engineering proceedings

MDPI

Proceeding Paper

Exploring Optimal Strategies for Small Hydro Power Forecasting: Training Periods and Methodological Variations †

Duarte Lopes, Isabel Preto * and David Freire

Enlitia, 4300-394 Porto, Portugal; duarte.lopes@enlitia.com (D.L.); david.freire@enlitia.com (D.F.)
* Correspondence: isabel.preto@enlitia.com
† Presented at the 10th International Conference on Time Series and Forecasting, Gran Canaria, Spain, 15–17 July 2024.

Abstract: This study investigates optimal training intervals for small hydro power regression models, crucial for accurate forecasts in diverse conditions, particularly focusing on Portugal's small hydro portfolio. Utilizing a regression model based on kernel density estimation, historical hourly production values, and calendar variables, forecasts are generated. Various approaches, including dynamic time warping (DTW), "K-Means Alike", and traditional K-means clustering, are assessed for determining the most effective historical training periods. Results highlight the "K-Means Alike" approach, which, with a 2-month training period, outperforms conventional methods, offering enhanced accuracy while minimizing computational resources. Despite promising results, DTW exhibits increased computational demands without consistent performance superiority.

Keywords: hydro power forecast; time series clustering; dynamical historical data; time-sequence; reference time series information; clustering

check for updates

Citation: Lopes, D.; Preto, I.; Freire, D. Exploring Optimal Strategies for Small Hydro Power Forecasting: Training Periods and Methodological Variations. *Eng. Proc.* **2024**, *68*, 27. https://doi.org/10.3390/engproc2024068027

Academic Editors: Olga Valenzuela, Fernando Rojas, Luis Javier Herrera, Hector Pomares and Ignacio Rojas

Published: 9 July 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The training period of a small hydro power forecast model is crucial, as its sensitivity depends on the variability of hydrological conditions influenced by factors such as rainfall, streamflow, and environmental dynamics [1]. Opting for the appropriate training duration is critical. A longer period provides a wider range of data for the model to learn from, potentially improving accuracy. In contrast, a shorter period may restrict exposure to various hydrological scenarios, diminishing predictive reliability. Moreover, transmission system operator (TSO) strategies impact hydroelectric power generation, affecting reservoir management and grid balancing. Considering these factors help to tailor forecasting models to specific system conditions, improving accuracy.

This study explores optimal training intervals for a regression model designed for small hydro power, essential for reliable forecasts under diverse conditions. It specifically targets Portugal's total small hydro portfolio.

2. Methodology

Using a regression model based on kernel density estimation (KDE) as a predictive model, the focus of the presented work was to explore if customized optimal historical data would produce better and more accurate results. KDE serves as a regressor that is fine-tuned using historical datasets leading up to the day of forecasting, as its predictive quality depends on the quality of the given history.

2.1. Standard KDE Approach

The standard use of the KDE model for hydro prediction depends on two major phases: the training phase and the testing phase. Given the dynamic nature of hydrological data and their real-time availability constraints, the KDE model is retrained with each

use. Hydro power values are predicted for the day ahead with 1 h granularity. To fully understand this predictive method, certain terms require clarification. Available date (AD) or (D) is the date at which the prediction is made; from this date, we establish the prediction date (PD) the date on which we are making the prediction. There are two main variables: power average history—which is the average hourly or quarter-hourly hydro power from all the available historic data until 2 days prior to our available date, D-2—and the daily forecast, which is the combination of the forecasted precipitation for the prediction date (D+1) and the daily average hydro power of the latest available day, established as being D-2. This is used to approximately estimate the daily average hydro power of the day we want to predict.

During training, the predictive model utilizes historical and hourly data, with daily average hydro power and specific hours as inputs, targeting hourly average hydro power. In the standard approach, the historical data used are limited to one month (30 days) prior to the two days before the available date (AD/D), due to real data availability limitations. While this approach provides satisfactory results and serves as a baseline to assess the performance of other methods, relying solely on the preceding month may be an oversimplification of complex hydrological patterns and strategies Therefore, alternative approaches, which customize and determine an optimal historical period, could enhance the accuracy of our models.

Finally, during the predicting/testing stage, the daily forecast for the prediction date (PD/D+1) is provided and the corresponding hours for outputting the hydro power value are specified. The KDE model interprets the historical daily average data and, using the daily forecast as an approximate estimation, predicts the hydro power value for each hour. The entire process of training and predicting is highlighted in Figure 1.

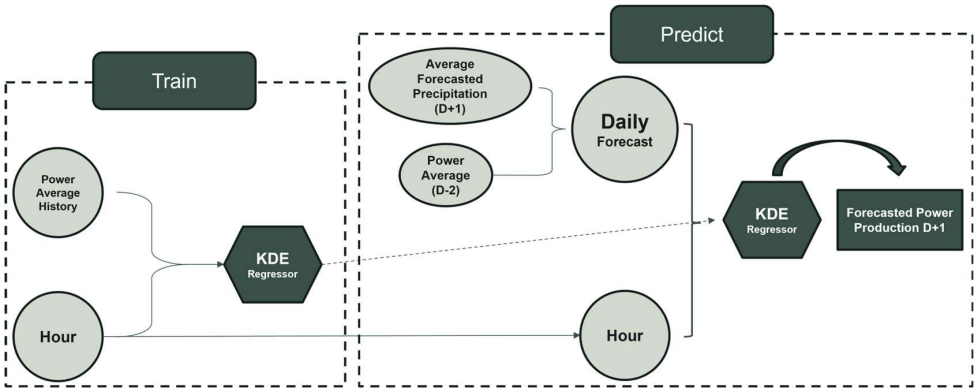


Figure 1. Graphic visualization of KDE model learning and predicting stages.

Using this standard approach, three other approaches were developed to optimize the training phase and achieve better overall performance. All three approaches focus on strategies to optimize our historical data so that they reflect the days on which our prediction data are most similar. If the historical data are more personalized to our prediction date, it is expected that the regressors will also improve. Each approach searches through all the points and defines the final training set using a different technique, but all consider two fundamental elements: AD Info—information related to the day on which the forecast is performed—and the available historical data, i.e., all the timestamps from the historical dataset that could potentially be included in our training set. The three developed approaches are as follows: d dynamic history, based on dynamic time warping (DTW); the K-means alike method, resembling K-means; and a final approach, traditional K-means, based on traditional K-means clustering.

2.2. DTW Dynamic History

The first approach was based on dynamic time warping (DTW) [2], a powerful technique used in time series analysis and pattern recognition to measure similarity between

two sequences that may vary in time or speed. Unlike traditional measures such as Euclidean distance, DTW aligns sequences in a nonlinear manner, allowing for more flexible matching. This is extremely helpful when dealing with sequences of varying lengths or when recognizing similarity between subsequences within a larger sequence. The DTW algorithm is able to take two sequences as input, and outputs a measure of similarity or dissimilarity between them, with a lower value indicating a higher degree of resemblance.

Utilizing this powerful tool, we aimed to use the DTW output value as our metric to select the best historical data for the specific available date (AD/D) on which we were to train our model and make predictions for the following day. The first step was to divide all the historical data into one-month time sequences, spaced out 2 weeks apart, totaling 24 time-sequences within one year of data. Next, acknowledging the potential limitations of daily data in adequately representing a specific period, we utilized data from the seven days leading up to two days before our forecast day as the reference time series information, or as we previously referred to it, our AD Info. This time sequence serves as our representation of the day we want to make predictions. Assuming some degree of stability across these days, we accepted the preceding seven days as the best means to characterize the information about our forecast day and, subsequently, the best way to identify the optimal training set. Finally, each of the one-month historical segments is evaluated against our AD Info using DTW as a metric, and the pair with the lowest DTW value is chosen as the optimal historical data training set.

This method, as the standard, not only requires our training set to have a length of one month (30 days), but also ensures a temporal dependency between the chosen points (as they are part of a continuous time sequence) and avoids sparsity. However, this approach offers a more personalized choice, which, in theory, assures that the points used during the training phase are much more like the ones we want to forecast. We maintain the underlying thesis that power production is directly correlated to meteorological variables, and different times of the year have specific characteristics. For example, if we are predicting particularly high-power production and a possibly rainy week in April, using a one-month sequence from a time of year that traditionally experiences more rainfall, like December or January, would produce better results than using the last 30 days before that week. For this purpose, we let the DTW decide and choose which one-month segment produces the lowest value and consequently shows the highest similarity with our AD Info.

Figure 2 shows how, from all the available historical data, we created our 24 one month-segments leading up to our specific available date (AD/D). Figure 2 shows a practical example of the use of DTW as our deciding metric. Using D as our available date (AD/D), we made up our AD Info using values from D-9 until D-2. Then, this seven-day time sequence was compared, using DTW, with all the one-month segments shown in Figure 2. Finally, the segment that output the lowest DTW value was chosen, which in this case was Segment 2.

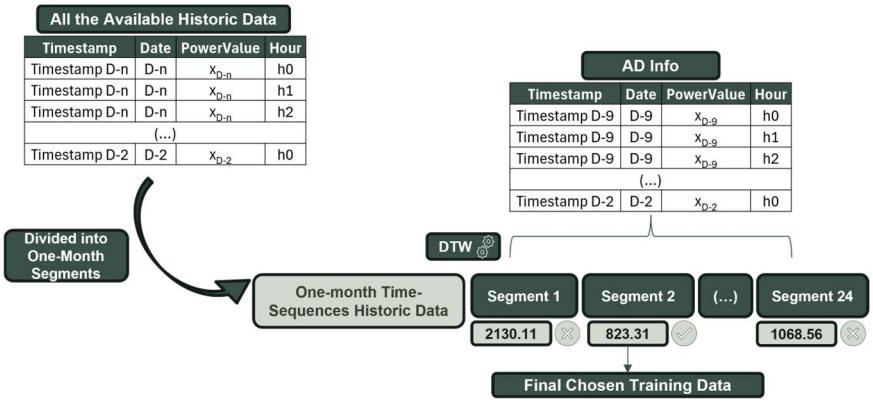


Figure 2. Graphic visualization of how DTW dynamic history operates.

2.3. K-Means Alike

The second approach is based on a modification to the traditional K-means algorithm [3] and attempts to select the available historical data using another premise. In contrast to the previous approach, K-means alike not only aims to choose the best time sequences but also intends to select the best points available without enforcing time continuity. Additionally, this method tests the optimal length for our training set because it not only aims to select the best and most similar points to the AD Info but also considers that one month of data might not be sufficient to achieve the best possible results.

Using the same reference time series information, denoted as AD Info, a centroid resembling a typical K-means centroid is calculated based on power production and the hour variable.

In contrast to the previous approach, the valuable information present in the AD Info could be translated into a centroid that summarized the days leading up to our forecasting day. Using Euclidean distance between this centroid and each point in the historical data, the n-nearest points closest to the centroid were identified. Subsequently, to optimize this process and determine the optimal length for our training history, this process was tested using different lengths, ranging from one to four months.

Assuming this process can accurately describe the information those data points hold about the forecasting day, it is only necessary to find the n most similar points to this center. This decision is made because, as highlighted before, the DTW method is very good at finding similarities between sequences, but if it were to be used to compare specific points, it would be outperformed by simpler methods like Euclidean distance.

As shown in Figure 3, once again, the period between D-9 and D-2 is used, but instead of computing a DTW distance directly, a centroid is created. Later, this same centroid is used to find the n closest points, which will comprise our training set, as described in Figure 3.

Finally, it is important to note that a version of this approach was also tested where instead of creating a centroid, a medoid was calculated using the same AD Info.

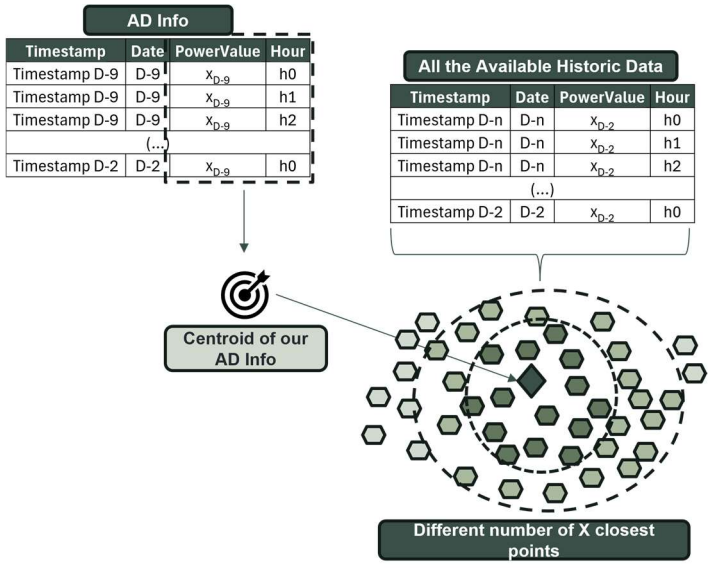


Figure 3. Graphic visualization of how K-means alike operates—creation of AD Info centroid and data point selection.

2.4. Traditional K-Means

For our last approach, we implemented a traditional K-means algorithm, which, in contrast with the previous approach, not only leverages the daily average of power but also incorporates the average daily precipitation as a variable. As previously stated, the value of hydro power is directly correlated with the value of precipitation. Therefore, we

defined this approach to identify not only the most similar days in terms of power but also regarding the forecasted value of precipitation for a given day.

For this method, adhering to how K-means is traditionally performed, we defined the number of specific clusters we wanted our data to be divided into. Just like with K-means alike, we aimed to test the possibility that one month may not be the ideal length for a training set. Therefore, the number of points used changed according to the number of clusters chosen (k). The K-means algorithm utilizes all the available data and the specified number of clusters to create k centroids and then assigns each data point to the closest centroid using the Euclidean distance. As expected, this method also does not enforce time continuity, and the selected points may come from different weeks and months. Finally, once the clusters are defined, the forecasted day—represented by a tuple of daily average power and precipitation—is assigned to a specific cluster, and all days corresponding to that cluster are included as the historical training data.

Although this approach is very similar in concept to K-means alike, it diverges from it in several ways. During the definition of our K-means space, instead of using the hourly or quarter-hourly hydro power values, we utilized tuples of daily average power combined with the daily average forecasted precipitation values. In contrast to what occurred using K-means alike, where each point represented one specific hour of one day, this time each point referred to a complete day. Furthermore, unlike the other approaches, this method does not incorporate the seven previous days leading to D-2; instead, it solely relies on the information from one day, specifically D-2. Because this method is based on the traditional algorithm of K-means, we wanted our AD Info (reference time series information) to be only one point in our 2D space. As depicted in Figure 4, our AD Info is only a tuple composed of the daily average power of D-2 and the daily average precipitation of D+1. Due to these two particularities, this method does not select points but instead selects days. This means that during this process, when we assigned our AD Info tuple to a cluster, we indicated that all the points from the days within that cluster would form our training set. This subtle distinction makes this method the only one that searches for typical days like our AD Info rather than for individual points associated with hours.

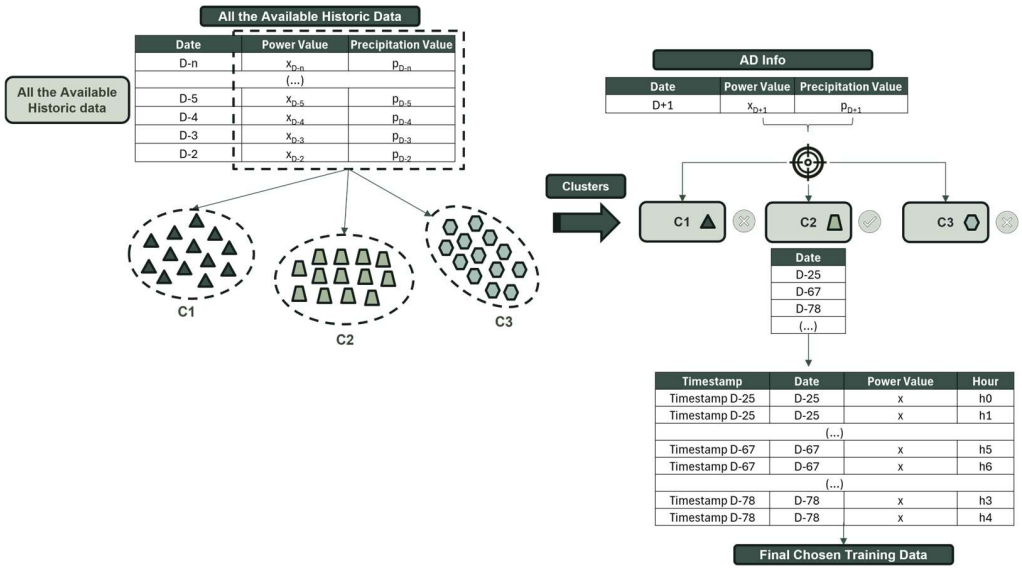


Figure 4. Graphic visualization of how traditional K-means works—creation of available data clusters and AD Info assignment.

Using the example described in Figure 4, we can observe that defining the clusters depends on the total number of clusters, which directly influences the number of samples in each segment—a higher number of clusters leads to smaller training sets. As previously stated, the K-means space is defined using the daily average values of both precipitation

and hydro power. In Figure 4, it is evident that after the different clusters are defined, we simply need to determine which cluster our AD Info point is assigned to. All the points (every hour) from the days present in the selected cluster will be used as the historical training data.

3. Results

3.1. Databases

All the approaches described in the previous section were developed and tested using three different NWP models. To check the viability of the methods presented, we not only compared the different approaches against each other and the standard, but we also used different weather-related data to assess if the results would hold using different data sources.

3.2. Results Standard vs. Dynamic Approaches

First, we defined NRMSE and MAPE as our major metrics to evaluate the overall performance of our approaches. Both these metrics are widely used in the forecast literature as they provide a comprehensive understanding of the forecasting model's performance, capturing both the magnitude and relative accuracy of predictions. Moreover, the training and predicting phases were both limited to enable a coherent analysis of the produced results. The available historical data ranged from 1 January to 26 October 2023; the predicting phase extended from 26 October 2023 to 2 February 2024. We only made prediction until this date to ensure that we would have the real values to evaluate the performance of each of our approaches with.

All the results are presented per approach, and in the last subsection, a comparative analysis is conducted to determine which approach had the best performance.

3.3. DTW Dynamic History

Table 1 shows the metrics referring to the DTW approach. As we can see, this approach demonstrates slightly better metrics for all the numerical weather prediction (NWP) models used.

Table 1. Result for standard KDE approach vs. DTW dynamic history for 3 different databases.

ID	MAPE	NRMSE
Standard Forecast NWP 1	8.16	10.57
DTW Dynamic History NWP 1	7.99	10.88
DTW Dynamic History NWP 2	7.87	10.55
DTW Dynamic History NWP 3	7.25	9.58

Even though one of the NWP models achieved the best result, overall, all NWP models yielded better values than our baseline, demonstrating the effectiveness of the latter. However, despite it achieving better results, this method has its drawbacks. DTW is very time-consuming, and comparing every segment to the AD Info takes a significant amount of time. Additionally, this approach requires the training time sequence to follow a temporal order, which can reduce the potential of reaching the optimal best-performing historical dataset.

3.4. K-Means Alike

The K-means alike approach also clearly outperforms the operational approach, reaching even better results than the DTW method. As the results shown in Table 2 indicate, using a 2-month length—1440 data points—yields significantly better results.

Table 2. Results for standard KDE approach vs. k-means alike for 3 different databases.

ID	MAPE	NRMSE
Standard Forecast Enlitia WRF	8.16	10.57
K-Means Alike NWP 1—1 Month Length	6.75	8.67
K-Means Alike NWP 1—2 Month Length	6.73	9.00
K-Means Alike NWP 1—3 Month Length	7.14	9.62
K-Means Alike NWP 1—4 Month Length	7.80	10.58
K-Means Alike NWP 2—1 MonthLength	6.99	8.82
K-Means Alike NWP 2—2 MonthLength	6.96	9.14
K-Means Alike NWP 2—3 MonthLength	7.16	9.63
K-Means Alike NWP 2—4 MonthLength	7.57	10.22
K-Means Alike NWP 3—1 MonthLength	6.46	8.33
K-Means Alike NWP 3—2 MonthLength	6.43	8.62
K-Means Alike NWP 3—3 MonthLength	6.81	9.24
K-Means Alike NWP 3—4 MonthLength	7.34	10.05

This analysis shows that the standard approaches not only fail to gather the best/most similar points to our reference data but also lack the optimal length to achieve the best possible results.

Lastly, these results show that when selecting the closest point to the AD Info, regardless of the temporal dependency, better results are obtained. This selection is based on each individual hourly point, leading to a very personalized training set. The centroid derived from the AD Info comprises the information about the previous seven days leading to D-2; then, the n closest points can be gathered from any month of the year, giving this method greater potential than employing the DTW distance between fixed one-month segments. The final results are very similar whether we use either a centroid or a medoid, with a slight advantage being observed when using the former.

3.5. Traditional K-Means

For this last set of results, instead of testing all the NWP models, due to data availability, we only performed the tests with database NWP 1. As depicted in Table 3, traditional K-means does not achieve the results produced by K-means alike. As we reduced the AD Info from the seven days leading up to D-2 to only information about D-2, and since we chose typical days and not specific data points, the results are not as good as before. Although we selected the same number of points, they are not as personalized as those from the previous approach.

Table 3. Result for standard KDE approach vs. traditional K-means.

ID	MAPE	NRMSE
Standard Forecast NWP 1	8.16	10.57
Traditional K-Means NWP 1—3 Clusters	8.16	11.01
Traditional K-Means NWP 1—6 Clusters	7.18	9.67
Traditional K-Means NWP 1—9 Clusters	7.70	9.88
Traditional K-Means NWP 1—12 Clusters	7.80	10.13

Nonetheless, traditional K-means could potentially yield better results if applied to different variables built as a function of precipitation, such as cumulative precipitation data, capturing trends within the dataset. This approach might enhance the identification of

'typical days', aiding in the selection of optimal training history. Moreover, both K-means alike and the traditional approach reach a similar optimal month length. Both approaches seem to reach training dimensions of two months of data, highlighting the existence of an ideal size for the training set.

3.6. Overall Results

The method that yielded the best results was K-means alike. This approach provided the best way to define the day we wanted to forecast—using the days leading up to D-2—and chose the training data based on point selection, implying much higher personalization than the other approaches. Moreover, K-means alike can choose any point from all the available points despite temporal dependency and delivers results much faster than the DTW method.

DTW exhibits promising results, albeit at the cost of increased computational time compared with baseline and K-means methodologies. Despite its computational demands, DTW does not consistently outperform other methods.

Traditional K-means tackles the problems using precipitation and the premise that selecting typical days could be beneficial. However, this method lacks the personalization that K-means alike offers and presents worse metrics. Nonetheless, this approach highlights that non-temporal dependency is better and outperforms both DTW and standard methods.

Overall, the results, presented in Table 4, are very positive and confirm the initial thesis that dynamically choosing our training set is undoubtedly beneficial for the development of better models.

Table 4. Overall results for standard KDE vs. all approaches.

ID	MAPE	NRMSE
Standard Forecast NWP 1	8.16	10.57
DTW Dynamic History NWP 1	7.99	10.88
K-Means Alike NWP 1—2 Month Length	6.88	9.24
Traditional K-Means NWP 1—6 Clusters	7.18	9.67

4. Conclusions

All the developed approaches enhance the results of the standard operational method. This evidences that a dynamic historical training set is beneficial. The provided results prove and highlight that more personalization leads to better results. Also, the results show that there is an optimal length for the number of points given to our model. Both traditional K-means and K-means alike results lead to the conclusion that a 2-month data set is ideal for reaching the best-performing model.

All the approaches described in this study demonstrate interesting results, emphasizing the importance of selecting appropriate historical data to achieve optimal outcomes. Proper clustering obviates the need for temporal dependencies in the data, instead of prioritizing the alignment of chosen days, and, more specifically, chosen data points, with the behavior of our AD (Available Date) Info.

The first conclusion on the use of different databases is that despite the data coming from different sources, the performance throughout the approaches was consistent. Although NWP3 data showed better overall performance regardless of the approach used, the results highlight that the data source only slightly influenced the final metrics and did not impact the overall analysis of our study.

Overall, this work shows that dynamically choosing data improves the quality of the final results. Also, different approaches show the advantages and disadvantages of different methods, but always produce exciting results compared with our baseline. Further enhancements can be achieved by reconsidering the characterization of our AD Info, the reference time series information, on the forecasted day. Exploring cumulative precipitation and seasonal clusters represents untapped potential for these methods to be refined.

Author Contributions: Conceptualization, D.L., D.F. and I.P.; methodology, D.L., I.P. and D.F.; software, D.L. and D.F.; validation, I.P. and D.L.; formal analysis, I.P. and D.L.; investigation, D.L.; writing—original draft preparation, D.L. and I.P; writing—review and editing, D.L. and I.P.; visualization, D.L.; supervision, I.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to company confidentiality matters.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Monteiro, C.; Ramirez-Rosado, I.J.; Fernandez-Jimenez, L.A. Short-term forecasting model for electric power production of small-hydro power plants. *Renew. Energy* **2013**, *50*, 387–394. [CrossRef]
2. Yadav, M.; Alam, M.A. Dynamic time warping (dtw) algorithm in speech: A review. *Int. J. Res. Electron. Comput. Eng.* **2018**, *6*, 524–528.
3. Ahmed, M.; Seraj, R.; Islam, S.M.S. The *k-means* Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics* **2020**, *9*, 1295. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.