International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Identifying the Dominant Features in Indonesia Smart Home Dataset by Interpreting Electrical Energy Consumption Prediction Results

Mochammad Haldi Widianto *[1], Alexander Agung Santoso Gunawan [2], Yaya Heryadi [1], Widodo Budiharto[2]

Submitted: 23/12/2023 **Revised**: 27/01/2024 **Accepted**: 08/02/2024

Abstract: Smart Home needs convergence between Machine Learning (ML) and IoT to make predictions, which means ML becomes the optimal prediction model for prediction and Interpretation. Electrical Energy Consumption is a critical problem that needs to be predicted and interpreted. The proposed study aims to find the dominant feature for the Indonesia Smart Home Dataset and prediction using K-Nearest Neighbors (KNN) with Hyperparameters (k and Distance Algorithm). The dominant feature is interpreted using SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME). The experiment's optimal prediction model is validated using error evaluation parameters such as RMSE, MSE, and MAE. The model results (k = 2 and Manhattan Distance) were obtained with RMSE = 0.158, MAE = 0.115, MSE = 0.025, and Manhattan Distance. Although LIME cannot interpret the feature as global, the dominant feature can be displayed globally using SHAP. The global interpretation SHAP result is that the "AC", "Washing Machine", "Lamp", "Water Pump", and "RiceCooker" must be reduced to reduce energy consumption. The KNN learning algorithm can build the model with (k=2 and Manhattan Distance) and SHAP model interpretation. Further research is needed to search for other hyperparameters based on search algorithms to maximize KNN performance.

Keywords: *Dominant Feature, K-Nearest Neighbors (KNN), Local Interpretable Model-Agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP), Smart Home*

1. Introduction.

In Southeast Asian countries, especially Indonesia, energy needs are more used in the electricity sector. The Steam Power Plant generates 66% of the demand for electrical energy. These things are produced by coal, which is a very unfriendly material. Furthermore, the steam power plant can change the world climate. Indonesia has not yet achieved clean energy for generating electrical energy. The highest consumption of electrical energy is in the household sector. Indonesian industry focuses on Small and Medium Enterprises (SME). As a result, the majority of residents work from home, causing households to have the most significant energy consumption [1].

Therefore, humans have to pay more attention to the need for new and renewable power generation technologies. In addition, another thing that can solve this problem is to create a predictive model for electricity consumption, saving electricity consumption. Smart Home can solve the large consumption of electrical energy in the household sector [2]–[4].

The concept of a Smart Home is still not recognized by the Community for its presence [5]. Smart Home can retrieve environmental data and electricity consumption using the Internet of Things (IoT). This device generates Smart Homes data through electrical energy consumption and environmental data [6],[7]. Usually, in Smart Homes, there is Artificial Intelligence (AI), which synergizes with IoT to become the Artificial Internet of Things (AIoT) [8]–[11]. This synergy is found in Smart Home, which is used to help in daily life. The part of AI that is suitable for predicting electricity consumption is Machine Learning (ML) [12], [13].

The applied Smart Home [14] must prioritize user convenience and personalization by utilizing the advantages of AIoT. In addition, the use of Smart Home for predicting electricity consumption in Indonesia is not maximized. However, the electricity consumption prediction is insufficient without interpreting prediction results [15]. Therefore, it is necessary to apply the proposed study of electricity consumption, especially in the Indonesia Smart Home Dataset, which predicts optimally and interprets the expected results [16], [17].

Several previous studies [18]–[20] have focused on getting an optimal predictive model, and there are few studies on interpreting the prediction results. In addition, accurate prediction and interpretation models must consider user convenience and personalization [21]. One possible model is by using K-Nearest Neighbors (KNN) [22]–[25].

Therefore, opportunities for this proposed study are hyperparameters that are rarely explored due to changes and

[1]*Computer Science Department, BINUS Graduate Program - Doctor of Computer Science, Bina Nusantara University, Jakarta, Indonesia, 11480*
[2]*Computer Science Department, School of Computer Science, Bina Nusantara University, Jl. K.H.Syahdan No. 9, Kemanggisan Palmerah, Jakarta, Indonesia, 11480*
** Corresponding Author Email: mochamad.widianto@binus.ac.id*

the Interpretation of the dominant features of the Indonesia Smart Home. The dominant feature interpretation can be found using Local Interpretable Model-Agnostic Explanations (LIME) [16] and SHapley Additive exPlanations (SHAP) [17].

The proposed study aims to take advantage of and find dominant feature from an optimal prediction model using the Indonesia Smart Home Dataset. The Data included Electrical Energy Consumption and environmental data collected from March 2022 to July 2022, which have been analyzed from research [26] and data can be downloaded in [27]. The data is limited in scope only in Indonesia by taking data on Electrical Energy Consumption and environmental data. The proposed study is built by a KNN learning algorithm with optimal search in the form of hyperparameter changes such as K (neighbors) and distance algorithm. Then, the model prediction results are interpreted using LIME and SHAP.

The goal of the proposed study to find dominant feature from the optimal model is explained in several key contributions following:

1. This proposed study presents a detailed research procedure to find the dominant feature of the Indonesia Smart Home Electrical Energy Consumption Dataset.
2. The study proposes a search for an efficient combination of algorithms for tuning the KNN learning algorithm by searching for the hyperparameter k and the distance algorithm.
3. The study validates combined results in a regression model using Mean Absolute Error (MAE)., Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).
4. The study elaborates on the results of the optimal prediction model with SHAP and LIME model interpretation.
5. The proposed study suggests reducing electricity consumption through the interpretation results for optimal prediction model.

The study is divided into the following four sections. Section 1 describes the background, problem description, scope, aim, and contribution. Section 2 describes the literature review. Section 3 describes the research method proposed study to build the optimal prediction model. Section 4 describes the experiment result and discussion. Section 5 describes the Conclusions and suggestions for further research.

2. Literature Review

As explained in the previous section, several possibilities can be explored more profoundly using ML to predict electricity consumption, especially in the Indonesia Smart Home Dataset. This section describes several previous studies to detect electrical energy consumption.

Study Dong *et al.*[18] 2018, The proposed Study for predicting electricity consumption using household appliances that can help improve demand-side planning and management, especially in electrical water heaters (EWH). The study has proved the adequacy and benefits of using ML algorithms. The study conducted a comprehensive evaluation of the ML algorithm. Three ML algorithms are used, namely (i) Random Forest (RF), (ii) Gaussian Naive Bayes (GNB), and (iii) Support Vector Machine (SVM). The study shows that SVM has the best performance, while RF has the worst performance.

Study Shapi *et al.* [19] 2021 The proposed Study aims to answer the problem of predicting electricity consumption in a Building Energy Management System (BEMS) by proposing a predictive model for energy consumption in a machine based on the Microsoft Azure cloud learning platform. The study uses three methodologies, SVM, Artificial Neural Network (ANN), and KNN, proposed for the prediction model algorithm in the case of Malaysia. Study results show that SVM gets the best results.

Study Khan *et al.*[20] 2021, The proposed Study related to machine learning to enhance the accuracy of electrical energy consumption predictions. The study predicts the overall learning curve error and the hybrid model using Jeju Island's energy consumption data in South Korea. One of the study contributions is assembling three machine learning models: XGBoost, Catboost, XGBoost, and Multi-Layer Perceptron.

Study Widianto *et al.*[15] 2023, The proposed study uses secondary datasets and builds prediction comparisons of various kinds of ML, such as Linear Regression (LR), Decision Tree (DT), XGBoost, and RF. This study contributed to the search for improved ML to predict electrical energy consumption and better ML interpretation. The interpretation model used in this study is LIME. This study shows that the XGBoost algorithm produces excellent predictions with a shallow error rate.

Most previous studies used a lot of ML because it takes less time and has low computation compared to Deep Learning (DL) such as ANN, etc. However, Previous research has several limitations and gaps that have not been resolved. Previous research still used a few primary data for Indonesia Smart Home Dataset, especially those using Smart Home as a data collection medium. Furthermore, Smart Home has a concept where comfort and personalization are highly prioritized by its users. KNN, by utilizing training data to solve problems, is considered capable of solving user issues based on the past. Then, previous research has not maximized the performance of the KNN hyperparameter, which is regarded as one of the determining factors for prediction. In addition, interpreting data to find dominant features is rarely solved. Furthermore, the latest research literature only uses LIME. In contrast, the interpretation results can provide suggestions for electricity consumption.

In response to the gaps and limitations of previous studies, this proposed study aims to find dominant feature from optimal prediction models that can be used optimally for Indonesia Smart Home Dataset conditions using the KNN learning algorithm, which maximizes its hyperparameter search and interprets it using LIME and SHAP.

3. Research Method

In this section, to better understand the optimal prediction model proposed by the study, the [system's study flow] was depicted as a diagram in Fig 1. The phases from Fig. 1 is described below:

1. The first phase is the introduction stage. The data is Smart Home consisting of electricity consumption and environmental data.
2. The second phase is data filtering. The data focuses on electrical energy data and environmental data.
3. The third stage is the data treatment stage, which will be divided into two parts, namely training data and testing data. At this stage, the data is divided into 75 days for training data and 1 day for testing data (forecasting).
4. The fourth phase deals with missing or non-numeric data. All data will be converted into numeric data.
5. In this stage, the data will be adjusted to a normal distribution.
6. The sixth phase builds or checks the data to prevent oversampling or undersampling.
7. The first hyperparameter setting is carried out at this stage: the number of neighbors (k) from 1 to 12. This aims to determine how many k matches for building the optimal prediction model.
8. The second hyperparameter is set using several distance algorithms (Manhattan, Minkowski, Euclidean, Cosine, Jaccarrd). The hyperparameter also looks for the optimal prediction model.
9. After obtaining several selected hyperparameters, this section learned all training data using the KNN learning algorithms. The following result is compared with the error.
10. The results in this section evaluate several measurement metrics, such as RMSE, MSE, and MAE.
11. Finally, the proposed study results will be interpreted or explain by how the KNN algorithm learning can work by providing dominant features using the SHAP Model Interpretation.

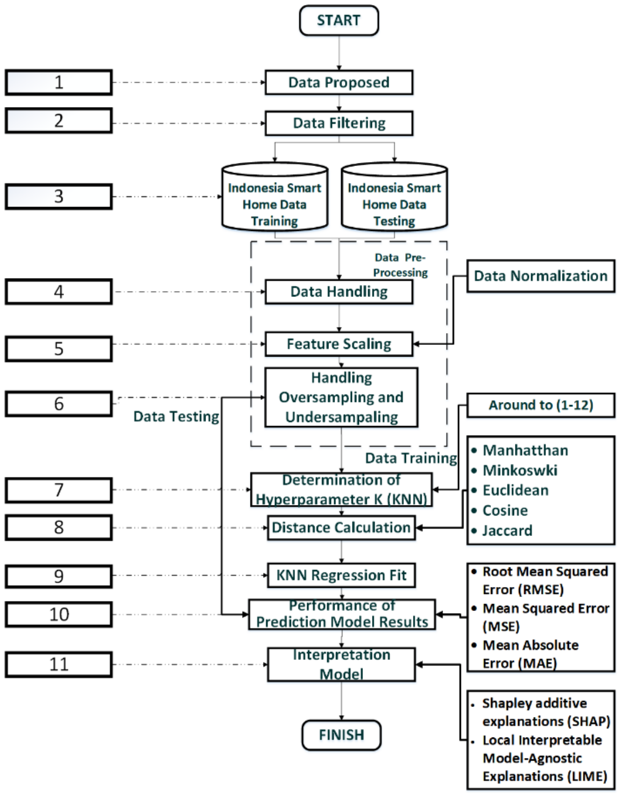

Fig. 1. Proposed Study Flow Chart of Prediction Model

3.1 Smart Home Design

The data taken was inspired by Smart Home research [28] related issues and recommendations for saving electricity by the Indonesian government [29]. Presented in Fig 2 is the Smart Home Design for the Indonesia dataset.

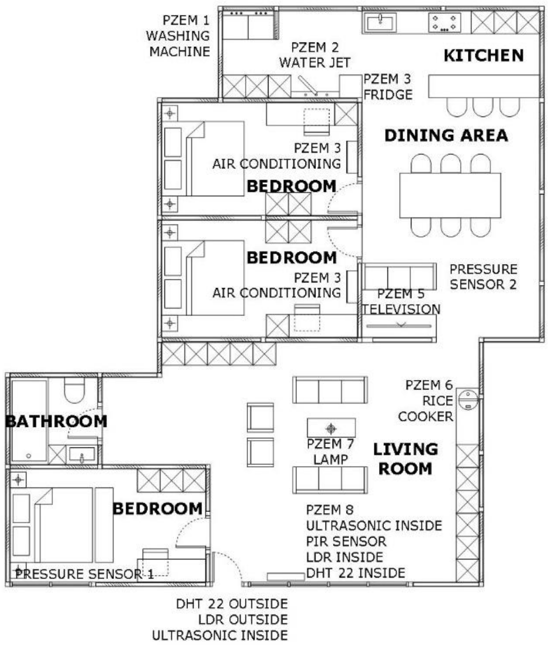

Fig. 2. Smart Home Design in Indonesia Dataset

Fig. 2 represents some of the most significant consumption and environmental influences that are a factor in energy consumption using an IoT sensor device (PZEM-004T). This sensor measures current, power, and energy voltage in cables. This sensor has 31 x 74 mm dimensions and a 33 mm diameter current transformer module.

Environmental condition devices utilize sensors such as the number of people (Ultrasonic HC-SR04). This tool has an operating range of 2cm – 400cm, with an accuracy of 3mm and a measuring angle of 30°.[30], temperature (DHT22), humidity (DHT22) sensor to detect temperature and humidity indoors and outdoors. The advantage that DHT22 has is a wide measurement range, namely 0 to 100% for humidity values and -40 degrees Celsius to 125 degrees Celsius for environmental light intensity.

(LDR Photoresistor) As an electronic component to measure light intensity in the surrounding environment, sofa pressure and bedroom pressure (Load Cell HX711) is a pressure sensor that measures an item's load. The author uses a capacity of 20 kg. All data is sent using a microcontroller (NodeMCU) board designed with ESP8266 inside, which functions as a Wi-Fi network connectivity with the microcontroller. This module is based on the Lua programming language and the Arduino IDE. These data are presented in Table 1. The authors use the dataset for predictions and Interpretation. The data contains 18 features, as shown in Fig 3, consisting of 8 electricity consumption features, and "**Use**" is the target attribute. Then, the other ten features come from the environment. These environmental factors are taken as the influence of electricity consumption. (note: AC (Air Conditioner) and LDR (Light Dependent Resistors).

All features that have been retrieved data are processed to see if there is an error in the data. The data is stored in the cloud by utilizing the MQTT protocol. All data that has been processed is stored in the form of a table data set file. This dataset is the result of primary smart data research [26]. The data is stored in the cloud, which can be accessed at [27]. This data is presented in Table 1.

Table 1. Datasets Information after Pre-Processing

No	Feature	Data Description
1	Use	All data on total electricity consumption on Smart Home
2	Ricecooker	Energy Consumption Data on the Ricecooker
3	Lamp	Energy Consumption Data on the Lamp
4	Television	Energy Consumption Data in Television
5	AC	Energy Consumption Data on AC
6	Refrigerator	Energy Consumption Data at Refrigerator
7	Washing Machine Energy	Energy Consumption Data on Washing Machine
8	Waterpump Energy	Energy Consumption Data on Waterpump
9	Bed Pressure	Environmental Data on Bed Pressure
10	Sofa Pressure	Environmental Data on Sofa Pressure
11	Outdoor Light	Environmental Data on Outdoor Light
12	Indoor Light	Environmental Data on Indoor Light
13	Outdoor Temperature	Environmental Outdoor Temperature Data
14	Indoor Temperature	Environmental Indoor Temperature Data
15	Outdoor Humidity	Environmental Outdoor Humidity data
16	Indoor Humidity	Environmental Indoor Humidity data
17	Indoor Object	Environmental Indoor Object data
18	People Count	Environment People Count data

Table 1 presents information about the dataset, where the **"use"** feature is the target feature. Data is also divided into 2 types, namely electrical energy consumption data and environmental data. These two data greatly influence the prediction of target features**.** (note: AC (Air Conditioner) and LDR (Light Dependent Resistors).

3.2 K-Nearest Neighbors (KNN)

The ML model suitable for Smart Home schemes is KNN, a personalized model because it can take solutions from past cases [21]. This study explored KNN based on distance algorithms (Euclidean, Minkowski, Manhattan, Cosine, Jaccard) using the equation (1) – (5).

$$\text{Minkowski: } d(x,y) = \left(\sum_{i=1}^{n}|x_i - y_i|^p\right)^{\frac{1}{p}} \tag{1}$$

$$\text{Euclidean: } d(x,y) = \left||x - y|\right| \tag{2}$$

$$\text{Manhattan: } d(x,y) = \sum_{i=1}^{n}|x_i - y_i| \tag{3}$$

$$\text{Cosine: } d(x,y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}} \tag{4}$$

$$\text{Jaccard} : d(x,y) \quad = J(x,y) = \left|\frac{x \cap y}{x \cup y}\right| \qquad (5)$$

Where:
i = 1 to n
p = positive integer

Equations (1)-(5), are used to find the best Hyperparameter performance. The number of neighbors (k) and the KNN distance algorithm are standard components researchers commonly use. Understandably, the best Hyperparameters result in a more optimal ML model.

3.3 Interpretation Model

LIME and SHAP interpret the KNN prediction model and look for dominant features. These interpretation models have the mathematical equations presented in equations (6) and (7) [16], [17]:

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} \mathcal{L}(f, g, \pi_x) + \mho(g) \qquad (6)$$

Where:

$\xi(x)$ = Interpretation of results based on data x

G = Interpretable model family

f = ML Complex Models

g = simple model of Interpretation

π_x = local neighbourhoods

$\mathcal{L}(f, g, \pi_x)$ = Base estimates on local neighbourhoods

$\mho(g)$ = manages the complexity of the simple replacement model.

$$\phi_i(f, x) = \sum_{z' \subseteq x'} \frac{|z'|!(F - |z'| - 1)!}{F!} \left[f_x(z') - f_x\left(\frac{z'}{i}\right)\right] \qquad (7)$$

Where:
ϕ_i = Shapley's score for feature i
f =Model BlackBox
x = Input Data
$z' \subseteq x$ = All input feature data
x' = Sample Input data
F = set of all features

LIME is a local interpretation of the model, where the model focuses on each feature compared to the selected prediction results. Then, the feature is made by local neighbors to see how many neighbors there are. The more features shown, the more dominant and vice versa [16], [17], [31]. On the other hand, SHAP is inspired by game theory, where a character gets playing exp according to his contribution in a battle. This is possible if SHAP works globally, in which case this model requires evaluating each feature to get a reasonable contribution result.

3.3 Error Evaluation

The experiment proposed study will be validated by error evaluation such as RMSE, MSE, and MAE with equations in (6)-(8) [32]:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}| \qquad (6)$$

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y})^2 \qquad (7)$$

$$\text{RMSE} = \sqrt{\text{MSE}} \qquad (8)$$

Where:

$\hat{y}$ = predicted value of y

$\bar{y}$ = mean value of y

In equations (6)-(8), this is used to evaluate the proposed research prediction model. The evaluation results are in the form of error values which produce different values for each error unit used. The smaller the error produced, the better the propose study.

4. Results and Discussion

In this section, the results of the research will be explained. The results of the study are divided into two parts. The first is the results of predictions on KNN with effects (k and the distance algorithm). Second, is the search result for the dominant features of the KNN prediction model using SHAP and LIME. The results of the two experiments focus on saving electricity consumption in Smart Homes

4.1. KNN Experiment Error Parameter Result

This section will discuss the prediction results evaluated with RMSE, MSE, and MAE. The results of this evaluation are used to see the best performance for predicting electricity consumption in the Indonesia Smart Home dataset. These results are shown in Table 2-4

Tabel 2. Result in KNN Based on RMSE

	K:1	K:2	K:4	K:6	K:8	K:10	K:11	K:12
Manhattan	**0.221**	**0.158**	**0.178**	**0.219**	**0.225**	**0.223**	**0.225**	**0.228**
Minkowski	0.273	0.252	0.291	0.301	0.306	0.311	0.311	0.314
Euclidean	0.268	0.246	0.286	0.298	0.301	0.306	0.306	0.313
Cosine	0.270	0.248	0.288	0.303	0.306	0.309	0.309	0.311
Jaccard	0.531	0.501	0.536	0.523	0.512	0.504	0.517	0.511

Tabel 3. Result in KNN Based on MAE

	K:1	K:2	K:4	K:6	K:8	K:10	K:11	K:12
Manhattan	**0.135**	**0.115**	**0.15**	**0.165**	**0.168**	**0.167**	**0.168**	**0.168**
Minkowski	0.158	0.167	0.212	0.208	0.219	0.24	0.241	0.251
Euclidean	0.152	0.162	0.207	0.201	0.213	0.233	0.238	0.248

Cosine	0.157	0.173	0.212	0.232	0.231	0.241	0.24	0.252
Jaccard	0.471	0.461	0.415	0.423	0.403	0.401	0.403	0.402

Tabel 4. Result in KNN Based on MSE

	K:1	K:2	K:4	K:6	K:8	K:10	K:11	K:12
Manhattan	**0.049**	**0.025**	**0.032**	**0.045**	**0.051**	**0.05**	**0.051**	**0.052**
Minkowski	0.075	0.064	0.085	0.091	0.094	0.097	0.097	0.099
Euclidean	0.072	0.061	0.082	0.089	0.091	0.094	0.094	0.098
Cosine	0.073	0.062	0.083	0.092	0.094	0.096	0.096	0.097
Jaccard	0.283	0.251	0.288	0.274	0.263	0.255	0.268	0.262

Tables 2-4 confirmed that Manhattan is the most efficient hyperparameter for the KNN learning algorithm. It is shown that the KNN learning algorithm's error results are low when using Manhattan, as measured by RMSE, MSE, and MAE. However, for each error, the parameter has a slight difference. At RMSE, MAE, and MSE, the value of k = 2 has the best results (0.158, 0.115 and 0.025).

Manhattan has a very minimum error value due to the nature of Indonesia Smart Home data, which has a multi-feature 18 in Fig.3. This makes it multi-dimensional in finding the closest training data to the testing data. This is a weakness for other distance algorithms, which are very susceptible to multi-dimensionality. Mathematically, the distance algorithm's formula in Manhattan is powerful in dealing with multi-dimensional cases [31].

Every distance algorithm (Manhattan, Minkowski, Euclidean, Cosine, and Chebshev) has the L_k norm, which is the matrix formula used to measure the distance between two points in a multi-dimensional space, as explained, it is this factor that systematically causes weaknesses in various kinds of errors when predicting [31].

4.2. The Result Interpretation Model

This section explains the KNN prediction results interpreted using SHAP and LIME to find the most dominant features in Indonesia Smart Home Datasets. The unique thing that will be different from the existing Indonesia Smart Home is that the culture and customs of each country are different. The LIME model interpretation results are presented in Fig. 3.

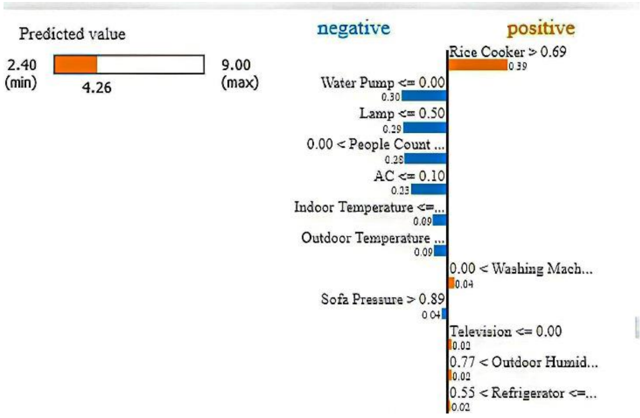

Fig. 3. LIME Result for Prediction Electrical Energy Consumption

LIME can only be used for certain prediction results. In Fig 3, the author uses LIME to find interpretation results on the prediction results of the KNN algorithm learning. The dominant features described by LIME are obtained because it mathematically uses local neighbors to see how many of those features are present when the author selects a prediction. If there are a lot of them, then the feature is dominant, and the most dominant feature, according to LIME, is "Ricecooker". Fig. 3 results have not been able to provide suggestions for reducing electricity consumption. Therefore, the experiment was continued using the SHAP presented in Fig 4 and 5.

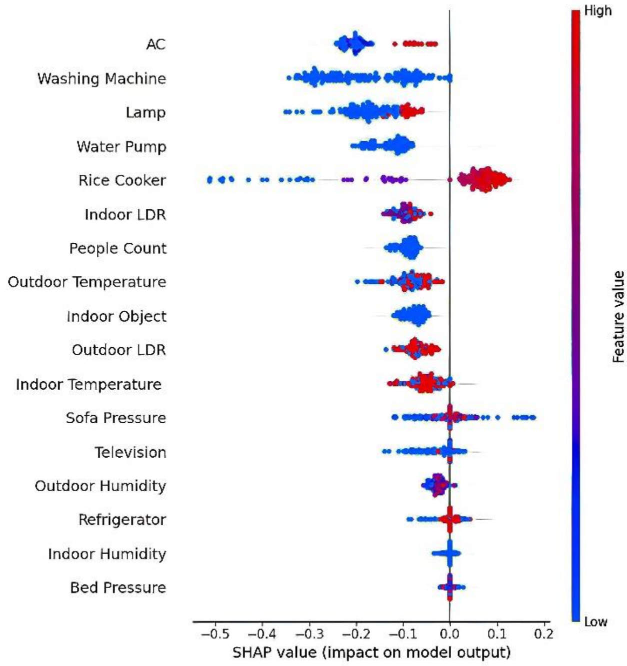

Fig 4. (BeeSwarm Plot) Feature Contribution to Prediction Globally

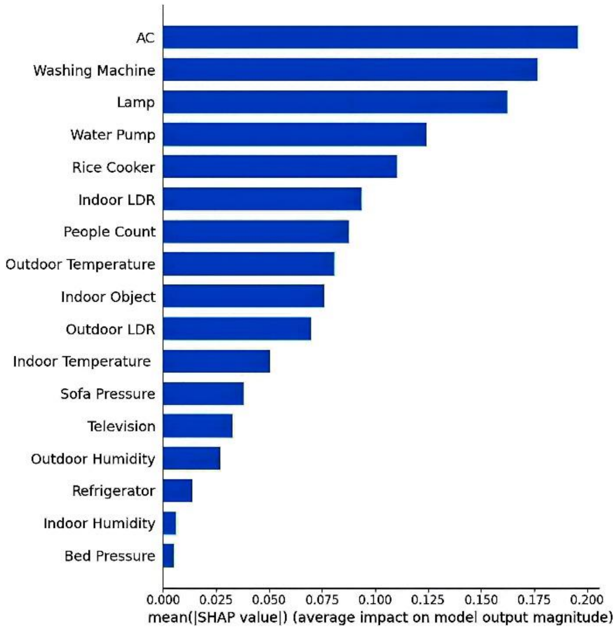

Fig 5. (Bar Plot / Absolute Value) Dominant Feature Contribution to Prediction Globally

The most prominent difference between SHAP and LIME. Interpretation model SHAP can be interpreted globally or comprehensively, which is not chosen by the predictors who want to interpret it. SHAP is global. The Interpretation results provide overall prediction results from training data and testing data. Fig 4, commonly called as a beeswarm, shows the herd of features that are very dominant in Indonesia Smart Home, namely "Rice Cooker, and AC," which have flocks that affect a high SHAP Value (a high SHAP Value represents a high predictive result and vice versa). Therefore, these features have a global influence on the high prediction of the KNN learning algorithm. Furthermore, this feature also has a flock when the prediction of the KNN learning algorithm is low. Therefore, in the context of these features, they must be saved or reduced.

Fig 5 shows how many features contribute to the prediction based on the SHAP value (not just the group or pool but the whole). Based on the results of Fig 5, "AC", "Washing Machine", "Lamp" , and “Water Pump” significantly affect the highest prediction results. The same results are described in Fig 4 and 5. Therefore, based on these results, the user can reduce electricity consumption at home by reducing the consumption of "AC", "Washing Machine", "Lamp", “Water Pump,” and “RiceCooker”. Then, the optimal model proposed is more likely to use SHAP because the results are interpreted globally.

4.3. Discussion

In this section, the author will discuss the results of the previous sub-sections, Namely, how the design of Smart Homes in each country is different. One of them is in Indonesia, which has a tropical climate. The plan was made based on energy-saving regulations from the central government and research on related smart home designs. However, the design of this Smart Home still has weaknesses, one of which is that the tools used are low-cost, and the accuracy of each tool has yet to be measured.

The success of the proposed study in finding the dominant feature from optimal prediction model is described:

1. The proposed study succeeded in presenting research procedures in finding the dominant feature in the optimal prediction model for the Indonesia Smart Home Electrical Energy Consumption Dataset.
2. The Proposed Study to find the optimal prediction model has succeeded in providing most efficient predictions for the KNN learning algorithm with the number of k=2 and Manhattan.
3. The proposed study obtained a small error value based on the error parameters resulting from RMSE, MSE, and MAE.
4. The Proposed Study succeeded in presenting the results of the LIME and SHAP interpretation models.
5. The proposed study succeeded in displaying features that need to be reduced to reduce electricity consumption.

Furthermore, the authors also present the optimal possibility of the model being applied to other cases. The culture or habits of people in Indonesia may differ from other countries, most similar to Southeast Asia. Especially those with many island patterns like Malaysia, the Philippines, and Vietnam. However, the optimal prediction model must still be tested experimentally by calculating the prediction results based on the error parameter. The author hopes to test the optimal model if there is a primary dataset in these countries.

5. Conclusion

This study proposes an Indonesia Smart Home Dataset, which is predicted using the KNN learning algorithm. In making predictions, model performance is based on a comparison of Hyperparameters (k and distance algorithm). Then, the results are interpreted using LIME and SHAP to understand the KNN prediction results and the basis for electricity savings based on the discovery of dominant features. The experimental results of the proposed study for the optimal prediction model are validated using prediction error parameters such as RMSE, MSE, and MAE. The prediction results were evaluated with the best hyperparameters obtained with k=2 and Manhattan Distance, resulting in experiments with error parameters such as RMSE (0.158), MAE (0.115), MSE (0.025), and Manhattan Distance. The proposed study interpretation used by LIME and SHAP is the result of searching for the best dominant features. Furthermore, interpretation results such as LIME only interpret specifically, not globally. However, a global interpretation of the Dominant Features can be presented with SHAP which shows that the dominant features from electrical energy consumption are found in "AC", "Washing Machine", "Lamp", "Water Pump" and “RiceCooker. Therefore, this feature needs to be emphasized to reduce electricity consumption. However, the future scope to obtain maximum prediction results is only limited to numerical data, and the data must come from IoT components. The data is very extensive and has more features, so optimization is needed to get the optimal prediction model. Therefore, further studies will be carried out in search of hyperparameter optimization such as Random Search, Bayesian Optimization, Grid Search, Particle Swarm Optimization (PSO), and Genetic Algorithm (GA) to get better results for Indonesia Smart Home Dataset.

References

[1] M. Yusuf, B. Surya, F. Menne, M. Ruslan, S. Suriani, and I. Iskandar, “Business Agility and Competitive Advantage of SMEs in Makassar City, Indonesia,” Sustainability, vol. 15, no. 1, Jan. 2023, doi: 10.3390/su15010627.

[2] Md. Shahjalal, Moh. K. Hasan, Md. M. Islam, Md. M. Alam, Md. F. Ahmed, and Y. M. Jang, “An Overview of AI-Enabled Remote Smart- Home Monitoring System Using LoRa,” in 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Feb. 2020, pp. 510–513. doi: 10.1109/ICAIIC48513.2020.9065199.

[3] W. Yan, Z. Wang, H. Wang, W. Wang, J. Li, and X. Gui, “Survey on recent smart gateways for smart home: Systems, technologies, and challenges,” Transactions on Emerging Telecommunications Technologies, p. e4067, 2020, doi: 10.1002/ett.4067.

[4] L. Mehdi, Y. Ouallou, O. Mohamed, and A. Hayar, “New Smart Home’s Energy Management System Design and Implementation for Frugal Smart Cities,” in 2018 International Conference on Selected Topics in Mobile and Wireless Networking (MoWNeT), Jun. 2018, pp. 149–153. doi: 10.1109/MoWNet.2018.8428865.

[5] Y. Chawla, A. Kowalska-Pyzalska, and W. Widayat, “Consumer willingness and acceptance of smart meters in Indonesia,” Resources, vol. 8, no. 4, Dec. 2019, doi: 10.3390/RESOURCES8040177.

[6] Z. N. Mohammad, F. Farha, A. O. M. Abuassba, S. Yang, and F. Zhou, “Access Control and Authorization in Smart Homes: A Survey,” in Tsinghua Science and Technology, Dec. 2021, pp. 906–917. doi: 10.26599/TST.2021.9010001.

[7] N. A. Prasetyo, A. G. Prabawati, and S. Suyoto, “Smart Home: Power Electric Monitoring and Control in Indonesia,” International Journal of Interactive Mobile Technologies (iJIM), vol. 13, no. 03, pp. 143–151, Mar. 2019, doi: 10.3991/ijim.v13i03.10070.

[8] H. Zhu, P. Tiwari, A. Ghoneim, and M. S. Hossain, “A Collaborative AI-Enabled Pretrained Language Model for AIoT Domain Question Answering,” IEEE Trans Industr Inform, vol. 18, no. 5, pp. 3387–3396, 2022, doi: 10.1109/TII.2021.3097183.

[9] A. K. Salama and M. M. Abdellatif, “AIoT-based Smart Home Energy Management System,” in 2022 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT), 2022, pp. 177–181. doi: 10.1109/GCAIoT57150.2022.10019091.

[10] Z. Zhang, T. He, M. Zhu, Q. Shi, and C. Lee, “Smart Triboelectric Socks for Enabling Artificial Intelligence of Things (AIoT) Based Smart Home and Healthcare,” in 2020 IEEE 33rd International Conference on Micro Electro Mechanical Systems (MEMS), 2020, pp. 80–83. doi: 10.1109/MEMS46641.2020.9056149.

[11] B.-W. Chen, Y.-S. Jhang, H.-T. Pai, S.-H. Wang, M.-H. Sheu, and T.-H. Chen, “AIoT-based Audio Recognition System for Smart Home Applications,” in 2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), 2021, pp. 1–2. doi: 10.1109/ICCE-TW52618.2021.9603103.

[12] F. Tang, Y. Kawamoto, N. Kato, and J. Liu, “Future Intelligent and Secure Vehicular Network Toward 6G: Machine-Learning Approaches,” Proceedings of the IEEE, vol. 108, no. 2, pp. 292–307, 2020, doi: 10.1109/JPROC.2019.2954595.

[13] I. Priyadarshini, S. Sahu, R. Kumar, and D. Taniar, “A machine-learning ensemble model for predicting energy consumption in smart homes,” Internet of Things, vol. 20, p. 100636, 2022, doi: 10.1016/j.iot.2022.100636.

[14] M. H. Widianto, A. A. S. Gunawan, Y. Heryadi, and W. Budiharto, “INTERPRETATION OF THE DOMINANT FEATURES IN THE PREDICTION RESULTS OF ELECTRICAL ENERGY CONSUMPTION USING SMART HOME DATASETS BASED ON KNN MACHINE LEARNING,” ICIC Express Letters, Part B: Applications, vol. 15, no. 4, 2024.

[15] M. H. Widianto, A. A. S. Gunawan, Y. Heryadi, and W. Budiharto, “Evaluation of Machine Learning on Smart Home Data for Prediction of Electrical Energy Consumption,” in 2023 International Conference on

Computer Science, Information Technology and Engineering (ICCoSITE), Feb. 2023, pp. 434–439. doi: 10.1109/ICCoSITE57641.2023.10127700.

[16] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, in KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, pp. 1135–1144. doi: 10.1145/2939672.2939778.

[17] S. M. Lundberg, P. G. Allen, and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017.

[18] J. Dong, J. Munk, B. Cui, P. R. Boudreaus, and T. Kuruganti, "Machine-learning model of electric water heater for electricity consumption prediction," in International High Performance Buildings Conference, 2018.

[19] M. K. M. Shapi, N. A. Ramli, and L. J. Awalin, "Energy consumption prediction by using machine learning for smart building: Case study in Malaysia," Developments in the Built Environment, vol. 5, p. 100037, 2021, doi: https://doi.org/10.1016/j.dibe.2020.100037.

[20] P. W. Khan, Y. Kim, Y. C. Byun, and S. J. Lee, "Influencing factors evaluation of machine learning-based energy consumption prediction," Energies (Basel), vol. 14, no. 21, Nov. 2021, doi: 10.3390/en14217167.

[21] B. López, "Case-based reasoning: a concise introduction," Synthesis lectures on artificial intelligence and machine learning, vol. 7, no. 1, pp. 1–103, 2013, doi: 10.2200/S00490ED1V01Y201303AIM020.

[22] C. Ma and Y. Chi, "KNN Normalized Optimization and Platform Tuning Based on Hadoop," IEEE Access, vol. 10, pp. 81406–81433, 2022, doi: 10.1109/ACCESS.2022.3195872.

[23] H. Younes, A. Ibrahim, M. Rizk, and M. Valle, "An Efficient Selection-Based kNN Architecture for Smart Embedded Hardware Accelerators," IEEE Open Journal of Circuits and Systems, vol. 2, pp. 534–545, 2021, doi: 10.1109/OJCAS.2021.3108835.

[24] W. Xing and Y. Bei, "Medical Health Big Data Classification Based on KNN Classification Algorithm," IEEE Access, vol. 8, pp. 28808–28819, 2020, doi: 10.1109/ACCESS.2019.2955754.

[25] S. Zhang, "Challenges in KNN Classification," IEEE Trans Knowl Data Eng, vol. 34, no. 10, pp. 4663–4675, 2022, doi: 10.1109/TKDE.2021.3049250.

[26] W. Hartanto, M. A. Hibatul, R. Y. Afif, and M. H. Widianto, "DEVELOPMENT OF IOT-BASED HOME ELECTRICITY USAGE MONITORING SYSTEM," Skripsi, Bina Nusantara University, Bandung, 2023.

[27] M. H. Widianto, "Smart Home Indonesia Dataset," May 2022. Accessed: May 26, 2023. [Online]. Available: https://docs.google.com/spreadsheets/d/1oteCvOo_d0UNAqipET_T6Lb9chUHS008/edit?usp=share_link&ouid=112500639100101980774&rtpof=true&sd=true

[28] G. Chimamiwa, M. Alirezaie, F. Pecora, and A. Loutfi, "Multi-sensor dataset of human activities in a smart home environment," Data Brief, vol. 34, Feb. 2021, doi: 10.1016/j.dib.2020.106632.

[29] ESDM, "Direktorat Jenderal EBTKE - Kementerian ESDM." Accessed: Jul. 16, 2022. [Online]. Available: https://ebtke.esdm.go.id/post/2020/05/11/2535/berbagai.kiat.hemat.energi.praktis.era.work.from.home

[30] G. Montazeaud et al., "Development of a low cost open-source ultrasonic device for plant height measurements," Smart Agricultural Technology, vol. 1, p. 100022, Dec. 2021, doi: 10.1016/J.ATECH.2021.100022.

[31] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the Surprising Behavior of Distance Metrics in High Dimensional Space," 2001.

[32] W. K. Adu, P. Appiahene, and S. Afrifa, "VAR, ARIMAX and ARIMA models for nowcasting unemployment rate in Ghana using Google trends," Journal of Electrical Systems and Information Technology, vol. 10, no. 1, p. 12, 2023, doi: 10.1186/s43067-023-00078-1.