Evolution of Smart Home Energy Management System Using Internet of Things and Machine Learning Algorithms

Tarana Singh[1], Arun Solanki[2] and Sanjay Kumar Sharma[3]
[1]*Gautam Buddha University, Department of Computer Science and Engineering, School of Information and Communication Technology, Greater Noida, India.*
[2]*Gautam Buddha University, Department of Computer Science and Engineering, School of Information and Communication Technology, Greater Noida, India.*
[3]*Gautam Buddha University, Department of Computer Science and Engineering, School of Information and Communication Technology, Greater Noida, India.*

Abstract

The cost and demand of energy are increasing day by day, leading the domain to find new and smart ways to monitor, control, and save energy. In the smart city, smart energy management systems help to resolve the energy management problem. Smart energy management systems cut the cost of energy in smart houses or buildings with their recommendations and predictions. This paper proposed a 5-layer architecture for a Home Energy Management System (HEMS), which collects real-time data; analyzes the patterns from the data, and further feed the patterns into the recommendation system to generate recommendations to save energy. A massive amount of data is collected using different sensors in the proposed architecture. This architecture has different layers, all of which are dedicated to performing specific tasks accordingly. The different preprocessing and Machine Learning (ML) techniques like Simple Linear Regression (SLR), Decision Tree Regression (DTR), Random Forest Regression (RFR), K-Nearest Neighbor Regression (KNNR), and Support Vector Regression (SVR) are used for data analysis. This study finds that Decision Tree regression (0.9999) and Random Forest Regression (0.9999) achieved good scores compared to the Simple Linear Regression (0.9901), K-Nearest Neighbor (0.9720), and Support Vector Regression (0.9966). The values of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for SLR (0.9900 and 0.0099), DTR (0.0439 and 0.0019), RFR (0.0427 and 0.0018), KNNR (0.0285 and 0.1690) and SVR (0.1394 and 0.0194). Thus, this paper concludes that the decision tree and random forest regression are having high scores and less error in comparison to the other algorithms. These regression techniques for data analysis can be used for the recommendation of energy management using the proposed architecture.

Key Words

Smart City, Energy management, Machine Learning, Internet of Things, Big Data.

1. INTRODUCTION

The effective use of energy in smart homes saves money, improves sustainability, and decreases the carbon impact on a wide scale [1, 2, 3]. The vast volume of data generated throughout a nation poses several data storage, management, and analysis issues. IoT and Big Data are logical solutions to these problems [4, 5]. IoT technologies may offer a ubiquitous computing platform for sensing, monitoring, and controlling the energy use of home appliances on a wide scale. This information is gathered utilizing a variety of wireless sensors put in residential units [6, 38]. Big Data technology may be used to collect simultaneously and analyze the data. Data is gathered, evaluated, and translated into usable information in reports, graphs, and charts that utilize predictive analytics and sophisticated technologies [7, 8, 9].

HEMS is a critical component of the smart grid environment because it enables load management among homeowners to reduce energy costs while flexibly supporting high penetration Renewable Energy Sources

ACI'22: Workshop on Advances in Computation Intelligence, its Concepts & Applications at ISIC 2022, May 17-19, Savannah, United States
EMAIL: taranasingh14@gmail.com (A. 1); ymca.arun@gmail.com (A. 2); sanjay.sharma@gbu.ac.in (A. 3)
ORCID: 0000-0002-9035-0068 (A. 1); 0000-0002-4359-406X (A. 2); 0000-0001-7918-4938 (A. 3)

(RES) at both the distant and local levels. Traditional household appliances, as well as developing ones such as Energy Storage Systems (ESS), Electrical Vehicles (EV) [11, 12], and others, must be considered in an efficient and cost-effective HEMS. The new appliances allow the HEMS to reduce costs, even more, minimize peak pressures, and overcome the volatility of RES production [13, 39, 40]. Unmanaged EV charging and discharging, for example, might increase the peak demand, lead to dangerous overload, and damage local distribution lines. To reduce the user's power cost and discontent, a convex programming home energy optimization framework including schedule-based appliances, battery-assisted appliances, and model-based appliances is proposed in a few research articles [14, 15]. There are several writers that have worked on Home Automation (HA) and energy conservation. In [45] a smart home energy management architectural proposal, the authors recommended that the sensors can be mounted for the detection and measurement of environmental data in their study. The microcontroller, which is the fundamental component of the design, receives the gathered data. The microcontroller is connected to various components of the architecture using various technologies such as Zigbee, X-bee shield, and so on. Internet connection, interpretation, and processing of data from various sensors transmit control signals to the architecture's appliances or actuators. It also delivers real-time environmental information to the website and handles requests sent from distant users via the webserver, according to the author. [47] suggested a home automation architecture that is part of the IoT application area and provides several potentials for building new beneficial applications. The author said that home automation (HA) is a collection of approaches for automating a house that incorporates technology into security, energy management, and welfare. Comfort is an important aspect of a home automation system since it encompasses all of the measures taken to enhance how occupants feel in their homes. [48] proposed a self-learning SHEMS architecture. The author said that their suggested system optimizes residential household units based on consumer comfort and energy cost, as well as reducing power supply overloading. For decision-making, the author employed a demand-side management system, a supply-side management system, price forecasting, price clustering, energy warning systems, and so on. The suggested system was verified by the author gathering real-time power usage data in Singapore. The author was doing a real-time case study. The structure of the smart home energy management system was addressed in [49]. The smart controller, which offers system management features such as logging, monitoring, and control, was mentioned by the author as a component of HEM. The smart microcontroller, according to the author, gathers real-time power consumption data from programmable and non-programmable appliances in order to apply effective demand management techniques.

Smart house IoT appliances [42, 43], Big Data [39], and Machine Learning Algorithms [37, 39] have limited capacities. As a result, more data handling choices must be included to properly gather, handle, and analyze massive amounts of information. Massive volumes of data are being collected and analyzed using big data analytics tools. researchers also allow for the appropriate interpretation and use of large volumes of sensor data. ML, on the other hand, is an artificial intelligence subfield that analyses algorithms and statistical models based on patterns and inferences that systems use to accomplish their goals. ML is also often used in real-time applications because to its viability and endurance. Simultaneously, ML addresses learning-related problems and discovers the background and characteristics of such difficulties in order to learn from them and enhance system performance. Finally, ML is classified into reinforcement, unsupervised, and supervised learning, is used to carry out tasks that need previously obtained data. The IoT paradigm, machine learning, big data technologies, and the application of these technologies in real-time are now the problems in the smart home space. The suggested architecture for energy management in smart homes utilizes IoT [40, 41], Bigdata [38, 39], and Machine Learning [39]. This study will assist in overcoming the obstacles that have previously been encountered. The proposed architecture enables real-time data collection, with the acquired information being saved in a Google document. This information is used for data analysis. This study uses five machine learning techniques to analyze the data i.e., SLR, DTR, RFR, KNNR, and SVR. The RMSE and MAE for the algorithms is also determined to validate these results.

The paper is divided into eight sections. The introduction to the domain is covered in the first section of the paper. The review of the related area is described in the second section of the paper. The proposed architecture's process flow is discussed in the third section of the paper. The fourth section of the paper discusses the proposed architecture of the energy management system. The data collection and preprocessing techniques are being discussed in the fifth section of the paper. The sixth section presents the evolution and analysis of the proposed system. The seventh section of the paper presents the discussion and visual representation of the results. The conclusion with future work is discussed in the eighth section of the paper, followed by the references.

2. Related Work

The smart city is a highly trendy topic these days, and it's drawing a lot of academics and experts to work in its many subdomains. There are many subdomains of smart cities, such as transportation, health care, education, agriculture, and so on, with energy management at the top of the list. Energy management is becoming a more substantial area to deal with these days. Energy is a vital necessity worldwide, whether for healthcare, transportation, education, or agriculture. Energy conservation is becoming a problem in every discipline at the same time. Many researchers have worked and continue to try to solve this challenge. Here are some of the research studies undertaken by various researchers to address these issues.

Sivaoragash et al. [16] focused on location-based optimum service selection for data management in smart grids using cloud computing. The author concentrated on the effective use of energy. According to the author, this is only achievable with the support of contemporary information technology. The author of this study offered a user-aware power regulation model for smart grids, as well as a location-based service selection strategy. The author then mentioned that they would also be introducing a secure and reliable solution. Liang et al. [17] discussed deep reinforcement learning for smart home energy control. The author analyzed the energy cost reduction challenge for smart homes in the absence of a building thermal dynamics model while keeping a pleasant temperature range in mind. The suggested Model's simulation results, based on real-world traces, indicate the algorithm's efficacy and unpredictability. Zafar et al. [18] proposed a study on the home energy management system and emphasized its principles, setup, and smart grid technologies. The author provided a thorough overview of the literature on home energy management systems, including references to key ideas, configurations, and supporting technologies. Dinh et al. [19] discussed a home energy management system that uses renewable energy and energy storage in the main grid and sells power. The author of this study developed an architecture with RES and ESS that considers the use of power from the main grid and electricity selling. Compared to earlier studies, the author can reduce residential energy consumption by 19.7% using the offered strategies. Designing, developing, and deploying an IoT-based smart energy management system was explored by Saleem et al. [44]. The author studied and examined several presentations, implementation, and validation of an IoT-based smart energy management approach, as well as the advantages of overcoming consumer energy management difficulties. To interact with any software-based smart solution, the author uses a variety of communication interfaces and protocols. For data collection and analysis, the author used Entrack software. The author validated the system in four distinct buildings. The case study analysis of this work demonstrates the system's efficiency.

Li et al. [48] proposed a HEMS, DSMS, SSMS unified for real-time operations of a smart home. The author discussed that the unified structure had some abilities such as worth prediction, value cluster, and energy aware structure which helped in enhancing its roles, it was done using ML techniques. The author discussed the proposed work with the self-energy HEMS infrastructure. Zekić-Sušac et al. [51] proposed a HEMS-IoT system which explores big data, and ML for household ease, protection, and saving energy. The author utilized the J48 ML model and Weka API to acquire consumer behavior and energy utilization outlines and categorize households with the respect to energy utilization. The author discussed their work with the help of the proposed architecture in the paper. Chouaib et al. [45] proposed a preposition of a general microcontroller-based HEMS architecture. The author explained that the proposed system aims to reduce the power utilization in smart homes, which was obtained by monitoring and regulating electrical home appliances. The author’s anticipated system has been used to proposal and apply an effective lighting system to reduce energy utilization in the smart home.

3. Process Flow Diagram

Figure 1 represents the process flow of the proposed architecture. The proposed architecture of HEMS collects real-time data. The collected data is being used for further analysis. The patterns of energy consumption are being analyzed using ML techniques in the administration layers of the proposed architecture. These observed patterns are further fed into the recommendation system in the service layer.

Figure 1: Process flow diagram of the proposed work

Following steps are carried out for proposed work.

Step 1: Data is collected in the device layer using different devices like sensors, controllers, actuators, gateways, etc.

Step 2: Preprocessing of the collected data is being done using different ML techniques.

Step 3: Using the preprocessed data obtained in step 2, the patterns of the energy consumption are obtained and different parameters which affect the energy consumption patterns are also analyzed.

Step 4: The patterns obtained in step 3 are fed into the recommendation system to generate the recommendations which can help to save energy.

Step 5: With the help of a pattern finder and the recommendation system in steps 3 and 4, energy-saving recommendations are being generated which helps in saving energy.

4. Proposed Architecture of Home Energy Management System

This paper proposed a new architecture for HEMS as shown in figure 2. This architecture has five layers, and each layer is dedicated to performing some specific tasks. The first layer of the architecture is the device layer dedicated to data collection from the smart home environment. The second is for the communication and information layer, which contains the information about the user, devices of the smart home, and the sensors. The third layer is the administration layer, which deals with the management of the information contained in the system with the pattern finder system. The fourth layer is the service and security layer dedicated to providing the services and security. The primary component of this layer is the recommendation system that utilizes the pattern data obtained from the previous layer and used those patterns for the recommendations. This layer is also having the security layer that performs the authentication and authorization. Few authors [45, 46, 47, 48, 49,50, 51] proposed the architecture of smart home energy management systems. The user interfaces layer at the end is interfaced with the users in real-time.

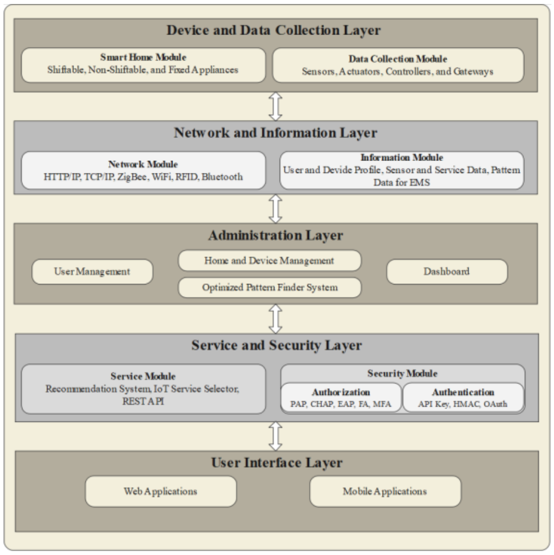


Figure 2: Layered Architecture of Energy Management system in Smart Home

4.1 Device Layer

This layer is divided into two sub-modules i.e., Smart home which consists of different types of devices in the smart home, and Data collection which contains different data collection devices.

4.1.1 Smart Home Module

HEMS is used to optimize the operating time of household appliances. HEMS makes the consumer's life easier by lowering electricity degeneracy and expense. Appliances are divided into three groups for optimization: shiftable, non-shiftable, and fixed appliances. Figure 3 depicts the classification of different appliances used in the smart home environment

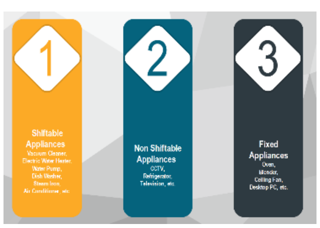


Figure 3: Types of Appliances used in Smart Homes or Buildings

4.1.2 Data Collection Module

IoT component mentioned in figure 4 i.e., gateways, sensors, actuators, and controllers permits data collection from numerous household appliances. The device layer also controls actuators and home automation devices.

Figure 4: Devices used for Data Collection in Smart Homes

4.2 Network and Information Layer

This layer is divided into two sub-modules i.e., network module and the information module.

4.2.1 Network Module

To choose the communication protocols for each home device, this layer evaluates components such as a collection of sensors, HTTP and TCP/IP, Bluetooth, WiFi, and 4G communication. The communication layer allows other levels in the architecture to communicate with one another. The HEMS communication layer utilizes Zigbee, TCP/IP, HTTP/IP, Bluetooth, WiFi, and other protocols/technologies.

4.2.2 Information Module

The information produced in the device layer is saved in this layer. The information layer, in particular, uses modules to handle five categories of data: user profile, device profile, sensor data, service data, and pattern data for the energy management system. The pattern data module monitors the data felt and gathered by the various smart home sensors using these five modules. This information is examined to find the best patterns in the data, which are then sent into the recommendation system.

4.3 Administration Layer

This layer executes and controls the tasks necessary to satisfy the application layer's user needs. The service layer ensures communication between the presentation and administration layers using the REST API, Recommendation system, and service selector. User management, Home and Device Management, Optimized Pattern Finder System, and Dashboard are the four categories of tasks done by the administration layer.

4.4 Service and Security Layer

This layer is separated into two sub-modules i.e., Service Layer and Security Layer.

4.4.1 Service Module

The application and management layer are linked by this layer. Additionally, this layer includes several REST APIs that enable users to access all HEMS features. IoT Service Selector, Recommendation System, and REST APIs are the primary components of this layer.

4.4.2 Security Module

This layer ensures data security and, as a result, the security and privacy of the gathered data from the device layer. This layer covers two security components: authorization (API Key, Basic Authorization, Hash Message Auth. Code, OAuth, and so on) and authentication (API Key, Basic Authorization, Hash-Based Message Authorization Code, OAuth, and so on) (Pass Authentication Protocol, THAP, EAP, SFA, TFA, MA, etc.).

4.5 User Interface Layer

This layer establishes a connection among the operator and the structure over a mobile application or web application.

Thus the proposed architecture has five layers which make system maintenance easier and increase scalability. The architecture aims to create an optimal pattern finder system that analyzes energy consumption patterns in the home environment. These discovered patterns are used in the service layer of recommendation systems. These recommendation systems provide customers with personalized recommendations and aid in reducing energy use.

5. Data Collection and Preprocessing:

The real-time data is collected using different sensors like temperature, humidity, pressure, gas, fire, laser object detection, voltage, electricity, etc. The data contains different parameters that affect the energy requirement in the smart homes' environment. A vast data set is required to implement the machine learning algorithm. The steps of data collection using the proposed framework are represented in figure 5(B). The design and application of the proposed system using the new architecture are shown in figure 5(A).

Step 1: In this IoT framework, the sensors are connected to the Arduino Uno.

Step 2: Arduino Uno is interfaced with NodeMCU and reads the input from sensors like fire, temperature, humidity, gas, voltage, and electricity sensors.

Step 3: NodeMCU collects the data and stores it in the Google Spreadsheet.

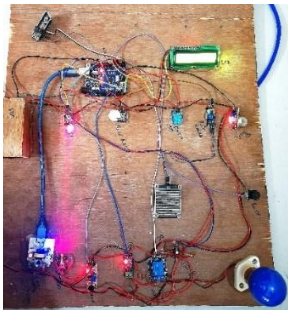


Fig 5(A)

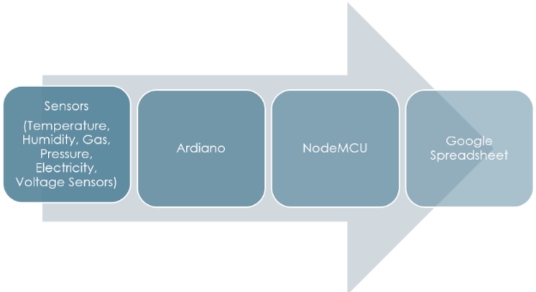


Fig 5(B)

Figure 5(A, B): Framework and Steps of Data Collection Process using the framework

Using the proposed system, a chunk of data is collected which is not feasible for ML implementation. So, this study used the data set available on online repositories [25] to implement and analyze the existing ML techniques on the proposed model. Using the proposed system, the sample of the collected dataset is represented in figure 6(A, B). The values obtained from different sensors are represented in figure 6 (A) with the parameters i.e., Date, Time, Voltage, Current, Gas. Figure 6 (B) represents the complete information of the collected dataset i.e., the data of the parameters, the total number of entries, the total number of rows and columns, etc.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 96453 entries, 0 to 96452
Data columns (total 12 columns):
Formatted Date              96453 non-null object
Summary                     96453 non-null object
Precip Type                 95936 non-null object
Temperature (C)             96453 non-null float64
Apparent Temperature (C)    96453 non-null float64
Humidity                    96453 non-null float64
Wind Speed (km/h)           96453 non-null float64
Wind Bearing (degrees)      96453 non-null float64
Visibility (km)             96453 non-null float64
Loud Cover                  96453 non-null float64
Pressure (millibars)        96453 non-null float64
Daily Summary               96453 non-null object
dtypes: float64(8), object(4)
memory usage: 8.8+ MB
```

Fig. 6(A)

```
Formatted Date              96429
Summary                        27
Precip Type                     2
Temperature (C)              7574
Apparent Temperature (C)     8984
Humidity                       90
Wind Speed (km/h)            2484
Wind Bearing (degrees)        360
Visibility (km)               949
Loud Cover                      1
Pressure (millibars)         4979
Daily Summary                 214
dtype: int64
```

Fig. 6 (B)

Figure 6(A, B):): Representation of the samples of the data collected using IoT

The preprocessing of the dataset is being done by removing all the missing and null values, and obtaining a feasible dataset for feeding into the ML models. The parameters are divided into the feature matrix and the prediction vector. The feature matrix (independent variables) has eight parameters: temperature, humidity, wind speed, visibility, pressure, summary, and precip type; on the other hand, the prediction vector (dependent variable) has one parameter that is apparent temperature. Here the ML model is being trained with the help of the feature matrix to predict the evident temperature according to the given parameters of the feature matrix.

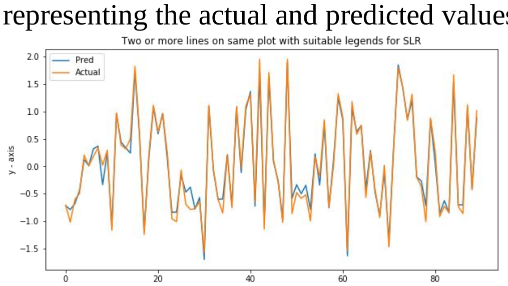
The dataset is separated into two parts i.e. training and testing sets in an 80-20% ratio, i.e., 80% of the dataset is the training set, and 20% of the data is the testing set. The different machine learning algorithms i.e. SLR, DTR, RFR, KNNR, SVR are being implemented. The ML models used in the proposed work are discussed in section 6.

6. Analysis of Machine Learning Techniques

Machine Learning is becoming the most powerful tool to face the challenges in technological development in the smart city. In this section, SLR, DTR, RFR, KNNR, SVR are implemented and analyzed for the recommendation of energy management in HEMS.

6.1 Simple Linear Regression

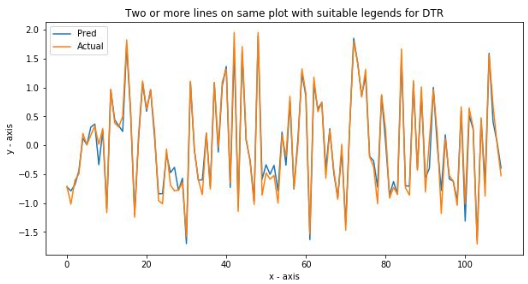
By fitting a line to the observed data, this Model estimates the connection between the variables. It's used to determine the asset of the link between two variables and the values of the dependent variables at a given value of the independent variable [21, 22, 23]. This ML model is implemented on the dataset to predict the apparent temperature. The score achieved by this model is 0.9901 given in table 1. Graph 1 is representing the actual and predicted values using the SLR model.



Graph 1. Representation of Actual and Predicted values using Simple Linear Regression

6.2 Decision Tree Regression

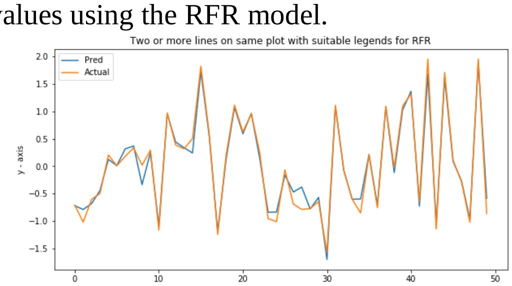
A decision tree constructs regression or classification models in the shape of a tree structure. It gradually cuts down a dataset into smaller and smaller sections while also developing an associated decision tree. A tree containing decision nodes and leaf nodes is the result. Each branch of a decision node represents a value for the property being examined. A decision on the numerical objective is represented by a leaf node. The root node is the highest decision node in a tree that corresponds to the best predictor. Both category and numerical data may be handled by decision trees [24, 25]. This ML model is implemented on the dataset to predict the apparent temperature. The score achieved by this model is 0.9999 given in table 1. Graph 2 is representing the actual and predicted values using the DTR model.



Graph 2. Representation of Actual and Predicted values using Decision Tree Regression

6.3 Random Forest Regression

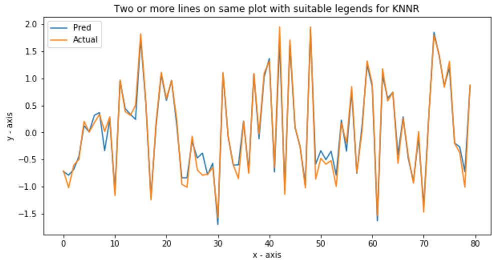
At training, random forests (RF) create many individual decision trees. To create the final forecast, the mode of the classes for classification or the mean prediction for regression, the predictions from all trees are combined. Ensemble approaches are named because they conclude based on a group of outcomes. With the use of several decision trees and a method called Bootstrap and Aggregation, often known as bagging, it can complete both regression and classification tasks. Instead, depending on individual decision trees, the main concept is to aggregate numerous decision trees to determine the outcome [26]. This ML model is implemented on the dataset to predict the apparent temperature. The score achieved by this model is 0.9999 given in table 1. Graph 3 is representing the actual and predicted values using the RFR model.



Graph 3. Illustration of Actual and Predicted values using Random Forest Regression

6.4 K-Nearest Neighbor Regression

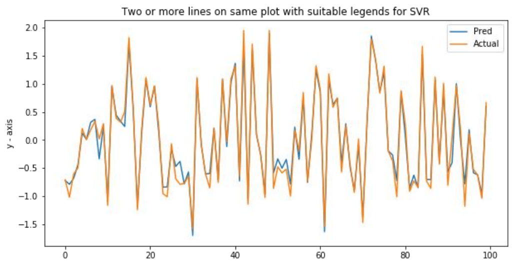
Both “classification and regression” both the issues may be solved with the KNN technique. The KNN forecasts the values of novel data points based on 'feature similarity.' This implies that a value is given to the new point depending on its similarity to the points in the training set. The distance between the new point and each training point must first be calculated. There are many ways of determining this distance, the most well-known of which are the Euclidian, Manhattan (continuous), and Hamming distances (for categorical) [27, 28]. This ML model is implemented on the dataset to predict the apparent temperature. The score achieved by this model is 0.9720 given in table 1. Graph 4 is representing the actual and predicted values using the KNNR model.



Graph 4. Illustration of Actual and Predicted values using KNNR

6.5 Support Vector Machine Regression

Kernel, Hyperplane, and verdict boundaries are three crucial factors in this machine learning model. Kernel supports in the exploration for a hyper-plane in a high-dimensional space while dropping the computation rate. As the size of the data grows, the computing cost will rise. A hyperplane splits the line between two data classes in a support vector machine. This line is used to forecast the constant output in SVR. The verdict boundary may be viewed as a separation line for simplification, with constructive instances on one side and undesirable instances on the other. The examples on this line may be considered as whichever constructive or undesirable [28]. This ML model is implemented on the dataset to predict the apparent temperature. The score achieved by this model is 0.9966 given in table 1. Graph 5 is representing the actual and predicted values using the SVR model.



Graph 5. Representation of Actual and Predicted values using Support Vector Regression

It is detected that the DTR and RFR can predict the apparent temperature with a high score i.e., 0.9999 (for both the models). The score for the rest of the ML models is 0.9720 (for KNNR), 0.9966 (for SVR), and 0.9901 (for SLR), which is less than the DTR and RFR scores. The RMSE and MAE values are also higher than the SLR, DTR, RFR, KNNR, and SVR are discussed in section 7.

7. Discussion and Visual Representation of the Results

The ML techniques i.e. SLR, DTR, RFR, KNNR, SVR discussed in section 6 used to analyze the data. The RMSE and MAE [31, 32, 33] values validate the prediction in comparison to actual values. The ML model has considered the best model for problem-solving if it has a high score and a low error value.

7.1 Root Mean Square Error

It's also known as root mean square deviation, most typically used to assess the accuracy of predictions. The standard deviation of the errors that occur while predicting a dataset is the RMSE given in equation 1. It uses Euclidean Distance to demonstrate how forecasts differ from actual values [34]. The RMSE values obtained by implementing the SLR, DTR, RFR, KNNR, and SVR are given in table 1.

Let N is the number of data points in the dataset, y_i is the i[th] measurements of the datapoint from the whole dataset, $y_{pred(i)}$ is the predictions for the i[th] measurement in the whole dataset. RMSE is the standard deviation of the forecast errors. It is being generally calculated to verify the experimental results of the regression analysis. The following steps are being carried out to calculate the RMSE.

Step 1: For all the predicted values, calculate the difference from the corresponding actual value, i.e., given in equation (a);

$$(\boldsymbol{y_i} - \boldsymbol{y_{pred(i)}}) \tag{a}$$

Step 2: Square the differences obtained in step 1 i.e., given in equation (b);

$$\left|\left|(\boldsymbol{y_i} - \boldsymbol{y_{pred(i)}})\right|\right|^{\mathbf{2}} \tag{b}$$

Step 3: Sum all the "squared differences" calculated in step 2, I.e., given in equation (c).

$$\sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \left|\left|(\boldsymbol{y_i} - \boldsymbol{y_{pred(i)}})\right|\right|^{\mathbf{2}} \tag{c}$$

Step 4: Calculate the average of the "sum of squared differences" derived in Step 3, which is called MSE or Mean Squared Error, i.e., given in equation (d).

$$\sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \left|\left|(\boldsymbol{y_i} - \boldsymbol{y_{pred(i)}})\right|\right|^{\mathbf{2}} /\boldsymbol{N} \tag{d}$$

Step 5: Finally take the square root of the values derived in step 4, which is RMSE, i.e., given in equation (1).

$$\textbf{RMSE} = \sqrt{\sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \left|\left|(\boldsymbol{y_i} - \boldsymbol{y_{pred(i)}})\right|\right|^{\mathbf{2}} /\boldsymbol{N}} \tag{1}$$

7.2 Mean Absolute Error

The error is the absolute difference between the real or actual values and the expected values. If the findings have a negative sign, it is ignored by the absolute difference.

MAE = Actual Values – Predicted Values (2)

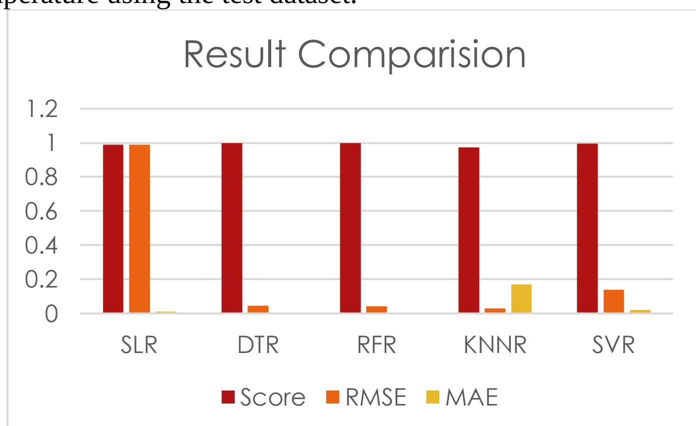
This function computes the average of this error over all samples in a dataset and returns the result given in equation 2. But, in certain cases, this value may not be the most important factor to consider when dealing with a real-life issue since the data that is being used to construct and assess the Model is the same, implying that the Model has never been exposed to genuine [36]. The MAE values obtained by implementing the SLR, DTR, RFR, KNNR, and SVR are given in table 1. As a result, it may perform well on visible data but may struggle or fail terribly when confronted with actual, hidden data.

Table 1:

Result Analysis of the different machine learning algorithms

Machine Learning Algorithm	Score	RMSE	MAE
Simple Linear Regression (SLR)	0.9901	0.9900	0.0099
Decision Tree Regression (DTR)	0.9999	0.0439	0.0019
Random Forest Regression (RFR)	0.9999	0.0427	0.0018
K-Nearest Neighbor Regression (KNNR)	0.9720	0.0285	0.1690
Support Vector Machine Regression (SVR)	0.9966	0.1394	0.0194

Graph 9 represents the Score, RMSE, and MAE analysis of the Machine Learning Models used on the dataset using SLR, DTR, RFR, KNNR, and SVR. These ML models are being implemented on the dataset to train the model on the given dataset and predict the apparent temperature for the test datasets. Graph 9 is the visual analysis and representation of the results of the models. This is concluded that the model that produces the highest score is DTR and RFR with the lowest error rate for predicting the apparent temperature using the test dataset.



Graph 9: Representation of the Score, RMSE, MAE investigation of the ML Models used on the dataset.

According to the discussion of the RMSE, and MAE values, given in table 1 and represented in figure 9. The best machine learning models for the data analysis in the proposed framework are random forest regression and decision tree regression as both models are having highest score value (0.99) and lowest RMSE (DTR: 0.0439, and RFR: 0.0427)and MAE (RMSE: 0.0019 and MAE: 0.0018) in comparison with the other Model's simple linear regression; the score is 0.9901, RMSE is 0.9900 and MAE is 0.0099, K-nearest neighbor score is 0.972, RMSE is 0.0285, MAE is 0.169 and support vector regression score is 0.9966, RMSE is 0.1394, MAE is 0.0194 obtained while implementing the ML models using the proposed architecture. After the whole study, it is observed that the apparent temperature gets affected (increase or decrease) because of the other parameters like temperature, humidity, pressure, wind, etc., and results in increased energy consumption. Thus, the proposed architecture is helpful to have a watch on the different parameters which cause high energy consumption. After that, this system will help manage the different parameters accordingly and reduce energy consumption.

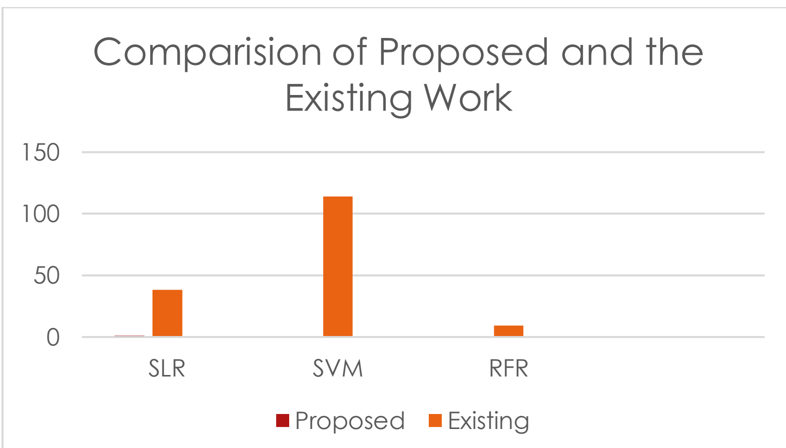
8. Comparison of the Proposed Work with Existing Work

In this section, the proposed work is being compared with the existing work. In 2020, Hoque et al. proposed a novel regression-based ensemble forecast system to forecast the energy requirement. The author addresses the enigma as regression modeling. The DS tool named RapidMiner is used for all data pre-processing computations. The author used the dataset available publicly from Kaggle, which was having 31 features gathered from different sensors mounted in smart homes. Out of all the features the author utilized 23 features for regression modeling. The model describes the 0.998 corrections between the features and their label and achieved 38.15 RMSE value [52]. The author also implemented several other ML models for regression such as SVM, RFR, etc., to apprehend their performance against the RMSE given in table 2. The author clearly stated that the their proposed work requires enhancement, particularly for RMSE value. The proposed HEMS is having much lower RMSE than the existing model. The RMSE values obtained in the proposed and the existing work are represented in Table 2 and Graph 10 represents the improvement in the proposed work and the existing work.

Table 2:

The RMSE values of the proposed and the Existing Work

Machine Learning Algorithm	RMSE(Proposed)	RMSE (Existing)
Simple Linear Regression (SLR)	0.9900	38.15
Support Vector Machine Regression (SVR)	0.1394	113.85
Random Forest Regression	0.0427	9.165



Graph 10: Representation of the comparison of the RMSE values of Proposed and Existing Work

A significant improvement is visible in Graph 10 in comparison of the existing work. The proposed work is having less RMSE which means the error in the prediction of the values is reduced significantly. Thus, the proposed work is improving the existing work as represented in section 7.

9. Conclusion with Future Work

At present, the energy requirement is growing day by day. With the increasing demand to meet the requirements of the consumers, the most prior concern of our administration/government is to look into the matter. This is possible to tackle this challenge with the help of modern technology. In a smart home, temperature, humidity, wind speed, air pressure, etc. are some primary parameters to predict the apparent temperature. The data of these parameters are being given to the machine learning models to predict the apparent temperature. In this paper, the regression models (DTR, RFR, SLR, KNNR, SVR) are applied to the dataset to train the ML model and to make the prediction of the apparent temperature, we obtain that the DTR and RFR are giving the highest score and lowest error rates represented in figure 9. This shows that these models are predicting the apparent temperature more accurately than the other models i.e., KNNR, SVR, and SLR. To conclude that analyzing and implementing the most affecting parameters which affect the increase or decrease of the demand of energy and consumption of energy in the proposed architecture can be done with the help of DTR and RFR. The existing architectures do not suggest the patterns which can be used for the recommendation system in HEMS. But in the proposed architecture, these systems are being introduced, which increase the system's impact and help to decrease the energy usages in the home environment. The future direction of the proposed work is, to use the real-time data of smart homes and optimized patterns of energy consumption in the obtained data using the machine learning algorithms on the administration layer. The obtained patterns will be provided to a personalized recommendation system on the service and security layer. This system will provide a recommendation to the users and will assistance in decreasing the energy usages as well as the bills of the residents in the smart homes. The aim of the system is to reduce the energy usages and electricity bills without interrupting the comfort zones of the residents.

10. References

[1] Sampathkumar, A., Murugan, S., Sivaram, M., Sharma, V., Venkatachalam, K., & Kalimuthu, M. (2020). Advanced Energy Management System for Smart City Application Using the IoT. Internet of Things in Smart Technologies for Sustainable Urban Development, 185-194.

[2] Cho, K., Jang, H., Park, L. W., Kim, S., & Park, S. (2019, January). Energy management system based on augmented reality for human-computer interaction in a Smart City. In 2019 IEEE International Conference on Consumer Electronics (ICCE) (pp. 1-3). IEEE.

[3] Kim, D., Kwon, D., Park, L., Kim, J., & Cho, S. (2020). Multiscale LSTM-Based Deep Learning for Very-Short-Term Photovoltaic Power Generation Forecasting in Smart City Energy Management. IEEE Systems Journal, 15(1), 346-354.

[4] Canizes, B., Pinto, T., Soares, J., Vale, Z., Chamoso, P., & Santos, D. (2017, June). Smart city: A GECAD-BISITE energy management case study. In International Conference on Practical Applications of Agents and Multi-Agent Systems (pp. 92-100). Springer, Cham.

[5] Canizes, B., Pinto, T., Soares, J., Vale, Z., Chamoso, P., & Santos, D. (2017, June). Smart city: A GECAD-BISITE energy management case study. In International Conference on Practical Applications of Agents and Multi-Agent Systems (pp. 92-100). Springer, Cham.

[6] Golpîra, H., & Bahramara, S. (2020). Internet-of-things-based optimal smart city energy management considering shiftable loads and energy storage. Journal of Cleaner Production, 264, 121620.

[7] Mariano-Hernández, D., Hernández-Callejo, L., Zorita-Lamadrid, A., Duque-Pérez, O., & García, F. S. (2021). A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis. Journal of Building Engineering, 33, 101692.

[8] Jasim, N. A., TH, H., & Rikabi, S. A. (2021). Design and Implementation of Smart City Applications Based on the Internet of Things. International Journal of Interactive Mobile Technologies, 15(13).

[9] Wang, D., Zhong, D., & Souri, A. (2021). Energy management solutions in the Internet of Things applications: Technical analysis and new research directions. Cognitive Systems Research, 67, 33-49.

[10] Lim, C., Cho, G. H., & Kim, J. (2021). Understanding the linkages of smart-city technologies and applications: Key lessons from a text mining approach and a call for future research. Technological Forecasting and Social Change, 170, 120893.

[11] Razmjoo, A., Nezhad, M. M., Kaigutha, L. G., Marzband, M., Mirjalili, S., Pazhoohesh, M., ... & Piras, G. (2021). Investigating smart city development based on green buildings, electrical vehicles and feasible indicators. Sustainability, 13(14), 7808.

[12] Al Sharif, R., & Pokharel, S. (2021). Smart City Dimensions and Associated Risks: Review of literature. Sustainable Cities and Society, 103542.

[13] Li, J., Dai, J., Issakhov, A., Almojil, S. F., & Souri, A. (2021). Towards decision support systems for energy management in the smart industry and Internet of Things. Computers & Industrial Engineering, 161, 107671.

[14] Francisco, A., Mohammadi, N., & Taylor, J. E. (2020). Smart city digital twin–enabled energy management: Toward real-time urban building energy benchmarking. Journal of Management in Engineering, 36(2), 04019045.

[15] Nikitas, A., Michalakopoulou, K., Njoya, E. T., & Karampatzakis, D. (2020). Artificial intelligence, transport and the smart city: Definitions and dimensions of a new mobility era. Sustainability, 12(7), 2789.

[16] Sivapragash, C., Padmanaban, S., Eklas, H., Holm-Nielsen, J. B., & Hemalatha, R. (2019). Location-based optimized service selection for data management with cloud computing in smart grids. Energies, 12(23). https://doi.org/10.3390/en12234517

[17] Yu, L., Xie, W., Xie, D., Zou, Y., Zhang, D., Sun, Z., Zhang, L., Zhang, Y., & Jiang, T. (2020). Deep Reinforcement Learning for Smart Home Energy Management. IEEE Internet of Things Journal, 7(4), 2751–2762. https://doi.org/10.1109/JIOT.2019.2957289

[18] Zafar, U., Bayhan, S., & Sanfilippo, A. (2020). Home Energy Management System Concepts, Configurations, and Technologies for the Smart Grid. IEEE Access, 1–1. https://doi.org/10.1109/access.2020.3005244

[19] Dinh, H. T., Yun, J., Kim, D. M., Lee, K. H., & Kim, D. (2020). A Home Energy Management System with Renewable Energy and Energy Storage Utilizing Main Grid and Electricity Selling. IEEE Access, 8, 49436–49450. https://doi.org/10.1109/ACCESS.2020.2979189

[20] Al-Ali, A. R., Zualkernan, I. A., Rashid, M., Gupta, R., & Alikarar, M. (2017). A smart home energy management system using IoT and big data analytics approach. IEEE Transactions on Consumer Electronics, 63(4), 426–434. https://doi.org/10.1109/TCE.2017.015014

[21] Mohapatra, B. N., & Panda, P. P. (2019). Machine learning applications to smart city. ACCENTS Transactions on Image Processing and Computer Vision, 5(14), 1.

[22] Maulud, D., & Abdulazeez, A. M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. Journal of Applied Science and Technology Trends, 1(4), 140-147.

[23] Kumar, H., Singh, M. K., Gupta, M. P., & Madaan, J. (2020). Moving towards smart cities: Solutions that lead to the Smart City Transformation Framework. Technological forecasting and social change, 153, 119281.

[24] Sampathkumar, A., Maheswar, R., Harshavardhanan, P., Murugan, S., Jayarajan, P., & Sivasankaran, V. (2020, July). Majority voting-based hybrid ensemble classification approach for predicting parking availability in smart city based on IoT. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-8). IEEE.

[25] Link of the Dataset used in this paper: https://www.kaggle.com/budincsevity/szeged-weather

[26] Shafiq, M., Tian, Z., Sun, Y., Du, X., & Guizani, M. (2020). Selection of effective machine learning algorithm and Bot-IoT attacks traffic identification for internet of things in smart city. Future Generation Computer Systems, 107, 433-442.

[27] Shahabi, H., Shirzadi, A., Ghaderi, K., Omidvar, E., Al-Ansari, N., Clague, J. J., ... & Ahmad, A. (2020). Flood detection and susceptibility mapping using sentinel-1 remote sensing data and a machine learning approach: Hybrid intelligence of bagging ensemble based on k-nearest neighbor classifier. Remote Sensing, 12(2), 266.

[28] Luckey, D., Fritz, H., Legatiuk, D., Dragos, K., & Smarsly, K. (2020, August). Artificial intelligence techniques for smart city applications. In International Conference on Computing in Civil and Building Engineering (pp. 3-15). Springer, Cham.

[29] Ahmad, W., Ayub, N., Ali, T., Irfan, M., Awais, M., Shiraz, M., & Glowacz, A. (2020). Towards short term electricity load forecasting using improved support vector machine and extreme learning machine. Energies, 13(11), 2907.

[30] Sivaram, M., Lydia, E. L., Pustokhina, I. V., Pustokhin, D. A., Elhoseny, M., Joshi, G. P., & Shankar, K. (2020). An optimal least square support vector machine-based earnings prediction of blockchain financial products. IEEE Access, 8, 120321-120330.

[31] Tong, Z., Miao, J., Tong, S., & Lu, Y. (2021). Early prediction of remaining useful life for Lithium-ion batteries based on a hybrid machine learning method. Journal of Cleaner Production, 317, 128265.

[32] Ahmad, F., Almuayqil, S. N., Mamoona, H., Shahid, N., Wasim Ahmad, K., & Kashaf, J. (2021). Prediction of COVID-19 cases using machine learning for effective public health management. Computers, Materials, & Continua, 2265-2282.

[33] Ahmad, A., Ostrowski, K. A., Maślak, M., Farooq, F., Mehmood, I., & Nafees, A. (2021). Comparative study of supervised machine learning algorithms for predicting the compressive strength of concrete at high temperature. Materials, 14(15), 4222.

[34] Angelaki, A., Singh Nain, S., Singh, V., & Sihag, P. (2021). Estimation of models for cumulative infiltration of soil using machine learning methods. ISH Journal of Hydraulic Engineering, 27(2), 162-169.

[35] Sahoo, A., Samantaray, S., & Ghose, D. K. (2021). Prediction of flood in Barak River using hybrid machine learning approaches: a case study. Journal of the Geological Society of India, 97(2), 186-198.

[36] Watt-Meyer, O., Brenowitz, N. D., Clark, S. K., Henn, B., Kwa, A., McGibbon, J., ... & Bretherton, C. S. (2021). Correcting weather and climate models by machine learning nudged historical simulations. Geophysical Research Letters, 48(15), e2021GL092555.

[37] Tarana Singh, Arun Solanki, Anand Nayyar, "Multilingual Opinion Mining Recommendation System Using RNN for Movie Domain" Accepted for publication in International conference in Computing, Communication and Cyber-security (IC4S) 12-13 October, 2019, Chandigarh.

[38] R. Rajput and A. Solanki, "Real Time Analysis of Tweets using Machine Learning and Semantic Analysis", International Conference on Communication and Computing Systems(ICCCS-2016), Taylor and Francis, at Dronacharya College of Engineering, Gurgaon, 9-11 September, pp. 687-692, 2016.

[39] Akash Tayal, Arun Solanki and Simarpreet Singh, " Integrated Frame work for Identifying Sustainable Manufacturing Layouts based on Big Data, Machine Learning, Meta-Heuristic and Data Envelopment Analysis ", Accepted for Publication in Journal Sustainable Cities and Society, Elsevier Publication, IF=5.2 SCIE.

[40] Solanki, A., & Nayyar, A. (2019). Green internet of things (G-IoT): ICT technologies, principles, applications, projects, and challenges. In Handbook of research on big data and the IoT (pp. 379-405). IGI Global.

[41] Nayyar, A., Rameshwar, R. U. D. R. A., & Solanki, A. (2020). Internet of Things (IoT) and the digital business environment: a standpoint inclusive cyber space, cyber crimes, and cybersecurity. The Evolution of Business in the Cyber Age, 10, 9780429276484-6.

[42] Singh, T., Solanki, A., & Sharma, S. K. (2021). Role of Smart Buildings in Smart City—Components, Technology, Indicators, Challenges, Future Research Opportunities. Digital Cities Roadmap: IoT-Based Architecture and Sustainable Buildings, 449-476.

[42] Singh, S. P., Solanki, A., Singh, T., & Tayal, A. (2021). Internet of Intelligent Things: Injection of Intelligence into IoT Devices. In Artificial Intelligence to Solve Pervasive Internet of Things Issues (pp. 85-102). Academic Press.

[43] M. U. Saleem, M. R. Usman and M. Shakir, "Design, Implementation, and Deployment of an IoT Based Smart Energy Management System," in *IEEE Access*, vol. 9, pp. 59649-59664, 2021, doi: 10.1109/ACCESS.2021.3070960.

[44] Chouaib, B., Lakhdar, D., & Lokmane, Z. (2019, March). Smart home energy management system architecture using IoT. In *Proceedings of the 9th International Conference on Information Systems and Technologies* (pp. 1-5).

[45] Al-Ali, A. R., Zualkernan, I. A., Rashid, M., Gupta, R., & AliKarar, M. (2017). A smart home energy management system using IoT and big data analytics approach. *IEEE Transactions on Consumer Electronics*, *63*(4), 426-434.

[46] Machorro-Cano, I., Alor-Hernández, G., Paredes-Valverde, M. A., Rodríguez-Mazahua, L., Sánchez-Cervantes, J. L., & Olmedo-Aguirre, J. O. (2020). HEMS-IoT: A big data and machine learning-based smart home system for energy saving. *Energies*, *13*(5), 1097.

[47] Li, W., Logenthiran, T., Phan, V. T., & Woo, W. L. (2018). Implemented IoT-based self-learning home management system (SHMS) for Singapore. *IEEE Internet of Things Journal*, *5*(3), 2212-2219.

[48] Zafar, U., Bayhan, S., & Sanfilippo, A. (2020). Home energy management system concepts, configurations, and technologies for the smart grid. *IEEE access*, *8*, 119271-119286.

[49] Dinh, H. T., Yun, J., Kim, D. M., Lee, K. H., & Kim, D. (2020). A home energy management system with renewable energy and energy storage utilizing main grid and electricity selling. *IEEE Access*, *8*, 49436-49450.

[50] Zekić-Sušac, M., Mitrović, S., & Has, A. (2021). Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities. *International journal of information management*, *58*, 102074.

[51] Hoque, M. S., Jamil, N., Saharudin, S. A., & Amin, N. (2020, August). Energy Usage Prediction for Smart Home with Regression Based Ensemble Model. In 2020 8th International Conference on Information Technology and Multimedia (ICIMU) (pp. 378-383). IEEE.