JCTA Journal of Computing Theories and Applications ISSN:3024-9104

Research Article

IoT-Based Home Electricity Monitoring and Consumption Forecasting using k-NN Regression for Efficient Energy Management

Apriandy Angdresey [1,*], Lanny Sitanayah [1], Zefanya Marieke Philia Rumpesak [1], and Jing-Quan Ooi [2]

[1] Department of Informatics Engineering, Universitas Katolik De La Salle, Manado 95253, Indonesia; e-mail: aangdresey@unikadelasalle.ac.id; lsitanayah@unikadelasalle.ac.id; 18013058@unikadelasalle.ac.id
[2] Department of Electrical and Electronics Engineering, Universiti Tunku Abdul Rahman, Selangor 43000, Malaysia; e-mail: jingquan03@1utar.my
* Corresponding Author: Apriandy Angdresey

Abstract: Electricity has emerged as an essential requirement in modern life. As demand escalates, electricity costs rise, making wastefulness a drain on financial resources. Consequently, forecasting electricity usage can enhance our management of consumption. This study presents an IoT-based monitoring and forecasting system for electricity consumption. The system comprises two NodeMCU micro-controllers, a PZEM-004T sensor for collecting real-time power data, and three relays that regulate the current flow to three distinct electrical appliances. The data gathered is transmitted to a web application utilizing the k-Nearest Neighbor (k-NN) algorithm to forecast future electricity usage based on historical patterns. We evaluated the system's performance using four weeks of electricity consumption data. The results indicated that predictions were most accurate when the user's daily consumption pattern remained stable, achieving a Mean Absolute Error (MAE) of approximately 1 watt and a Mean Absolute Percentage Error (MAPE) ranging from 1% to 1.7%. Additionally, predictions were notably precise during the early morning hours (3:00 AM to 8:00 AM) when k=6 was employed. This study demonstrates the effectiveness of integrating IoT-based systems with machine learning for real-time energy monitoring and forecasting. Furthermore, it emphasizes the application of data mining techniques within embedded IoT environments, providing valuable insights into the implementation of lightweight machine learning for smart energy systems.

Keywords: Data mining; Internet of Things; k-NN algorithm; Prediction; Sensors.

Received: July, 10th 2025
Revised: August, 4th 2025
Accepted: August, 5th 2025
Published: August, 7th 2025

1. Introduction

Electricity is a significant source of energy. It leads to many new designs and discoveries of electronic and electrical devices that support human activities in their daily life. These include household appliances, office equipment, industrial machines, electric trains, public lighting, and heating systems, among others. Electricity has always become a basic need for human life. With a population of approximately 250 million people, Indonesia ranks fourth in the world. It plays an important role in regional and international markets as a major producer and consumer of electrical energy. Accordingly, the government's plans to expand access to electricity have decreased the number of people without electricity, from approximately 100 million in 2000 to around 23 million in 2016, despite a population increase of almost 25% [1]. As consumers, people often underestimate electricity use in their everyday lives. For example, turning on electrical and electronic devices when not in use. Constantly turning on electronic and electrical devices can increase the risk of higher electricity bills and a short circuit, potentially leading to a fire. Hence, it is necessary to predict electricity consumption so that we can plan to consume it wisely, which can decrease unnecessary risks.

Along with the current and rapid development of technology that leads us to Industry 4.0, there exists a kind of technology with a concept where ordinary objects can send and receive data over a network without direct human assistance. We refer to this technology as the Internet of Things (IoT). This technology aims to expand the advantages of a continuous Internet connection, which can help people access information about objects that may not be readily available [2]. As for capabilities such as remote control and data sharing, these also extend to real-world objects. For example, equipment, living things, collections, electronics, and many more are connected to networks through embedded devices that can sense [3]. IoT development is based on an advanced combination of Internet, microelectromechanical, wireless, and QR Code technologies [4]. However, it encourages people to continue thinking creatively.

In some cases, people only need a computer or a smartphone connected to the Internet to access the information. For example, an application for health and safety using IoT, where wearable sensors can send and receive data to a gateway through a LoRa network. They form a heterogeneous IoT platform with a medical signal-sensing network using Bluetooth. Once the hazard is detected, sensor nodes will warn and notify the end-user [5]. By leveraging IoT to obtain data-driven insights, monitoring activities generate large amounts of data, necessitating the extraction of relevant information. This is where data mining becomes invaluable, as it extracts the essential patterns within the data to yield meaningful insights [6].

Data mining and analysis processes use statistical, mathematical, and artificial intelligence methods. Many practical applications use the combination of IoT technology and data mining methods [7]. For example, the Scikit-Learn-based k-means model can be used for weather clustering and sensor anomaly detection to analyze weather data. The results of the data analysis showed that it is not impossible to gather important information from complex data sets [8]. Furthermore, with the increasing dominance of IoT applications, a significant amount of geospatial information has been accumulated from various sources. To facilitate developers in creating larger spatial data applications, it is essential to develop up-to-date technologies that efficiently handle large amounts of spatial data. A platform for efficiently storing, extracting, processing, and analyzing geospatial big data was proposed in [9]. In [10], [11], data obtained from a sensor device were used to identify the depth of potholes on the road surface using the k-means algorithm. In [12], [13], data collected by sensors at entrance and exit gates were used to calculate the number of vehicles in a parking area. The C4.5 algorithm was implemented in this application to distinguish detected objects.

Several supervised learning methods are widely used in data mining: regression and classification. Regression returns a numerical target for each sample, similar to supervised learning methods, while classification assigns a label from two or more distinct groups or classes to each sample. The k-Nearest Neighbor (k-NN) algorithm is a method that classifies objects based on the nearest training data to those objects. Learning data is described in a multidimensional space, where each dimension correlates with a specific characteristic or feature of the data [14]. The k-NN algorithm can be utilized to forecast a value. To predict a value, it uses the value of the nearest neighbor, which is the result of regression calculated based on the previously determined distance of k [15]–[17]. Predictions of future values can be made from the results of these calculations.

Multiple linear regression has been applied in various prediction techniques, such as rice production [18]. In [19], the long short-term memory to predict household electricity consumption. By implementing IoT technologies and data mining techniques, we aim to effectively address existing challenges related to electricity consumption. The objective of this project is to develop an IoT-based monitoring and forecasting system for electricity consumption. Our contribution includes a monitoring sensor device equipped with three relays for controlling electrical appliances, two NodeMCUs, and a PZEM-004T sensor for capturing power data. The k-NN algorithm is utilized in a web-based application to predict electricity consumption based on historical data. This study contributes to data mining by demonstrating a real-world implementation of k-NN in sensor-driven forecasting. It showcases an effective machine learning deployment within an IoT-based energy monitoring system. Part of this work has been published as a conference paper in [20].

Considering the characteristics of IoT-based systems that necessitate lightweight and rapid-response algorithms, this study evaluated multiple machine learning methods — specifically k-NN, Linear Regression (LR), and Random Forest (RF) — for predicting electricity consumption. k-NN was selected as the primary approach due to its simplicity, effectiveness in managing non-linear relationships, and its suitability for real-time applications with constrained computational resources.

2. Literature Review

Technological developments are growing and inevitable. As a breakthrough, the IoT has become an inseparable part of human life [2]. In IoT, some ordinary objects are embedded with sensor devices to connect, communicate, exchange data, and control via an Internet connection. Devices in IoT are called smart devices as they have the Machine-to-Machine (M2M) communication capabilities [4]. Smart devices are designed to assist people in accomplishing various tasks. Through this technology, objects can be monitored and controlled. There is a microcontroller in every object. It sends data collected by one or more sensors to an application where people can view information regarding the object [21]. People can also interact with the object by sending commands through the application, and an actuator is responsible for carrying out those commands [22]. The IoT requires data processing, data storage, and analysis of results as its three fundamental requirements to transform raw data into actionable information. In its applications, IoT brings a lot of benefits. Besides facilitating people's work, IoT is also helpful in various fields [23], [24], such as the agricultural sector [3], [25]. In [3], the authors developed an IoT application for smart farming to increase yields and provide organic agriculture. The authors proposed the use of remote sensing for agricultural parameters and a control system for greenhouse farming to regulate light, temperature, soil moisture, and CO2 levels. The windows or doors of a greenhouse can be controlled based on soil moisture levels. In addition, the timely, accurate, and practical acquisition, dissemination, and utilization of information may improve the management level of agriculture and support the rapid development of the agricultural economy. This can be achieved by effectively combining intelligent agriculture with IoT technology to develop an intelligent agricultural IoT system based on heterogeneous network technology, which can have a positive impact on agricultural development [25].

IoT technology is currently being applied in the medical and health sector, especially during the recent COVID pandemic. RFID and IoT applications have been developed to intelligently track patient health information, such as weight, height, temperature, and pulse [26]. In IoT, contactless sensors measure or check patients' body parameters, including weight, height, temperature, and SpO_2. RFID tags enable doctors to periodically check a patient's health condition, which is collected through contactless sensors [23]. Furthermore, IoT-based connectivity between vehicles such as ambulances and surrounding hospitals can simultaneously support real-time data collection from patients to multiple hospitals. EMTs can assess for damage and expedite doctors' time to begin treating patients upon arrival [26]. Another medical application is to assist deaf people in communicating by utilizing sensors to detect the intensity of sound when they speak and facilitate speech-to-text transcription, thereby transcribing verbal communication from another person [27], [28]. Moreover, IoT-based systems can also monitor physiological and environmental conditions, improving workplace safety by combining multiple sensors to detect physiological and ecological parameters [21]. When dangerous situations are detected, sensor nodes provide users with adequate warning and notification mechanisms [5], [29].

Constructing an IoT ecosystem requires smart devices and various other supporting elements. Data mining is a method for identifying patterns in extensive data sets to provide valuable insights [30]. Typically, this method is found in machine learning and statistics. Data mining methods were developed due to the increasing complexity of computer work. With data mining, collecting and selecting data is more practical. A data mining method is sometimes adjusted before applying it according to the user's needs. There are various methods in data mining. One of the primary techniques is clustering or segmentation, which involves grouping a class into multiple segments based on specific attributes. The selection of these attributes must align with the similarities shared among different classes. For instance, in paper [10], the authors utilized the k-means algorithm to cluster potholes on the road.

Another data mining method is classification, a process that identifies the definition of characteristic similarity within a group or class. Data mining classification is one of the most commonly used methods. This method estimates the class for an object with an unknown label. In data mining, data can be gathered through surveys, data gathering from sensor networks, crawling data from social media, or scraping documents from websites. Then, the collected data will be converted into a form that can be processed using effective methods or algorithms, producing new information or knowledge [7]. In [31]–[33], a large-scale network traffic security anomaly detection utilized a hybrid intrusion detection model that combined abuse and anomaly detection technologies.

The k-NN is a simple, non-parametric, lazy machine learning algorithm. K-NN is easy to implement for solving both regression and classification problems. It belongs to the supervised learning type. The number of parameters used for non-parametric algorithms, such as k-NN, is flexible and typically increases as more data becomes available [34]. All training data is used in the testing phase to speed up the training process. The k-NN algorithm assumes that similar data points exist in proximity or in terms of neighbors' distances, which means that data with similar characteristics are close to each other. K-NN utilizes available data to classify a new case according to distance functions or similarity measures. Then, the latest case is grouped into a class where most of the neighboring data are located [35]. The study in [16] utilized k-NN to classify a heart disease dataset, predicting heart disease with an accuracy of 69% and a prediction speed of approximately 5,600 observations per second. In paper [36], the authors investigated an active control strategy for detecting anomalies and faults in high-capacity optical networks. Bit error rate data are gathered over 230 km of fiber optic links in the CANARIE network. The k-NN algorithm classifies bit error rates with an accuracy of 97.8%. In [17], the authors estimated stock prices, where a group of stock price information was first synthesized as training data and then used as input to the k-NN model for learning.

In IoT, almost every device or machine can become an intelligent machine. Hence, IoT has greatly influenced all aspects of human life. IoT is pervasive and ubiquitous in society, having a transformative and constructive impact on various application fields. Various studies have proposed similar implementations, investigating IoT-based electricity monitoring systems that utilize machine learning algorithms. For example, methods such as k-NN and C4.5 have been applied for real-time classification and prediction of electricity consumption, demonstrating effectiveness in enhancing energy efficiency and control within smart homes [37], [38]. The increasing number of IoT-connected devices has generated large amounts of data. This brings a perfect overlap between big data creation and IoT [8], [21], [30]. In addition to being huge in size and volume generated by IoT, data must be highly accurate and vary according to its form and quality. Consequently, intelligent data analysis to gain valuable insights is a challenging task. Intelligent data analysis is a crucial prerequisite for maximizing the business value of IoT and realizing its substantial market potential [8]. Numerous studies have explored the intersection of IoT and data mining. For instance, the authors in [39] introduced a dynamic data mining framework designed to process sensor data by developing a sensor data mining model applicable to dynamic change processes. In this framework, each sensor network environment is regarded as a distinct physical system. Historical changes in sensor data are gathered and mined to train the parameters. Relationships between different sensor network environments are identified by analyzing the relationships between various physical system parameters.

In a study [40], the authors explored an air pollution dataset collected by the pollution control board, analyzing the proportionality of air pollutants. They also estimated the influence of several environmental parameters, including humidity, wind speed, and temperature, on different pollutants, such as SO_2, PM10, CO, NO_2, and NO. In addition, each sensor device was meticulously designed and implemented with consideration of the physical phenomena the system aims to analyze. In a study [41], two IoT-based monitoring and recommendation systems, specifically for aquariums and ornamental plants, were developed using several sensors to detect various supporting parameters. These systems utilized web-based applications to provide real-time recommendations to users regarding the condition of the monitored objects, employing the C4.5 classification algorithm. Moreover, in [10], [11], the authors utilized the k-means algorithm for clustering road potholes. A sensor device, installed on a vehicle, collected data, and the Google API was used for real-time monitoring and displaying information.

3. Design and Implementation

In this section, we present an in-depth overview of the system architecture designed for monitoring and forecasting home electricity consumption. The discussion encompasses both hardware components, such as sensors, microcontrollers, and data processing units, as well as software elements, including data processing algorithms, user interface design, and forecasting methods. This comprehensive examination aims to demonstrate how the system is constructed and implemented to achieve accurate, efficient, and user-friendly performance.

3.1. System Design

Fig. 1 depicts our IoT-based monitoring and forecasting system. The proposed system follows an end-to-end data pipeline for efficient electricity monitoring and consumption forecasting. The sensor device utilizes two NodeMCUs. The first NodeMCU connected to the PZEM-004T sensor receives real-time power consumption data from electrical devices. This NodeMCU sends the acquired data to the web application via the HTTP protocol, where the data is saved in a database for further processing. A user can also send commands to the sensor device through the application to turn on and off electrical devices. The second NodeMCU receives commands from the application and forwards them to the relays responsible for disconnecting or connecting electrical current to the electrical devices.

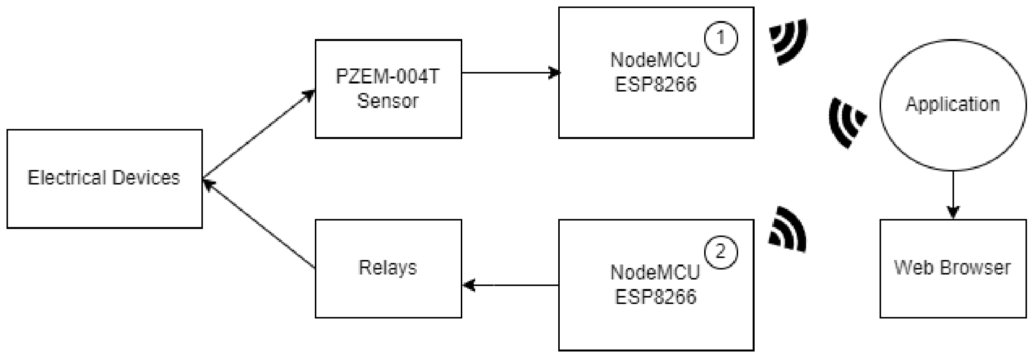

Figure 1. IoT-based monitoring and forecasting system for electricity consumption.

3.2. Hardware Design

Figure 2 illustrates the hardware design of the sensor device. It has two NodeMCUs that are connected to PZEM-004T and relays. The first NodeMCU linked to PZEM-004T is responsible for reading power data. In contrast, the second one connected to relays is responsible for disconnecting or connecting the electric current to electrical devices. In this design, we provide three relays, allowing us to connect to three electrical devices.

The configuration of pins for the first NodeMCU is shown in Table 1. PZEM-004T's RX pin is connected to NodeMCU's D5 pin, and the TX pin is connected to the D6 pin. Moreover, the VCC and GND pins of the two components are connected accordingly. The configuration of pins for the second NodeMCU is shown in Table 2. NodeMCU's D1, D2, and D3 digital pins are connected to the output pins of the three relays. The GND and VCC pins on the NodeMCU are also connected to the GND and VCC pins of each relay.

Table 1. The configuration of pins for the first NodeMCU.

NodeMCU	PZEM-004T
D5	RX
D6	TX
VCC	VCC
GND	GND

Table 2. The configuration of pins for the first NodeMCU.

NodeMCU	Relay 1	Relay 2	Relay 3
D1	Out	-	-
D2	-	Out	-
D3	-	-	Out
VCC		VCC	VCC
GND	VCC	GND	GND

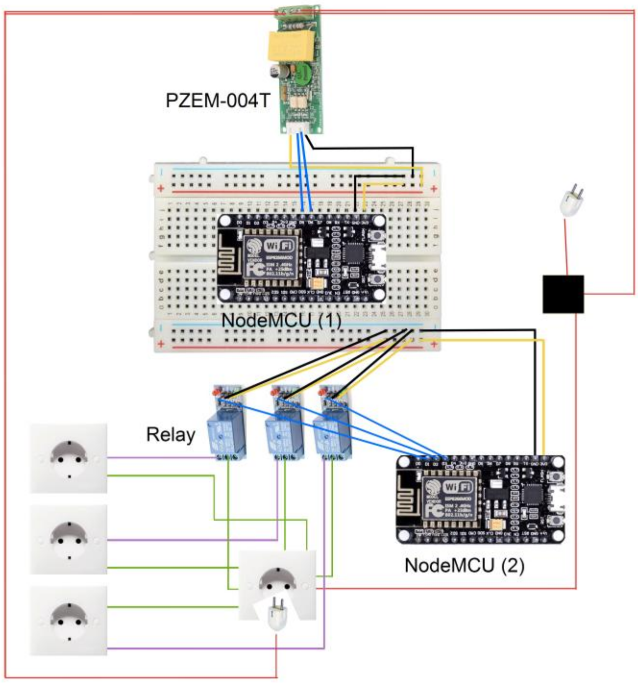

Figure 2. Hardware design illustration.

3.3. Software Design

The software in the IoT-based monitoring and forecasting system consists of two parts. The first is a program that collects data at the sensor device, and the second is a web application that performs algorithm computation. The flowchart, which shows an end-to-end system diagram, is given in Figure 3. Firstly, the program reads power data from the PZEM-004T sensor. The data are transmitted wirelessly via the Internet to the web application, where they are recorded in a database. This activity repeats in a loop until the device is powered off. The algorithm calculation begins after the application gets data from the sensor device. It then loads the dataset and defines the value of k. After that, the Euclidean distances are calculated for all records. The results are sorted in ascending order before the k nearest distances are chosen from the sorted list. Finally, the algorithm classifies the test points according to the number of similarities in the selected k. The classification results are displayed in the application after the data has been classified.

4. Data Processing and Classification Method

This section presents the preprocessing steps and exploratory analysis performed on the electricity consumption dataset before model implementation. The objective is to ensure the quality, relevance, and interpretability of the data used for training and evaluating the k-NN regression model. The analysis includes data cleaning, feature selection, a classification method, and visualization of consumption patterns to guide the evaluation strategy.

4.1. Data Cleaning and Preparation

The dataset consists of 672 hourly electricity consumption records collected over four consecutive weeks. Each record includes a timestamp (*tanggal*), hour (*jam*), power consumption in watts (watt), and day of the week (*id_hari*, *nm_hari*). To enhance the data quality, all records with watt values equal to zero were excluded, as these typically indicated either inactive usage periods or potential sensor anomalies. After cleaning, 452 valid records observations remained for analysis. Categorical encoding was applied to the day of the week, while the hour feature was already represented in numeric format. Given that both features, *jam* (hour of the day) and *id_hari* (day of the week), show similar numeric ranges, normalization was deemed unnecessary. This approach ensured that the distance-based k-NN algorithm could function effectively without any bias due to feature scaling.

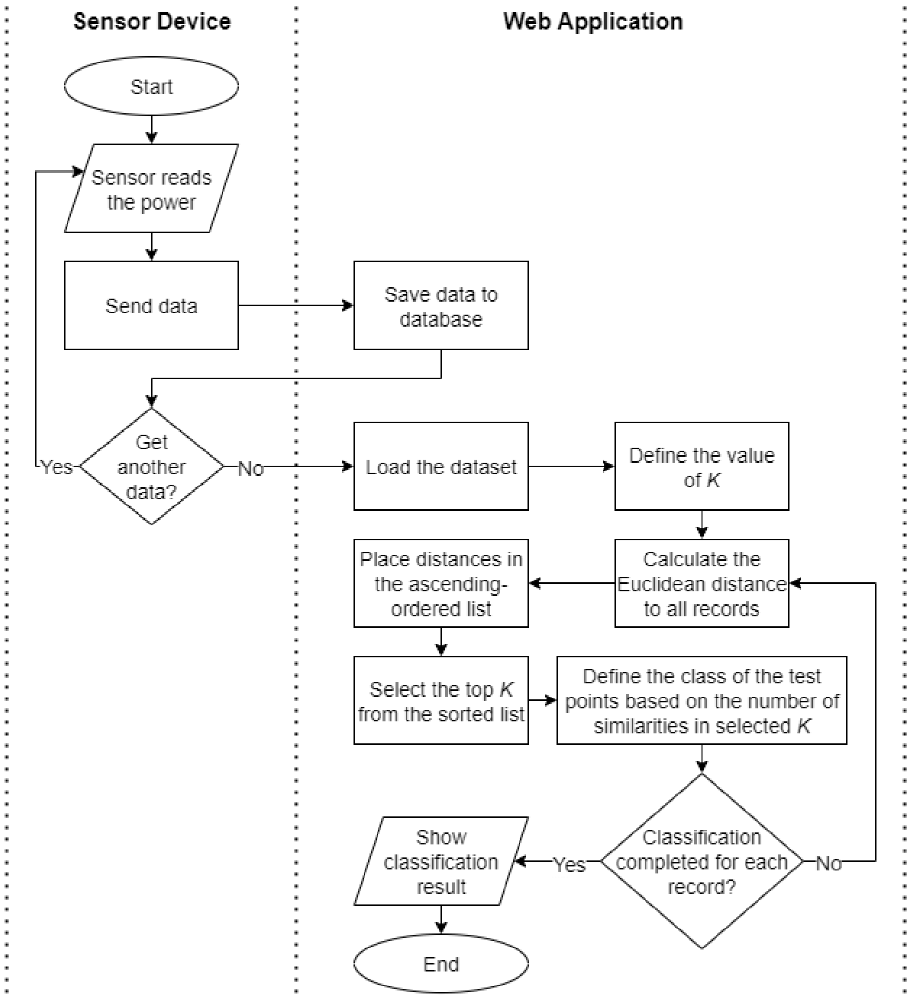

Figure 3. End-to-end system diagram.

4.2. Feature Selection

Two input features were selected for the prediction task: *jam* (hour of the day) and *id_hari* (day of the week). These features effectively capture the temporal patterns of electricity consumption that align with daily and weekly routines. An initial statistical analysis revealed that wattage values ranged from 30.0 to 119.8, with a mean of 38.28 and a standard deviation of 11.55, indicating a moderate level of variability in usage. No extreme outliers were detected beyond the expected operating limits of household appliances. However, differences in variance between specific hours of the day suggest the presence of distinct usage patterns, which informed the definition of time-based evaluation categories.

4.3. Exploratory Data Analysis (EDA)

An exploratory analysis revealed distinct temporal patterns in electricity consumption. As shown in Figure 8, power usage is typically elevated during the early morning hours (12:00 AM – 2:00 AM) and late evening (9:00 PM – 11:00 PM). Meanwhile, the standard deviation is lowest between 3:00 AM and 8:00 AM, indicating a more consistent usage pattern during this period, likely due to fewer user activities. In contrast, from 3:00 PM to 8:00 PM, consumption becomes more erratic, as evidenced by a higher standard deviation, reflecting more variability in user activity. Based on these observations, we defined two time-based evaluation categories: the Stable Consumption Period (3:00 AM – 8:00 AM, characterized by low variance) and the Fluctuating Consumption Period (3:00 PM – 8:00 PM, characterized by high variance). These categories were used to evaluate the predictive performance of the k-NN model under varying levels of temporal variability, thereby evaluating its robustness in both consistent and dynamic usage conditions.

4.4. Classification Method

The k-NN algorithm is a supervised, non-parametric, and instance-based learning method. It is widely applied for both regression and classification tasks due to its simplicity and effectiveness. Unlike parametric algorithms, k-NN does not require a prior assumption about the data distribution. Instead, it determines the output for a new instance based on the similarity to existing labelled instances in the training set. To classify or predict a query instance, the algorithm calculates the distance between the query point and all points in the training dataset. The k closest neighbors are then selected, and the majority class (for classification) or the average value (for regression) among these neighbors determines the predicted outcome. The distance between two data points is commonly measured using the Euclidean distance, which for an n-dimensional feature space is defined as Equation (1).

$$dis(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \tag{1}$$

Where p_i, q_i are the values of the i-th feature for the query instance and the training instance, respectively, and n denotes the number of features. For one-dimensional cases, the distance is calculated using a single feature, while in higher dimensions, multiple features are considered simultaneously. The selection of the parameter k is critical, as it influences the balance between bias and variance. A smaller k tends to capture more local patterns but is sensitive to noise, while a larger k provides smoother decision boundaries but may overlook finer variations.

To evaluate the performance of the k-NN regression model, forecasting accuracy is assessed using error metrics such as Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). MAE measures the average magnitude of errors in the predictions without considering their direction, while MAPE expresses the error as a percentage relative to the actual values. They are defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|A_i - F_i| \tag{2}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|A_i - F_i|}{A_i} \tag{3}$$

Generally, an MAPE value below 10% indicates excellent forecasting accuracy, 10–20% indicates good accuracy, 20–50% indicates reasonable accuracy, and values above 50% indicate poor forecasting performance. These metrics ensure that the predictive model not only minimizes errors but also provides interpretable results for practical applications.

5. Performance Evaluation

In this section, we elaborate on the performance evaluation of the IoT-based monitoring and forecasting system. We provide details about our experimental setup before presenting the results and discussing the findings.

5.1. Experimental Setup

The sensor device in our IoT-based monitoring and forecasting system uses a PZEM-004T sensor to retrieve power data once every hour. The data was transmitted to be saved in the database. In this system, we utilized the Hypertext Transfer Protocol (HTTP) as the data transmission protocol. We collected data on our electricity consumption habits for four consecutive weeks, resulting in 672 data records used as training data. We processed the data using the k-NN algorithm to predict future electricity usage. An end-user could interact with the system to view the prediction results through the web application. After that, to evaluate the performance of the prediction resulting from the k-NN algorithm, we employed data analysis techniques in forecasting, utilizing several test criteria, including the value of k, the split of the training data, and the time category.

5.2. Experimental Results and Discussion

Before evaluating the prediction performance of the system across various time categories, we first compared three models: k-NN, LR, and RF, utilizing multiple train-test split ratios. The evaluation employed MAE, MAPE, RMSE, and R^2 as the performance metrics, as summarized in Table 5. The results show that k-NN consistently outperformed the other two models in terms of both accuracy and generalizability. Based on these findings, we have selected k-NN as the primary algorithm for further analysis.

Table 3. Model performance metrics by split ratio.

Split Ratio	Model	MAE	MAPE	RMSE	R^2
60:40	k-NN	4.709092	0.1136682	8.267177	0.2270535252
	Linear Regression	6.171881	0.1499366	9.660156	0.0004904555
	Random Forest	5.150051	0.1254872	8.359346	0.1607799548
70:30	k-NN	5.048615	0.1307958	7.042575	0.1908075095
	Linear Regression	5.718191	0.1487897	7.502663	0.0011367855
	Random Forest	5.246889	0.1349243	7.464166	0.0695669844
80:20	k-NN	5.186490	0.1083448	11.126017	0.2997381140
	Linear Regression	7.199014	0.1576920	12.754845	0.0762584473
	Random Forest	6.266536	0.1330076	12.345499	0.1156378577

As shown in Figure 4, the k-NN model achieved the lowest MAE across all data split ratios (60:40, 70:30, and 80:20), indicating its superior accuracy in estimating power consumption. Conversely, LR consistently resulted in the highest MAE, while RF showed moderate performance but remained less accurate than k-NN. Moreover, Figure 5 presents a comparison of MAPE, where k-NN attained the smallest percentage errors, particularly in the 80:20 split, demonstrating robustness even with reduced training data. In contrast, LR has the highest MAPE, underscoring its limited ability to model non-linear patterns.

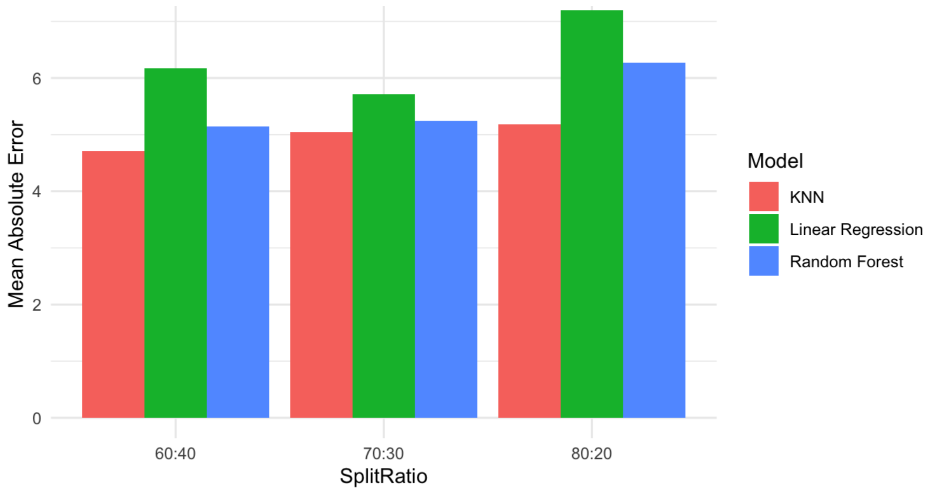

Figure 4. MAE comparison by model and split ratio.

Additionally, the Root Mean Square Error (RMSE) results shown in Figure 6 reinforce the advantage of k-NN. Its RMSE values remained the lowest, especially under the 70:30 and 60:40 split ratios. Particularly, both LR and RF showed an increase in error magnitudes as the test set proportion rose to an 80:20 ratio. Moreover, Figure 7 shows the R^2 scores, which indicate the proportion of variance explained by each model. K-NN achieved the highest R^2 across all split ratios, with a peak of 0.30 under the 80:20 configuration ratio. In comparison, both LR and RF demonstrated significantly lower R^2 values, indicating a weaker model fit. Collectively, these results confirm k-NN as the most reliable model among those tested. Therefore, we selected it for further evaluation under different time categories and conditions of variability in the following analyses.

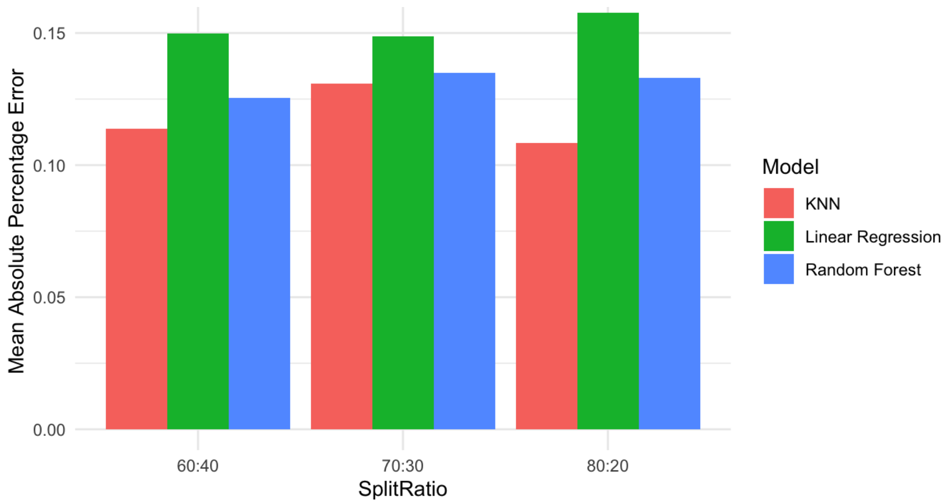

Figure 5. MAPE comparison by model and split ratio

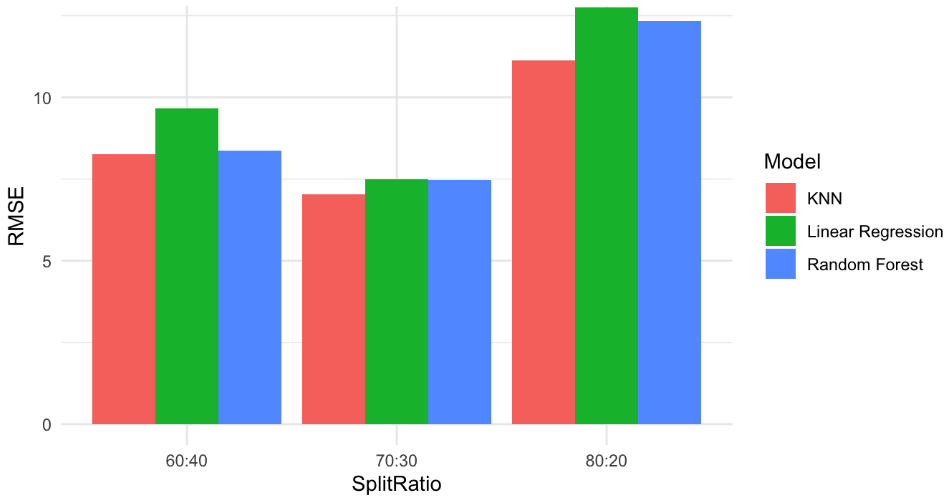

Figure 6. RMSE Comparison by Model and Split Ratio.

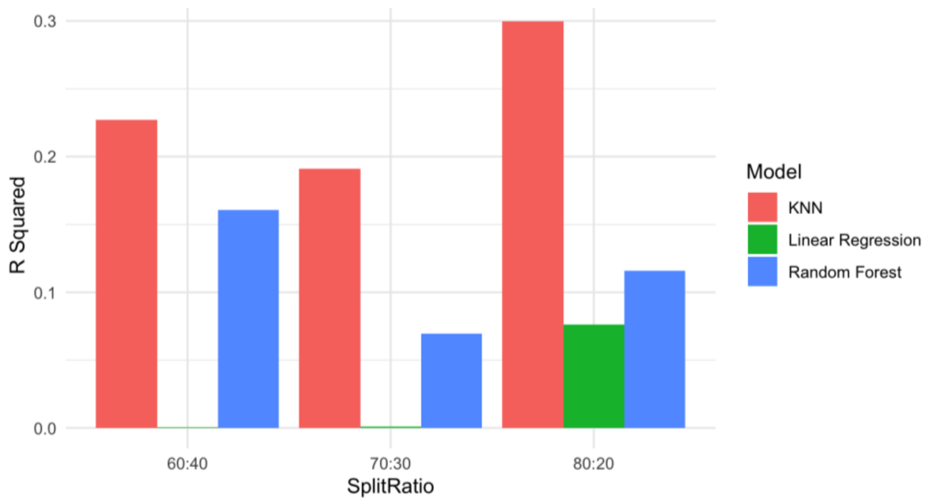

Figure 7. R^2 Comparison by Model and Split Ratio.

The diagram in Figure 8 presents the hourly average power consumption and standard deviation for a four-week period from 12:00 AM to 11:00 PM. The lowest variability is observed between 3 AM and 8 AM, representing a period of low variability, while the highest fluctuations occur between 3 PM and 8 PM, identified as a period of high variability. These time categories are used to evaluate the performance of the prediction system. The high variability in power consumption occurs between 3:00 PM and 8:00 PM, indicated by the elevated standard deviation values, and reflects more dynamic and unpredictable user activity during the afternoon and early evening hours. This variability presents challenges for the k-NN algorithm in identifying consistent patterns or neighbors within the training data. This results in increased prediction error during these periods, as shown by increased MAE and MAPE values. Therefore, understanding these fluctuations is crucial for selecting the appropriate prediction strategy and optimizing model performance across different time categories.

The data set of 672 rows is split into three sets with ratios of 80:20, 70:30, and 60:40 for training and testing purposes. The value of k is varied between 2 and 6. To determine the optimal value of k in the k-NN model, a manual parameter tuning was conducted by testing k values from 2 to 6. The evaluation used MAE and MAPE metrics, tested across three different data splitting ratios (60:40, 70:30, and 80:20), and two distinct time categories with different consumption stability levels. The best k was selected based on the lowest average MAE.

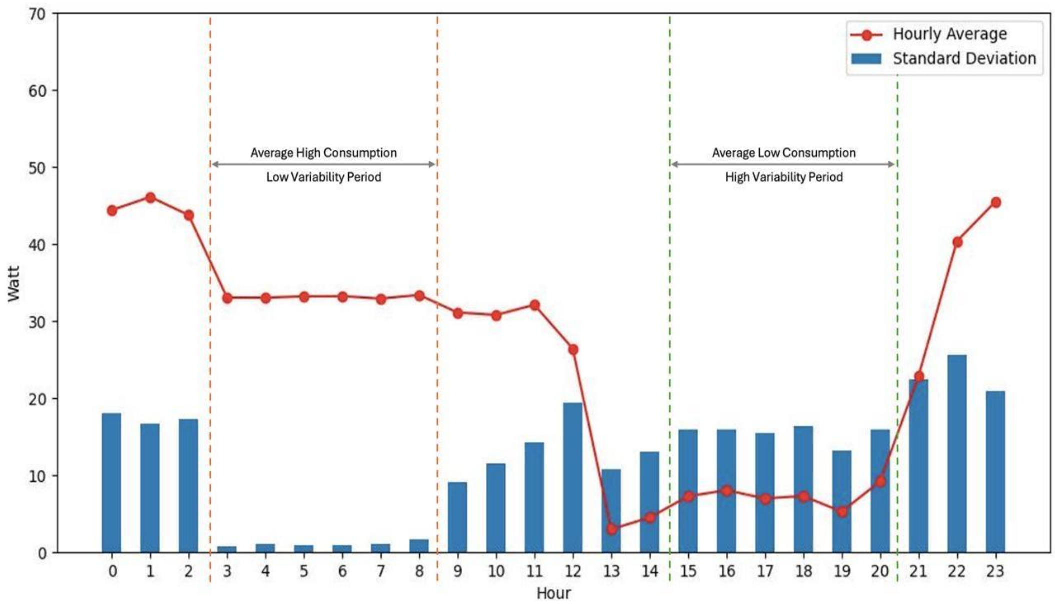

Figure 8. Average power consumption.

The simulation is executed ten times for each test criterion, and the average is taken. The prediction accuracy, as measured by MAE, is displayed in Figure 9, while MAPE is shown in Figure 10. The evaluation results indicate that the MAE of the electricity consumption forecast system is approximately 1 Watt, the MAPE of around 1-1.7%, and k=6 shows the best forecast performance. This result is closely linked to the data characteristics observed during the time window from 3 AM to 8 AM, as shown in Figure 8. During this period, the standard deviation of electricity consumption is significantly lower, indicating more stable and predictable user behavior. The low variability reduces the risk of outliers and increases the similarity among data points, allowing the k-NN algorithm to identify more relevant historical neighbors. As a result, this contributes to a more stable and accurate prediction process, particularly when k = 6, since the neighborhoods consist of highly representative samples.

Furthermore, we perform a comparable evaluation for a different time category, specifically from 3 PM to 8 PM. We have selected this time range because the standard deviations have been consistently high during this period, as shown in Figure 8. The results of the second evaluation are presented in Figure 11 and Figure 12, which show the MAE and MAPE, respectively. As expected, the MAE and MAPE are higher due to the high standard deviations. In this experiment, the MAE is approximately 5-6 watts, the MAPE is around 22%-32%, and k=2 yields the best results. These findings suggest that the predictions are more accurate when a person's daily electricity consumption habits are consistent.

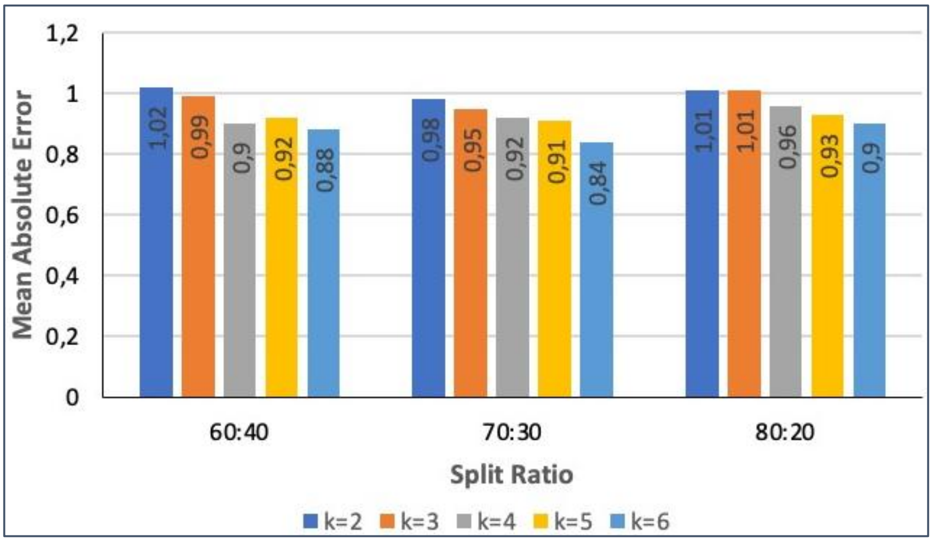

Figure 9. MAE for the 3 AM to 8 AM time category.

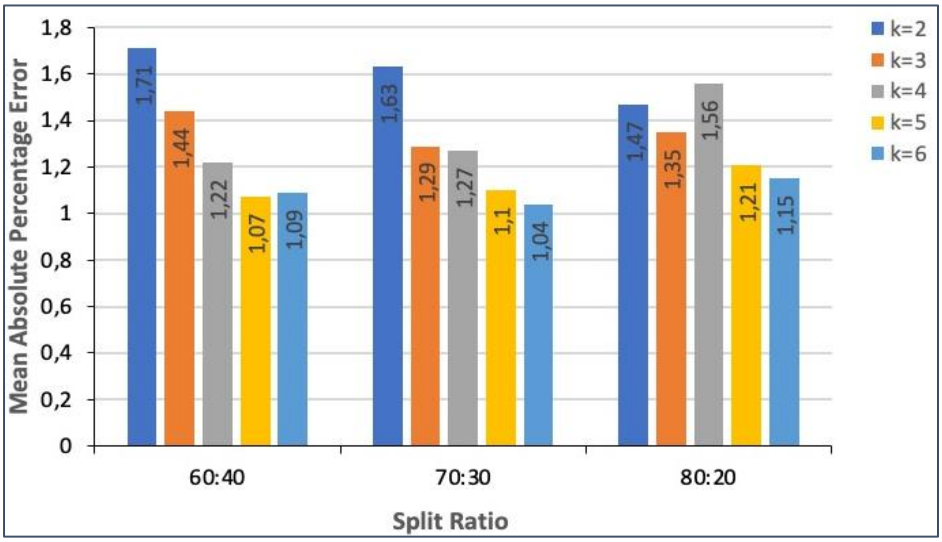

Figure 10. MAPE for the 3 AM to 8 AM time category.

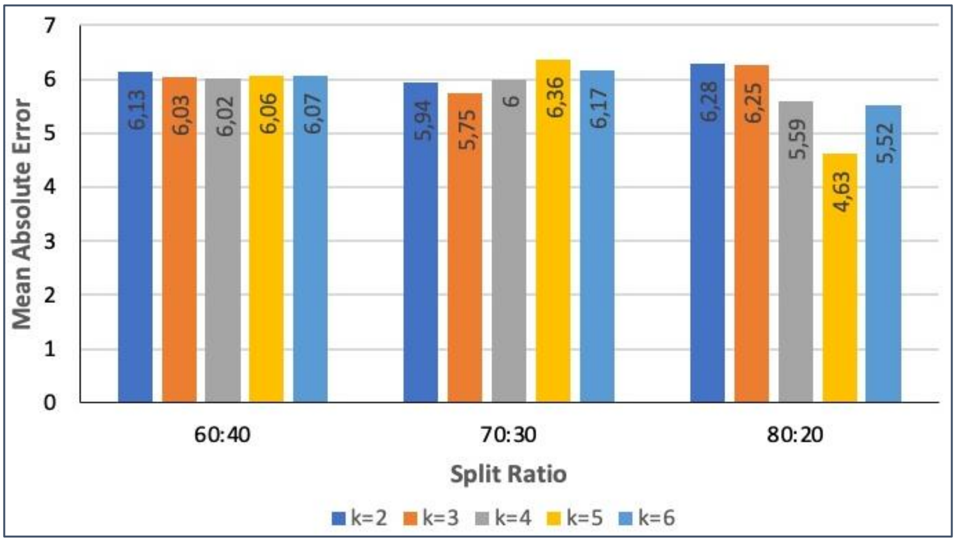

Figure 11. MAE for the 3 PM to 8 PM time category.

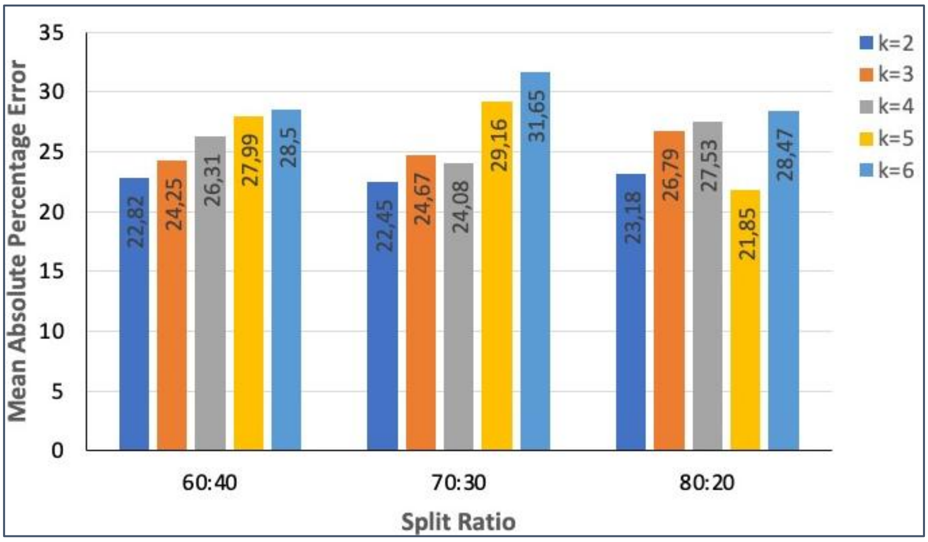

Figure 12. MAPE for the 3 PM to 8 PM time category.

6. Conclusions

This study proposes an IoT-based system for monitoring and forecasting household electricity consumption using machine learning. This system integrates two NodeMCUs, three relays to control the current to three electrical devices, and a PZEM-004T sensor that regularly reads the power data from electrical devices. The collected data is transmitted to a web application, where a k-NN regression model is applied to forecast short-term electricity usage, leveraging daily and weekly time features. This approach demonstrates the viability of integrating lightweight machine learning with embedded IoT technologies. The experimental results show that the system achieves high prediction accuracy and consistent performance, particularly during periods of stable electricity consumption. These findings validate the effectiveness of simple algorithms such as k-NN for real-time forecasting in environments with limited computational resources. It is essential to note that this study focuses on aggregate household electricity consumption, without distinguishing between individual appliances. Although the system includes three relays to control various devices, the model predictions are based on combined usage data rather than device-specific analyses. A limitation of this study is the lack of an ablation analysis to determine the specific contribution of each relay or device. Conducting such an analysis, in which each relay is selectively deactivated to assess its impact on consumption and prediction accuracy, could provide valuable insights. However, this aspect falls outside the scope of the current study, presenting a promising opportunity for future exploration. Additionally, we aim to refine the model by expanding its prediction targets to encompass variables such as kilowatt-hour (kWh) consumption and estimated electricity costs. Furthermore, we will investigate the integration of more advanced time-series forecasting methods to improve long-term accuracy while maintaining responsiveness for real-time applications.

Author Contributions: Conceptualization: A.A. and L.S.; Methodology: A.A., L.S., and J.-Q.O.; Software: Z.M.P.R.; Validation: A.A. and L.S.; Formal Analysis: A.A., L.S., and J.-Q.O.; Investigation: A.A.; Resources: Z.M.P.R.; Data Curation: A.A. and J.-Q.O.; Writing—Original Draft Preparation: A.A., L.S., Z.M.P.R., and J.-Q.O.; Writing—Review and Editing: A.A. and L.S.; Visualization: A.A., L.S., and Z.M.P.R.; Supervision: A.A. and L.S.; Project Administration: A.A. and L.S.; Funding Acquisition: None. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

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

References

[1] Ministry of Energy and Mineral Resources Republic of Indonesia, *Handbook of Energy & Economic Statistics of Indonesia 2021*. 2021. [Online]. Available: https://www.esdm.go.id/assets/media/content/content-handbook-of-energy-and-economic-statistics-of-indonesia-2021.pdf

[2] I. Rafiq, A. Mahmood, S. Razzaq, S. H. M. Jafri, and I. Aziz, "IoT applications and challenges in smart cities and services," *J. Eng.*, vol. 2023, no. 4, Apr. 2023, doi: 10.1049/tje2.12262.

[3] S. Pallavi, J. D. Mallapur, and K. Y. Bendigeri, "Remote sensing and controlling of greenhouse agriculture parameters based on IoT," in *2017 International Conference on Big Data, IoT and Data Science (BID)*, Dec. 2017, pp. 44–48. doi: 10.1109/BID.2017.8336571.

[4] F. Meneghello, M. Calore, D. Zucchetto, M. Polese, and A. Zanella, "IoT: Internet of Threats? A Survey of Practical Security Vulnerabilities in Real IoT Devices," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8182–8201, Oct. 2019, doi: 10.1109/JIOT.2019.2935189.

[5] F. Wu, T. Wu, and M. R. Yuce, "Design and Implementation of a Wearable Sensor Network System for IoT-Connected Safety and Health Applications," in *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, Apr. 2019, pp. 87–90. doi: 10.1109/WF-IoT.2019.8767280.

[6] A. Shi-Nash and D. R. Hardoon, "Data Analytics and Predictive Analytics in the Era of Big Data," in *Internet of Things and Data Analytics Handbook*, Wiley, 2017, pp. 329–345. doi: 10.1002/9781119173601.ch19.

[7] C. C. Aggarwal, *Data Mining*. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-14142-8.

[8] N. Afshan and R. K. Rout, "Machine Learning Techniques for IoT Data Analytics," in *Big Data Analytics for Internet of Things*, Wiley, 2021, pp. 89–113. doi: 10.1002/9781119740780.ch3.

[9] L. Wu, J. Leng, and S. Wu, "Intelligent Perception and Data Mining Technology of City Information Modeling CIM in Urban Renewal," in *2024 3rd International Conference on Data Analytics, Computing and Artificial Intelligence (ICDACAI)*, Oct. 2024, pp. 627–632. doi: 10.1109/ICDACAI65086.2024.00120.

[10] L. Sitanayah, A. Angdresey, and E. Kristalino, "A Real-Time Application for Road Conditions Detection based on the Internet of Things," *JOIV Int. J. Informatics Vis.*, vol. 6, no. 3, p. 629, Sep. 2022, doi: 10.30630/joiv.6.2.891.

[11] L. Sitanayah, A. Angdresey, and E. Kristalino, "A sensor-based application for road conditions detection," in *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, Nov. 2020, pp. 651–652. doi: 10.1145/3384419.3430437.

[12] L. Sitanayah, A. Angdresey, and J. W. Utama, "A Low Cost Vehicle Counting System Based On The Internet of Things," *EPI Int. J. Eng.*, vol. 4, no. 1, pp. 14–20, Sep. 2021, doi: 10.25042/epi-ije.022021.03.

[13] L. Sitanayah, A. Angdresey, and J. W. Utama, "An Internet of Things-based Vehicle Counting System," in *Proceedings of the 2021 International Conference on Embedded Wireless Systems and Networks*, 2021, pp. 175–176. doi: 10.5555/3451271.3451292.

[14] M. A. Araaf, K. Nugroho, and D. R. I. M. Setiadi, "Comprehensive Analysis and Classification of Skin Diseases based on Image Texture Features using K-Nearest Neighbors Algorithm," *J. Comput. Theor. Appl.*, vol. 1, no. 1, pp. 31–40, Sep. 2023, doi: 10.33633/jcta.v1i1.9185.

[15] E. Brilliandy, H. Lucky, A. Hartanto, D. Suhartono, and M. Nurzaki, "Using Regression to Predict Number of Tourism in Indonesia based of Global COVID-19 Cases," in *2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, Sep. 2022, pp. 310–315. doi: 10.1109/AiDAS56890.2022.9918731.

[16] C. C, "Prediction of Heart Disease using Different KNN Classifier," in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, May 2021, pp. 1186–1194. doi: 10.1109/ICICCS51141.2021.9432178.

[17] Q. Yunneng, "A new stock price prediction model based on improved KNN," in *2020 7th International Conference on Information Science and Control Engineering (ICISCE)*, Dec. 2020, pp. 77–80. doi: 10.1109/ICISCE50968.2020.00026.

[18] J. A. Ingio, A. S. Nsang, and A. Iorliam, "Optimizing Rice Production Forecasting Through Integrating Multiple Linear Regression with Recursive Feature Elimination," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 2, pp. 96–108, Aug. 2024, doi: 10.62411/faith.2024-17.

[19] I. W. A. Suranata, I. N. K. Wardana, N. Jawas, and I. K. A. A. Aryanto, "Feature engineering and long short-term memory for energy use of appliances prediction," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 19, no. 3, p. 920, Jun. 2021, doi: 10.12928/telkomnika.v19i3.17882.

[20] A. Angdresey, L. Sitanayah, and Z. M. P. Rumpesak, "An Electricity Consumption Monitoring and Prediction System Based on The Internet of Things," in *2022 Seventh International Conference on Informatics and Computing (ICIC)*, Dec. 2022, pp. 1–6. doi: 10.1109/ICIC56845.2022.10007020.

[21] A. Angdresey and J.-Q. Ooi, "Analyzing Multimodal Sensory Signals Using Unsupervised Machine Learning," in *2023 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, Oct. 2023, pp. 301–306. doi: 10.1109/IC3INA60834.2023.10285741.

[22] M. Weyrich and C. Ebert, "Reference Architectures for the Internet of Things," *IEEE Softw.*, vol. 33, no. 1, pp. 112–116, Jan. 2016, doi: 10.1109/MS.2016.20.

[23] K. V. Sahukara, M. B. Ammisetty, G. S. K. G. Devi, S. Prathyusha, and T. S. Nikhita, "COVID-SAFE: IoT Based Health Monitoring System using RFID in Pandemic Life," in *2021 IEEE International Conference on RFID Technology and Applications (RFID-TA)*, Oct. 2021, pp. 203–206. doi: 10.1109/RFID-TA53372.2021.9617312.

[24] N. Alsbou, D. Price, and I. Ali, "IoT-Based Smart Hospital using Cisco Packet Tracer Analysis," in *2022 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, Jun. 2022, pp. 1–6. doi: 10.1109/IEMTRONICS55184.2022.9795743.

[25] Y. Lu, J. An, and S. Shi, "Research on Smart Agriculture IoT System Based Heterogeneous Networking Technology," in *2021 IEEE 4th International Conference on Information Systems and Computer Aided Education (ICISCAE)*, Sep. 2021, pp. 485–488. doi: 10.1109/ICISCAE52414.2021.9590756.

[26] J. Baiense, E. Zdravevski, P. Coelho, I. M. Pires, and F. J. Velez, "Driving Healthcare Monitoring with IoT and Wearable Devices: A Systematic Review," *ACM Comput. Surv.*, vol. 57, no. 11, pp. 1–38, Nov. 2025, doi: 10.1145/3731595.

[27] A. Angdresey, I. V. Masala, V. D. Kumenap, M. G. Sumampouw, K. A. Dame, and I. D. R. Riady, "A Communication Assistant Application for the Deaf," in *2021 Sixth International Conference on Informatics and Computing (ICIC)*, Nov. 2021, pp. 1–6. doi: 10.1109/ICIC54025.2021.9633011.
[28] K. R. Senthil, J. B, S. B, and N. T, "A Novel System For Blind, Deaf And Dumb People Assistant Using Raspberry Pi," in *2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon)*, Nov. 2022, pp. 1–6. doi: 10.1109/NKCon56289.2022.10126949.
[29] L. Sitanayah, H. R. M. Joseph, and J. B. Sanger, "A Low-Cost Hydroponic Monitoring System with Internet of Things and Fuzzy Logic," *J. Comput. Theor. Appl.*, vol. 2, no. 3, pp. 441–455, Feb. 2025, doi: 10.62411/jcta.12059.
[30] A. C. Onal, O. Berat Sezer, M. Ozbayoglu, and E. Dogdu, "Weather data analysis and sensor fault detection using an extended IoT framework with semantics, big data, and machine learning," in *2017 IEEE International Conference on Big Data (Big Data)*, Dec. 2017, pp. 2037–2046. doi: 10.1109/BigData.2017.8258150.
[31] L. Zhang, J. Zhang, Y. Chen, and S. Lao, "Hybrid Intrusion Detection Based on Data Mining," in *2018 11th International Conference on Intelligent Computation Technology and Automation (ICICTA)*, Sep. 2018, pp. 299–301. doi: 10.1109/ICICTA.2018.00074.
[32] A. Aneja, S. Sharma, P. Thapar, and S. Tiwari, "Optimizing Network Intrusion Detection with Hybrid DTRJ Model: A Data Mining Approach," in *2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT)*, Aug. 2024, pp. 1–5. doi: 10.1109/ICEECT61758.2024.10739005.
[33] J. P. Ntayagabiri, Y. Bentaleb, J. Ndikumagenge, and H. El Makhtoum, "OMIC: A Bagging-Based Ensemble Learning Framework for Large-Scale IoT Intrusion Detection," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 4, pp. 401–416, Feb. 2025, doi: 10.62411/faith.3048-3719-63.
[34] K. Cao, J. Wu, Q. Huang, and Y. Gan, "Optimization Study of KNN Classification Algorithm on Large-Scale Datasets: Real-Time Optimization Strategy Based on Balanced KD Tree and Multi-threaded Parallel Computing," in *2023 4th International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI)*, Aug. 2023, pp. 423–427. doi: 10.1109/ICHCI58871.2023.10277807.
[35] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient kNN Classification With Different Numbers of Nearest Neighbors," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 29, no. 5, pp. 1774–1785, May 2018, doi: 10.1109/TNNLS.2017.2673241.
[36] S. Allogba and C. Tremblay, "K-Nearest Neighbors Classifier for Field Bit Error Rate Data," in *2018 Asia Communications and Photonics Conference (ACP)*, Oct. 2018, pp. 1–3. doi: 10.1109/ACP.2018.8596133.
[37] T. Van Nguyen, "Electricity Generation Cost Optimization Based on Lagrange Function and Local Search," *J. Comput. Theor. Appl.*, vol. 1, no. 2, pp. 104–114, Nov. 2023, doi: 10.33633/jcta.v1i2.9354.
[38] N. K. Mon, "Optimizing Cloud Computing Performance by Integrating the Novel PSBR Service Broker Policy and Load Balancing Algorithms," *J. Comput. Theor. Appl.*, vol. 2, no. 2, pp. 212–221, Oct. 2024, doi: 10.62411/jcta.11221.
[39] Y. Yin, L. Long, and X. Deng, "Dynamic Data Mining of Sensor Data," *IEEE Access*, vol. 8, pp. 41637–41648, 2020, doi: 10.1109/ACCESS.2020.2976699.
[40] M. Shakir and N. Rakesh., "Investigation on Air Pollutant Data Sets using Data Mining Tool," in *2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference on*, Aug. 2018, pp. 480–485. doi: 10.1109/I-SMAC.2018.8653757.
[41] A. Angdresey, L. Sitanayah, T. V. N. Kairupan, and T. M. I. Sumajow, "Implementation of the C4.5 Algorithm in the Internet of Things Applications," 2024, pp. 53–74. doi: 10.1007/978-3-031-47590-0_4.