that (further) refinement of the data set does not bring advantages, and/or considers the cluster partitions satisfactory, the programme is terminated.

D. Test data

The known data set, as described in Section II-D1 may be replicated by the code found in [30], Appendix III. The AMS-data sets with and without the customer segment information, as described in Sections II-D2 and II-D3 cannot be given out as they are classified as sensitive information by the Norwegian Ministry of Petroleum and Energy [31].

1) Known data set: The known data set consisted of four known, visually distinct and identifiable time series groups, with a minimum of 50 time series per group. The following time series groups were included:

- A square wave time series set with a varying amplitude, with no phase shift, and a constant frequency.
- A sine wave time series set with a slight phase shift between some of the sine wave time series, no varying amplitude, and a constant frequency.
- A set of time series made available by `tslearn` characterised by a peak and then a drop, and finally stabilising to the starting value.
- A set of time series made available by `tslearn` characterised by a sloping increase in level from one stable level to another higher level.

2) AMS-data set without customer segment information: The DSO Lnett provided an AMS-data set which did not include customer segment information. The data set consisted of 378 customer IDs with associated consumption measurements in kWh taken on an hourly basis. Standardisation of the data was performed to achieve amplitude and offset invariance, and is achieved by recomputing the time series such that the standard deviation is equal to one, and the mean is equal to zero.

3) AMS-data set with customer segment information: A data set that included customer segmentation information was also provided by Lnett, comprised of nine different customer segment groups, and 103 IDs. This was a smaller data set than the first one provided by Lnett, but was considered sufficient to provide indicative results.

E. Parameter tuning process

The parameters considered for tuning are `K` and `n_init` for the clustering algorithm, and the DTW boundary for the outlier analysis methods. A `K` with a good CVI score should be chosen, while also being large enough to provide useful partitions. `n_init` should be chosen large enough to provide `tslearn KShape` algorithm result stability.

1) Data refinement: Before the programme was tested, the AMS-data had to be processed to a format which was appropriate for the subsequent steps of the programme: a .txt file where all AMS time series were sorted by ID row-wise and time column-wise.

2) Performance test on known data set without outlier analysis: The steps 1) to 6) of the customer segmentation model were tested to assess how effective the Silhouette and K-Shape algorithms are in evaluating the optimal number of clusters `K_optimal` in a time series data set, and how the data set is partitioned when `K_optimal` is given. This was done using the known data set as described in Section II-D1. Five tests were performed on this data set:

1) A Base Case test on the known data set where the programme was tested to see if it would correctly segment the time series groups.
2) A test to see if the programme accuracy suffered when the number of time series in a group in the known data set was significantly decreased.
3) A test to see if the programme clustering algorithm was truly time invariant.
4) A test to see whether the programme accuracy would be altered when significantly decreasing the time-length of the time series groups in the known data set.
5) A test to see whether the programme accuracy was affected by adding randomness ("random noise") to the known data set.

These tests with the known data set may be replicated with the code in [30], Appendix III.1-III.5 respectively.

3) Programme assessment on known data set with outlier analysis: Next, the programme including outlier analysis functionality was tested on the known data, to test the efficiency of the proposed outlier analysis methods. This test may be replicated with the code in [30], Appendix IV using the data set given in Appendix III.5.

F. Programme assessment using AMS-data without customer segment information

The developed programme was first tested on an AMS-data set provided by the DSO Lnett, which did not include customer segment information. Outlier analysis was performed on the data set. Then, the developed programme was tested on the inlier data set to quantify the effect of removing the most "disturbing" outliers. This test may be replicated with the code found in [30], Appendix IV.

G. Programme assessment using AMS-data with customer segment information

Following this assessment of the first data set, the programme was tested on an AMS-data set which contained customer segment information, as described in Section II-D3. This was done to compare the results of the programme with respect to the current DSO customer segmentation practices. This test may be replicated with the code found in [30], Appendix IV.

III. RESULTS AND DISCUSSION

A. Results on comparison of clustering algorithms

As described in Section II-A2, a test between shape-based k-Shape with DTW and non shape-based k-Means with Euclidean distance clustering algorithms was performed with the AMS-data described in Section II-D2, to test the initial hypothesis that k-Shape would provide preferable results when used with AMS-data. It was clear from the test results that the k-Shape algorithm produces better quality clusters when measured with the CVI described in Section II-A3.

B. Results from assessment of programme on known data set without outlier analysis:

From tests performed on a known data set as described in Section II-E2, the following performance was observed:

- *Given a data set of multiple, well-defined time series groups, the developed programme:* Is able to correctly assess the number of distinct time series groups in the data set; correctly cluster the time series groups given the calculated optimal number of time series groups; produce shape-based time series clusters and cluster centroids; is not prone to significant inaccuracy given variations in group sparseness; is not prone to significant inaccuracy given variations in length of the time series when calculations are performed within the `KShape` algorithm's stability zone; produces clusters and cluster centroids which appear to be stable over time.
- *Given a data set of multiple, well-defined time series groups, the `KShape` algorithm:* Displays reduced stability given increased sparseness of a time series group in a data set; displays reduced stability given a set of shorter time series.
- Given sufficient randomness in a data set of multiple, well-defined time series groups, the programme will calculate varying optimal partitions depending on the nature of the added random data.
- In a data set of multiple, well-defined time series groups with added randomness, the programme is able to produce cluster centroids with approximately the same shape as it would without the random data added.

C. Results from assessment of programme on known data set with outlier analysis:

From tests performed on a known data set with outlier analysis functionality included as described in Section II-E3 it was observed that outlier analysis makes the `KShape` algorithm more stable, improving the runtime of the algorithm, but the user of the programme has to visually or in other ways be able to distinguish between outliers and inliers.

D. Results from assessment of programme on AMS-data without customer segment information:

The following performance of the programme was observed, as described in Section II-F when tested on AMS-data which did not include customer segment information:

- The AMS-data set achieved overall much lower Silhouette scores for each `K`, when compared to analysis on the known data set.
- The highest cluster compactness and distinctness was achieved for `K`=2 in both the initial/original and inlier AMS-data sets.
- When inspecting the `K` for which the second-highest Silhouette score is achieved in the inlier data set (`K`=7), the programme produces substantial clusters.
- Flat time series was shown to disturb the `KShape` algorithm.
- After the flat time series was removed from the data set, the stability of the `KShape` algorithm for all `K` improved.
- Outlier analysis was shown to be useful when analysing AMS-data.

E. Results from assessment of programme on AMS-data with customer segment information:

The developed customer segmentation programme, when tested as described in Section II-G, produced segments with a higher Silhouette score (Silhouette score of at best 0.066 and at worst -0.015) than the standard DSO customer segmentation method does (Silhouette score of -0.032). The developed customer segmentation programme produces clusters with an improved quality (compactness and distinctness).

F. Discussion

Machine learning has become an important area of research, which has resulted in increased accessibility of machine learning algorithms and methods. One such library, which can be applied to time series data is `tslearn`. Only `tslearn` was found to operate with distance metrics suitable for time series.

The computer programme synthesised for shape-based (i.e. amplitude, offset and time invariant) DSO customer segmentation based on AMS time series data was, on a known/defined data set, shown to produce results that were known to be correct beforehand. When tested on the first AMS-data set, the programme showed a tendency to prefer segmenting based on similar periodicity of the time series. AMS-data was more challenging for the algorithm to cluster than the known data set, as it contained less distinct time series groups and to some extent more irregular data. This was clear from the number of iterations needed to reach algorithm stability (i.e. 40 iterations on the known data set, to 10000 for the AMS-data set). Outlier analysis was shown to improve the programme performance by removing irregular (i.e. flat) time series. When testing the developed programme for the known data set, the runtime of the clustering algorithm was acceptable. However, when analysing the AMS-data set, the runtime of the programme increased dramatically due to the increased number of iterations needed to reach algorithm stability. The large number of required iterations proved challenging for the available hardware (a standard laptop with 4.00 GB RAM and an Intel(R) Pentium(R) CPU 4405U @ 2.10GHz 2.11 GHz processor), with a runtime of $\approx$10 hours for 10000 iteration. The extended runtime may be due to the programming language that the customer segmentation programme is written in, i.e. Python. A lower-level language would likely significantly improve the programme runtime.

The developed customer segmentation programme was shown to produce a better partition compactness and distinctness of the second AMS-data set than by the standard DSO method, when measured with a CVI.

IV. Conclusions

Through this study, we were able to illustrate that opportunities and benefits related to the use of AMS-data can be achieved by applying the existing, openly available advanced data analysis tools, and be easily realised by DSO personnel with a power systems background and an information technology interest. In this work, a relatively user friendly framework based on openly available advanced data analysis tools has been developed for customer segmentation. The developed computer programme aims to segment DSO customers with a shape-based clustering approach, with the use of AMS time series of electricity consumption data. Experience with the programme indicates that runtime may be an issue when analysing large AMS-data sets.

Some potential improvements to the developed programme have also been identified. One of these is the potential for the programme to include amplitude and offset information in the analysis. Another area for improvement is the language in which the programme is written. Python is a relatively high-level coding language, and the efficiency of the programme and thereby calculation time may be improved if a lower-level language such as C++.

References

[1] EPRI Technical Update Report, “Exploring Customer Segmentation for Utility Energy Efficiency Programs,” May 2008. [Online]. Available: https://www.epri.com/research/products/1016386

[2] J. Kwac, J. Flora, and R. Rajagopal, “Household energy consumption segmentation using hourly data,” *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 420-430, Jan. 2014.

[3] S. Haben, C. Singleton, and P. Grindrod, “Analysis and clustering of residential customers energy behavioral demand using smart meter data,” *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 136-144, Jan. 2016.

[4] Y. Yu, G. Liu, W. Zhu, F. Wang, B. Shu, K. Zhang, N. Astier, and R. Rajagopal, “Good consumer or bad consumer: Economic information revealed from demand profiles,” *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 2347-2358, May 2018.

[5] Y. Yuan, K. Dehghanpour, F. Bu, and Z. Wang, “A data-driven customer segmentation strategy based on contribution to system peak demand,” *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 4026-4035, Sep. 2020.

[6] Norwegian Energy Regulatory Authority. [Online]. Available: https://www.nve.no/stromkunde/smarte-strommalere-ams (in Norwegian). [Accessed: 02 February 2021]

[7] Norwegian Ministry of Petroleum and Energy. [Online]. Available: https://lovdata.no/dokument/SF/forskrift/1999-03-11-301 (in Norwegian). [Accessed: 02 February 2021]

[8] Report by the Norwegian Water Resources and Energy Directorate. [Online]. Available: http://publikasjoner.nve.no/rme_rapport/2020/rme_rapport2020_06.pdf (in Norwegian). [Accessed: 02 February 2021]

[9] Report by the Norwegian Water Resources and Energy Directorate. [Online]. Available: http://publikasjoner.nve.no/rme_hoeringsdokument/2020/rme_hoeringsdokument2020_01.pdf (in Norwegian). [Accessed: 02 February 2021]

[10] O. Dzobo, K. Alvehag, C. T. Gaunt, and R. Herman, "Multi-dimensional customer segmentation model for power system reliability-worth analysis," *International Journal of Electrical Power and Energy Systems*, vol. 62, pp. 532-539, Nov. 2014.

[11] [Online]. Available: https://www.elvia.no/ (in Norwegian). [Accessed: 02 February 2021]

[12] [Online]. Available: https://www.lysenett.no/ (in Norwegian). [Accessed: 02 February 2021]

[13] [Online]. Available: https://nett.bkk.no/ (in Norwegian). [Accessed: 02 February 2021]

[14] [Online]. Available: https://keras.io/. [Accessed: 02 February 2021]

[15] [Online]. Available: https://scikit-learn.org/stable/. [Accessed: 02 February 2021]

[16] [Online]. Available: https://www.tensorflow.org. [Accessed: 02 February 2021]

[17] A. Tureczek, "Structured literature review of electricity consumption classification using smart meter data," *Energies*, vol. 10, pp.584-603, Apr. 2017.

[18] C. Flath, D. Nicolay, T. Conte, C. van Dinther and L. Filipova-Neumann, "Cluster analysis of smart metering data: An implementation in practice," *Business and Information Systems Engineering*, vol. 4, pp.31-39, Feb. 2012.

[19] F. McLoughlin, A. Duffy, and M. Conlon, "A clustering approach to domestic electricity load profile characterisation using smart metering data," *Applied Energy*, vol. 141, pp. 190-199, Mar. 2015.

[20] W. Wang, N. Yu, B. Foggo, J. Davis, and J. Li, "Phase identification in electric power distribution systems by clustering of smart meter data," *Proc. 15th IEEE International Conference on Machine Learning andApplications (ICMLA)*, Anaheim, California, USA, Dec. 2016, pp. 259-265.

[21] B. Yildiz, J. I. Bilbao, J. Dore, and A. Sproul, "Household electricity load forecasting using historical smart meter data with clustering and classification techniques," *Proc. IEEE Innovative Smart Grid Technologies Conference - Asia (ISGT Asia)*, Singapore, May 2018, pp. 873-879.

[22] F. L. Quilumba, W. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 911-918, Mar. 2015.

[23] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 159-165, Feb. 1978.

[24] Z. Zhang, R. Tavenard, A. Bailly, X. Tang, P. Tang, and T. Corpetti, "Dynamic time warping under limited warping path length," *Information Sciences*, vol. 393, pp. 91-107, Feb. 2017.

[25] J. Paparrizos and L. Gravano, "k-Shape: Efficient and accurate clustering of time series," ACM SIGMOD Conference, Melbourne, Australia, Jun. 2015,vol. 45, pp. 69-76.

[26] R. Tavenard, "tslearn Documentation". [Online]. Available: https://readthedocs.org/projects/tslearn/downloads/pdf/stable/. [Accessed: 02 February 2021]

[27] F. M. Alvarez, A. Troncoso, J. C. Riquelme, and J. S. Aguilar Ruiz, "Energy time series forecasting based on pattern sequence similarity," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no.8, pp. 1230-1243, Aug. 2011.

[28] F. M. Alvarez, A. Troncoso, J. C. Riquelme, and J. S. Aguilar Ruiz, "LBF: A labeled-based forecasting algorithm and its application to electricity price time series," *Proc. 8th IEEE International Conference on Data Mining*, Pisa, Italy, Dec. 2008, pp. 453-461.

[29] P. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53-65, Nov. 1987.

[30] K. Walstad, "Electric utility customer segmentation from advanced metering systems data," M.S. thesis, Department of Electric Power Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, Jun. 2020. [Online]. Available: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2778229. [Accessed: 22 August 2022]

[31] Norwegian Ministry of Petroleum and Energy. [Online]. Available: https://lovdata.no/dokument/SF/forskrift/2012-12-07-1157 (in Norwegian). [Accessed: 10 February 2021]