A Big data Approach in Demand Response Management in Smart Grid Prophet Model

***Abstract*—Smart Grids (SG) generate extensive data sets regarding the system variables viz., demand and supply. These extremely large data sets are known as Big data. Hence, preprocessing of this huge data and integration become critical steps in the load forecasting process. The precise prediction of load is the main concern while balancing the demand and supply in SG. Many techniques were devised for load forecasting using machine learning methods such as Deep-learning Models. However, in case of large data sets, only a few models provide good performance viz. Autoregressive Integrated Moving Average (ARIMA). But, this approach is difficult as it takes a minimum of 50 observations for taking up an evaluation. The Prophet technique is used in the prediction of future demand response based on the past data which is in the form of a time series. The technique is useful even if a few values in the time series are not available. Furthermore, the procedure is not affected by fluctuations, trends and abnormal variations, as well. Automatic model fitting approach is adopted for its effective performance. ARIMA model and Prophet model were used for forecasting using various evaluation matrices and demand response management was achieved with two data sets. The results show effectiveness of the Prophet model in the demand Demand response (DR).**

***Index Terms*—ARIMA model, Big data, Demand response management, Energy prediction, Prophet model, Smart Grid, Load Forecasting**

I. Introduction

SMART grid differs from the traditional electricity system which involves flow of energy in one way. It enables the real time collection of data in the transmission and distribution network facilitating monitoring of electricity for efficient energy management. It involves technologies such as data acquisition, control, automation and communication, which work together in the grid to respond to the ever-changing demands of consumers. The flexibility can be achieved by making the customers to follow demand response programs. However, a number of challenges are involved in implementing these programs in a real-time environment. Demand side management (DSM) in electrical power system is one of the solutions to these challenges by shifting the flexibility of the power system to the consumer side.

Besides load forecasting, there are two kinds of demand responses (DR) which take place in the power sector: non-dispatchable and dispatchable demand responses. Dispatchable demand response deals with customer appliances, directly. In some cases, the utility directs the consumer regarding cutting down of the air conditioner or heater load during peak demand periods, and thereby reducing the cost. So, the consumer will be directed only when the utility can forecast and predict the peak load beforehand. The problem arises when the forecasting model shows errors. The non-dispatchable demand response or the retail price-responsive demand is when the customer has the liberty to decide whether to cut down his consumption. It is based on the retail rate design and does not remain fixed. This includes dynamic pricing programs. In this case, the issue is redundant and in most cases the increase in prices have no perceptible effect on the consumption pattern of consumers. There is a vast amount of data which is derived from the SG setup, on a second-by-second basis. However, there are various challenges associated with load forecasting using this data. First, the data which is mostly accumulated are unstructured as they are gathered from a wide area containing a number of households. Second, the data collected is interlinked. Therefore, extraction of a particular data is found difficult for real-time applications. To solve these issues, a number of works were carried out on load forecasting and demand response management using Big data.

For managing the demand-response of energy, authors have discussed about the ARIMA model [1]. This model has the capability to lag its forecast errors by itself. It captures the consumed energy in the grid. In this paper, ARIMA model is used to detect abnormalities. The authors have used automated fitting methods. However, it is not suitable for electricity consumption where there is high variation in consumption behaviour. Subsequenly, the occupancy levels which can improve energy prediction is highlighted. Their accumulated data is related to the premises of a residence [2]. Here, the authors collected data on the network activity and the consumption data of the consumers, on a daily basis. They used ARIMA model to forecast with accuracy. It is acknowledged that the measurements of the constructed residence has a remarkable explanatory variable. In a similar work, authors elaborated on a short to medium term load prediction model [3]. In this paper, they highlighted the Big data prediction approach for smart homes.

The authors have used Big data in [4] for energy prediction. In a similar work [5] data management system was used for forecasting. This model can help the customers to manage energy and reduce grid failure. They developed a model to make it economical and effective for better load distribution. In [6], authors have evaluated an hourly demand profile, which is effectively trained. Here, they have used hybrid model of the Bi-directional Long-Short-Term Memory (BLSTM) and the ARIMA model for prediction of energy. In another work, the ARIMA model is used in all the phases of a time series data. In another work the authors have dealt with short-term wind power prediction [7] where ARIMA model is used for better performance [8]. Here, the authors have described issues with real-time price forecasting method. In [9], the authors have randomized a consumer algorithm for managing demand response [10]. Moreover, authors have designed an optimized demand response for efficient management of supply and demand in SG.

A. Related work

Energy management is the most crucial part in an SG environment. For managing supply and demand of energy, various big data analytics approaches were performed well in SG. Safhi *et al.* discussed about load forecasting which is based on Big data. There are various prediction techniques discussed in the literature for managing demand and supply gap. ARIMA model is a very important analytical method where time series data is used. This model is trained to understand the data set, in an efficient way. ARIMA model has a great role in the prediction of the future trends of a time series. Krishna *et al.* discussed about ARIMA model where it captures the process of energy consumption in the power system. Moreover, it has the capability to check the validity of the consumed energy of the system.

The above model was unable to capture the experiment-based characteristics in a proper way. In order to overcome this limitation the Prophet model was proposed by some authors. In the work carried out [11], the authors have used K-means clustering with ARIMA model to obtain DR from an area in their university. They have used data of two academic years viz., 2014-2015. First, they obtained the predicted electricity consumption data for 2016. The electricity consumption forecast was done and DSM variabales were obtained thereafter. In [12], the authors have used two datasets viz., one dataset comprising the electricity consumption of three months (February 2013-April 2013) and the other relating to the four years (July 2009 - June 2013). On both these datasets, dynamic demand response was carried out using six different prediction models including ARIMA model and the results were finally compared.

The authors in [13] have carried out work on applications of technologies involving Big data such as online monitoring of renewable energy systems and wind turbines, based on ultra-sphere model. The third application was backup and recovery in electric power generation process. The authors were of the view that big data technologies are very essential for building and handling Smart Grids. However, their research lacked the use of big data in SG. In [14], the paper deals with the analysis of big data for renewable energy resources. It further aims at presenting techniques for Demand response management (DRM) using big data. The technology used to carry out the DRM is a simulator. While issuing DR, the simulator studies the big data and identifies correlation among it. The paper basically aims at using the technology and how it would be fruitful in future with advanced features added to it. However, the research has not been made using different Machine-learning techniques for carrying out the DR. By optimising the programming, it would become easier for advancing the technology, in the near future.

The research in [15] was carried out to solve two main issues while handling big data. First, the authors used expansion K-SVD sparse representation technique to extract hidden patterns of electricity consumption. Second, it can also be used to compress the data and store it efficiently. Further, they used a series of other computational techniques like SVM, PCA etc. to classify the consumers into various groups. So, basically they worked on efficient compression of data and its extraction. However, their research lacked obtaining a DR from the extracted data. In [16], the authors have used big data technology to recognise the patterns of electricity consumption by consumers. They have further shifted the peak load by studying the consumption patterns. They used K-means algorithm for achieving the final clusters. However, they did not predict the future energy consumption patterns of the consumers.

The work in [17] was based on obtaining DR on electricity consumption data of one month only. The authors have used the big data technologies like Apache Spark to analyse the data to obtain DR and thereby have successfully curtailed the air conditioner and heating loads. In [18], the authors have created various priority lists varying from consumer to consumer. They have further analysed various aspects such as time, need and power consumed while creating the priority lists. However, they have not done DSM using the priority lists. Furthermore, they did not predict the future electricity consumption by the consumers.

Almajarouee *et al.* introduced techniques based on peak load where long-term forecasting provided results, accurately. They have tried to save cost and time by using this model [19]. Many authors discussed about Prophet model approach for the demand of energy in SG. They tried to improve energy generation, consumption and forecasting. In another paper, authors discussed about data cleaning algorithm [20]. Here, authors have highlighted the errors in the models and various issues such as Benchmark-related algorithms. In a different work, authors defined the smart energy management which can maintain quality of life for the consumer [21]. In this paper, authors paid attention towards minimization of energy by developing an efficient model.

The energy consumption was managed [22] by using operational approach techniques. This technology is designed for proper communication with SG. The Energy Management System (EMS) has a great role in managing demand and supply [23]. However, the data sample taken to study were small datasets. In another paper, authors have highlighted the multiple homes. The problem formulation is carried out via multiple-knapsack problems [24]. In this paper, authors discussed about the importance of the external variables of consumed energy. They defined the state of the art energy load forecasting method. They have also defined the challenges which are involved with big data. Table I shows, at a glance, the various research work done in this area.

TABLE I: Works related to Energy Forecasting for Demand Response

S.No.	Author(s), Year	Techniques	Description	Application/Domain	Reference
1.	Newsham *et al.*[2010]	ARIMA model	To collect data related to the total occupied building, They installed wireless sensors in the building at eastern zone at Ontario	Energy forecasting	2
2.	Krishna *et al.*[2015]	ARIMA model forecasting method	They have used automatic model fitting methods	Energy prediction	1
3.	Asaleye *et al.* [2017]	Decision support tool	Used for renewable energy microgrids (DSTREM)	To ascertain daily energy consumption	6
4.	Luo *et al.*[2018]	An innovative hybrid RTP forecasting model	RTP model has been used for analysing customer conducts. They got more benefit by scheduling the use of home appliances	Real-time forecasting	9
5.	Sendric *et al.*[2019]	Data cleaning algorithm	To solve problems of ambiguous data from Big data wireless sensor networks. They used these to clean data in smart cities	Cleaning the ambiguous data	14
6.	Gupta *et al.* [2019]	Short-term Wind Power Prediction (WPP) to improve efficiency of power systems	Hybrid ARIMA-GARCH model	Forecasting of Wind Power	15
7.	Amelec *et al.*[2019]	ARIMA model method	In this paper, Euclidean distance (ED) is used which is not capable of provididng better results than ARIMA model.	Energy prediction	5
8.	Liu *et al.*[2020]	Big data analytics	Big data analytics in running Smart cities	To build Smart cities	4
9.	Wang *et al.*[2020]	hybrid model based on the ARIMA model In this paper, Bi-directonal(Bi-LSTM) model and Bayesian optimization (BO) model	The ARIMA model can tackle the linear part of the time series data. Bi-LSTM can handle the non-linear features The hybrid model provided a good prediction tool	The installation or replacement of electrolytic capacitors	7
10.	Almazrouee *et al.*[2020]	Prophet model	Prophet model used and forecasting accuracy compared with Holt–Winter's model.	To predict Long-term peak loads	12

B. Motivation

ARIMA model is mainly used by professionals who have prior knowledge of the intricacies of the model. If a single parameter in the equation is incorrect, the entire result will get affected. But, Prophet model uses a Bayesian curve fitting method and does not require prior knowledge of datasets. It automatically finds seasonal trends from the data. Prophet model is made by incorporating seasonal trends such as holidays and weekends whereas ARIMA model incorporates both seasonal and non-seasonal trends with time series data. It provides great precision compared to any other method.

C. Contribution

The ARIMA model requires expert knowledge as a prerequisite to make use of it. In addition, it is not flexible in use and is non-automatic. The Prophet model overcomes all the aforesaid limitations and is a powerful tool for prediction. It gives precise results with non-seasonal trends and also incorporates non-linear trends with datasets. The contribution of the proposed work is as follows to:

- optimize the parameters for Prophet model for better performance.
- apply preprocessing techniques to clean the data.
- reduce the abnormal data for prediction of consumption of energy in forecasting.
- compare and analyse ARIMA model and Prophet model with different performance metrics.

D. Organization

In this paper, Section II defines the detailed description of the methodology and work flow of our proposed model. Section III describes the performance of the models along with evaluation parameters. Subsequently, the Section IV is devoted to the discussion of the dataset and the Prophet model. Finally, the conclusion of this work is stated in Section V.

II. Methodology

Forecasting is the process of predicting the future values using the available past data which is in the nature of an ordered times series spread over time. Further, The plot can be univariate or multivariate. The data is then split in such a way that the maximum amount of data is put into the training set as compared to the validation set. All these things lead to a forecasting model. Hence, a number of different prediction models have their own techniques and choosing the right model is crucial depending on the data and situation. Some standard examples are straight line, moving average, simple linear regression, multiple linear regression, naive forecasting and many more. The Prophet model is an open source model for forecasting first used by Facebook. It is more effective than the present traditional models based on time series. Further, in a traditional time series model there are certain problems, as mentioned below.

- Time interval between data has to be the same throughout, while this is not a problem in Prophet model.
- Day with NA (nil data) is not allowed, while this is not an issue in Prophet model.
- Seasonality with multiple periods is difficult to handle, while in Prophet model, this problem is handled by default.
- Parameter tuning by an expert is necessary in ARIMA model, while in Prophet model there is a default setting, from which parameters are easily interpreted. Prophet model is extensively used in various fields, for forecasting with a large range of data.

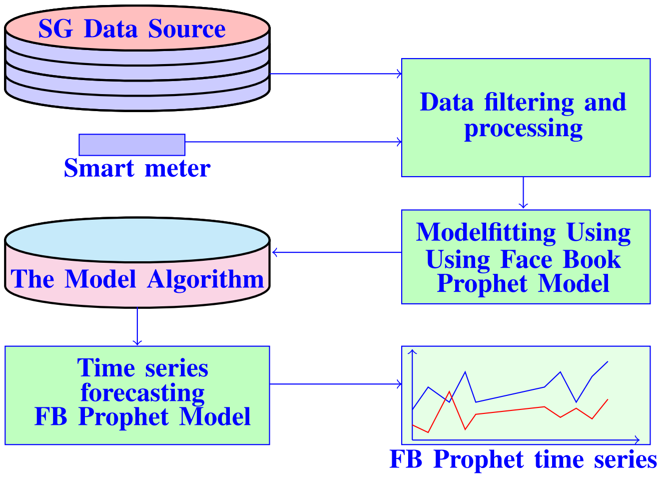


Fig. 1: The Workflow diagram of the Prophet model for the prediction of demand response

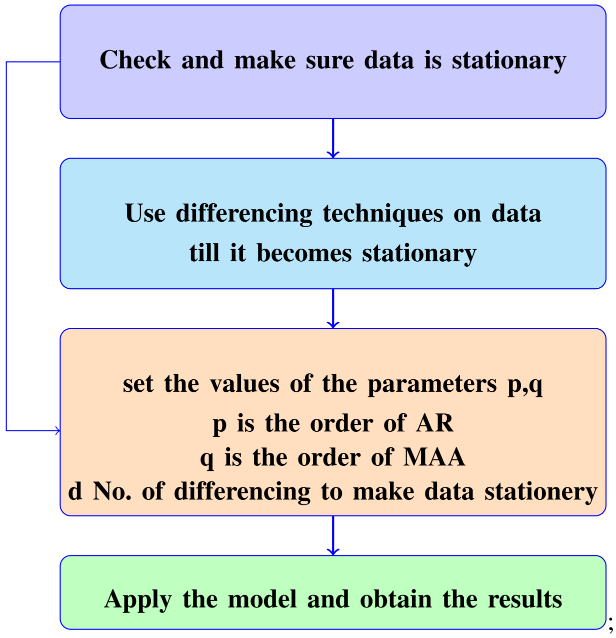


Fig. 2: The Workflow of ARIMA

The workflow diagram of the the Prophet model is shown in Fig. 1 wherein a high volume of data is obtained from the Smart grid and the Smart meters connected to the electric power system. The data consist of extraneous values, noises etc and hence the data needs to be filtered and the relevant data extracted from the mass of data. Subsequent to processing, model fitting is done and then used in the Prophet model using the Face Book method. The ProfThe workflowet model generates reliable forecast in the form of time series values for use in the Demand response process.

III. Mathematical Modelling

A. Modelling of ARIMA

The ARIMA model process data in a time series, for making prediction. The ARIMA model is used in case of both linear and multiple-regression models. Multiple-regression model refers to prediction of outcomes of dependent variables which are based on variables of independentent variables.

The model is generally referred to as ARIMA (p, d, q) where p, d and q are zero or positive numbers. ARIMA model makes use of a stationary time series. By using multiple linear regression model, it can work over non-stationary time series data. The values of p, q and d can be found using auto-ARIMA. The process seeks to identify the most optimal parameters for ARIMA model settling on a single-fitted model. The process works by conducting differencing tests to determine the order of differencing 'd' and then fitting the models within the ranges of defined start p, max p, start q, and max q. The parameters p, q and d were set to (4, 1, 1). Finally, the model was trained on 2014-2015 data to obtain a prediction for 2016 consumption data.

The ADF (Augmented Dickey Fuller) test is useful in detecting the unit root in a series, and thereby helping us to understand whether the series is stationary or not. Here, the null and alternate hypothesis state that if a series has a unit root, it fails to reject the null hypothesis which says that the unit has a root. Then, the series is non-stationary. This means that the series can be linear, stationery, or difference stationary. The experimental results show that the ARMA (AutoRegressive Moving Average model) is based on real data which is a stationary time series. In the flow chart shown in Fig. 2, three features of the stationary data have been selected. Here, the first characteristic is the mean which is constant, the second one is the variance which is also constant and the third characteristic is the co-variance where the signal of past data, at different times, is constant.

Here, the daily stationary signal does not meet the first condition, but satisfies the second and third conditions. The moving average component of ARMA is set for change mean, and therefore, the first condition is not important for the appropriate ARMA for a given time series. Later, the process of residual checking is completed.

If conditions are satisfied, then the process is stopped or otherwise continued. In Fig. 2, the procedure of the ARIMA model is shown. Here, the power consumption datasets were collected and then the data is preprocessed. Subsequently, any abnormal data present in the datasets are eliminated. After selection of features, the important features are extracted by using classifier (Support Vectors Machine) and at last the predicted energy. On the basis of lagged data, future prediction is decided. In the above model, the equations are based on an autoregressive function. It is a function where the current value is generated based on the immediately preceding value. In the second process, the current value is generated based on the last two values. An AR (0) process would imply that there are no dependencies among the terms in the equations.

The aforesaid term predicts certain errors known as moving averages. The time series is differenced to make the data stationary. There are many models such as Random-walk model, Random-trend model, Autoregressive model and the Exponential-smoothing model. All are special cases of ARIMA model. The time series representing the electricity consumption of a single consumer, at time t, is given by the value of Y_t. The ARIMA model is discussed as in Fig. 2 which represents the flow chart of the model where the parameters are indicated. Here, the values of parameters are selected and residual checking is done. The residual value is differenced between observed value and the predicted value. ARIMA model aims to predict power.

$$Y_t = c + \in t + \sum_{i=1}^{q} \alpha_i X_{t-i} + \sum_{k=1}^{r} \beta_k \in (t-k) \tag{1}$$

Y_t indicates the consumption of energy at time t, c indicates obstruction of the signal at q the previous point X_{t-i} with linear coefficient α_i. The ARMA handles daily stationary signals that have no meaning of constant. It assumes Gaussian noise as in $\in_t$ and compounds q over time periods. Here, It contributes linearly to the signal $t-k$ with coefficient β_k.

$$f(x) = x^2 + 2x + 1 \tag{2}$$

In Eq. (2)$f(x)$, is the dependent variable and it indicates the prediction of the energy consumption using a time series data. The values x^2, $2x$ are independent variables and define the first order differencing for making a stationary data into a non-stationary data.

The prediction of the energy consumption in a time series data is described as follows:

$$\begin{aligned} Z_{t1} = \alpha_0 - \psi_1(z_{t1}-1) - \psi_2(z_{t1}-2) - ... - \psi_z n_t 1 - p \\ + \epsilon - \alpha_1(\epsilon_{t1}-1) - \alpha_2(\epsilon_{t_1}-2) - ... - \alpha \epsilon t 1 - q \end{aligned} \tag{3}$$

Where at time t_1, z_{t1} and ϵ_{t1} are the predicted values and the random error of data $\psi(z_{t_1}-1)\cdots p$ indicates the model parameter, $\alpha_1 ... q$ indicates the model parameter, p and q are represented by the autoregressive and moving average orders. Eq. (3) shows some important cases of the ARIMA models, If q_2 = 0, then Eq. (3) becomes an AR model of order p_2, and when $p_2 = 0$, the model decreases to a MA model to work with order q_2. The past data is the main basis in the prediction of energy by ARIMA model.

The general forecasting equation is:

$$\begin{aligned} z_t = \mu + \phi 1(z_t - 1) + \ldots + \phi p z_t - p \\ - \theta 1(e_t - 1) - \ldots - \theta q e_t - q \end{aligned} \tag{4}$$

Box and Jerkins have introduced the moving average parameters (θ) having negative values in the equation. Hence, the actual numbers are used in the equation and there is no ambiguity, as the output was read by us at the time of use of this software. These parameters are denoted by AR(1), AR(2)...AR(N) and MA(1), MA(2)....MA(N). To recognise suitable ARIMA model for Z_1, the order of differencing viz., (d_2) is to be decided. It is very important to make the series spatial so that the characteristics of seasonality can be removed. If prediction of the differenced next series is constant, then we have applied Random-trend model. Here, the series is autocorrelated and the errors show the number of AR terms ($p_2 \geq 1$) and number MA ($q_2 \geq 1$). These are also needed in the equation. To determine the values of p_2, d_2 and q_2 is the best way for a given time series.

There are many types of non-seasonal ARIMA models that are discussed as follows: ARIMA1, 0, 0) model is denoted as the first order autoregressive model. If the series is stationary and autocorrelated, then it can be forecast as a multiple of its own previous value, plus a constant. The forecasting equation, in this case is as below.

$$Z_t = \mu + \phi_1 Z_{1t} - 1 \tag{5}$$

In Eq. (5), Z_1 is less developed data on itself by one period. This is an ARIMA1, 0, 0) + constant model. If the mean value of Z_1 is zero, then the constant value will not be sufficient. If the slope coefficient ϕ is positive and less than 1, Z_1 is stationary. If the value of next time period value is predicted to be ϕ times it creates a great distance from the mean, as this is a time value. If ϕ_1 is negative, it predicts a mean level with alteration of the signs. It also predicts that Z_1 will be below the mean of the next period if it is above the mean at that time.

In a 2nd order autoregressive model ARIMA2, 0, 0), there is $Y(t-2)$ term on the right, as well as, on the left and depends on the signs and magnitudes of the coefficients. It describes a system where the mean level is of a sinusoidal wave pattern. It is like the motion of a mass on a spring that is subjected to random shocks. If autoregressive coefficient is equal to 1, it is a series with infinitely slow mean which returns to the previous state. The equation for this model can be written as follows:

$$z_t - Z_1 t - 1 = \mu \tag{6}$$

or equivalently

$$z_t = \mu + Z_1 t + 1 \tag{7}$$

where the constant term is the mean which changes periodically (i.e., the long-term drift) Z_1. This model could be fitted as a no-intercept regression model in which the first differencing of y is the dependent variable as it includes only a nonseasonal difference and a constant term. It is "ARIMA0,1,0) with a constant." The Random-walk without drift model would be ARIMA0,1,0) without any constant. ARIMA1,1,0) is used as differenced with first-order in autoregressive model. Here, autocorrelated means that the errors are found as in Random walk model. Then the problem can be settled by adding one past data of the dependent variable to the forecast equations i.e., by regressing the first difference of z on itself lagged by one period. The forecast equation is:

$$Z_t - (Z_1 t - 1) = \mu + \phi_1(Z_1 t - 1) - (Z_1 t - 2) \tag{8}$$

$$yt - (Z_1 t - 1) = \mu \tag{9}$$

B. The Modeling of Prophet

Prophet model is based on the additive type model developed by Facebook for forecasting The Prophet model. It is a time series predictive method where the aim is to predict power in SG. For this purpose, appliances are categorised as: interruptible, non-interruptible and base appliances. Power categorization of:

- Interruptible Appliances

$$F_{in} = \sum_{t-1}^{U} \sum_{ineIN} \sigma_{in} * rw_{in}(t) \tag{10}$$

where F_{in} is the power consumption of appliances, $ineIN$ indicates Interruptible Appliance, σ_{in} indicates power rating, U is the total time slot and $rw_{in}(t)$ is the state of each Interruptible Appliance at time slot t.

$$rw_{in}(t) = \sum \begin{cases} 0 & \text{if appliances are off} \\ 1 & \text{if appliances are on} \end{cases} \tag{11}$$

- Non-interruptible Appliances

$$F_{in} = \sum_{t-1}^{U} \sum_{niNI} \sigma_n i * rw_{in}(t) \tag{12}$$

Where F_{in} is power consumption of appliances, $niNI$ indicates Non-Interruptible Appliance, σ_{in} indicates power rating, U is total time slot and $rw_{in}(t)$ is the state of each Non-Interruptible Appliances at time slot t.

$$rw_{in}(t) = \sum \begin{cases} 0 & \text{if appliances are off} \\ 1 & \text{if appliances are on} \end{cases} \tag{13}$$

- Base appliances are similar to fixed appliances which do not have flexibility of operation. The pattern of consumption of energy and operational period of appliances cannot be changed. It is important that these appliances must be 'ON' when user wants to switch them ON such as home appliances viz. TV, Fridge and other devices.

$$F_b = \sum_{t-1}^{U} \sum_{beB} \sigma_B * rw_B(t) \tag{14}$$

Where F_b represents total energy consumption, B is the base appliance at time t, rw_B is each base appliance, σ_B is the power rating.

Since Prophet model works over data trends, holidays and seasonal data provide complex features. Seasonality is input on the basis of day, week and year. Prophet model where consumption is represented by time series is a data method expressed, as follows,

$$Z(t) = Y(t) + S(t) + H(t) + E(t) \tag{15}$$

$Z(t)$ indicates the consumption, $Y(t)$ represents the data trend function, $S(t)$ indicates the seasonal data, $H(t)$ indicates the holiday-based data and $E(t)$ represents the errors.

The trend function of Prophet model $H(t)$ is highlighted by a piecewise linear growth model. It is also called a Saturation-growth model. The maximum load data does not show a saturating growth which is a piecewise linear growth model represented, as follows:

$$h(t) = (l + a(t)^T \delta)t + (n + a(t)^T \sigma) \tag{16}$$

Here l is the growth rate, δ indicates rate adjustment, n is an offset parameter and σ is the change point.

$$b_j(t) = \begin{cases} 1 & \text{if } t \geq U_k \\ 0 & \text{otherwise} \end{cases}$$

Where b_j is the output and U_k is the change point.

The seasonality function is manifested by the following equation:

$$T(t) = \sum_{n=1}^{L} (b_n cos(2 * p * in * t/q) \\ + d_n sin(2 * p * in * t/q) \tag{17}$$

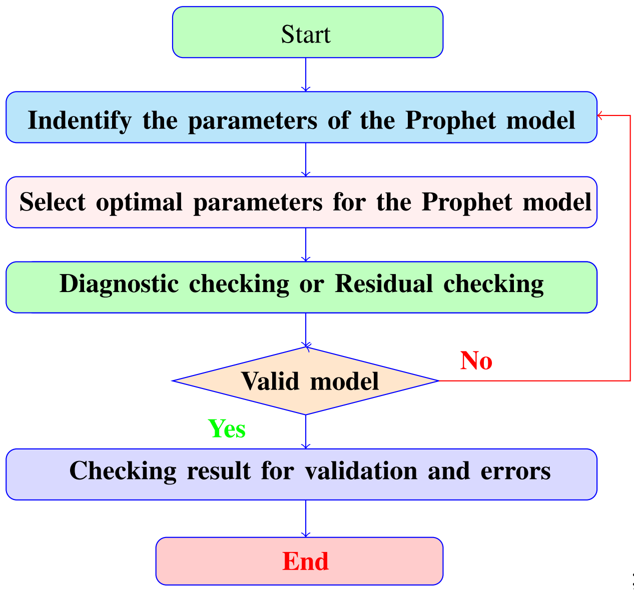
In Eq. (17), $T(t)$ is the seasonality function. Here, the time series multiperiod seasonality method is used. The Fourier series is applied to the daily and seasonal changes. Therefore, the seasonality function is discussed as:

$$A_t = [1(t \in E_1), \cdots, 1(t \in C_m)] \tag{18}$$

Here, A_t indicates matrix of regresses, E indicates holidays and 1 represents the holidays parameter.

$$H_t = A(t)l \tag{19}$$

In Eq. (19), H_t indicates holidays and l indicates corresponding change in the forecast. It produces estimates of unknown variables.



;

Fig. 3: The flow chart of Prophet Model

IV. RESULTS AND DISCUSSION

A. Description of Datasets

This section provides the description of the datasets. Table II shows the consumer preference regarding the appliance taken with the reason for selection. Table IV shows the rating of each of the appliances. The dataset taken is hourly time series and forecasting was done using Facebook's Prophet model and ARIMA model. Energy consumption has unique characteristics. First, the electricity consumption for the year 2016 was predicted using ARIMA model. However, to make use of ARIMA model, it is to be made sure that the data is stationary.

TABLE II: Preference matrix for different devices

S.No.	Time/Reason	Priority									
1	00:01-03:00	Fridge	Air conditioner	Tubelight	Microwave oven	Sandwich Maker	Dishwasher	Hair dryer	Vacuum cleaner	Washing machine	Clothes dryer
	Reason	Fridge is essential for use all the 24hrs. The other items are taken according to priorities during the course of the day									
2	3:01-6:00	Fridge	Air conditioner	Tubelight	Dish washer	Microwave oven	Sandwich maker	Hair dryer	Vacuum cleaner	Washing machine	Clothes dryer
	Reason	Dishwasher is placed as the fourth priority here. Some people sleep early and wake up early.									
3	6:01-9:00	Fridge	Tubelight	Dish washer	Hair dryer	Sandwich maker	Microwave oven	Washing machine	Vacuum cleaner	Clothes dryer	Air conditioner
	Reason	It is time to go to school and office in the morning.									
4	9:01-12:00	Fridge	Washing machine	Vacuum cleaner	Clothes cryer	Microwave oven	Sandwich maker	Hair dryer	Dish washer	Air conditioner	Tubelight
	Reason	People leave home for office by 9:00am. Thereafter, the priority for people at home is to clean the house and wash clothes.									
5	12:01-15:00	Fridge	Clothes dryer	Air conditioner	Microwave oven	Dishwasher	Hair dryer	Vacuum cleaner	Washing machine	Tubelight	Sandwich maker
	Reason	It is the lunch time, so Microwave oven is used. Besides, the ambient temperature is high. So, air conditioner is also used.									
6	15:01-18:00	Fridge	Dish washer	Air conditioner	Tubelight	Microwave oven	Sandwich maker	Washing machine	Clothes dryer	Hair dryer	Vacuum cleaner
	Reason	It is time to take an afternoon nap and wake up for tea in the evening									
7	18:01-21:00	Fridge	Tubelight	Microwave oven	Hair dryer	Air conditioner	Sandwich maker	Dishwasher	Vacuum cleaner	Washing machine	Clothes dryer
	Reason	Now, time for people to return from office. take bath and use the hair dryer. After dinner they relax using Air conditioner.									
8	21:01-24:00	Fridge	Tubelight	Air conditioner	Microwave oven	Dishwasher	Hair dryer	Sandwich maker	Vacuum cleaner	Washing machine	Clothes dryer
	Reason	Some people return late from the office, and usually have their dinner at 9:00pm.									

TABLE III: Smart Grid Dataset

Date: 01/01/2014	
Time	**Energy kWh/half-hour**
00:00.0	0.488
30:00.0	0.449
00:00.0	0.424
30:00.0	0.439
00:00.0	0.291
30:00.0	0.262
00:00.0	0.308
30:00.0	0.138
00:00.0	0.404

TABLE IV: Home appliances rating used in housholds

S.No.	**Name of the appliance**	**Rating (KWh)**
1	Fridge	0.2
2	Tubelight	0.055
3	Air conditioner	4
4	Microwave	1.7
5	Dishwasher	1.5
6	Hair dryer	1
7	Sandwich maker	1
8	Vacuum cleaner	1.4
9	Washing machine	0.5
10	Clothes dryer	2.5

A stationary time series data has its mean, variance and other statistical properties constant over a given time frame. The data used here is shown in Table III. It is a collection of data from the source [25]. The aforesaid data help in the visualisation and analysis of the changes. In addition, the future trend of the variables under analysis can be predicted using Machine-learning Algorithms. The entire dataset used is from smart meter energy user data from London households. It had half-hour time stamps and generated energy data in kWh/half-hour. This data comprised 1 million values, from which our final dataset comprising 1,49,999 values was extracted. The dataset used for analysis, is a dataset with 100000 values and a test dataset with 49999 values. A sample of the final data set is shown in Table V. A total 24 appliances were taken into consideration for the analysis.

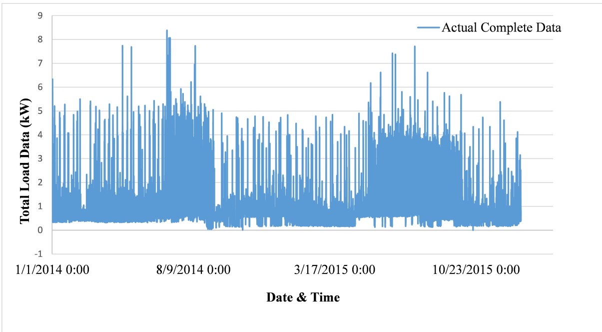


Fig. 4: A stationary time series data has its mean, variance and other statistical properties constant over the given time frame.

B. Prediction using ARIMA model

The combined dataset is shown in Fig. 4 which consists of the data of the years 2014 to 2015. In the beginning, it was predicted based on the electric consumption for the year 2016 using the ARIMA model. Before applying ARIMA model, It is to be made sure that the data is stationary. With the graph plotted in Fig. 4 it can be observed the data is not in stationary mode. The differencing between the present time period and the previous period gives us the first differencing value as shown in Fig. 6 The obtained values are then plotted to check if the statistical properties are constant or not. If still not constant, second differencing is obtained using the first differencing values.

Fig. 6 also depicts the data undertaken and autocorrelation present in the datasets. The method of differencing needs to be applied, repeatedly, till the data obtained becomes stationary. The checking of the correlation between the data with its past values is called autocorrelation. For this, the autocorrelation

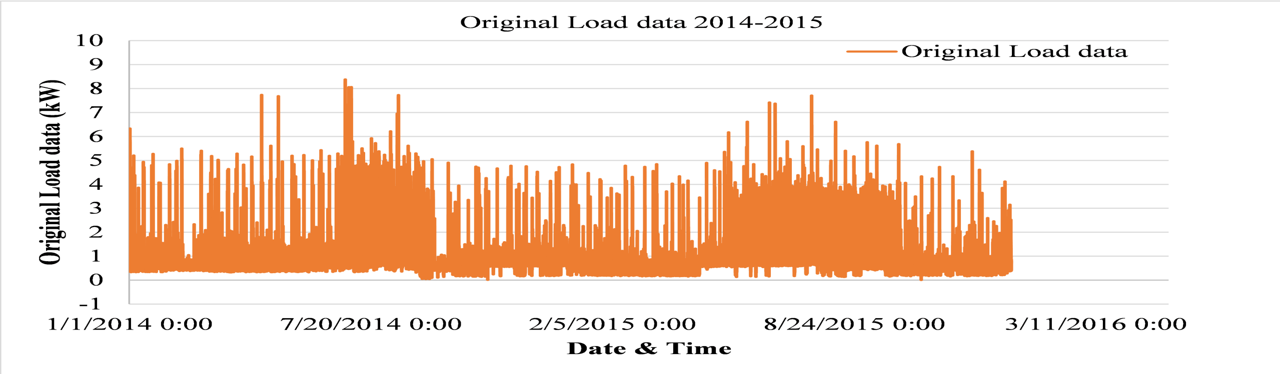


Fig. 5: Original Load Data for 2016

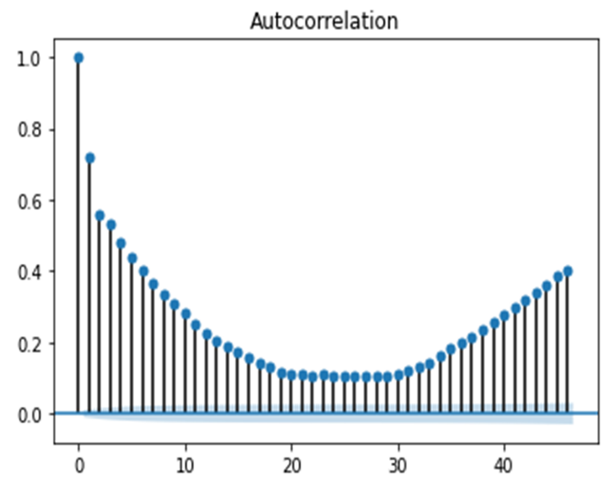


Fig. 6: The first differencing applied to data-Autocorrelation.

function plot (ACF) is used. The plot shows the correlation between various points . The correlation coefficient is plotted on the x-axis with the number of lags on the y-axis. ACF plot is mainly used to determine which one among them is to be used as data. AR model is used in case of positive autocorrelation at lag1 while MA model is used for negative correlation at lag 1.

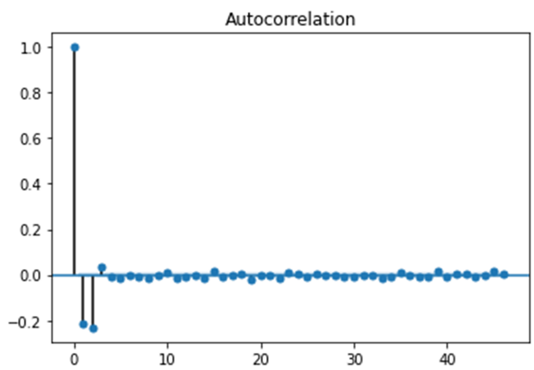


Fig. 7: Autocorrelation present in the datasets.

When differencing is done, the data obtained generally oscillate. This means that the subject data has achieved the

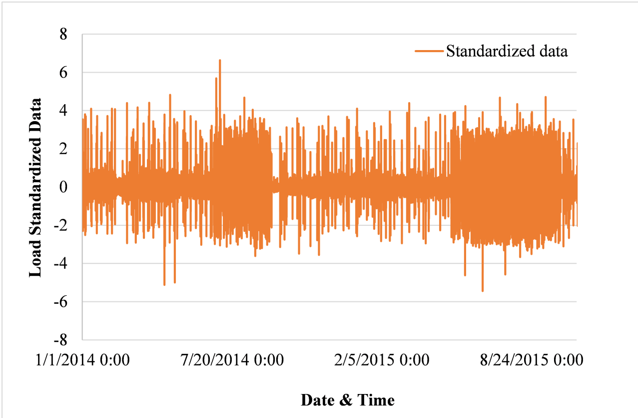


Fig. 8: The standardized data with a prediction of day ahead

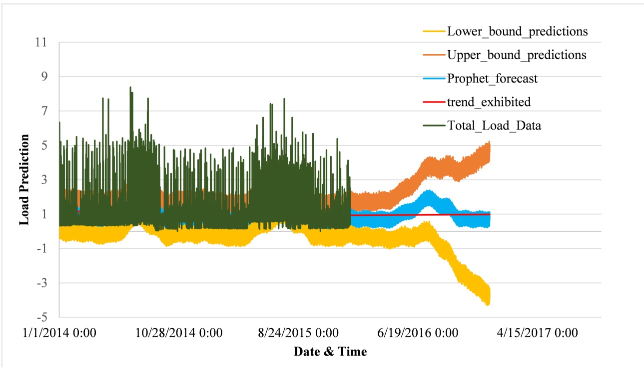


Fig. 9: Prophet forecast analysis between actual and predicted data.

constant values of statistical variables.

That is, the data is now stationary and ready to be worked upon using ARIMA model. Subsequent to making the data stationary, the parameters (p, q, d) are as follows:

- p: order of the AR term,
- q: order of the MA term and
- d: number of differencing required to make the time series stationary

In Fig. 7, the blue line represents the consumption data of predicted values for the year 2016. After extensive compu-

TABLE V: Comsumption dataset from London households

Date: 04/10/2016									
Units of all the columns are in Kw except column 3 in KwH per half-hour, E represents base 10 with its superscript on the right side									
Time	**Energy**	**KWH/hh**	**Genneration**	**AC**	**Furnace**	**Cellar Lights**	**First floor lights**	**Dining room**	**Microwave**	**Total Load**
30:00.0	0.08	0.04	5.78E≈05	0.009531111	0.005335556	0.000125556	0.011175	0.004299444	0.004733	3.10E-01
00:00.0	0.109	0.0545	0.001534444	0.364338333	0.005522222	4.33E-05	0.003514444	0.003588889	0.004445	0.72753611
30:00.0	0.113	0.0565	0.001846667	0.417988889	0.005503889	4.44E-05	0.003527778	0.003521667	0.004396	0.625756112
00:00.0	0.41	0.205	0.001744444	0.410652778	0.005556111	5.94E-05	0.003498889	0.003403889	0.004262	0.78331889
30:00.0	0.054	0.027	3.00E≈05	0.017152222	0.005301667	0.000118889	0.003693889	0.003915	0.004407	1.99E-01
00:00.0	0.043	0.0215	0.000442222	0.12696	0.005415	5.44E-05	0.003626667	0.003812778	0.004398	0.497510554

tation, these three values were set to be (1,0,4). Finally, the model was trained on the data relating to the years 2014-2015 to obtain a prediction for 2016. The data can be considered as stationary if it has constant amplitude and oscillates in the given time frame. In the case of a signal with noise, ARIMA model acts as a filter to extract the data of the signal from the given system. The extracted signal is then worked upon to carry out future predictions. For fitting the data on the ARIMA model, it is important to make the data stationary using the differencing technique to obtain a standardised data as shown in Fig. 8.

C. Prediction using Prophet model

The electricity consumption data of 2014 to 2015 was used to work out with the help of Facebook Prophet Model. The forecast for the data of 2016 was done using the model's inbuilt the predictions. Fig. 5 depicts the data of 2014 to 2015. The model was trained on this data to predict the electricity data for a later year. Fig. 9 summarizes all the computations carried out in the process. The predicted values are in a quite good sync with the original data, and hence forecast was made for the year 2016. There could be certain variations in the real data of 2016, and therefore a margin of upper and the lower bound were created. Fig. 9 indicates the trend of the predicted data. Fig. 10, shows the trend exhibited by the predicted data for the electricity consumption for a year, the seasonality is taken into account. The various components of the Prophet model included were weekly, yearly and daily components and the trend exhibited by the predicted values. Fig. 10 shows the predicted and actual values for 2016. The values predicted, as shown in Table VI with the original 2016 data, are compared for both the ARIMA and the Prophet model. In case of Prophet model, the mean square error is 0.67546 and the mean absolute error is 0.5308. The same is 1.06877 and 0.6239, respectively for ARIMA model. It is seen that the root mean square error and mean absolute percentage error of Prophet model are significantly less than those of ARIMA model. Hence, after comparing the two models, it has been found that the DR from Prophet model has better overall performance.

TABLE VI: Performance comparison of Arima and Prophet for different Evaluation parameters

Model	MSE	RMSE	MAE	MAPE
ARIMA	1.06877	1.03381	0.6239	1.1932
Prophet	0.67546	0.82186	0.5308	1.0399

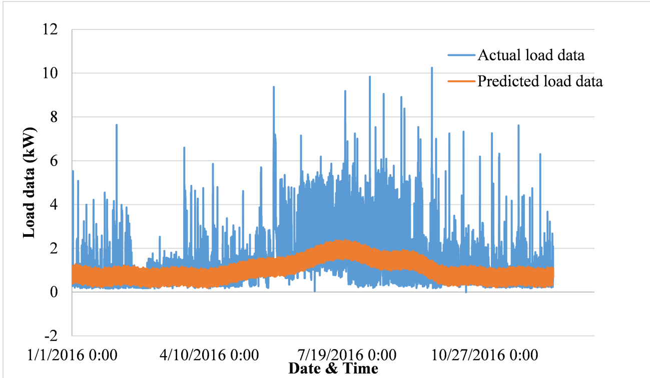


Fig. 10: Comparison between real data and predicted data Facebook Prophet Model.

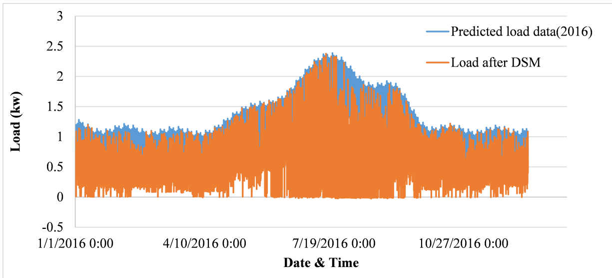


Fig. 11: Demand Side Management on predicted data

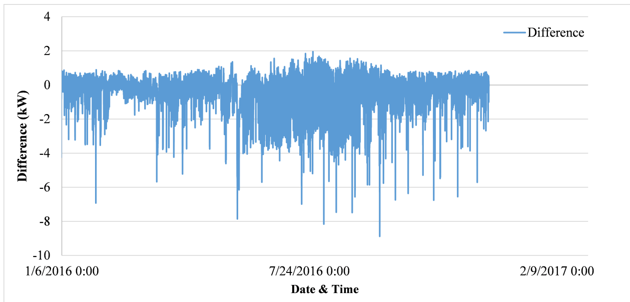


Fig. 12: The difference between actual and predicted data

Facebook's Prophet Model is selected for carrying out DSM in the subject research. The data was fed to the appliances at home and the appliances swiched off one-by-one on the basis of priority of usage. Fig. 11 shows the DSM on the predicted values. The demand response was carried out between the total load and the generated electricity. For the DSM, the difference between the original and the predicted values of the year 2016 is calculated and presented as shown in Fig. 12. The trend, over different time lengths, is shown in Fig. 13 and can be useful for controlling the electric consumption of the appliances at households.

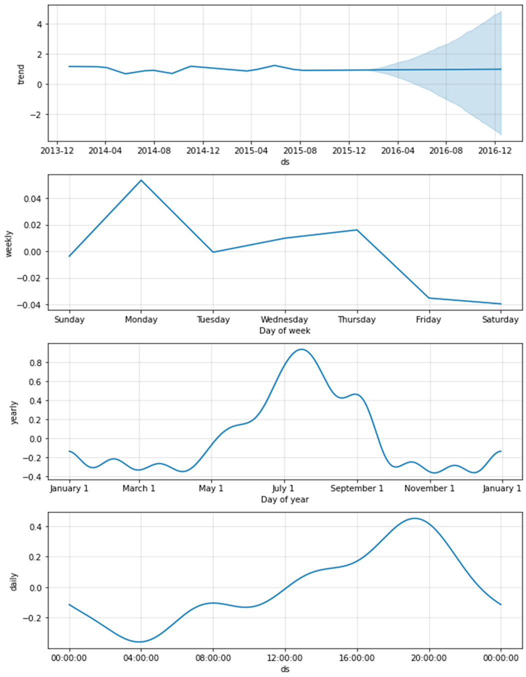


Fig. 13: Charts (top to bottom): Trend, Weekly, Yearly and Daily-datasets used in the Prophet model

V. Conclusion

The forecasting of demand and supply of electrical energy is essential and critical in order to achieve efficiency and econmomy in the running of an SG. The time series data available for the purpose in a real life situation is, at times, discontinuous and the same has to be the basis for a forecast in the absence of alternatives. The ARIMA model is seen providing a good forecast data provided the data available is complete. But the same is highly expensive as its computation requires computer systems with memory and computational speeds of a very high order. The Prophet model while offsetting all the above limitations of the ARIMA model generates a reliable forecast even in the absence of a few values in the data. Therefore, it is felt that the Prophet model is preferable compared to ARIMA in a real life situation. There is a need to find ways to improve on the Prophet model due its fexibiliy, economy and reliability.

References

[1] V. Badrinath Krishna, R. Iyer, and W. Sanders, "Arima-based modeling and validation of consumption readings in power grids," 05 2016, pp. 199–210.

[2] G. Newsham and B. Birt, "Building-level occupancy data to improve arima-based electricity use forecasts," *BuildSys'10 - Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, 01 2010.

[3] S. Chatterjee, S. Dey, and M. Dasgupta, "A solution to demand response with soft-computing techniques," in *2020 Fourth International Conference on Inventive Systems and Control (ICISC)*, 2020, pp. 122–126.

[4] H. Liu, *Smart Cities: Big Data Prediction Methods and Applications*. Springer Singapore, 2020.

[5] A. Viloria, R. Prieto Pulido, J. Guiliany, J. Ventura, H. Palma, J. Torres, O. Bilbao, and O. Lezama, *Analyzing and Predicting Power Consumption Profiles Using Big Data*, 11 2019, pp. 392–401.

[6] D. Asaleye, M. Breen, and M. Murphy, "A decision support tool for building integrated renewable energy microgrids connected to a smart grid," *Energies*, vol. 10, p. 1765, 11 2017.

[7] Z. Wang, J. Qu, X. Fang, H. Li, T. Zhong, and H. Ren, "Prediction of early stabilization time of electrolytic capacitor based on arima-bi_ lstm hybrid model," *Neurocomputing*, vol. 403, 04 2020.

[8] A. Gupta, K. Sharma, A. Vijayvargia, and R. Bhakar, "Very short term wind power prediction using hybrid univariate arima-garch model," 12 2019, pp. 1–6.

[9] X. Luo, X. Zhu, and E. Lim, "A hybrid model for short term real-time electricity price forecasting in smart grid," *Big Data Analytics*, vol. 3, 10 2018.

[10] R. Pal, C. Chelmis, M. Frincu, and V. Prasanna, "Towards dynamic demand response on efficient consumer grouping algorithmics," *IEEE Transactions on Sustainable Computing*, vol. 1, no. 1, pp. 20–34, 2016.

[11] A. Y. Bishnu Nepal, Motoi Yamaha, "Electricity load forecasting using clustering and arima model for energy management in buildings," in *Wiley Online Library*, 2019, pp. 1–6.

[12] C. C. Saima Aman, Marc Frincu, "Prediction models for dynamic demand response-requirements, challenges, and insights," in *IEEE conference*, 2019, pp. 1–6.

[13] J. Zhan, J. Huang, L. Niu, X. Peng, D. Deng, and S. Cheng, "Study of the key technologies of electric power big data and its application prospects in smart grid," in *2014 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 2014, pp. 1–4.

[14] T. Y. Ku, W. K. Park, and Choi, "Demand response operation method on energy big data platform," in *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*, 2018, pp. 823–825.

[15] Y. Wang, Q. Chen, C. Kang, Q. Xia, and M. Luo, "Sparse and redundant representation-based smart meter data compression and pattern extraction," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2142–2151, 2017.

[16] X. Zhang, D. Li, M. Cheng, and P. Zhang, "Electricity consumption pattern recognition based on the big data technology to support the peak shifting potential analysis," in *2014 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 2014, pp. 1–5.

[17] V. A. H. Vajjala, "A novel solution to use big data technologies and improve demand response program in aggregated residential houses," in *2016 IEEE Conference on Technologies for Sustainability (SusTech)*, 2016, pp. 251–256.

[18] S. Singh and A. Yassine, "Mining energy consumption behavior patterns for households in smart grid," *IEEE Transactions on Emerging Topics in Computing*, vol. 7, no. 3, pp. 404–419, 2019.

[19] A. Almazrouee, A. Almeshal, A. Almutairi, M. Alenezi, and S. Alhajeri, "Long-term forecasting of electrical loads in kuwait using prophet and holt-winters models," *Applied Sciences*, vol. 10, p. 5627, 08 2020.

[20] B. Sandrić, M. Marcelić, I. Topolovac, and M. Jurčevic, "Survey of data cleaning algorithms in wireless sensor networks," in *2019 2nd International Colloquium on Smart Grid Metrology (SMAGRIMET)*, 2019, pp. 1–4.

[21] B. Hrnjica and A. D. Mehr, *Energy Demand Forecasting Using Deep Learning*. Cham: Springer International Publishing, 2020, pp. 71–104. [Online]. Available: https://doi.org/10.1007/978-3-030-14718-1_4

[22] G. Grigoraş, G. Cârţină, and E.-C. Bobric, "Trends and directions for energy saving in electric networks," in *Management of Technological Innovation in Developing and Developed Countries*, H. Sun, Ed. Rijeka: IntechOpen, 2012, ch. 1. [Online]. Available: https://doi.org/10.5772/35516

[23] J. Joseph W. Forbes, "System, method, and apparatus for electric power grid and network management of grid elements," in *US2014/0039699A1*, 2014, pp. 13/563,535.

[24] S. Iqbal, M. F. Bari, and M. Rahman, "Solving the multi-dimensional multi-choice knapsack problem with the help of ants," vol. 6234, 09 2010, pp. 312–323.

[25] online https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london households.