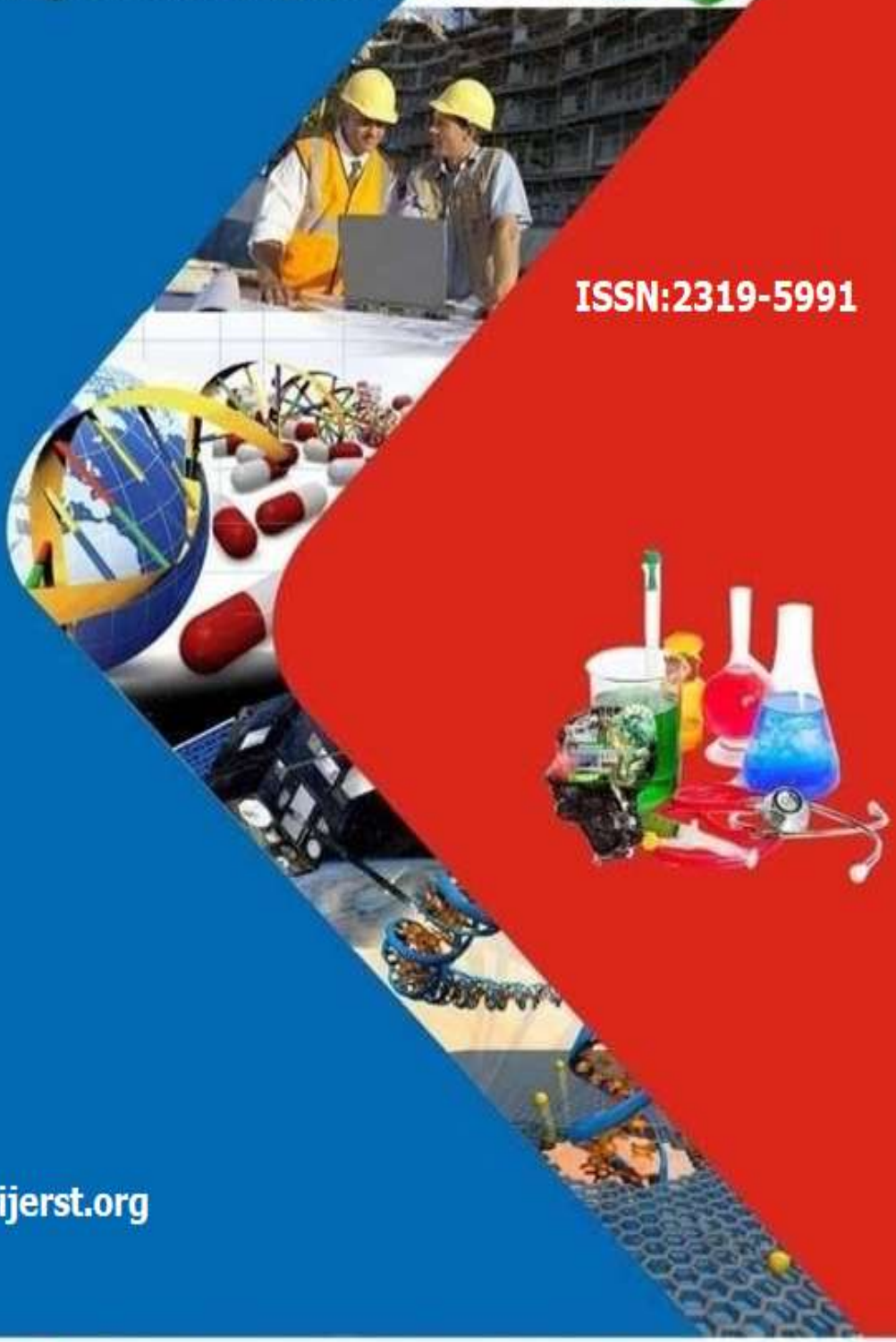
International Journal of Engineering Research and Science & Technology

IJERST

ISSN:2319-5991



www.ijerst.org

E-mail: editor@ijerst.org or ijerst.editor@gmail.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

IOT BASED SMART METER DATA ANALYSIS USING MACHINE LEARNING CLASSIFIERS FOR ENERGY CONSUMPTION PREDICTION

Mr. S Sundeep[1], B.Yashwika[2], B.Kiran Kumar[2], R.Satya Dharma Teja[2], K.Shreyas[2]

[1]Assistant Professor, [2]UG Student, [1,2] Department of Computer Science and Engineering(DS)

Sree Dattha Group of Institutions, Sheriguda, Hyderabad, Telangana

ABSTRACT

The emergence of Internet of Things (IoT) technology has revolutionized various sectors, including the energy industry. In particular, IoT-based smart meters have gained significant attention due to their ability to provide real-time data on energy consumption. These smart meters offer a plethora of data, which can be leveraged to optimize energy usage, improve efficiency, and reduce costs. One promising avenue is the application of machine learning techniques to analyze smart meter data and predict energy consumption patterns accurately. Traditionally, energy consumption monitoring relied on manual meter readings, which were often infrequent and prone to human errors. This approach limited the ability to obtain timely insights into energy usage patterns, leading to inefficiencies in energy management. Moreover, traditional systems lacked the capability to adapt to changing consumption patterns or provide proactive suggestions for optimization. The challenge is the sheer volume and complexity of the data generated by IoT devices, which often requires sophisticated analytical techniques to extract meaningful insights. Additionally, traditional regression-based approaches may not fully capture the nonlinear relationships and temporal dynamics present in energy consumption data. The system aims to address these challenges by leveraging machine learning classifiers for energy consumption prediction based on IoT smart meter data. By integrating advanced analytics with IoT technology, the proposed system offers several key advantages. Firstly, machine learning algorithms can capture complex patterns and relationships within the data, leading to more accurate consumption predictions. Secondly, real-time analysis enables proactive decision-making and optimization strategies. Thirdly, by employing techniques such as anomaly detection, the system can identify and mitigate potential issues, thereby enhancing efficiency and reliability.

Keywords: Smart Meters, Energy Consumption, Machine Learning, Real-time Data, Consumption Prediction, Anomaly Detection, Energy Optimization, Data Analytics, Nonlinear Patterns.

1.INTRODUCTION

1.1 History

The history of energy consumption monitoring traces back to the early days of electricity distribution when manual meter readings were the norm. These readings, performed by utility personnel, were infrequent and prone to errors, leading to inefficiencies in energy management. Over time, advancements in technology led to the introduction of automated meter reading (AMR) systems, which improved data collection but still suffered from limitations such as lack of real-time monitoring capabilities.

The emergence of IoT technology marked a significant turning point in energy monitoring and management. IoT-based smart meters, equipped with sensors and communication capabilities, revolutionized the way energy consumption data was collected and analyzed. These smart meters

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

provided real-time insights into energy usage patterns, enabling both consumers and energy providers to make informed decisions about energy consumption and distribution.

As IoT technology continued to evolve, the capabilities of smart meters expanded exponentially. Today, smart meters not only provide real-time data on energy consumption but also offer a plethora of additional information, such as voltage levels, power quality, and grid stability. This wealth of data has opened up new possibilities for optimizing energy usage, improving efficiency, and reducing costs through advanced analytics and machine learning techniques.

2.LITERATURE SURVEY

Electric Power is the most flexible and broadly utilized type of vitality and worldwide request is developing persistently. In present day life all individuals are customer of Electric power. Electricity can be utilized to feel comfort at home, to cool, to warm, light them, wash garments, cook to eat, to engage and different purposes Currently electric energy distribution and deployment with in smart environment fairly and intelligently faces different challenges. Forecasting customer's electric energy consumption manages and handle challenges that results from currently unbalanced distribution of smart electric energy. Forecasting the electricity consumption by applying different machine learning mechanisms and models is the best approach to save energy as well as economy. Accurate forecasting will empower utility suppliers to design extra assets and furthermore take control activities to adjust the electricity supply and demand. Forecasting electric utilization is an imperative assignment to give insight to smart grid. It includes prediction of maximum power usage of appliance, peak demand and customers level of life style. Different researchers tried to address challenges of current electric consumption forecasting systems in which some of them are described below. Lines, Jason, et al [5] Grouping of household unit devices by power utilization profiles by utilizing 15 minute settled time interim of smart meter information, order every household device. Smart meter following and recording of the entire homes of utilization and transmit back to the organizations. Effectively see every device utilization detail and in view of that how to arranging the household device was the point of the paper. The main contribution of the paper was tending to time serious classification issue through determining arrangement of highlights that portray the example of utilization and the measure of energy utilized when a device is on. The capacity to consequently recognize the kind of appliance gives experiences into the breakdown of the household unit utilization design and get the possibility for giving functional reaction of the consumer ,both As far as minimizing their utilization and done issue identification. For grouping correctness they posed applying diverse machine learning classification like k-Nearest neighbour, Support vector machine, and C4.5 and Random Forest. At last the outcome demonstrated that with a week after week profile they could precisely separate between classes of device set of enlightening highlights and random forest or nearest neighbour classifier were a critical systems for precisely give week after week utilization detail however irregular random forest was the better machine learning algorithm by inferring with exactness of over 80% for chilly and 72% for screen amass dataset. Ning Lu, et al [6] shown breaking down information qualities, removing key information signature and incorporating data of 15 minute information from every single pertinent datum source. The point of the paper was to recognize conceivable information recommendation from smart meter estimation to infer a key information marks for different target applications. Autocorrelation and cross-correlation are procedures used to extricate these key information marks to portray the typicality and variation from the norm use of electric power. The examined result demonstrates that an information signature database can be worked for various time determination informational collections for the smart meter information data management system; at that point various applications can get to the information marks to recognize anomalies in smart operation,

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

customers vitality utilizations, and grid control and correspondence systems . Adrian Albert et al [7] Smart meter driven division: What your utilization says in regards to you. The paper contribute in various angles whether the individual worldly utilization is sensibly depicted, contrast and group client agreeing with their comparability design and use the derived qualities of utilization to anticipate exogenous attributes (socioeconomic, apparatus stock, and so forth.). As per the investigation arranging customers in view of their closeness was difficult, however they proposed an information driven system that permits predicting user qualities utilizing using only consumption attributes as gathered from the utilization time arrangement. Hidden Markov Model (HMM) and another clustering methods unearthly HMM method to section user's way of life and properties by utilizing information from clients Smart meter information, customers utilization data and their statistic, family unit, and apparatus stock qualities. The outcome uncovers that there is both a lot of consistency in use's state arrangement and impressive variety for gathering certain way of life and appliance stock attributes. Dynamic model of the time arrangement as caught by HMM examination can serve as significant prompts in directing custom fitted mediation. Tomas Ząbkowskie et al. [8] Short term electricity power forecasting utilizing singular smart meter information. The paper points in exact expectation short term electric power stack utilization with twenty four hours on singular smart meter information from different machine learning techniques connected for estimating for past work at long last they chose utilizing MLP neural system and Support Vector Machine (SVM) on 24 hours customers smart meter information to estimate individuals short term power utilization. The outcome was both machine learning approaches effectively perform great forecast with slightest blunder and satisfactory exactness. Wei Yu et al. [9] examined Statistical Displaying and Machine Learning Based energy Use forecasting in smart grid. The paper exhibited detail of smart grid vitality utilization estimating used to adjust request supply and compelling administration and control of vitality in the smart grid. Creating measurable demonstrating examination to infer factual vitality appropriation vitality use. Estimating the vitality in smart grid is essential thing for service organizations for arranging extra vitality source, producing new vitality. Generally the Shapiro Wilk test and Quantile-Quantile plot typicality test were connected to explore the measurable dispersion of vitality utilization and the machine learning based methodologies. Those machine learning based methodologies incorporate standard Radial Basis Function (RBF), support vector machine (SVM, the Least Squares (LS) based SVM, and the Backward Propagation Neural Network (BPNN) were created to lead the precise forecasting of vitality utilization. The effectiveness of the created approaches was approved through real world informational index. At long last their result demonstrates that the vitality utilization can be to a great extent approximated with a Gaussian distribution and the SVM based machine learning methodologies could precisely determining the vitality use. Sudha Gupta et al [10] support Vector Machine Based Proactive Cascade Prediction in Smart Grid Using Probabilistic Framework. The principle commitment of the examination was to catch the pith of the falling disappointment utilizing probabilistic structure and combination of SVM machine learning tool. To build an expectation administer rule which would have the capacity to predict the situations of the power outage as right on time as could be allowed. The proposed model was confirmed by applying the IEEE 30 transport test-bed framework. Finally prediction of cascade disappointment using SVM machine learning approach empowered to predict cascade preceding before huge power outages happen that utilized for proactive cascade prescient in arranging and maintenance of smart grid early cautioning framework.

Kui Wu et al [11] A Machine Learning Approach to Meter Placement for Power Quality Estimation in Smart Grid is another investigation made concerning electric power consumption. The point of paper was lessening the cost of PQ observing in power network. The fundamental contribution of the examination was a system demonstrate for PQ estimation, in view of the device inactive highlights

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

that are found out from a certifiable data set, a savvy entropy-based methods and a Bayesian Network (BN)-based way to deal with tackle the meter situation issue. Nonetheless, it is accounted for that it was not a simple undertaking, since the power quality estimation devices are expensive and fiscally hard to screen all aspects of energy organize. They portrayed a BN-based algorithm for choosing areas for setting power meters in a power network. The approach uses Monte Carlo (MC) examining and probabilistic inference approaches used to recognize areas in the power grid which displayed unpredictable PQ occasions. At long last, the proposed arrangements fundamentally settled the vulnerability of PQ values on monitored control joins. Md. Sumon Shahriar et al [12] Urban Sensing and Smart Home Energy Optimization is another machine learning approach research work. The objective of the work was integrating machine learning techniques with data from different urban sensing sensors to exhibit that energy optimization application for smart home. Optimizing the energy obtained from various sources from different urban sensing sensors for effective utilization of energy in smart home was one of the aspects of the researchers to save both energy and economy. M5P algorithms is the machine learning algorithms integrated with the collected dataset for prediction of solar energy for efficient energy management. The outcome of the study shows that combining machine learning algorithms with public data resulted in improving the classification accuracy of traditional models. The study indicates that integrating power features with additional context feature of public data provides 96% of classification accuracy. David Walkera, et al [13] examined forecasting household water use with information gathered amid the progressing water venture. Determining household water utilization from consumer's water smart meter information was the objective of the investigation. As indicated by the paper forecasting the genuine use of water was used to determine the deficiency of water asset and furthermore used for spillage identification. The ultimate result of the work portrayed two technique a manufactured neural system (Artificial Neural Network) to predict whenever ventures of water use. In next strategy models were prepared using an evolutionary algorithm, with characterizations decided experimentally. The paper additionally acknowledged customers can deal with their assets and economy adequately if appropriate forecasting of water asset is applied. Joseph Siryani, et al [14] System utilizing Bayesian Belief Networks for Utility Effective Management and Operations utilizing Bayesian Belief Networks is an investigation made on utility administration and operation with help of machine learning procedures. They attempted to show nonspecific predictive examination outline work for choice of effective utility administration operation. Diminishing the utility cost during operation, support and enhancing the general framework administration and operational proficiency, execution and consumer loyalty was the primary objectives of the examination work. Damminda et al [15] additionally outlined smart electric power meter information knowledge for future vitality frameworks. They introduced exhaustive study of smart electric power meters and their usage concentrating on key parts of metering process, the distinctive partner interests and technologies used to fulfil partner interests. Most generally used metering insight exercises, key difficulties and fate of smart metering was profoundly talked about. Machine learning methodologies of support vector machine, principal component analysis (PCA) and fuzzy rationale techniques were likewise in located in similar approach.

Stephen Haben et al [16] Examination and clustering of residential customers vitality behavioural request using smart meter information is additionally one of the contributors of current research area. The paper experiences profound investigation of customers smart meter information to decide the peak demand and major source of inconstancy in their conduct. The information was analysed in how it is arranged into four time arrangement and 10 unmistakable behavioural of client in light of their request and source of changeability were found by Finite Mixture Model clustering method. The

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

model was at long last assessed using existing bootstrapping methods and the outcome uncovered the grouping was done in the sensible way. Samuel Idowu et al. [17] applied machine learning: Forecasting heat stack in district heating framework. Information driven approach determining and investigation of both space and water warm vitality use of DHS (water and space warmer around 90% of Sweden people groups using this innovation) was the commitment of the examination. The load figure models were produced using supervised machine learning procedures, specifically, support vector machine, regression tree, feed forward neural system, and different linear regression. They examine was used four key properties the open-air temperature, historical values of heat load, time factor factors and physical parameters of region warming substations as its input. The forecasts models were delivered and assessed using information observed from 10 area warming substations serving five multi-family apartment and five business structures. The models are assessed with shifting estimate horizons of consistently from1up to 48h.The applied techniques demonstrate that support vector machine, feed forward neural system and multiple linear regression were more appropriate machine learning strategies with bring down execution error than the regression tree. Among those machine learning techniques Support vector machine approach demonstrates better than average expectation result minimum standardized root mean square mistake of 0.07 for a forecast horizon of 24h.By understanding that each building has its own particular use design conduct notwithstanding working with similar classes. Thierry Zufferey et al [18] Among the current investigations determining of Smart Meter Time Series in view of Neural Networks (NN) utilizing Support Vector Machine (SVM) or more refined ANNs like Recurrent Neural Network (RNN) and Long Short-Term Memory estimating utility utilization from different sorts smart meter profile. Utilizing single smart meter information or individual is hard to predict and couldn't give successful outcome. Then again by aggregating diverse smart meter information and by dividing in light of time and ANN the work is finished. Standardized Root Mean Square Error (NRMSE) and the Mean Absolute Percentage Error (MAPE) are strategies used to assess the execution of forecasting algorithm. The last outcome uncovers that enhanced exactness has been acquired from commercial and industrial loads rather than individual meter data. Martinez-Pabon et al.[19] smart meter information examination for excellent customer determination demand response programs .Selecting customer for demand response program is one parts of concentrate in which smart meter information is investigated in light of customers utilization information and way of life choosing customers which who was qualified for Demand reaction program was the point of paper. Smart Meter Data Analytics for Optimal Customer Selection in Demand Response Programs is among such examinations done in late time. These have been utilized for enhancing age limit and search for different options. Specifically, when appeal is available, the cost of power is higher than amid off-top hours. Utilities encourage clients to move their utilization examples to low-peak hours and advantage from the motivations. The Paper was profoundly contributes predicting qualification to take an interest in DR programs utilizing load utilization attributes of customer. This approach was unique in relation to other such huge numbers of studies. This was the first occasion when they were connected R-programming language for smart meter information investigation. What's more, utilized both order and grouping systems for effectively enlisted customers taking an interest demand response program those techniques were an agglomerative progressive clustering use for determination of the correct number of groups. And further more for best mix of cluster used K-implies. Besides they connected four distinctive machine learning algorithms (K-Nearest Neighbour, choice tree, counterfeit neural system and arbitrary timberland.). To predict enlistment in DR programs in view of household units' power stack profile shape. Those chose customers would have been prepared keeping in mind the end goal to move their peak hour utilization peak on getting to be peak off by utilizing continuous smart

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

meter information. At long last irregular random forest up being an important strategy for the investigation of brilliant meter information for expansive data sets, with an exactness of 95.1% Joseph Siryani et al [20].

3.PROPOSED SYSTEM

3.1 Overview

The proposed system aims to leverage IoT smart meters and machine learning techniques to optimize energy consumption and improve efficiency. Here's an overview of the project code:

- **Data Collection**: The system collects real-time data from IoT smart meters installed in various locations, such as households or commercial buildings. This data includes information about energy consumption patterns, which is crucial for analysis.
- **Data Preprocessing**: Once the data is collected, it undergoes preprocessing to clean and prepare it for further analysis. This step may involve handling missing values, normalizing the data, and converting it into a suitable format for machine learning algorithms.
- **Machine Learning Model**: The preprocessed data is then fed into machine learning models, such as KNN and Random Forest Model. These models are trained on historical energy consumption data to learn patterns and relationships.
- **Energy Consumption Prediction**: After training, the machine learning models are capable of predicting future energy consumption patterns based on the input data. This prediction is essential for proactive decision-making and optimization strategies.
- **Optimization Strategies**: The system utilizes the predicted energy consumption patterns to implement optimization strategies. These strategies may include load balancing, demand response, and scheduling energy-intensive tasks during off-peak hours to reduce costs and improve efficiency.
- **Anomaly Detection**: Anomaly detection techniques are applied to identify any abnormal energy consumption patterns that deviate significantly from the predicted values. This helps in detecting potential issues, such as equipment malfunction or energy theft, and taking appropriate actions.
- **Decision Making**: Based on the analysis results and anomaly detection, the system makes informed decisions to optimize energy usage and ensure reliable energy supply. These decisions may involve adjusting energy distribution, deploying additional resources, or alerting utility suppliers and consumers about potential issues.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

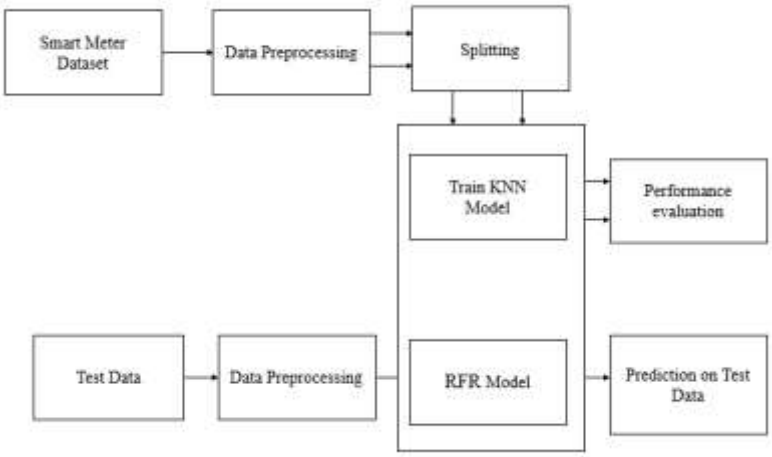


Fig. 1: Block Diagram of Proposed System.

3.2 Data Preprocessing

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So, for this, we use data pre-processing task. A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data pre-processing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

- Getting the dataset
- Importing libraries
- Importing datasets
- Finding Missing Data
- Encoding Categorical Data
- Splitting dataset into training and test set

Importing Libraries: To perform data preprocessing using Python, we need to import some predefined Python libraries. These libraries are used to perform some specific jobs. There are three specific libraries that we will use for data preprocessing, which are:

Numpy: Numpy Python library is used for including any type of mathematical operation in the code. It is the fundamental package for scientific calculation in Python. It also supports to add large, multidimensional arrays and matices. So, in Python, we can import it as:

import numpy as nm

Here we have used nm, which is a short name for Numpy, and it will be used in the whole program.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

Matplotlib: The second library is matplotlib, which is a Python 2D plotting library, and with this library, we need to import a sub-library pyplot. This library is used to plot any type of charts in Python for the code. It will be imported as below:

import matplotlib.pyplot as mpt

Here we have used mpt as a short name for this library.

Pandas: The last library is the Pandas library, which is one of the most famous Python libraries and used for importing and managing the datasets. It is an open-source data manipulation and analysis library. Here, we have used pd as a short name for this library. Consider the below image:

```
# importing libraries
import numpy as nm
import matplotlib.pyplot as mtp
import pandas as pd

```

Handling Missing data: The next step of data preprocessing is to handle missing data in the datasets. If our dataset contains some missing data, then it may create a huge problem for our machine learning model. Hence it is necessary to handle missing values present in the dataset. There are mainly two ways to handle missing data, which are:

- By deleting the particular row: The first way is used to commonly deal with null values. In this way, we just delete the specific row or column which consists of null values. But this way is not so efficient and removing data may lead to loss of information which will not give the accurate output.
- By calculating the mean: In this way, we will calculate the mean of that column or row which contains any missing value and will put it on the place of missing value. This strategy is useful for the features which have numeric data such as age, salary, year, etc.

Encoding Categorical data: Categorical data is data which has some categories such as, in our dataset; there are two categorical variables, Country, and Purchased. Since machine learning model completely works on mathematics and numbers, but if our dataset would have a categorical variable, then it may create trouble while building the model. So, it is necessary to encode these categorical variables into numbers.

3.3 Random Forest Model

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

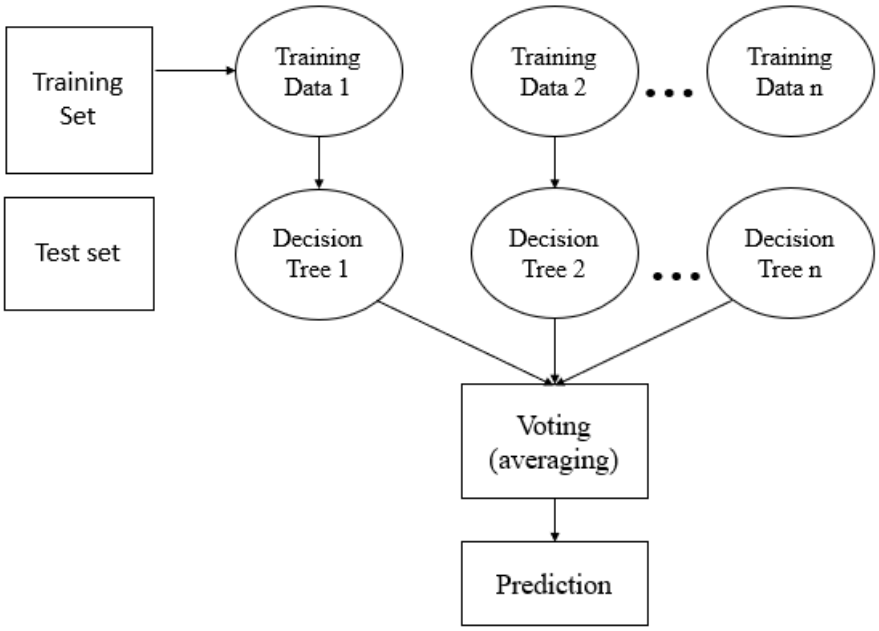


Fig. 2: Random Forest algorithm.

3.4.1 Random Forest algorithm

Step 1: In Random Forest n number of random records are taken from the data set having k number of records.

Step 2: Individual decision trees are constructed for each sample.

Step 3: Each decision tree will generate an output.

Step 4: Final output is considered based on Majority Voting or Averaging for Classification and regression respectively.

3.4.2 Important Features of Random Forest

- **Diversity**- Not all attributes/variables/features are considered while making an individual tree, each tree is different.
- **Immune to the curse of dimensionality**- Since each tree does not consider all the features, the feature space is reduced.
- **Parallelization**-Each tree is created independently out of different data and attributes. This means that we can make full use of the CPU to build random forests.
- **Train-Test split**- In a random forest we don't have to segregate the data for train and test as there will always be 30% of the data which is not seen by the decision tree.
- **Stability**- Stability arises because the result is based on majority voting/ averaging.

3.4.3 Assumptions for Random Forest

Since the random forest combines multiple trees to predict the class of the dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output. Therefore, below are two assumptions for a better Random Forest classifier:

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

- There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.
- The predictions from each tree must have very low correlations.

Below are some points that explain why we should use the Random Forest algorithm

- It takes less training time as compared to other algorithms.
- It predicts output with high accuracy, even for the large dataset it runs efficiently.
- It can also maintain accuracy when a large proportion of data is missing.

3.4.4 Types of Ensembles

Before understanding the working of the random forest, we must look into the ensemble technique. Ensemble simply means combining multiple models. Thus, a collection of models is used to make predictions rather than an individual model. Ensemble uses two types of methods:

Bagging– It creates a different training subset from sample training data with replacement & the final output is based on majority voting. For example, Random Forest. Bagging, also known as Bootstrap Aggregation is the ensemble technique used by random forest. Bagging chooses a random sample from the data set. Hence each model is generated from the samples (Bootstrap Samples) provided by the Original Data with replacement known as row sampling. This step of row sampling with replacement is called bootstrap. Now each model is trained independently which generates results. The final output is based on majority voting after combining the results of all models. This step which involves combining all the results and generating output based on majority voting is known as aggregation.

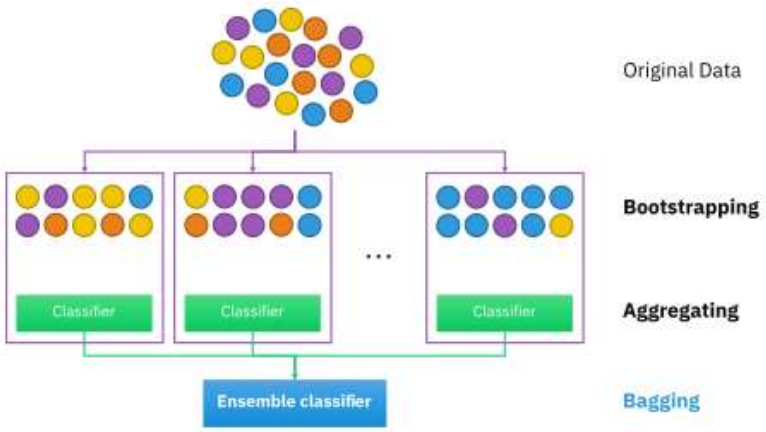


Fig. 3: RF Classifier analysis.

Boosting– It combines weak learners into strong learners by creating sequential models such that the final model has the highest accuracy. For example, ADA BOOST, XG BOOST.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

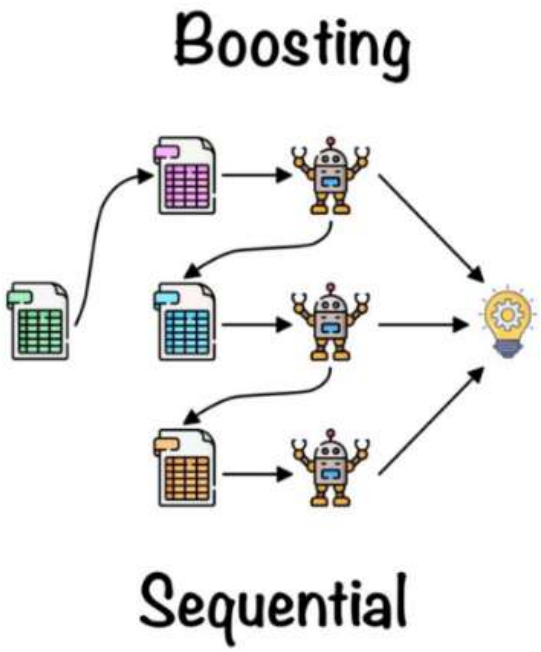


Fig. 4: Boosting RF Classifier.

4.RESULTS AND DESCRIPTION

10.1 Implementation Description

Here's an implementation description for the proposed system:

- **Data Collection:** Implement modules to collect data from IoT smart meters installed in various locations. Use appropriate communication protocols (e.g., MQTT, HTTP) to gather data from smart meters. Develop mechanisms to handle data transmission errors and ensure reliable data collection.
- **Data Preprocessing:** Create preprocessing pipelines to clean and transform raw data collected from smart meters. Perform data validation, filtering, and normalization to ensure data quality and consistency. Handle missing values, outliers, and noise to improve the accuracy of downstream analysis.
- **Machine Learning Model Training:** Choose suitable machine learning algorithms (e.g., regression, neural networks) for energy consumption prediction. Split the preprocessed data into training and validation sets for model training and evaluation. Tune hyperparameters and optimize model performance using techniques like cross-validation and grid search.
- **Energy Consumption Prediction:** Deploy trained machine learning models to predict future energy consumption based on historical data. Implement real-time prediction pipelines to generate forecasts for different time intervals (e.g., hourly, daily). Develop mechanisms to monitor model performance and retrain models periodically to adapt to changing consumption patterns.
- **Optimization Strategies:** Design optimization algorithms to minimize energy consumption and maximize efficiency. Incorporate predictive models and real-time data to dynamically adjust energy usage and scheduling. Implement feedback mechanisms to validate the effectiveness of optimization strategies and make adjustments as needed.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

— **Anomaly Detection:** Develop anomaly detection algorithms to identify unusual patterns or deviations in energy consumption data. Utilize statistical methods, machine learning techniques, or rule-based systems to detect anomalies. Trigger alerts or notifications when anomalies are detected to prompt further investigation or intervention.

10.2 Dataset Description

— **DateTime:** This column represents the timestamp or date-time information associated with each data point. It indicates the date and time when the energy usage and other measurements were recorded.

— **TotalUsage:** This column represents the total energy usage at the corresponding timestamp, measured in units like kilowatt-hours (kWh) or another appropriate energy unit.

— **Month:** This column indicates the month corresponding to each data point, represented as numeric values (e.g., 1 for January, 2 for February).

— **TemperatureF:** This column represents the temperature in degrees Fahrenheit at the time of measurement. It provides information about the ambient temperature, influencing energy consumption patterns.

— **Humidity:** This column represents the relative humidity at the time of measurement, affecting energy usage, especially in heating, ventilation, and air conditioning (HVAC) systems.

— **Hour_y:** This column indicates the hour component of the timestamp, representing the hour of the day when the data point was recorded, typically in a 24-hour format.

— **Minute_y:** This column indicates the minute component of the timestamp, representing the minute of the hour when the data point was recorded.

— **Day_y:** This column represents the day of the week corresponding to each data point, represented as numeric values (e.g., 1 for Monday, 2 for Tuesday).

— **Weekend:** This column indicates whether the corresponding day is a weekend or not, represented as a binary variable (e.g., 1 for weekend, 0 for non-weekend).

— **Holiday:** This column indicates whether the corresponding day is a holiday or not, also represented as a binary variable (e.g., 1 for holiday, 0 for non-holiday)

10.3 Results Description

The figure 1 presents a sample snapshot of the Smart meter Energy dataset. It displays a portion of the dataset, showcasing the structure and format of the data, including columns such as DateTime, TotalUsage, Month, TemperatureF, Humidity, Hour_y, Minute_y, Day_y, Weekend, and Holiday. The figure 2 illustrates the correlation between different variables/features in the Smart meter Energy dataset. It provides insights into how each variable relates to others, which is crucial for understanding the dataset's characteristics and potential predictive power. The figure 3 showcases the dataset after undergoing preprocessing steps. It likely includes data cleaning, transformation, and normalization processes to prepare the dataset for model training and evaluation. The figure 4 displays the performance metrics (such as accuracy, precision, recall, etc.) of a K-Nearest Neighbors (KNN) model. It assesses how well the KNN model performs in predicting energy consumption patterns

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

based on the dataset. The figure 5 presents the prediction outcomes of the KNN model visually. It compares the actual energy consumption values with the predicted values generated by the KNN model, providing insights into the model's predictive accuracy. This figure 6 showcases the performance metrics (similar to Figure 4) but for a Random Forest Regression (RFR) model. It evaluates the RFR model's performance in predicting energy consumption patterns based on the dataset. Similar to Figure 5, this figure 7 visualizes the prediction outcomes of the RFR model. It compares the actual energy consumption values with the predicted values generated by the RFR model, offering insights into the model's predictive accuracy.

	DateTime	TotalUsage	Month	TemperatureF	Humidity	Hour_y	Minute_y	Day_y	Weekend	Holiday
0	01-01-2016 00:00	19.843233	1	50.0	63.0	0	0	6	0	0
1	01-01-2016 01:00	18.462483	1	49.8	63.0	1	60	6	0	0
2	01-01-2016 02:00	17.414167	1	48.9	61.0	2	120	6	0	0
3	01-01-2016 03:00	15.914683	1	48.6	61.0	3	180	6	0	0
4	01-01-2016 04:00	19.195933	1	47.7	63.0	4	240	6	0	0

Fig. 5: Presents the Sample dataset of the Smart meter Energy dataset.

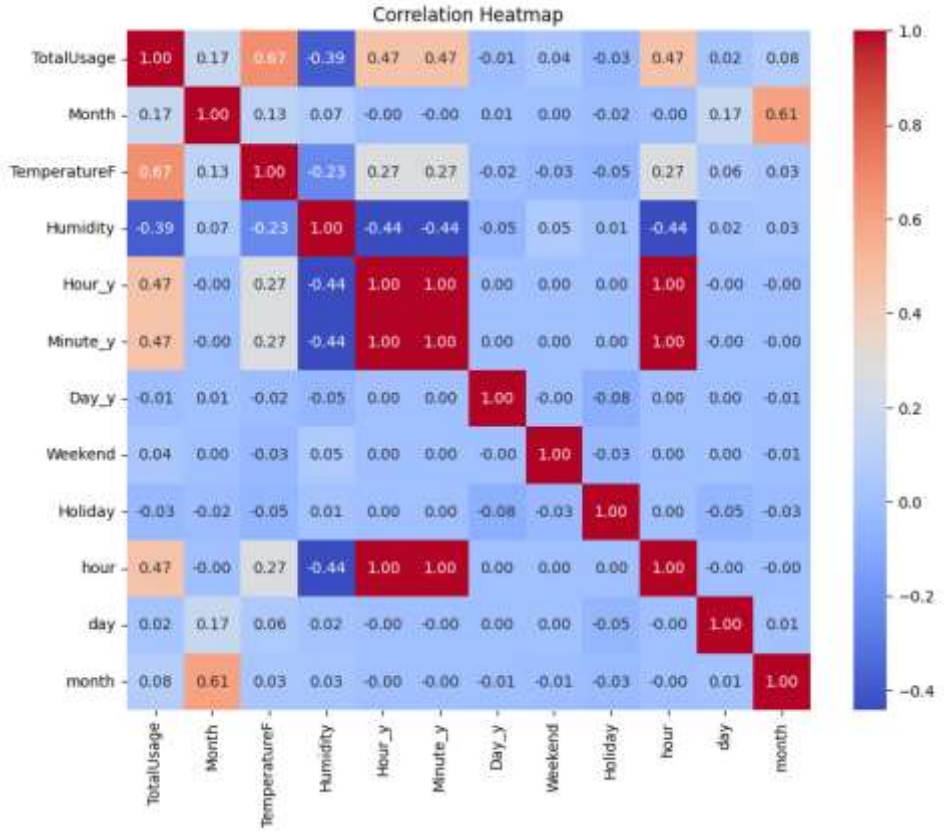


Fig. 6: Presents the correlation of Smart meter Energy dataset.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

	TotalUsage	Month	TemperatureF	Humidity	Hour_y	Minute_y	Day_y	Weekend	Holiday	hour	day	month
0	19.843233	1	50.0	63.0	0	0	6	0	0	0	1	1
1	18.462483	1	49.8	63.0	1	60	6	0	0	1	1	1
2	17.414167	1	48.9	61.0	2	120	6	0	0	2	1	1
3	15.914683	1	48.6	61.0	3	180	6	0	0	3	1	1
4	19.195933	1	47.7	63.0	4	240	6	0	0	4	1	1
...	...	...	...	...	...	...	...	...	...	...	...	...
17424	23.331300	12	38.5	90.0	13	780	1	1	0	13	31	12
17425	25.814400	12	37.2	84.0	14	840	1	1	0	14	31	12
17426	29.153450	12	36.1	84.0	15	900	1	1	0	15	31	12
17427	30.285350	12	35.4	84.0	16	960	1	1	0	16	31	12
17428	33.841383	12	34.2	84.0	17	1020	1	1	0	17	31	12

Fig. 3: Shows the Pre-processed Dataset.

```
K-Nearest Neighbors Regressor Mean Squared Error: 22.3776
K-Nearest Neighbors Regressor Mean Absolute Error: 3.5147
K-Nearest Neighbors Regressor R^2 Score: 0.8969
```

Fig 4: Shows a performance metrics of of a KNN model.

```
RandomForestRegressor Mean Squared Error: 14.1437
RandomForestRegressor Mean Absolute Error: 2.7893
RandomForestRegressor R^2 Score: 0.9348
```

Fig 6: Shows a performance metrics of of a RFR model.

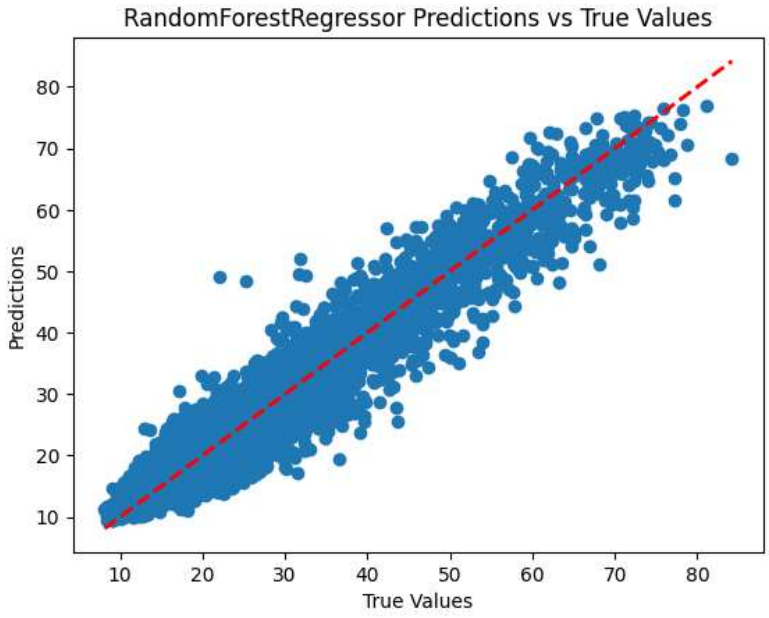


Fig. 8: Prediction graph of RFR model.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

	DateTime	Month	TemperatureF	Humidity	Hour_y	Minute_y	Day_y	Weekend	Holiday
0	01-01-2016 00:00	1	50.0	63	0	0	6	0	0
1	01-01-2016 01:00	1	49.8	63	1	60	6	0	0
2	01-01-2016 02:00	1	48.9	61	2	120	6	0	0
3	01-01-2016 03:00	1	48.6	61	3	180	6	0	0
4	01-01-2016 04:00	1	47.7	63	4	240	6	0	0
5	01-01-2016 05:00	1	46.9	63	5	300	6	0	0

Fig. 9: Presents the uploading of test dataset.

	Month	TemperatureF	Humidity	Hour_y	Minute_y	Day_y	Weekend	Holiday	hour	day	month	predicted
0	1	50.0	63	0	0	6	0	0	0	1	1	17.710996
1	1	49.8	63	1	60	6	0	0	1	1	1	17.711560
2	1	48.9	61	2	120	6	0	0	2	1	1	17.113530
3	1	48.6	61	3	180	6	0	0	3	1	1	17.266835
4	1	47.7	63	4	240	6	0	0	4	1	1	18.353047
5	1	46.9	63	5	300	6	0	0	5	1	1	18.352340

Fig 9: Proposed RFR model Prediction on uploaded test dataset.

	Algorithm Name	MSE	MAE	R2_Score
0	KNN Regressor	22.377616	3.514750	0.896883
1	Random Forest Regressor	14.143676	2.789258	0.934826

Fig. 10: Performance metrics of all models.

This figure 8 illustrates the process of uploading a test dataset for evaluation. It shows the interface or steps involved in uploading the dataset to assess the performance of the trained models. The figure 9 presents the predictions made by the proposed Random Forest Regression (RFR) model on the uploaded test dataset. It visualizes the model's predictions compared to the actual energy consumption values from the test dataset. The figure 10 consolidates and compares the performance metrics of all models evaluated in the study. It provides a comprehensive overview of how each model performs in predicting energy consumption patterns, facilitating comparison and decision-making regarding model selection.

5.CONCLUSION

The integration of machine learning techniques with IoT smart meter data holds immense promise for revolutionizing energy management and sustainability efforts. By accurately predicting energy consumption patterns, organizations can optimize resource allocation, reduce costs, and mitigate environmental impacts.

Looking ahead, the future scope of research in this area is vast. One avenue for future exploration is the development of more sophisticated machine learning algorithms capable of handling the complexities of smart meter data, such as temporal dependencies, seasonality, and outliers. Additionally, there is a need for research focusing on the integration of renewable energy sources and

energy storage systems into predictive models. By considering factors such as weather conditions, solar irradiance, and battery storage capacity, predictive analytics can help optimize the integration of renewables into the grid and maximize energy efficiency.

Furthermore, the application of predictive analytics in conjunction with emerging technologies such as blockchain and edge computing holds promise for enhancing data security, privacy, and scalability in IoT-based energy management systems. In essence, the journey towards more sustainable and efficient energy management practices is ongoing, and the integration of machine learning with IoT smart meter data represents a significant step forward in this endeavor. By continuing to innovate and collaborate across disciplines, we can unlock new possibilities for a greener and more sustainable future.

REFERENCES

[1] Jin, Jiong, et al. IEEE Internet of Things Journal 1.2 (2014): An information framework for creating smart city through internet of things

[2] Al-Ali, A. R. (Energy Procedia 100 (2016): "Internet of things role in the renewable energy resources

[3] Haghi, Armin, and Oliver Toole.” CS229 Course paper (2013). "The use of smart meter data to forecast electricity demand.

[4] Kwac, Jungsuk, June Flora, and Ram Rajagopal. IEEE Transactions on Smart Grid 5.1 (2014) "Household energy consumption segmentation using hourly data."

[5] Lines, Jason, et al. Learning-IDEAL 2011 (2011) Classification of household devices by electricity usage profiles." Intelligent Data Engineerin and Automated.

[6] Ning Lu, Pengwei Du, Xinxin Guo, and Frank L. Greitzer (Transmission and Distribution Conference and Exposition (T&D), 2012 IEEE PES) 16.Smart meter data analysis.

[7] Adrian Albert and Ram Rajagopal (IEEE Transactions on Power Systems28.4 (2013) Smart Meter Driven Segmentation: What Your Consumption Says About You.

[8] Tomasz Ząbkowskie, et al. (Procedia Computer Science35 (2014), Elsevier) Short term electricity forecasting using individual smart meter data.

[9] Wei Yu et al. (Applied Computing Review 15.1 (2015): ACM). Towards Statistical Modelling and Machine Learning Based Energy Usage Forecasting in Smart Grid.

[10] Sudha Gupta et al, (IEEE Transactions on Industrial Electronics 62.4 (2015). Support Vector Machine Based Proactive Cascade Prediction in Smart Grid Using Probabilistic

[11] Kui Wu et al. (IEEE Transactions on Smart Grid 7.3 (2016) A Machine Learning Approach to Meter Placement for Power Quality Estimation in Smart Grid.

[12] Md. Sumon Shahriar, M. Sabbir Rahman (Proceedings of the 2015 on Internet of Things towards Applications. Urban Sensing and Smart Home Energy Optimisations: AMachine Learning Approach.

[13] David Walkera, Enrico Creacoa et al. (Procedia Engineering 119 (2015, Elsevier). Forecasting Domestic Water Consumption from Smart Meter Readings using Statistical Methods and Artificial Neural Networks.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

[14] Joseph Siryani, Thomas Mazzuchi, Shahram Sarkani (2015 IEEE First International Conference) Framework using Bayesian Belief Networks for Utility Effective Management and Operations.

[15] Damminda Alahakoon, Xinghuo Yu (IEEE Transactions on Industrial Informatics 12.1 (2016): Smart Electricity Meter Data Intelligence for Future Energy Systems: A Survey

[16] Stephen Haben, Colin Singleton, and Peter Grindrod IEEE transactions on smart grid 7.1 (2016). Analysis and clustering of residential customers' energy behavioral demand using smart meter data.

[17] Samuel Idowu ∗, Saguna Saguna, et al. (Energy and Buildings 133 (2016, Elsevier). Applied machine learning: Forecasting heat load in district heating system

[18] Thierry Zufferey, Andreas Ulbig. et al. (Data Analytics for Renewable Energy Integration. Springer, 2016) Forecasting of Smart Meter Time Series Based on Neural Networks.

[19] Martinez-Pabon, Madeline, Timothy Eveleigh, and Bereket Tanju E7 (2017) Smart Meter Data Analytics for Optimal Customer Selection in Demand Response Programs.

[20] Joseph Siryani, Bereket Tanju, Timothy Eveleigh, (IEEE Internet of Things Journal 4.4 (2017)) A Machine Learning Decision-Support System Improves the Internet of Things' Smart Meter Operation