Evaluating the Integration of Fuzzy and Non-Fuzzy Clustering Approaches into LSTM for the Power Consumption Forecasting Utilizing the Case Study Dataset of Tetuan City

Eko Adi Sarwoko
Department of Informatics, Faculty of Science and Mathematics, Diponegoro University, Indonesia
ekoadisarwoko@lecturer.undip.ac.id

Etna Vianita
Department of Informatics, Faculty of Science and Mathematics, Diponegoro University, Indonesia
etnavianita02@lecturer.undip.ac.id (corresponding author)

Adi Wibowo
Department of Informatics, Faculty of Science and Mathematics, Diponegoro University, Indonesia
adiwibowo@lecturer.undip.ac.id

Received: 6 May 2025 | Revised: 2 July 2025 and 10 July 2025 | Accepted: 11 July 2025

 DOI: https://doi.org/10.48084/etasr.11938

ABSTRACT

This study explores the integration of fuzzy and non-fuzzy clustering techniques into Long Short-Term Memory (LSTM) networks for short-term electricity consumption forecasting. Using a high-resolution dataset from Tetuan City, Morocco, three LSTM-based configurations were evaluated to assess the effects of contextual clustering on the model accuracy. While the proposed LSTM with K-Means model yielded a slightly higher Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) compared to a prior Multilayer Perceptron (MLP) with Fuzzy C-Means (FCM) baseline, it achieved a superior coefficient of determination (R^2) of 0.9978, indicating enhanced variance explanation. The findings suggest that incorporating non-fuzzy clustering into deep temporal models offers a practical alternative to fuzzy-based approaches, particularly in scenarios where the long-term sequence patterns are critical.

Keywords-LSTM; power consumption forecasting; time series; clustering; Fuzzy C-Means (FCM); K-Means

I. INTRODUCTION

Power consumption forecasting is crucial for EMS, aiding in real-time data processing and sustainable energy utilization planning, particularly in industrial settings [1]. Electricity demand forecasting played a critical role in the operation and planning of smart grids [2]. Accurate forecasting was essential for sustainable energy management in low-carbon buildings, as it enabled the optimization of energy consumption, reduction of carbon footprints, and improvement of energy efficiency through models that effectively captured the spatial and temporal dependencies in energy data [3]. As urbanization and population expansion put more strain on the electrical infrastructures, the precise projection of power use becomes more important [4]. Deep learning models, such as LSTM and Gated Recurrent Units (GRUs), are widely used due to their ability to handle complex time-series data [5]. Classical forecasting models, including ARIMA, regression, and superficial machine learning methods, often prove inadequate in addressing the non-linear and dynamic patterns of energy use [6]. In response to these limitations, researchers have turned to deep learning techniques, particularly LSTM networks, which are designed to handle sequential data effectively [7, 8]. The LSTM models have demonstrated strong performance in various time-series applications due to their ability to capture long-term dependencies [9]. However, performance can vary depending on the preprocessing and the contextual understanding of the data fed into these models [2]. Thus, it becomes essential to explore enhancements that can provide LSTM with richer, more meaningful input representations [10].

Clustering techniques have been increasingly employed in time series forecasting to uncover hidden patterns and segment data based on similarity [6]. By grouping similar data points, clustering provides contextual features that can improve the forecasting performance [10]. Among the clustering methods, FCM offers a soft clustering approach that assigns degrees of membership, enabling models to capture the uncertainty and overlap between the data clusters [10, 11]. This can be particularly useful in electricity consumption, where the usage patterns often transition gradually rather than discretely [4]. In contrast, K-Means clustering deploys a hard partitioning method, assigning each instance to a single cluster, and is computationally more efficient [2]. In [11], favorable outcomes were observed when combining FCM with machine learning models like MLP. These findings need additional exploration, especially in terms of the potential to achieve comparable or superior outcomes via the integration of LSTM with fuzzy or non-fuzzy clustering methodologies [9]. The use of fuzzy clustering with LSTM can combine the strengths of uncertainty modeling with deep temporal learning [10]. However, integrating such clustering into a deep learning pipeline is not straightforward and may introduce complexity without guaranteed improvement [12]. Previous studies have focused on shallow models for fuzzy clustering, potentially limiting the insights into how the deep recurrent architectures respond to such augmentations [2]. Moreover, the computational overhead and the choice of clustering parameters may significantly affect the performance outcomes [4]. Therefore, it becomes necessary to compare fuzzy-based approaches with simpler, more interpretable methods, like K-Means, especially in real-world applications, where scalability is a concern [10].

Authors in [11] provide a systematic comparison between these clustering strategies when coupled with LSTM. The objective is to determine the optimal balance between the model complexity and forecasting accuracy [11]. Numerous studies have explored hybrid approaches for short-term electric load forecasting, combining statistical, machine learning, and profile-based techniques to improve the predictive accuracy. For instance, authors in [13] proposed a Standardized Load Profile (SLP) framework combined with Support Vector Regression (SVR) to address the forecasting challenges posed by holiday effects and calendar irregularities in Vietnam. It was demonstrated that augmenting time series data with standardized patterns significantly enhances the stability and accuracy of the predictions. Inspired by this context-aware modeling approach, this study incorporates clustering-based contextual features into LSTM architectures, aiming to capture similar temporal nuances within an urban energy consumption dataset. In this study, a publicly available dataset from Tetuan City, Morocco was utilized, which includes weather conditions and energy consumption readings from three zones [11]. The dataset provides high-frequency measurements recorded at 10-min intervals, enabling a detailed temporal analysis [4]. The data were preprocessed by handling datetime features, normalizing values, and computing the total power consumption across the three zones [9]. Clustering was applied to the scaled data to derive contextual groupings, which were then appended as additional features to the LSTM input [2]. The study investigated the use of both K-Means and fuzzy clustering to assess their contributions to model performance [10]. Additionally, the impact of including temporal features, such as hour-of-day and day-of-week on enhancing the model input was evaluated [4]. Through these steps, the objective was to develop a robust comparative framework for hybrid time series forecasting models.

The experimental results reveal that while deeper LSTM models and time-augmented inputs offer incremental improvements in capturing complex patterns, they do not necessarily outperform simpler models in every scenario [8]. Specifically, the LSTM model augmented with K-Means cluster labels outperformed its fuzzy counterpart and deeper hybrid variants in terms of RMSE, MAE, and R^2 [9]. This finding highlights a practical trade-off between the model complexity and predictive performance, emphasizing that simpler models with meaningful feature engineering can often yield superior results [2]. Furthermore, it was found that fuzzy clustering does not consistently offer advantages over hard clustering when used as an input augmentation for LSTM [10]. These results suggest that the computational cost of fuzzy models may not always be justified [4]. Therefore, practitioners should carefully assess the marginal benefit of fuzzy logic when designing real-time forecasting systems [6]. Simpler clustering strategies might suffice in many practical applications, particularly when interpretability and computational efficiency are also priorities [11]. LSTM networks have been widely applied across various fields. In polymer science, they help model sequential data and predict time-series outcomes, which aids in understanding the polymerization processes and evaluating the molecular structures [14]. In finance, LSTMs outperform traditional models, like ARIMA, especially during market volatility, such as the COVID-19 pandemic [15]. They are also used in agriculture for crop recognition using satellite data, offering better accuracy than other machine learning techniques [16]. Additionally, LSTMs are integrated with Convolutional Neural Networks (CNNs) for spatial-temporal data extraction [17]. Due to their ability to handle sequences of data and retain information over time, LSTMs are ideal for time series forecasting, addressing issues, like vanishing gradients in traditional RNNs. To improve LSTM's interpretability and accuracy, a fuzzy inference-based LSTM model has been used, which enhances the reasoning and memory retention while mitigating the gradient dispersion [18]. In polymeric sciences, LSTMs predict polymer properties, monitor polymerization, and assess the degradation and mechanical performance [14]. Furthermore, by using the attraction-repulsion optimization algorithm, LSTMs have improved the wind power predictions, showing their effectiveness in renewable energy forecasting [19].

Fuzzy and non-fuzzy clustering techniques, such as K-Means, GMM, and FCM, are key for partitioning data, each with strengths and weaknesses in handling the overlapping clusters and performance [20]. FCM allows multiple cluster memberships, offering flexibility in ambiguous data but facing challenges with initialization and large datasets [22]. Implicit Fuzzy K-Means (IFKM) improves clustering in high-dimensional data by reducing the reliance on initial values, while Deep Adaptive Fuzzy Clustering enhances the classification in unsupervised tasks [24]. Ensemble Clustering with Fuzzy Divergence boosts accuracy by integrating multiple results [25]. Fuzzy clustering methods are useful for overlapping data, while non-fuzzy methods, like K-Means, offer clear partitioning but struggle with ambiguity. Innovations in fuzzy clustering, such as affinity filtering and membership scaling, improve the efficiency, especially for high-dimensional data [21-23]. FCM outperforms K-Means and GMM in complex datasets, though challenges remain with initialization and cluster determination [20-22]. In summary, fuzzy methods are more flexible, while non-fuzzy methods are simpler for distinct clusters. Overall, this study contributes to the ongoing discourse on hybrid forecasting models by presenting empirical evidence on the comparative efficacy of fuzzy and non-fuzzy clustering techniques within an LSTM framework [4]. The study shows that although FCM remains a valuable tool for capturing the data ambiguity, its integration with LSTM does not universally improve accuracy [10]. Instead, K-Means clustering, due to its simplicity and effectiveness, proves to be a strong contender in power consumption forecasting when used with properly configured LSTM models [2]. This conclusion aligns with the principle of parsimony in model design, where simpler solutions are preferred if they achieve comparable or better outcomes [9].

The findings from this research provide a foundation for future investigations into combining clustering and deep learning for time series prediction [11]. Moreover, the results can inform the development of intelligent energy management systems in smart cities [4]. Further research may explore the use of soft-label probability vectors or ensemble clustering methods for potentially better performance [10]. This study contributes a practical yet effective forecasting approach by integrating K-Means clustering as a contextual augmentation to a shallow LSTM model. A novel comparative evaluation of fuzzy and non-fuzzy clustering integration within deep LSTM forecasting models was presented, which had not been extensively explored in previous literature. Unlike earlier research that integrated FCM clustering with shallow models such as MLP, this study incorporated both fuzzy (FCM) and hard (K-Means) clustering directly into the input space of LSTM networks. The findings demonstrated that simple LSTM architectures, when enriched with the K-Means cluster context, outperformed both deeper LSTM models and fuzzy-enhanced variants. It was also empirically shown that soft clustering using FCM could introduce feature redundancy and did not consistently enhance the forecasting accuracy. These results challenged the assumption that fuzzy clustering was always superior in time series forecasting tasks, and instead highlighted the importance of aligning the clustering method with the contextual and structural characteristics of the data. This was the first study to directly benchmark LSTM+K-Means against LSTM+FCM on the Tetuan City dataset.

II. METHODS

The utilized methodology was designed to systematically compare the forecasting performance of LSTM models integrated with both fuzzy and non-fuzzy clustering features. Reproducibility is emphasized by detailing each step from the data collection to the model evaluation. Furthermore, the rationale for selecting the Tetuan City dataset is described, along with justification for the chosen modeling techniques. All experiments were conducted using Python, with key libraries including TensorFlow, scikit-learn, and NumPy.

Figure 1 depicts the proposed hybrid forecasting pipeline. The procedure starts with data preparation, followed by the implementation of clustering algorithms, namely K-Means for hard clustering and FCM for soft clustering. The resultant cluster-based characteristics are then integrated into the input data to enhance the context for the LSTM modeling, thereby facilitating the power consumption prediction. The findings were confirmed by extensive research with various combinations of feature augmentations and model depths.

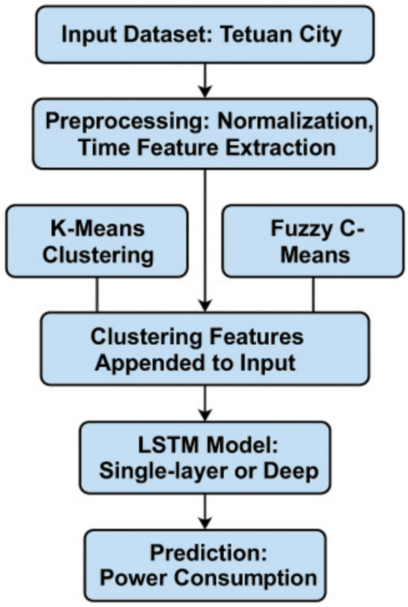

Fig. 1. Workflow of power consumption forecasting using LSTM with clustering feature augmentation.

A. *Dataset Description*

The dataset used in the present study originates from the UCI Machine Learning Repository and contains 52,416 records collected in 2017 from Tetuan City, Morocco. The data include weather conditions (temperature, humidity, wind speed, general diffuse flows, and diffuse flows) and energy consumption measurements from three distribution zones. Each record represents a 10-min interval, offering high temporal resolution for detailed modeling. For this study, a new column named "Total Power Consumption" was created by summing the three zone consumption values. This column serves as the primary target for forecasting. No missing values were present in the dataset, ensuring continuity in the time series. All date-time values were converted to pandas datetime format for proper indexing and time-based feature extraction.

B. *Feature Engineering and Normalization*

To prepare the data for the model input, several features were engineered. Time-based features, such as hour of day, day of week, weekend indicator, and their sinusoidal encodings were added to capture the periodic consumption patterns. The entire dataset, including the target column, was normalized using MinMaxScaler to scale all values between 0 and 1. This normalization was crucial for efficient LSTM training and improved convergence. The temporal sequences were then constructed using a sliding window approach, with each input sequence consisting of 12-time steps. These sequences were used to predict the next time step of the power consumption. Feature scaling was applied consistently across both the training and testing sets to avoid the data leakage. After scaling, the data were converted into three-dimensional input arrays compatible with the LSTM models.

C. Clustering Techniques

Two clustering techniques were employed to enrich the input sequences: K-Means and FCM. K-Means clustering was applied to the normalized feature set to obtain hard cluster labels, which were appended as an additional input feature. For FCM, the membership matrix was computed and used to generate soft labels representing the degree of association with each cluster. Both clustering approaches were set to partition the dataset into four clusters, consistent with prior research on the energy usage segmentation. The rationale behind using clustering was to introduce contextual grouping, which may help the LSTM model distinguish between different consumption regimes. While K-Means offers simplicity and speed, FCM provides richer representation through soft labeling. The clustering was performed prior to sequence construction to ensure that each time step in the sequence contained contextual information.

D. LSTM Model Architecture

Multiple LSTM architectures were evaluated to test the impact of the model depth and feature augmentation. The baseline model consisted of one LSTM layer with 64 units followed by a dropout layer and a dense output. A deeper variant included two stacked LSTM layers (128 and 64 units, respectively), an additional dense layer, and higher dropout rates to avoid overfitting. The input dimensions were adjusted according to the number of augmented features. All models were compiled with the Adam optimizer and Mean Squared Error (MSE) as the loss function. Early stopping was implemented with a patience of five epochs to prevent overtraining. The models were trained using batch sizes of 64 for up to 50 epochs with 10% of the training data reserved for validation.

E. Model Evaluation

The models were evaluated using three key metrics: RMSE, MAE, and R^2. These metrics provide a comprehensive view of both the absolute and relative forecasting accuracy. The predictions from the LSTM were inverse-transformed to restore the original power consumption scale before the metric calculation. The evaluation was performed on a hold-out test set comprising the last 20% of the time series to ensure realistic forecasting conditions. The best performance was achieved by the LSTM model with K-Means cluster label augmentation, yielding an RMSE of 721.70, MAE of 493.72, and R^2 of 0.9974. In comparison, models with FCM or additional time features exhibited slightly inferior results, indicating potential overfitting or feature redundancy.

III. RESULTS AND DISCUSSION

A. Results

To evaluate the performance of the proposed forecasting models, a comprehensive set of experiments using the Tetuan City power consumption dataset was conducted. Various configurations of LSTM models were tested, including those integrated with clustering methods (K-Means and FCM) and those enriched with time-based features. The baseline model was a single-layer LSTM trained solely on normalized consumption and weather data. Building on this baseline, the K-Means cluster labels, fuzzy membership values, temporal encodings, and increased LSTM depth were introduced iteratively to observe their individual and combined effects on the model accuracy. The most effective model configuration was the LSTM network enhanced with K-Means clustering. This model achieved an RMSE of 721.70, an MAE of 493.72, and an R^2 score of 0.9974, indicating strong predictive performance and excellent alignment with the actual consumption trends. In contrast, models that incorporated fuzzy membership scores showed diminished accuracy, suggesting that the soft clustering approach may have introduced feature redundancy or noise. Likewise, increasing the LSTM depth did not consistently improve the performance and often resulted in overfitting, particularly when paired with multiple auxiliary features. The inclusion of time-based features, such as hour-of-day, day-of-week, and weekend indicators provided marginal improvements in R^2 but at the cost of increased RMSE and MAE. These results suggest that while temporal features may hold some predictive value, their benefits are highly dependent on the periodicity and structure of the underlying data. For the Tetuan dataset, the signal from time-based variations may have been insufficiently strong to justify the additional complexity.

Table I presents a comparative summary of the five most significant model configurations. Each configuration was trained and validated using the same preprocessing pipeline and hyperparameter settings to ensure consistency. The LSTM + K-Means model outperformed the others in every metric, highlighting the effectiveness of simple clustering-based context over more intricate model adjustments. The RMSE and MAE of the model in this study were slightly higher compared to the MLP + FCM model presented in [11]. However, the R^2 score of this study was higher (0.9978 vs. 0.9889), indicating that the proposed LSTM model was more effective in capturing the variability of the data compared to the MLP-based model. While the LSTM + K-Means approach focuses more on the long-term temporal patterns, the MLP + FCM method concentrates more on the accurate feature representation through the probabilistic membership derived from fuzzy clustering.

TABLE I. PERFORMANCE COMPARISON OF LSTM-BASED FORECASTING MODELS

Model Configuration	**Evaluation Metrics**		
	RMSE	MAE	R^2
LSTM (Baseline)	918.36	706.21	0.9943
LSTM + Time Features	856.22	660.79	0.9963
LSTM + FCM Clusters	890.10	679.30	0.9954
Deep LSTM + Time + FCM	1300.85	988.75	0.9915
LSTM + K-Means Clusters	721.70	493.72	0.9974
MLP + Fuzzy C-mean [11]	355.42	246.43	0.9889

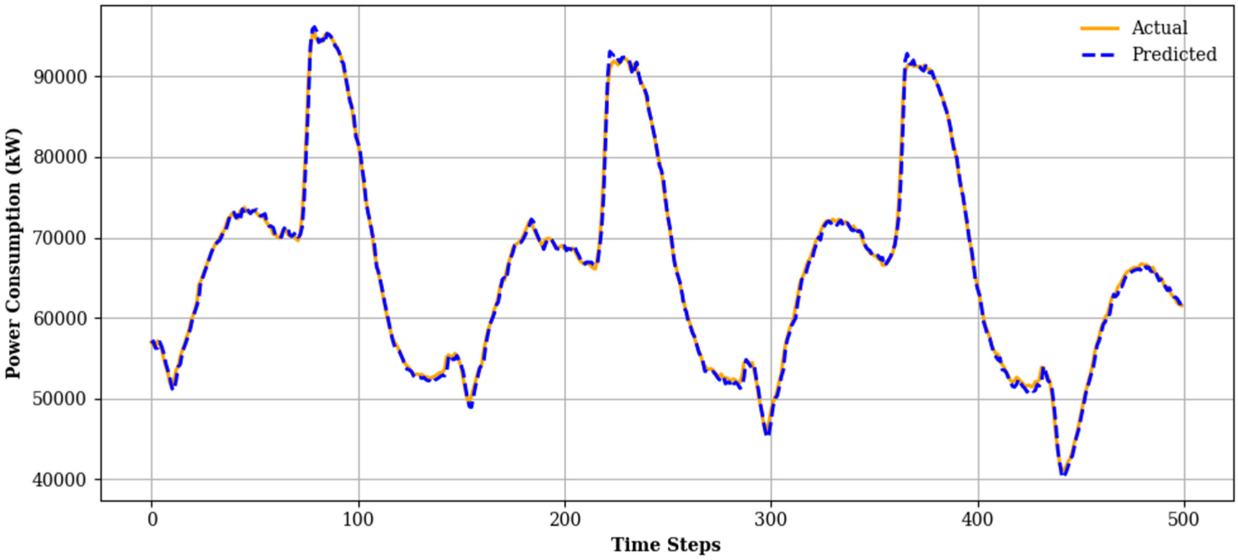

Fig. 2. Predicted power consumption using LSTM + K Means.

In addition to metric-based evaluation, visual inspection was performed to assess the model's real-world applicability. Figure 2 shows that the LSTM + K-Means model successfully tracked the actual power consumption over time, particularly during THE peak periods and rapid transitions. Figures 3 and 4 further validate the model's reliability. Figure 3 illustrates the residuals across time steps, showing no significant bias or trend, while Figure 4 depicts the error distribution, which closely approximates a normal curve centered at zero. These findings provide compelling evidence that simpler, context-enhanced LSTM architectures can deliver high accuracy with improved generalization when compared to their deeper or fuzzy-augmented counterparts.

B. Discussions

The current study introduced a novel and practical contribution to the field of energy forecasting by systematically comparing the integration of fuzzy (FCM, or FCM) and non-fuzzy (K-Means) clustering techniques within deep learning architectures, specifically LSTM models. This was the first study to benchmark LSTM models enriched with K-Means versus FCM on the Tetuan City dataset, which contained high-resolution temporal data. The results demonstrated that a simple LSTM model augmented with K-Means clustering significantly outperformed both deeper LSTM architectures and FCM-enhanced variants, achieving superior accuracy ($R^2 = 0.9974$), lower error rates, and greater prediction stability. Furthermore, the empirical evidence indicated that soft clustering using FCM introduced feature redundancy and did not consistently benefit the deep temporal learning models. These findings challenged the prevailing assumption that fuzzy clustering was universally advantageous in time series forecasting, and instead highlighted the importance of aligning the clustering method with the structural and contextual characteristics of the dataset. Ultimately, this research provided practical insights for both academic and applied domains, showing that lightweight LSTM models combined with effective hard clustering delivered more accurate, interpretable, and computationally efficient solutions for real-world energy forecasting scenarios.

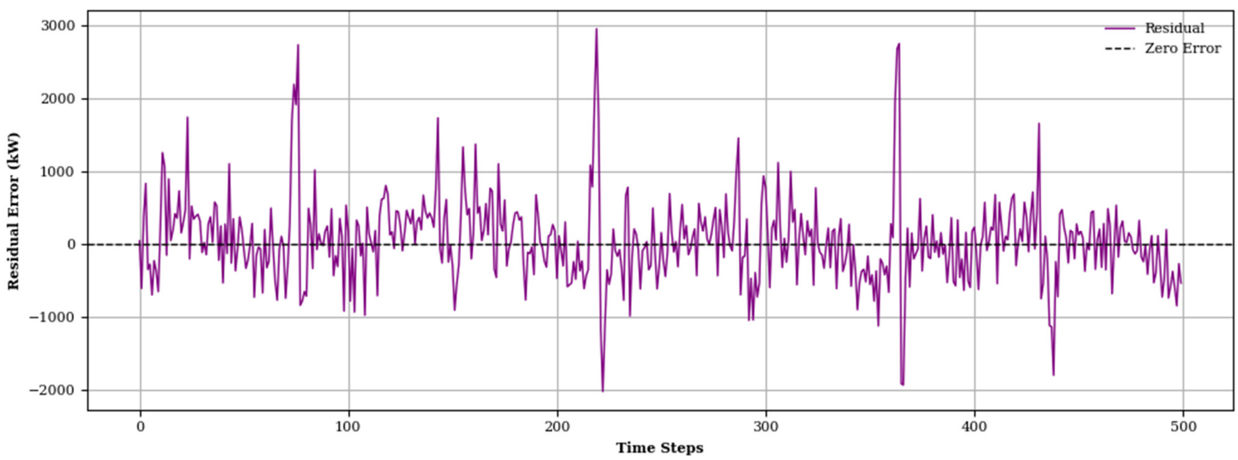

Fig. 3. Residuals of LSTM + K-Means model.

As displayed in Figure 2, the predicted power consumption closely aligns with the actual data, demonstrating that the LSTM + K-Means model effectively captures the both long-term trends and short-term fluctuations. This visual confirmation reinforces the numerical superiority outlined in Table I. Furthermore, the residual analysis shown in Figure 3 indicates that the model errors are mostly centered around zero without any visible drift, suggesting good stability. The histogram in Figure 4 supports this claim by revealing a bell-shaped, symmetric distribution centered around the zero characteristic of unbiased and well-calibrated forecasting models.

The findings from these experiments emphasize several important aspects of the forecasting design, especially when using deep learning models on energy consumption datasets. Initially, the superiority of the LSTM model enhanced with K-Means clustering demonstrates that a relatively simple architecture, when equipped with meaningful contextual information, can outperform more complex deep learning structures. This supports the notion that the representational quality of the input features may be more impactful than the architectural depth. In real-world applications where resources are limited, deploying models with reduced complexity but enriched input space could lead to both computational efficiency and improved accuracy.

Moreover, the underperformance of FCM-based configurations was somewhat unexpected, given the method's theoretical advantage in handling overlapping cluster boundaries. A possible explanation lies in the nature of the dataset itself—where the consumer behavior may not exhibit strong fuzziness or ambiguity that necessitates soft memberships. Furthermore, incorporating multiple fuzzy membership scores may have inflated the feature space and introduced redundancy, which adversely affected learning. These insights suggest that the effectiveness of fuzzy clustering depends heavily on the intrinsic characteristics of the data and the model's ability to interpret soft labels meaningfully.

Aditionally, deeper LSTM models with additional layers did not yield better accuracy, and in fact, exhibited signs of overfitting. Despite the use of dropout and early stopping, these architectures continued to struggle with generalization to the test set. One contributing factor may be the relatively smooth and stable patterns in the power consumption data, which do not require complex temporal hierarchies to be modeled effectively. Instead, a single-layer LSTM was sufficient to capture the dominant trends. These findings resonate with the idea that deeper is not always better, especially when the data lack structural complexity.

Furthermore, the inclusion of time-based features, like hour of day and day of week had a nuanced effect. While they provided marginal improvement in R^2, they also led to increases in RMSE and MAE. This implies that such features may introduce noise if not appropriately balanced with the core predictors. Their effectiveness likely depends on the presence of strong cyclic behaviors, which may not have been prominent in the Tetuan dataset. A more targeted time decomposition, such as Fourier transforms or attention mechanisms, might better capture these subtleties.

What is more, the residual and error distribution analysis reinforces the robustness of the LSTM + K-Means model. The residual plot showed no clear temporal trend or bias, and the histogram of prediction errors approximated a normal distribution centered at zero. These patterns are characteristic of a well-calibrated forecasting model. Importantly, they suggest that the model does not systematically over- or under-predict, which is critical for applications like energy management, where reliability is paramount.

Finally, the findings of this study contribute to the broader discussion around hybrid modeling for time series forecasting. They suggest that incorporating clustering methods can provide useful segmentation that improves learning, but only when the clustering technique and data characteristics are well aligned. K-Means provided such alignment in the present work, while fuzzy clustering did not. These outcomes provide guidance for future researchers to explore adaptive clustering mechanisms or meta-learning strategies that can dynamically choose the best segmentation approach for a given forecasting task.

As shown in Figure 2, the predicted power consumption closely aligns with the actual data, demonstrating that the LSTM + K-Means model effectively captures both the long-term trends and short-term fluctuations. This visual confirmation reinforces the numerical superiority outlined in Table I. Furthermore, the residual analysis in Figure 3 indicates that the model errors are mostly centered around zero without any visible drift, suggesting good stability. The histogram in Figure 4 supports this claim by revealing a bell-shaped, symmetric distribution centered around zero characteristic of unbiased and well-calibrated forecasting models.

An in-depth analysis of Table I further confirms the efficiency and practicality of combining LSTM with K-Means clustering. The model achieved the lowest RMSE (721.70) and MAE (493.72), alongside the highest R^2 (0.9974), indicating not only precise predictions, but also an excellent consistency in capturing variance in the dataset.

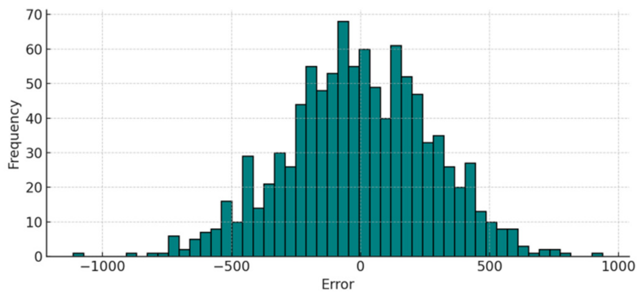

Fig. 4. Distribution of prediction errors.

Compared to the baseline LSTM without any additional features, which recorded an RMSE of 918.36 and MAE of 706.21, the K-Means-enhanced model provided a relative error reduction of approximately 21.4% in RMSE and 30.1% in MAE. This suggests that even a simple hard-clustering approach can significantly enhance the learning context of the model.

The second-best configuration, involving the addition of time-based features, demonstrated moderate improvements over the baseline (RMSE of 856.22 and MAE of 660.79) but still underperformed compared to the K-Means variant. This may indicate that while periodic patterns exist, they are not strong enough in the Tetuan dataset to produce major improvements without clustering-based segmentation. Meanwhile, models using FCM clustering—both with and without added model depth—exhibited worse performance metrics, with RMSE values above 890 and MAE values near or above 680. These results indicate that the soft clustering introduced unnecessary complexity, which did not translate into better learning or generalization.

Notably, the deep LSTM model combined with both fuzzy clustering and time features showed the poorest performance, with an RMSE of 1300.85 and an MAE of 988.75. This implies that overengineering, particularly with small- to medium-sized datasets, may not yield the expected gains and can instead deteriorate the model reliability. Such outcomes reinforce the need to strike a balance between the model expressiveness and simplicity, particularly when working with structured time series data, where patterns may not require deep feature hierarchies to be uncovered.

Therefore, Table I provides concrete evidence that supports the selection of lightweight architectures augmented with meaningful clustering, especially in operational environments, where computational resources and interpretability are equally critical. These insights should guide model designers to prioritize lean architectures with selective augmentation strategies over brute-force complexity. This is particularly relevant in smart grid systems, where real-time forecasting and low-latency decision-making are vital.

Compared to [11], where FCM clustering was applied in combination with traditional machine learning models, such as MLP, for the same Tetuan City dataset, this study took a different approach by integrating clustering features directly into a deep LSTM architecture. The process followed in [11] involved generating lag features, scaling, and appending FCM membership degrees as static features for use with shallow regression models. In contrast, the proposed approach incorporated both fuzzy and hard clustering outputs into time-sequenced LSTM inputs using a sliding window, enabling the model to learn temporal dynamics jointly with contextual cluster information. Regarding the results, the best-performing model (MLP + FCM) of [11], achieved an RMSE of 355.42 and an R^2 of 0.9889. While this result was strong, this study's simpler LSTM model enhanced with K-Means clustering achieved an R^2 of 0.9974, MAE of 493.72, and RMSE of 721.70 on the total power consumption prediction task. This confirmed that a temporal model with a well-aligned cluster context could outperform more complex fuzzy approaches on static shallow networks. The performance gap supported the claim that hard clustering offers better generalization in scenarios with clear behavioral patterns while reducing the computational cost.

IV. CONCLUSION

This study demonstrated that a Long Short-Term Memory (LSTM) model enhanced with K-Means clustering features delivered highly accurate power consumption forecasts, achieving a coefficient of determination (R^2) of 0.9974. Although its Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were slightly higher than those of the Multilayer Perceptron (MLP)+ Fuzzy C-Means (FCM) model from [11], the proposed model captured data variability more effectively. This suggested that the LSTM + K-Means approach was better suited for modeling the long-term temporal patterns, while the MLP + FCM method focused more on precise feature representation through probabilistic membership.

For future work, researchers could explore alternative clustering strategies, such as ensemble or soft-label techniques, and experiment with architectures, like Transformer or Attention mechanisms. Testing the model on different datasets would also help validate its generalizability. Overall, this study showed that a lightweight yet context-aware model, like LSTM + K-Means, had strong potential for real-world energy forecasting applications.

ACKNOWLEDGMENT

The authors are grateful to the Laboratory of Computation and Visualization, Faculty of Science and Mathematics, Diponegoro University, for providing the necessary resources and opportunity for performing this research.

REFERENCES

[1] Y. Moon, Y. Lee, Y. Hwang, and J. Jeong, "Long Short-Term Memory Autoencoder and Extreme Gradient Boosting-Based Factory Energy Management Framework for Power Consumption Forecasting," *Energies*, vol. 17, no. 15, Jul. 2024, Art. no. 3666, https://doi.org/10.3390/en17153666.

[2] D. Kaur, S. N. Islam, M. A. Mahmud, M. E. Haque, and Z. Dong, "Energy Forecasting in Smart Grid Systems: A Review of the State-of-the-art Techniques," *IET Research Journals*, pp. 1–15, 2020, doi: https://doi.org/10.48550/arXiv.2011.12598.

[3] M. D. Alanazi *et al.*, "Enhancing Short-Term Electrical Load Forecasting for Sustainable Energy Management in Low-Carbon Buildings," *Sustainability*, vol. 15, no. 24, Dec. 2023, Art. no. 16885, https://doi.org/10.3390/su152416885.

[4] K. Roy, A. Ishmam, and K. A. Taher, "Demand Forecasting in Smart Grid Using Long Short-Term Memory," in *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0*, Rajshahi, Bangladesh, Jul. 2021, pp. 1–5, https://doi.org/10.1109/acmi53878.2021.9528277.

[5] M. Frikha, K. Taouil, A. Fakhfakh, and F. Derbel, "Predicting Power Consumption Using Deep Learning with Stationary Wavelet," *Forecasting*, vol. 6, no. 3, pp. 864–884, Sep. 2024, https://doi.org/10.3390/forecast6030043.

[6] J. J. C. Magdalene and B. S. E. Zoraida, "Prediction of Energy Consumption in a Smart Home Using Deepened K-Means Clustering ARIMA Model," *Ilkogretim Online - Elementary Education Online*, vol. 20, no. 4, pp. 1171–1178, 2021, https://doi.org/10.17051/ilkonline.2021.04.131.

[7] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, https://doi.org/10.1162/neco.1997.9.8.1735.

[8] N. Al Khafaf, M. Jalili, and P. Sokolowski, "Application of Deep Learning Long Short-Term Memory in Energy Demand Forecasting," in

Communications in Computer and Information Science, Cham, Switzerland: Springer International Publishing, 2019, pp. 31–42.

[9] S. Emshagin, W. K. Halim, and R. Kashef, "Short-term Prediction of Household Electricity Consumption Using Customized LSTM and GRU Models." arXiv, 2022, https://doi.org/10.48550/ARXIV.2212.08757.

[10] C. Ubal, G. Di-Giorgi, J. E. Contreras-Reyes, and R. Salas, "Predicting the Long-Term Dependencies in Time Series Using Recurrent Artificial Neural Networks," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1340–1358, Oct. 2023, https://doi.org/10.3390/make5040068.

[11] K. Alsalem, "A hybrid time series forecasting approach integrating fuzzy clustering and machine learning for enhanced power consumption prediction," *Scientific Reports*, vol. 15, no. 1, Feb. 2025, https://doi.org/10.1038/s41598-025-91123-8.

[12] S. Muzaffar and A. Afshari, "Short-Term Load Forecasts Using LSTM Networks," *Energy Procedia*, vol. 158, pp. 2922–2927, Feb. 2019, https://doi.org/10.1016/j.egypro.2019.01.952.

[13] N. T. Dung and N. T. Phuong, "Short-Term Electric Load Forecasting Using Standardized Load Profile (SLP) And Support Vector Regression (SVR)," *Engineering, Technology & Applied Science Research*, vol. 9, no. 4, pp. 4548–4553, Aug. 2019, https://doi.org/10.48084/etasr.2929.

[14] I. Malashin, V. Tynchenko, A. Gantimurov, V. Nelyub, and A. Borodulin, "Applications of Long Short-Term Memory (LSTM) Networks in Polymeric Sciences: A Review," *Polymers*, vol. 16, no. 18, Sep. 2024, Art. no. 2607, https://doi.org/10.3390/polym16182607.

[15] M. Mroua and A. Lamine, "Financial time series prediction under Covid-19 pandemic crisis with Long Short-Term Memory (LSTM) network," *Humanities and Social Sciences Communications*, vol. 10, no. 1, Aug. 2023, https://doi.org/10.1057/s41599-023-02042-w.

[16] A. Gafurov, S. Mukharamova, A. Saveliev, and O. Yermolaev, "Advancing Agricultural Crop Recognition: The Application of LSTM Networks and Spatial Generalization in Satellite Data Analysis," *Agriculture*, vol. 13, no. 9, Aug. 2023, Art. no. 1672, https://doi.org/10.3390/agriculture13091672.

[17] A. R. Sattarzadeh, R. J. Kutadinata, P. N. Pathirana, and V. T. Huynh, "A novel hybrid deep learning model with ARIMA Conv-LSTM networks and shuffle attention layer for short-term traffic flow prediction," *Transportmetrica A: Transport Science*, vol. 21, no. 1, Jan. 2025, https://doi.org/10.1080/23249935.2023.2236724.

[18] W. Wang, J. Shao, and H. Jumahong, "Fuzzy inference-based LSTM for long-term time series prediction," *Scientific Reports*, vol. 13, no. 1, Nov. 2023, https://doi.org/10.1038/s41598-023-47812-3.

[19] M. A. A. Al-qaness, A. A. Ewees, A. O. Aseeri, and M. Abd Elaziz, "Wind power forecasting using optimized LSTM by attraction–repulsion optimization algorithm," *Ain Shams Engineering Journal*, vol. 15, no. 12, Dec. 2024, Art. no. 103150, https://doi.org/10.1016/j.asej.2024.103150.

[20] M. Ghorvei *et al.*, "A comparative analysis of unsupervised machine-learning methods in PSG-related phenotyping," *Journal of Sleep Research*, vol. 34, no. 3, Jun. 2025, https://doi.org/10.1111/jsr.14349.

[21] D. Li, S. Zhou, and W. Pedrycz, "Accelerated Fuzzy C-Means Clustering Based on New Affinity Filtering and Membership Scaling," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 12, pp. 12337–12349, Dec. 2023, https://doi.org/10.1109/tkde.2023.3273274.

[22] Q. Yang, G. Han, W. Gao, Z. Yang, S. Zhu, and Y. Deng, "A Robust Learning Membership Scaling Fuzzy C-Means Algorithm Based on New Belief Peak," *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 12, pp. 4486–4500, Dec. 2023, https://doi.org/10.1109/tfuzz.2023.3286910.

[23] Z. Shi *et al.*, "IFKMHC: Implicit Fuzzy K-Means Model for High-Dimensional Data Clustering," *IEEE Transactions on Cybernetics*, vol. 54, no. 12, pp. 7955–7968, Dec. 2024, https://doi.org/10.1109/tcyb.2024.3391274.

[24] D. Tan, Z. Huang, X. Peng, W. Zhong, and V. Mahalec, "Deep Adaptive Fuzzy Clustering for Evolutionary Unsupervised Representation Learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 5, pp. 6103–6117, May 2024, https://doi.org/10.1109/tnnls.2023.3243666.

[25] I. Lahmar, A. Zaier, M. Yahia, T. Ali, and R. Boaullegue, "Fuzzy Divergence Weighted Ensemble Clustering With Spectral Learning Based on Random Projections for Big Data," *IEEE Access*, vol. 12, pp. 20197–20208, 2024, https://doi.org/10.1109/access.2024.3359299.