*Article*

# Random Forest Based Power Sustainability and Cost Optimization in Smart Grid

**Danalakshmi D [1], Łukasz Wróblewski[2,*], Sheela A[3], A. Hariharasudan[4] and Mariusz Urbański[5]**

[1] GMR Institute of Technology, Rajam, India; danalakshmi.d@gmrit.edu.in
[2] Faculty of Applied Sciences, WSB University in Dąbrowa Górnicza, Zygmunta Cieplaka 1c, 41-300 Dąbrowa Górnicza, Poland; lwroblewski@wsb.edu.pl
[3] Kongu Engineering College, Erode, India; a_sheela@kongu.ac.in
[4] Faculty of English, Kalasalingam Academy of Research and Education, Krishnankoil 626128, India; dr.a.hariharasudhan@gmail.com
[5] Road and Bridge Research Institute, ul. Instytutowa 1, 03-302 Warszawa, Poland, murbanski@ibdim.edu.pl
* **Correspondence**: lwroblewski@wsb.edu.pl

**Abstract:** Presently power control and management play a vigorous role in information technology and power management. Instead of non-renewable power manufacturing, renewable power manufacturing is preferred by every organization for controlling resource consumption, price reduction and efficient power management. Smart grid efficiently satisfies these requirements with the integration of machine learning algorithms. Machine learning algorithms are used in a smart grid for power requirement prediction, power distribution, failure identification etc. The proposed Random Forest-based smart grid system classifies the power grid into different zones like high and low power utilization. The power zones are divided into number of sub-zones and map to random forest branches. The sub-zone and branch mapping process used to identify the quantity of power utilized and the non-utilized in a zone. The non-utilized power quantity and location of power availabilities are identified and distributed the required quantity of power to the requester in a minimal response time and price. The priority power scheduling algorithm collect request from consumer and send the request to producer based on priority. The producer analysed the requester existing power utilization quantity and availability of power for scheduling the power distribution to the requester based on priority. The proposed Random Forest based sustainability and price optimization technique in smart grid experimental results are compared to existing machine learning techniques like SVM, KNN and NB. The proposed random forest-based identification technique identifies the exact location of the power availability, which takes minimal processing time and quick responses to the requestor. Additionally, the smart meter based smart grid technique identifies the faults in short time duration than the conventional energy management technique is also proven in the experimental results.

**Keywords:** Smart Grid; Random Forest; Internet of Things; Power management; Machine Learning; Smart Meter; Priority Power Scheduling.

## 1. Introduction

Nowadays, agriculture, manufacturing, healthcare, automobile and electrical sectors operation depends on automation. Likewise, automation depends on information technology (IT) and frequent power supply. Non-frequent power supply leads to poor automation control and management. Hence, IT management is integrated into power supply management. In existing techniques, both were controlled and managed independently. Thus, power utilization, power availability, price of processing and fault identifications are difficult [1]. Hence, information and power sustainability techniques are integrated through a Smart Grid (SG). Government is also supporting the implementation of renewable energy sources [2] and suggests to improve the usage of natural resources in an efficient manner [3]. The effective implementation of renewable energy sources helps the countries considered as vulnerable segment who are living without proper energy supply [4] as well as the country like Ukraine [5]. Smart grid consists of smart meters, load control switches, distribution boards, renewable energy resources, energy storage, computers, etc. These devices are integrated with IT for quick and efficient management of electricity demand. The SG's are used to manage renewable resources like biomass, solar, wind and tidal, etc. due to several factors for which the consumers adopt the renewable energy [6]. The power wastages are reduced through SG and gap between the producer and consumers is also reduced. To improve the efficiency of SG, the Machine Learning (ML) algorithms are integrated into SG [7]. The quality of power should be maintained using the technologies like dynamic voltage regulator in SG [8].

Presently, the ML algorithms play a vital role in the agriculture, manufacturing, healthcare, automobile and electrical sectors. The major roles of ML algorithms are classification, clustering, prediction, regression etc. In electrical sector, ML algorithms are used in SG in different aspects like improving power utilisation, identifying false data injection, reducing power and system failure, quick recovery, price reduction, secure power transmission, energy management, high and low power utilization prediction etc. [9]. The ML algorithms are helpful to manage SG based Internet of Things (IoT) also. The IoT devices collect power from various sources and are stored in SG. The collected power distributed the consumer with the help of the ML classification algorithms [10]. The ML algorithms are also used to schedule the power between smart cities. The scheduling algorithms collect information like power demand, price, and power availability for allocating the power to the required place in an efficient way [11].

The integration of ML and SG leads to numerous benefits like responding to consumer in a quickest time period, predicting consumer requirements and analysing consumer behavior etc. [12]. The ML algorithms are also used for optimizing load between industries. Load balance during the peak demand influences the energy market price [13][14]. In load forecasting error minimization process, power is efficiently distributed to consumers through ML. The complex SG monitoring process is done by Smart Meter (SM) and SM is used for load balance [15]. The ML algorithm collects information from the IoT devices and analysis the load and forecasting information. The ML and IoT embedded algorithms takes lower processing price and produce highest accuracy rate [16]. Through these analyses, the ML algorithms play a vital role in SG is identified. Hence, the ML algorithm-based power sustainability and price optimization technique is proposed in this work.

The remaining part of this paper is organized as follows: In section II, the existing ML techniques used in SG management are analysed deeply with its merits and limitations. In section III, the proposed RF based power sustainability and price optimization technique in SG management is discussed with necessary architecture and algorithms. Section IV, the proposed technique experimental results are compared with the existing techniques and section V, the proposed technique is concluded with the future enhancements.

## 2. Related Works

Hossein Taherian et al. [17] proposed the optimal dynamic pricing technique-based SG system. The optimal energy management algorithm is used to analyse the customer power utilization in different resources. Afterwards, the customer's behaviours are analysed to identify the electricity usage of the particular time period. The customer past and present power utilization is easily identified. But the customer future requirement is not analysed. Hence, this technique is not preferable for future data requirement analysis. Sharmila. P et al. [18] discussed the hybrid of ML and data analytics techniques in smart energy management. The clustering technique was used for analysing power utilization. The power area is divided into a number of zones and regions for analysis. This analysis provides better power distribution but unable to handle emergency demands. Renugadevi. N et al. proposed the IoT and ML based SG technique. The well-defined in-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2022 doi:10.20944/preprints202201.0138.v1

frastructure was used for future power demands analysis through IoT devices. The implementations of well-defined infrastructure in all sites are difficult. Hence, an alternate technique is required to predict the present and future power requirement analysis [19].

Fabiano Pallonetto et al. [20] used the ML models for the construction of SG. The demand response algorithm was the hybrid algorithm, such as the rule-based and predictive-based approaches were integrated to identify demands. These two algorithms are commonly used to predict the power demand and utilization of the consumer. The dynamic control and monitoring technique was initiated and not completed. Marijana Zekic-susac et al. [21] used the ML algorithm-based SG management in public sectors. The IoT devices are fixed in every organization and power consumption was collected in a particular period. The collected data was analyzed by ML algorithms for predict the future power requirements. Based on this analysis, the power distribution is reconstructed in the future. This reconstruction of power distribution required higher price than the scheduling process. Hongbo Zou et al. [22] proposed the reinforcement learning-based optimal solution for facilitating the trivial computational searching process. The reinforcement technique minimized the investment and management price efficiently. This technique analyzes the demands in an hourly manner only, not lesser than the hours.

Tanveer Ahmad et al. [23] discussed the medium and long-term energy demands. The artificial neural networks, multivariate linear regression and adaptive boosting models are combined together for finding the energy demands. The energy demand was analyzed in the time duration of one month, seasonal and one year time periods. This technique analyzed the future energy requirements. The working complexity of this technique is higher than other techniques. Imtiaz Parvez et al. [24] proposed online power disturbance detection by using a Support Vector Machine (SVM) model. Using the SM, the power distribution was continuously monitored for efficient power distribution and management. The accuracy rate of SVM technique depends on training data. Enrico De Santis et al. [25] discussed cluster-based power optimization in an SG organization. The fuzzy membership function and clustering evolutionary techniques were combined together for identifying the faults in SG. Only a two-class classification technique was used for the classification and clustering process. The two-class classification is not enough for producing an accurate result.

Ding Li et al. [26] designed the optimal customer decisions on SG management technique. The micro-grid was used to collect resources from home appliances and managed by SG. Q-Learning technique was used for dynamic decision making. This Q-learning technique produces better results than other ML algorithms. The consumer side was discussed, not from the producer side. Sardar Ali et al. [27] proposed the ML based power quality estimation in SG. The entropy-based measurements and Bayesian network models were used for monitoring the power links. This monitoring technique was used for observing the un-monitoring links. Hence, the power values were calculated perfectly. But in real-life applications, power wastage is occurring in monitoring links also. Salahuddin Azad et al. [28] discussed the ML based transformation in SG. The ML algorithms were used for intelligent decision-making and response to customer requests, unexpected changes in power distribution, etc. Similarly, the wireless technology-based power distribution and attacks such as intrusion detection, malicious activities identification and other security issues were addressed.

Zaib Ullah et al. [29] deeply discussed artificial intelligence and ML based smart cities using SG. The author focused on power shortage problems, drawbacks of existing algorithms in the SG and improvement procedure of current SG based smart city applications. The combination of artificial intelligence and ML algorithms produces outstanding results than the individual algorithms. The time complexity of the integrated algorithm is higher than traditional algorithms. Muralitharan et al. [30] proposed the Neural network-based energy demand prediction in SG. The conventional neural network was used to find the energy demand on a consumer side, and the particle swarm optimization technique was used to find energy availability on the producer side. The conventional neural network was used for finding future energy demands. The accuracy

 doi:10.20944/preprints202201.0138.v1

rate of conventional techniques is lesser than modern algorithm. Mohamed A. Mohamed et al.[31] discussed the energy management technique in SG. With the help of sensors and SM, the power consumption was automatically monitored for optimal power management process. Similarly, the author discusses the reduction of greenhouse effect, energy efficiency improvement, price reduction and power utilization minimization. Kasun S.Perera et al. [32] focused on both small scale and large scale power utilizations and productions. Similarly, the future power generation resources were discussed with device location and configuration. These techniques were discussed through different ML algorithms. Based on the above analysis the following limitations are identified in the existing techniques.

Research Gap of existing techniques:

- Future power requirement and emergency period power requirements were not analysed.
- Dynamic control and monitoring are not yet completed.
- Reconstruction of infrastructure leads to higher expenditure.
- Accuracy rate depends on training data of the algorithm.

The random forest-based power sustainability and price optimization in smart grid is proposed in this work overcomes these limitations. In a proposed technique, the **first gap** is overcome by the priority scheduling algorithm; the **second** and **third gap** is fulfilled with SM in a SG system without infrastructure reconstruction. Finally, the accuracy rate of training set is improved by a random forest technique. These points are discussed in the upcoming section.

## 3. Random Forest Based Smart Grid Optimization

Random Forest (RF) is a supervised learning algorithm used for classification and regression processes. The ensemble learning algorithm is a foundation of RF technique for efficiently solving complex problems. In RF classification, the numbers of decision tree results are aggregated and the average is calculated to improve the prediction accuracy rate. A larger number of sub-tree results lead to improve the accuracy of prediction and resolve the over and under fitting problems. The Information Gain, Gini Index, Decision Tree and Bootstrap aggregations are worked together to improve the accuracy of RF algorithm results. The benefits of RF are listed as follows:

- RF takes minimal time for the analysis of large size data.
- RF produces higher accuracy rate than other ML algorithms.
- RF algorithm processed both continuous and categorical data in a perfect way.
- The missing values are efficiently handled without normalization.

Due to these benefits the RF algorithm is proposed in the SG power sustainability and optimization process. The contributions of the proposed system are:

- Identifying the low power and high-power utilization zones through SM's and RF technique.
- Identifying the power requirement in an advance and applying priority power scheduling algorithm for effective utilization of power without increasing the cost of production or reconstruction of infrastructure.
- To resolve faults quicker than the traditional grid technique.

Figure 1 shows the proposed RF based power sustainability and cost optimization in a SG. Terms which are used in the proposed system is listed in Table 1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2022 doi:10.20944/preprints202201.0138.v1

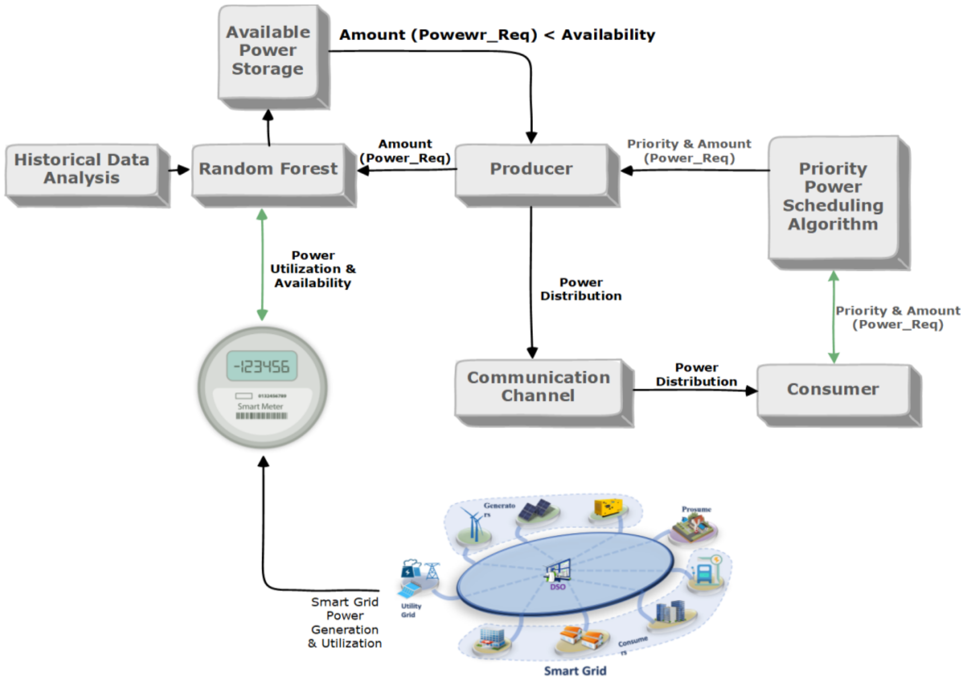


**Figure 1**. System Architecture of Random Forest based Smart Grid Optimization

**Table 1.** Terms used in proposed system

| Terms | Abbreviations |
|---|---|
| SG | Smart Grid |
| IoT | Internet of Things |
| ML | Machine Learning |
| RF | Random Forest |
| SVM | Support Vector Machine |
| SM | Smart Meter |
| MW | Mega-Watt |
| KW | Kilo-Watt |
| $H_P$ | High Priority |
| $M_P$ | Medium Priority |
| $N_P$ | Normal Priority |
| SMBill | Smart Meter Bill |
| AP | Available Power |
| GP | Generated Power |
| UP | Utilized Power |
| AT | Access Time |
| PA | Power Access |
| PU | Power Utilization |
| TAP | Total Available Power |

The roles of above-mentioned architecture are discussed as follows:

**Consumer –** The consumer sends the unit of power request to the priority power scheduling algorithm along with type of priority (Medical emergency = High Priority ($H_P$), Government work = Medium Priority ($M_P$), Festival and other works = Normal priority ($N_P$)). The priority type is assigned by the producer.

**Priority Power Scheduling Algorithm (PPSA)** - Depends on the consumer request, the PPSA schedules the request and sends the request along with the priority and unit of power requirement to the producer. Likewise, the PPSA fixes the threshold (th) value for

 doi:10.20944/preprints202201.0138.v1

every consumer power utilization. If the power consumption exceeds the 'th', the request is processed later; else, the request is processed depends on priority. The scheduler plays a major role in the power optimization process. The major issue of priority scheduling is starvation. This starvation issue in PPSA is avoided by an ageing factor. When more number of higher priority requests are raised, the lesser priority request unable to consumer the power. Thus, the aging factor is assigned to each request based on request arrival. If the waiting time exceeds the assigned waiting time, the lower priority request consumes the power. Equation 1 is used for measuring Access Time ($A_T$) of the power request. If the request is $H_P$, access granted immediately, else if $M_P$, check the other request and grand the access, otherwise wait for a particular period till the power is available.

$$A_T = \begin{cases} (Priority = H_P) & Access\ Grand \\ (Priority = M_P) & Check\ Other\ Req \\ (Priority = N_P) & Wait \end{cases} \tag{1}$$

**Producer** – The producer has all the information about the SG, such as the quantity of power utilized in a particular zone, available quantity power in a particular zone and future power requirement etc. Based on this information, the power requirement is analyzed in RF and supplied to the consumer through a communication channel if the power request is lesser than power availability; else needs to wait until the requested power is available. Equation 2 is used to check the Power Access ($P_A$).

$$P_A = \begin{cases} Power_{Req} < A_P & Access\ Grand \\ Power_{Req} \geq A_P & Wait \end{cases} \tag{2}$$

**Smart Grid** – SG collects power utilization information from the consumer through communication channel and collects the Available Power ($A_P$) from various power sources like solar, wind, biomass, tidal, etc. The SM plays a major role in the SG. The advanced energy meter collects power request information from a consumer along with consumer details and sends the collected details to RF classifier by using the specialized IoT devices. The IoT sensor communicated through a communication channel for collecting the information by specialized devices. The SG working principle is based on digital, two-way communication which is done by sensors, self-monitoring, self-healing and multiple consumer choice-based techniques. The typical function of SM is two-way communication, data collection, recording and storing, load control, security, display and billing. The benefits of SM are to reduce processing time, better power management during peak times, improve load management and highly efficient use of resources. Similarly, SM shows consumer power utilization habits, accurate billing, better electrical equipment usage and reducing $CO_2$ emission. Equation 3. is used to measure the power consumption of customers using SM.

$$SM_{Bill} = \sum_{i=1}^{n} (End_{Time} - Start_{Time}) \tag{3}$$

The $SM_{Bill}$ depends on the start and end time of a particular time period of all the devices is connected in to SM. Where '*i*' is the number of devices connected to the SM for the particular time period. Based on the power utilization, the $A_P$ is identified for the particular time period. Equation 4 is used for measuring the $A_P$ of a particular node such that Utilized Power ($U_P$) is subtracted from the $G_P$.

$$A_P = G_P - U_P \tag{4}$$

**Random Forest** – RF based on ensemble classifier, which contains number of decision trees and the output of RF depends on individual trees output. In RF technique, the trees are trained by the randomly selected subset data from a large dataset and made the summarization to get the final decision as the better prediction. The particular zone power utilization is given to RF to classify the high and low power utilization. The classified power stored in a power storage location for the distribution of power to requestor. If $P_U$ is lesser than the threshold value, the consumer utilizes the power in future. Otherwise, the $P_U$ is stopped for that customer till power is available. Equation 5 is used for the calculation of Power Utilization ($P_U$). Based on equation 5, the lower and higher power utilization is classified.

$$P_U = \begin{cases} P_U < threshold & Low\ P_U \\ P_U \geq threshold & High\ P_U \end{cases} \tag{5}$$

Using equation 4 the $A_P$ is calculated for one particular branch. Equation 6 is used to find the total amount of $TA_P$ in a specific zone, calculated by the sum of all branches $A_P$ from node 1 to *n*.

$$TA_P = \sum A_{P_1} + A_{P_2} + \cdots \ldots + A_{P_n} \tag{6}$$

Based on $TA_P$ the $A_P$ for a specific period is calculated for requested power distribution. Equations 7 and 8 are used to find the Total $P_U$ ($TP_U$) of a specific consumer and $A_P$ access limit. $TP_U$ is the sum of Existing $P_U$ ($EP_U$), Current $P_U$ ($CP_U$). The $TP_U$ is subtracted from the Power Access Limit ($PA_L$) to get the Balance $P_U$ ($BP_U$).

$$TP_U = \sum (EP_U + CP_U) \tag{7}$$

$$BP_U = PA_L - TP_U \tag{8}$$

By using SM, the resources are monitored in a dynamic manner. If any fault is identified on a producer or consumer side, the SM immediately sends it to the producer. Based on the immediate identification, the producer chooses the alternate resources for power generation and distribution. Hence, the fault identification and solution are done most quickly. Based on the proposed technique, the SG power generation, power utilization, power storage is maintained perfectly, and consumer request is satisfied in the quickest time is proved.

Algorithm 1. The RF based power sustainability and cost optimization technique in a SG.

**Algorithm 1: Random Forest based Smart Grid Optimization**

**Input:** $U_P$, $Power_{Req}$, $A_P$, Priority, $G_P$

**Output:** Access Grand or not

**Procedure:**

Step 1: To collect consumer request and priority

Step 2: Analyze the Priority of the request

**If** (Priority= $H_P$)

Access Grand

**Else if** (Priority = $M_P$)

Check another request, if no existing request, Allow

**Else**

Assign scheduling number to distribute power

Step 3: Send request to producer to check $A_P$

Step 4: Check the $A_P$ through RF using Equation 6

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2022 doi:10.20944/preprints202201.0138.v1

**If** (Power_Req < $A_P$)

Access Grand

**Else**

Wait till $A_P$ > Power_Req

Step 5: Using equation 5 to measure the $P_U$.

**If** ($P_U$ < threshold)

Allow to utilize the power

**Else**

Stop utilization of power

Step 6: Using equation 3 to measure the bill amount of the consumer $U_P$.

Based on Algorithm 1, the power request, utilization and availability are managed in an efficient manner. The experimental results of the proposed system are discussed in further section.

## 4. Results

In this section, the proposed technique's experimental results will be discussed with the dataset, implementation, and comparison details. The implementation is done in a python language along with necessary libraries like Keras and Tensor-Flow for the implementation of ML algorithms. The experimental results are taken by the analysis of energy load dataset. The energy load dataset having the data from 2002 to 2018 energy consumption in an hourly manner with 116,189 data instances as a whole. From these instances, 110,000 data instances are taken for training, and 6189 instances are taken for testing [1]. Energy consumption is divided into two categories such as high-power utilization and low power utilization. The high-power utilization range starts from Mega-Watt (MW) power utilization per hour. The lower power utilization is lesser than MW, such as the Kilo-Watt range (KW). In a proposed system, the accuracy rate is compared among machine-learning algorithms like SVM, K Nearest Neighbor (KNN), Naïve Bayes (NB) and RF. The accuracy rate of the ML algorithm depends on the training and test data ratio. In the proposed technique, 80 percentages are taken for training and 20 percentages are taken for testing. In a proposed RF based prediction, the high and low power utlization is classified by a two class prediction confusion matrix. The accuracy rate is calculated from table 2 and equation 9. Table 3 shows the prediction accuracy.

.

Table 2. Two class prediction confusion matrix

| | Predicted | |
|---|---|---|
| | Low | High |
| Actual | A | B |
| | C | D |

$$Accuracy = \frac{A + D}{A + B + C + D} \tag{9}$$

Table 3 High and low power prediction accuracy of RF

| Training and Testing Data Splitting ratio (80:20) | | | | 10 Fold Cross validation | | | |
|---|---|---|---|---|---|---|---|
| High power utilization Zones Historical statistics | | | | High Power Utilization Zones RF Prediction Accuracy | | | |
| Total Zones | Correctly Predicted | Incorrectly Predicted | Accuracy | Total Zones | Correctly Predicted | Incorrectly Predicted | Accuracy |
| 80 | 77 | 3 | 95.71% | 290 | 270 | 20 | 92.85% |
| High power utilization Zones Historical statistics | | | | High Power Utilization Zones RF Prediction Accuracy | | | |
| Total Zones | Correctly Predicted | Incorrectly Predicted | Accuracy | Total Zones | Correctly Predicted | Incorrectly Predicted | Accuracy |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2022 doi:10.20944/preprints202201.0138.v1

| 60 | 55 | 5 | 90.00% | 130 | 121 | 9 | 92.50% |
|---|---|---|---|---|---|---|---|
| High power utilization Zones Historical statistics | | | | High Power Utilization Zones RF Prediction Accuracy | | | |
| Total Zones | Correctly Predicted | Incorrectly Predicted | Accuracy | Total Zones | Correctly Predicted | Incorrectly Predicted | Accuracy |
| 140 | 132 | 6 | 95.00% | 420 | 391 | 29 | 92.75% |

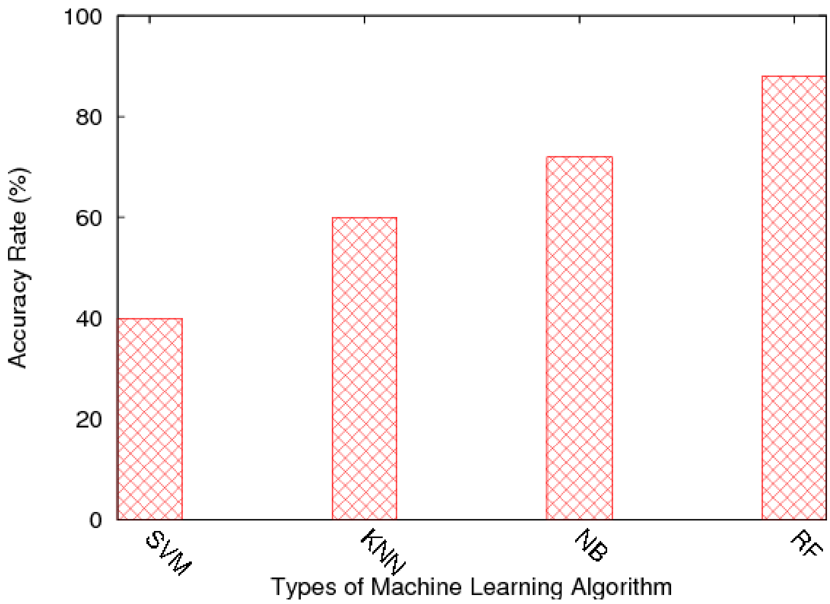


**Figure 2.** Machine Learning Algorithm Based Power Availability Verification

The accuracy calculation is discussed in table 3 is followed for other classification also. Figure 2 shows the accuracy rate of different ML algorithms. The horizontal axis shows the type of algorithm, and the vertical axis shows each algorithm's accuracy rate. It clearly shows that the proposed RF based algorithm produced the highest accuracy rate than the other ML algorithms like SVM, K-nearest neighbor (KNN) and Naïve Bayes (NB). The accuracy rate and training time of SVM depends on data size. When a large sized data is considered of evaluation the training time is also increased. Likewise, the accuracy rate of SVM depends on the hyperplane equation values. Similarly, the KNN classification accuracy depends on the K-Value. The accuracy rate of RF depends on majority voting of sub-tree values, and hence, the accuracy rate is higher than other ML algorithms. Thus, the proposed RF based SG power sustainability algorithm produced better results is proven.

 doi:10.20944/preprints202201.0138.v1

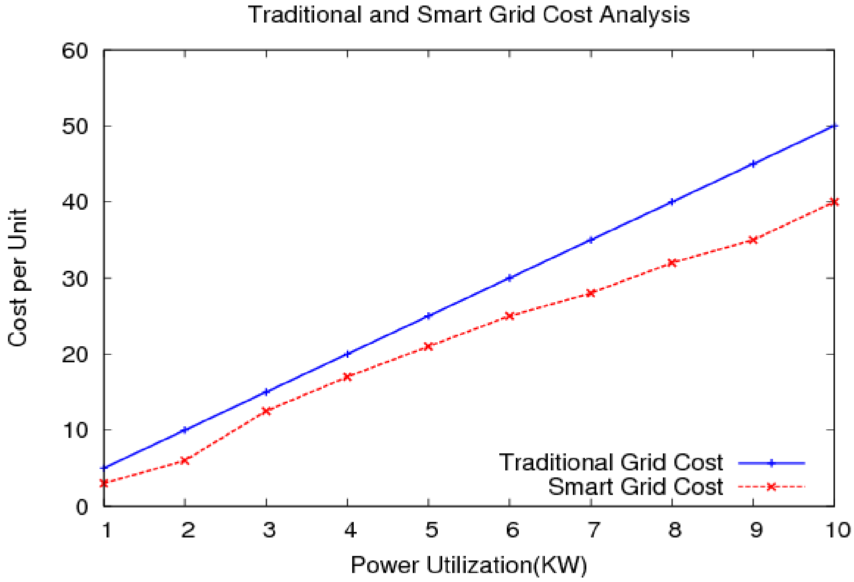


**Figure 3.** Cost analysis of the proposed system and traditional system

Figure 3 shows the cost analysis of the proposed RF based SG optimization technique with a traditional grid cost. The SG technique calculates the power utilization accurately and based on this utilization, the cost is also calculated accurately by using SM. The following code is used to measure the power consumption of each zone in a specific time period.

```
Set_global_assignment –name POWER_ESTIMATION_START_TIME "<start_time>"
Set_global_assignment –name POWER_ESTIMATION_END_TIME "<end_time>"
```

Power consumption is measured in three different timings like working day, holiday and night time. Working hour's power consumption is larger than the night and holiday. The SM dynamically measures the power utilization. Thus, the digital meter is used for measuring the utilization every few seconds. So, the accuracy rate is higher than the traditional analog grid. Hence, the price of SG is lesser than the traditional price analysis. Table 4 shows the power consumption measurement of the proposed system and traditional system. In general, based on the power utilization range, the price per unit is varied such that the price of the power utilization also varies from one user to another. Depending on the power request, the minimum power demand is fixed as a threshold value. Based on historical data analysis, the minimum power requirement for the individual consumer is 20KW. Hence in the proposed technique, the minimum power allocated to the specific user is equal to 20KW is fixed as the threshold value. If the power utilization is above 20 KW, then the price of utilization will be increased by 10 percentages in a traditional method. Whereas, in the proposed technique, the power utilization rate is measured accurately. Hence, the rate of power utilization is reduced in the proposed technique.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2022 doi:10.20944/preprints202201.0138.v1

**Table 2.** Power consumption measuring of SM and Traditional Grid

| Time (seconds) | Proposed SM (watts) | Traditional Grid (watts) |
|---|---|---|
| 1 | 8.12 | 9 |
| 2 | 12.20 | 13 |
| 3 | 18.52 | 19 |
| 4 | 19.10 | 20 |
| 5 | 20.15 | 21 |
| 6 | 29.67 | 30 |
| 7 | 31.89 | 32 |
| 8 | 38.22 | 39 |
| 9 | 42.2 | 43 |
| 10 | 42.8 | 43 |

Figure 4 shows the power utilization region analysis such as high power utilization region and low power utilization region. It shows that the power distribution range between different priority based requirement such as how much power utilized in a high priority range and how much power utilized in a low priorty range based on the threshold value. If the power utilization rate is higher than the threshold value, then it is identified as the high power utilized zone; otherwise based on power utilization rate, the power zones are identified. The threshold value for the proposed technique is fixed as 20 KW. If the utlization rate is higher than 20 KW, then that zone is identified as higher power utilization zone. Similarly, the other zones are also identified. Through this analysis, the producer easily identifies whether the consumer able to access the power in future or not. This process reduces the decision making time of power distribution. Based on this analysis the power utilization is clearly identified.

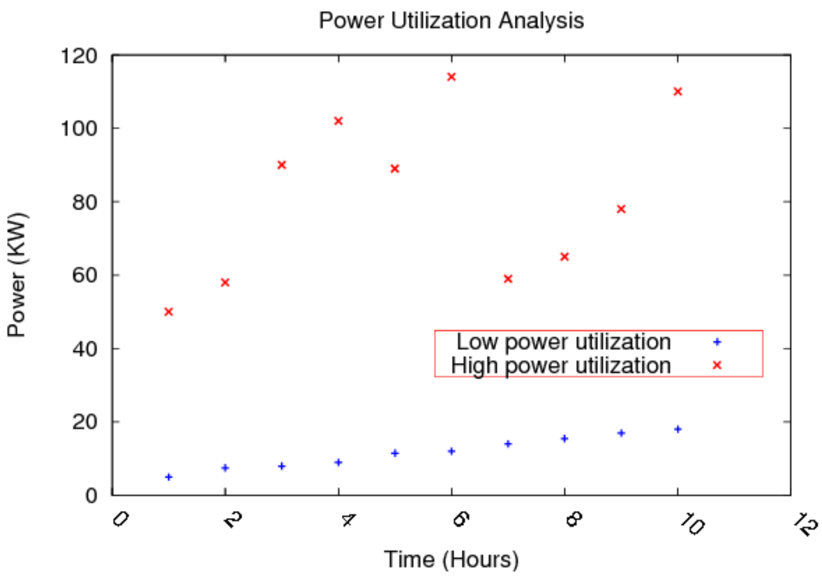


**Figure 4.** Power Utilization analysis

Figure 5 (a), (b), (c) and (d) shows the comparison between the ML algorithms such as SVM, KNN, NB and the proposed RF respectively. The horizontal axis shows the power availability for a particular time in hours and vertical axis shows the power availability in KW. The proposed RF algorithm clearly identifies the power availability for a particular time period with the help of SM. The power availability is compared to consumer request and allocates the requested power if the availability is greater than the demand. Otherwise the priority is analyzed. The power availability which is calculated

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2022 doi:10.20944/preprints202201.0138.v1

by the proposed RF algorithm at different time is shown in Figure 5 (d). It is found that the power availability at the 8th hour is 500 KW which is obtained by the proposed RF method. At the same time, other methods such as SVM, KNN and NB yields 168KW, 447KW and 176 KW respectively. It is proven that the maximum power availability is obtained by the proposed RF method. At time 8 hours, 40% of total power request and 95% of high priority power request are satisfied using RF based method, but SVM, KNN and NB satisfies the 33.6%, 34.2% and 35.2% of total power request and 85%, 89% and 92% of high priority power request at the same mentioned time. Similarly, in the other timings, the proposed RF based method satisfies the high priority power request at optimal time compared to SVM, KNN and NB.

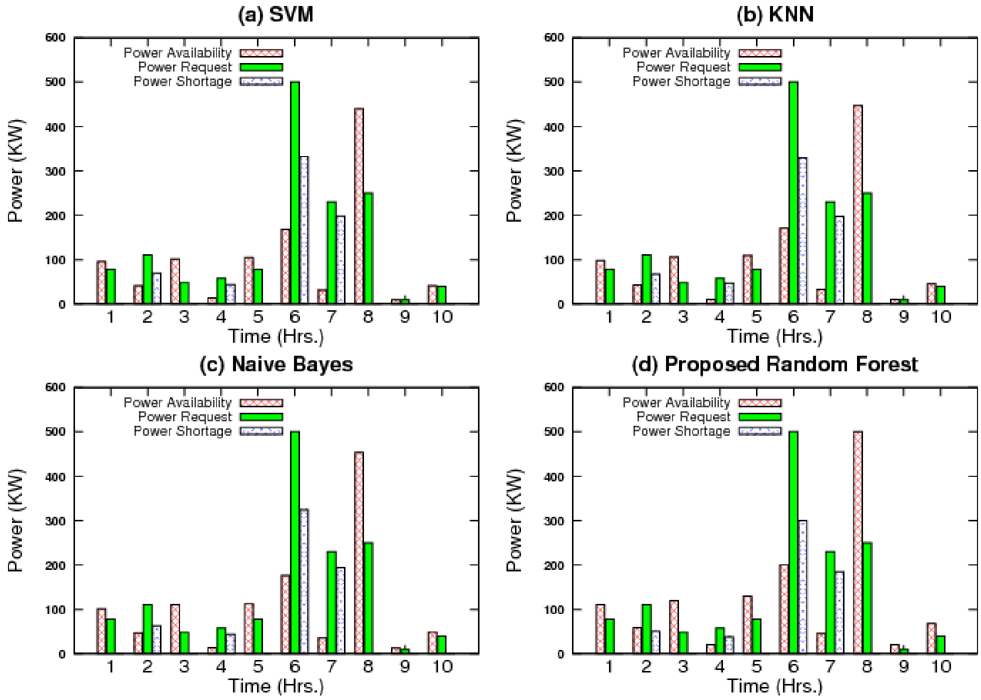


**Figure 5.** Power Request and distribution analysis

If the request falls under $H_P$, then the request is allocated immediately by stopping the $N_P$ process. Otherwise the power will be distributed in future. Based this analysis, the power distribution is performend in an efficient way.

Figure 6 shows the fault identification time between traditional and SM based SG. The SM dynamically identifies the fault location and uses the GSM module to communicate the location of fault and producer. After identifying the fault, the alarm message is immediately sent to the producer without delay through the GSM module. The alarm message is being continuously sent until the producer responds. In the proposed technique, the fault is identified with the predefined threshold value by using hidden markov model. The general fault types like load loss, generator outage, generator ground, single phase transmission line outage and thress phase transmission line outage are identified. In these type of faults, loasd loss, generator ground and generator outages are not change the SG topology structure but, the single and three phase faults are change the topology. When a power flow deviations or topology change occurs, the current power flow is compared with the threshold value and topology structure is compared to existing topology structure. If any deviations occur, the SM checks each device functioning for find the abnormality. After identifying the fault, the alarm message is transferred to the producer and the device details to the producer. Thus the proposed technique dynamically identifies the power deviation. Hence, the identification time is very less than the traditional method. Through this analysis, the proposed SG based optimization technique with lesser fault identification time is proven.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2022 doi:10.20944/preprints202201.0138.v1

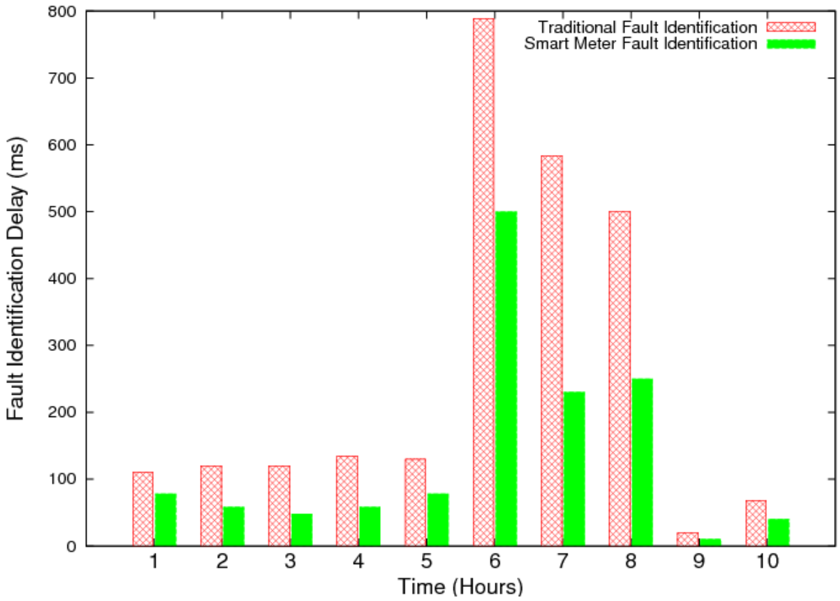


**Figure 6**. Fault Identification Time

## 4. Discussion

The proposed Random Forest based power sustainability and cost optimization technique in smart grid provides better results in different aspects. In the first aspect, the low and high-power utilization zones are identified for efficient power distribution to power requester with the help of power scheduling algorithm. The power scheduling algorithm analyses the type of power request and distributes the power to the requester based on the power availability. Thus, the requester demand is satisfied based on the power availability. Through this task, the power sustainability is maintained for all requesters. In the second aspect, the cost optimization is achieved through smart metering technique. When compared to traditional cost calculation, the proposed technique measures the power consumption accurately. Based on this process, the exact cost is calculated on the consumed power. Thus, the cost optimization is achieved in the proposed technique. Finally, the fault is identified at the earliest. When a fault identification and resolution is rectified in an earliest manner, it increases the power production and distribution. Thus, the power sustainability and cost optimization are achieved.

Towards the power availability, the proposed random forest-based algorithm performs better compared to the other ML algorithms such as SVM, KNN and NB due to the majority voting of sub-tree values. The power utilization is analyzed region wise towards high and low utilization areas by which the producer can be able to easily identify the consumers' requirement. This further reduces the time to take decision during power distribution. The RF based algorithm is very much useful towards calculating the power availability by which the power is allocated.

False identification is essential in SG through which the producer could be able to monitor the faults to take necessary action. This is performed by the SM which monitors the device details continuously. When compared to the traditional system, the time required for identifying the fault is very less. The results of the proposed method are compared with the traditional grid system.

Based on the above-mentioned analysis, the proposed RF based power sustainability and cost optimization in SG technique is used to identify the power availability for future, cost analysis between traditional and SG technique and power utilization region identifications. Through this data the power distribution is performed in an efficient manner with lesser control and maintenance cost to the requestor without increasing the infrastructure cost. Hence, the proposed technique provides better results than traditional grid and ML algorithm is proved by an efficient power distribution process.

## 5. Conclusions and Future work

The proposed Random Forest based Smart Grid optimization technique is used to manage the power utilization in an efficient manner when compared to the existing power sustainability techniques. The existing power sustainability techniques do not focus on future power demand and dynamic control and monitoring. Similarly, the existing machine learning algorithms like support vector machine, naïve based technique, *k*-nearest neighbor techniques are unable to predict the power availability accurately. Additionally, the traditional power management technique was unable to find the exact power availability, and also unable to schedule the power request in a dynamic manner. In the traditional technique-based power utilization, the cost is higher than the current estimation and fault identification time is also slow. To overcome these issues, the random forest based smart grid optimization technique in smart grid is proposed in this work. In the proposed technique, the consumer request is transferred to priority power scheduling algorithm for allocating the power based on requester priority demand. Afterwards, the request is transferred to producer for satisfying the request. Now, the request is verified in the smart grid to check the power availability. If power availability is high and request is low, immediately the power distribution is performed; otherwise, the system checks the availability through random forest-based classification process. These classification processes identify the low power utilization zone for availing the unutilized power on a specific time and distribute to producer. Based on this process the power utilization is managed in an efficient manner. The smart meter-based billing system produces the exact cost of the utilized power. When compared to traditional cost calculation technique, the proposed cost identification technique takes lesser cost. Similarly, in the proposed technique the resource faults are identified in an early stage by the GSM module. This GSM module sends the fault device details along with location to producer for quicker response. Thus, the proposed system reduced power utilization cost; improve power utilization in a better way.

In future work, the proposed random forest technique can be integrated with the convolution neural network for improving the prediction accuracy with micro zone identification in an exact manner for reducing the power wastage in a greater ratio. Similarly, the proposed technique identifies only the location of fault. In future, the type of faults in a specific device can be identified using IoT devices.

**Author Contributions:** Concepts and methodology of the work in this paper is done by DD. Third author AS has done the work related to validation. Literature review and formal analysis are done by second and fourth author. Draft preparation, editing and visualization are done by fifth author.

**Data Availability Statement:** In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. Please refer to suggested Data Availability Statements in section "MDPI Research Data Policies" at https://www.mdpi.com/ethics. You might choose to exclude this statement if the study did not report any data.

**Conflicts of Interest:** "The authors declare no conflict of interest."

## References

1. Jamil, F.; Iqbal, N.; Imran; Ahmad, S.; Kim, D. Peer-to-Peer Energy Trading Mechanism Based on Blockchain and Machine Learning for Sustainable Electrical Power Supply in Smart Grid. *IEEE Access* **2021**, *9*, doi:10.1109/ACCESS.2021.3060457.
2. Smirnova, E.; Kot, S.; Kolpak, E.; Shestak, V. Governmental support and renewable energy production: A cross-country review. *Energy* **2021**, *230*, doi:10.1016/j.energy.2021.120903.
3. Haseeb, M.; Kot, S.; Iqbal Hussain, H.; Kamarudin, F. The natural resources curse-economic growth hypotheses: Quantile–on–Quantile evidence from top Asian economies. *J. Clean. Prod.* **2021**, *279*, doi:10.1016/j.jclepro.2020.123596.
4. Smirnova, E.; Szczepańska-Woszczyna, K.; Yessetova, S.; Samusenkov, V.; Rogulin, R. Supplying energy to vulnerable segments of the population: Macro-financial risks and public welfare. *Energies* **2021**, *14*, doi:10.3390/en14071834.
5. Sabishchenko, O.; Rębilas, R.; Sczygiol, N.; Urbański, M. Ukraine energy sector management using hybrid renewable

energy systems. *Energies* **2020**, *13*, doi:10.3390/en13071776.

6. Wall, W.P.; Khalid, B.; Urbański, M.; Kot, M. Factors influencing consumer's adoption of renewable energy. *Energies* **2021**, *14*, doi:10.3390/en14175420.
7. van Kooten, G.C. Economic analysis of feed- in tariffs for generating electricity from renewable energy sources. In *Handbook on Energy and Climate Change*; 2013.
8. Danalakshmi, D.; Prathiba, S.; Ettappan, M.; Krishna, D.M. Reparation of voltage disturbance using PR controller-based DVR in Modern power systems. *Prod. Eng. Arch.* **2021**, *27*, doi:10.30657/pea.2021.27.3.
9. Nawaz, R.; Akhtar, R.; Shahid, M.A.; Qureshi, I.M.; Mahmood, M.H. Machine learning based false data injection in smart grid. *Int. J. Electr. Power Energy Syst.* **2021**, *130*, doi:10.1016/j.ijepes.2021.106819.
10. Dharmadhikari, S.C.; Gampala, V.; Rao, C.M.; Khasim, S.; Jain, S.; Bhaskaran, R. A smart grid incorporated with ML and IoT for a secure management system. *Microprocess. Microsyst.* **2021**, *83*, doi:10.1016/j.micpro.2021.103954.
11. Awan, N.; Khan, S.; Rahmani, M.K.I.; Tahir, M.; Alam, N.M.D.; Alturki, R.; Ullah, I. Machine Learning-Enabled Power Scheduling in IoT-Based Smart Cities. *Comput. Mater. Contin.* **2021**, *67*, doi:10.32604/cmc.2021.014386.
12. Omitaomu, O.A.; Niu, H. Artificial Intelligence Techniques in Smart Grid: A Survey. *Smart Cities* **2021**, *4*, doi:10.3390/smartcities4020029.
13. Szkutnik, J.; Jakubiak, D. New trends in consumption management of electric energy. *Polish J. Manag. Stud.* **2012**, *5*.
14. Mikita, M.; Kolcun, M.; Špes, M.; Vojtek, M.; Ivančák, M. Impact ff Electrical Power Load Time Management at Sizing And Cost of Hybrid Renewable Power System. *Polish J. Manag. Stud.* **2017**, *15*, doi:10.17512/pjms.2017.15.1.15.
15. Ungureanu, S.; Topa, V.; Cziker, A. Industrial load forecasting using machine learning in the context of smart grid. In Proceedings of the 2019 54th International Universities Power Engineering Conference, UPEC 2019 - Proceedings; 2019.
16. Mukherjee, A.; Mukherjee, P.; Dey, N.; De, D.; Panigrahi, B.K. Lightweight sustainable intelligent load forecasting platform for smart grid applications. *Sustain. Comput. Informatics Syst.* **2020**, *25*, doi:10.1016/j.suscom.2019.100356.
17. Taherian, H.; Aghaebrahimi, M.R.; Baringo, L.; Goldani, S.R. Optimal dynamic pricing for an electricity retailer in the price-responsive environment of smart grid. *Int. J. Electr. Power Energy Syst.* **2021**, *130*, doi:10.1016/j.ijepes.2021.107004.
18. Sharmila, P.; Baskaran, J.; Nayanatara, C.; Maheswari, R. A hybrid technique of machine learning and data analytics for optimized distribution of renewable energy resources targeting smart energy management. In Proceedings of the Procedia Computer Science; 2019; Vol. 165.
19. Renugadevi, N.; Saravanan, S.; Naga Sudha, C.M. IoT based smart energy grid for sustainable cites. *Mater. Today Proc.* **2021**, doi:10.1016/j.matpr.2021.02.270.
20. Pallonetto, F.; De Rosa, M.; Milano, F.; Finn, D.P. Demand response algorithms for smart-grid ready residential buildings using machine learning models. *Appl. Energy* **2019**, *239*, doi:10.1016/j.apenergy.2019.02.020.
21. Zekić-Sušac, M.; Mitrović, S.; Has, A. Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities. *Int. J. Inf. Manage.* **2021**, *58*, doi:10.1016/j.ijinfomgt.2020.102074.
22. Zou, H.; Tao, J.; Elsayed, S.K.; Elattar, E.E.; Almalaq, A.; Mohamed, M.A. Stochastic multi-carrier energy management in the smart islands using reinforcement learning and unscented transform. *Int. J. Electr. Power Energy Syst.* **2021**, *130*, doi:10.1016/j.ijepes.2021.106988.
23. Ahmad, T.; Chen, H. Potential of three variant machine-learning models for forecasting district level medium-term and long-term energy demand in smart grid environment. *Energy* **2018**, *160*, doi:10.1016/j.energy.2018.07.084.
24. Parvez, I.; Aghili, M.; Sarwat, A.I.; Rahman, S.; Alam, F. Online power quality disturbance detection by support vector machine in smart meter. *J. Mod. Power Syst. Clean Energy* **2019**, *7*, doi:10.1007/s40565-018-0488-z.
25. De Santis, E.; Rizzi, A.; Sadeghian, A. A cluster-based dissimilarity learning approach for localized fault classification in Smart Grids. *Swarm Evol. Comput.* **2018**, *39*, doi:10.1016/j.swevo.2017.10.007.
26. Li, D.; Jayaweera, S.K. Machine-Learning Aided Optimal Customer Decisions for an Interactive Smart Grid. *IEEE Syst. J.* **2015**, *9*, doi:10.1109/JSYST.2014.2334637.

 doi:10.20944/preprints202201.0138.v1


27. Ali, S.; Wu, K.; Weston, K.; Marinakis, D. A Machine Learning Approach to Meter Placement for Power Quality Estimation in Smart Grid. *IEEE Trans. Smart Grid* **2016**, *7*, doi:10.1109/TSG.2015.2442837.
28. Azad, S.; Sabrina, F.; Wasimi, S. Transformation of smart grid using machine learning. In Proceedings of the 2019 29th Australasian Universities Power Engineering Conference, AUPEC 2019; 2019.
29. Ullah, Z.; Al-Turjman, F.; Mostarda, L.; Gagliardi, R. Applications of Artificial Intelligence and Machine learning in smart cities. *Comput. Commun.* 2020, *154*.
30. Muralitharan, K.; Sakthivel, R.; Vishnuvarthan, R. Neural network based optimization approach for energy demand prediction in smart grid. *Neurocomputing* **2018**, *273*, doi:10.1016/j.neucom.2017.08.017.
31. Mohamed, M.A.; Eltamaly, A.M.; Farh, H.M.; Alolah, A.I. Energy management and renewable energy integration in smart grid system. In Proceedings of the International Conference on Smart Energy Grid Engineering, SEGE 2015; 2015.
32. Perera, K.S.; Aung, Z.; Woon, W.L. Machine Learning Techniques for Supporting Renewable Energy Generation and Integration: A Survey. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **2014**, *8817*, doi:10.1007/978-3-319-13290-7_7.