A Hybrid Energy Optimization Approach for Home Appliances

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering

in

Computer Science and Engineering

Submitted by

Jasmeet Kaur

(**Roll no: 801632016**)

Under the supervision of:

Dr. Anju Bala

Assistant Professor, CSED

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY
PATIALA-147004

June 2018

Certificate

I hereby certify that the work which is being presented in the thesis entitled, *"A Hybrid Energy Optimization Approach for Home Appliances"*, in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Computer Science and Engineering* submitted in Computer Science and Engineering Department of Thapar Institute of Engineering and Technology (Deemed to be University), Patiala, is an authentic record of my own work carried out under the supervision of *Dr. Anju Bala* and refers other researchers work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

Jasmeet Kaur

Signature:

(Jasmeet Kaur)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

Anju Bala

Dr. Anju Bala

Assistant Professor,

CSED

Abstract

Energy Management refers to analyzing and saving the energy by monitoring and controlling the energy profiles. Energy Management has gained its importance in every sector such as transportation sector, agriculture sector, residential sector and industrial sector. Energy Management in dwellings has become the recurrent issue nowadays due to an increase of electrical appliances with the upcoming new technologies. The energy consumption of these appliances depends on various factors such as climatic conditions, the number of occupants and their behaviour, the usage of appliances in the homes etc. The energy demand has increased in the residential sector and the energy suppliers are not able to fulfil the need of energy which is leading to power cuts. Further, the appliances emit the harmful radiation and leading to greenhouse gas emissions. So, there is the need to predict the energy consumption in the residential sector and to optimize the energy in such a way such that energy supplier would be able to supply the electricity in dwellings.

Though many researchers have worked on the energy consumption and its optimization, a very few have taken into account the climatic conditions for its prediction. Hence, the motive of this dissertation is to predict and to optimize the energy consumption for the suppliers. For this, it was important to recognize the hidden patterns in which the appliances are being used in the house. For finding the hidden patterns, principal component analysis with K-means clustering was performed and three clusters were formed which consist of the appliances having high energy consumption, moderate energy and the continuous energy consumption. The prediction models were implemented for predicting the energy consumption at appliance level after integrating the cluster according to different climate conditions such as temperature, wind speed, visibility etc. Finally, the hybrid optimization was performed for minimizing the value of energy consumption. This dissertation will be benefits to the energy suppliers for providing the electricity demand to the end user and also will be benefits to the households for completion for their daily activities.

Acknowledgements

I would like to express my deep gratitude to my supervisor **Dr. Anju Bala** for their invaluable advice and encouragement at every step of my M.E. program. Without her unfailing support and belief in me, this thesis would not have been possible. Their contribution to this thesis goes well beyond their role as an academic supervisor and includes constant support on a personal level without which this journey may never have been completed. And for this, I am truly grateful.

I would like to express my gratitude to the Professor and Head, CSED **Dr. Maninder Singh** for his constant motivation and encouragement.He sets high principles for his students and motivates and guides them to meet those principles.

Before ending I would like to thank my parents and friends for their love, motivation, support and blessings. They have been a constant source of love, concern, support and strength for me all these years.

Finally, I would like to thank the management of Thapar Institute of Engineering and Technology for providing me a great opportunity for learning, not just in academics but also in many other creative things.

I sincerely regret any inadvertent omissions. With my heartiest thanks to all.

Jasmeet Kaur

Table of Contents

List of Figures

List of Tables

List of Notations

X^i	number of principal components
A_i	number of appliances
R^2	coefficient of determination
err	error rate
p_i	particle update
v_i	velocity update
w	inertia of weight
d	number of parameters
f_i	food source
y_i	current source
D	reduced dataset
k	number of clusters
P_i	probability
Y_i	standard variable
F_j	common factor
P	total power

List of Abbreviations

ABC_ANN	Tuning of ANN with ABC
ABC	Artificial Bee Colony
AKMA	Adaptive K-Means Algorithm
ANN	Artificial Neural Network
BN	Bayesian Network
BPSO	Binary Particle Swarm Optimization
CA	Classification Algorithm
CS	Cluster Splitting
DEMS	Demand-side Energy Management
EMM	Energy Management Model
EWMA	Exponentially Weighted Moving Averages
FOA	Fruit Fly Optimization Algorithm
FOAGRNN	Generalized Regression Neural Network with FOA
GA	Genetic Algorithm
GAMs	General Additive Model using Splines
GHEMS	Green Home Energy Management
HC	Hierarchical Clustering
HEMS	Home-side Energy Management
knnr	K-nearest Neighbour Regression
LC	Load Clustering
LDA	Linear Discriminant Analysis
LR	Linear Regression
LSO	Load scheduling optimization
MILP	Mixed Integer Linear Programming
MLDM	Multidimensional Linear Discriminate Method
MLP	Multi-layer Preceptron
MSC	Mean Shift Clustering
OLS	Ordinary Least Squares
PCA	Principal Component Analysis

PCs	Principal Components
PLS	Partial Least Square Discriminant Analysis
PSO_ANN	Tuning of ANN with PSO
PSO	Particle Swarm Optimization
PSO	Particle Swarm Optimization
R^2	Coefficient of Determination
RM	Regression Method
RMSE	Root Mean Square Error
SEM	Supply-side Energy Management
SFOASVR	Seasonal SVR with Fruit Fly Optimization
SRSVRCABC	Seasonal Recurrent SVR with Chaotic Artificial Bee Colony
SVMs	Support Vector Machine
SVR	Support Vector Regression

Chapter 1

Introduction

This chapter is about the overview of energy prediction and its management methods. In this, different methods are explained for predicting the energy and different types of machine learning algorithms are discussed.

Energy consumption and its management have gained importance with the intensifying energy cost and its environmental effects. The usage of energy has been required in every sector such as residential sector, the transportation sector, agriculture sector and industrial sector. With the rise of the technology, the usage of electronic appliances has been increased. People are using electronic appliances in every field for the basic activities. The residential sector is using most of the electronic appliances for carrying out their daily activities with has been leading to rising in electronic bills and the greenhouse gas emission. The residential sector contributes to the 40% of CO_2 emission [3]. There have been various indicators that contribute to the usage of appliances in residential sector i.e. Climate, number of occupants and their behaviour, geographical conditions etc. Climatic

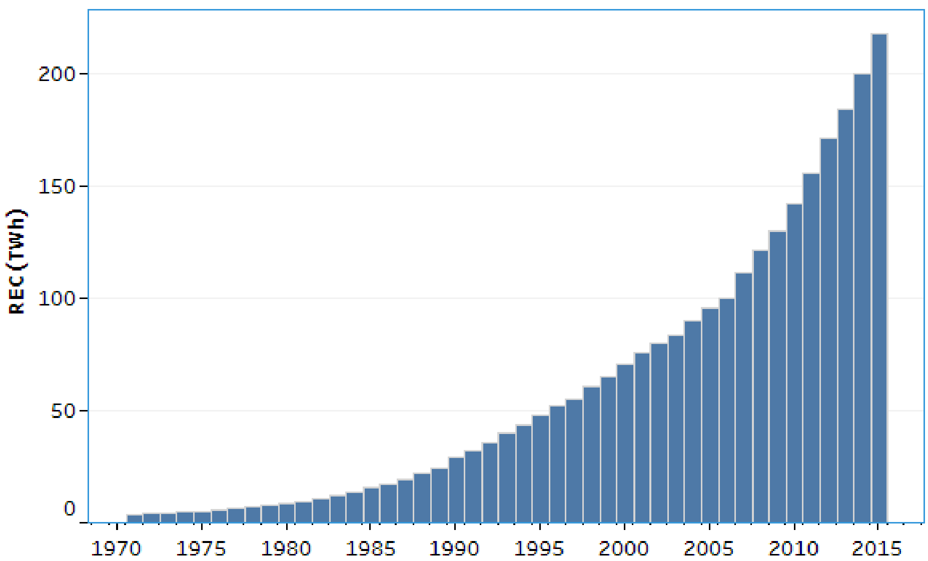


Figure 1.1: Residential Electricity Consumption Trend [2]

conditions have an adverse effect on the usage of electronic appliances in homes. The usage of appliances and the climate are an inter-related process. Residential sector energy consumption includes different residential appliances, cooking activities, space heating and cooling appliances. It is becoming difficult for the energy supplier to supply the electricity to the households all the time for fulfilling the basic demand for the households. The Figure 1.1 embellishes about the electricity consumption in the residential sector. The consumption in the residential sector has increased from 200 TWh to 250 TWh in the year 2015. From the Figure 1.1 it can be clearly inferred that electricity demand in the residential sector is ever increasing. So, it is important to predict the energy consumption in the residential sector and to optimize the energy usage so that the energy supplier should able to provide the minimum demand of the households for the accomplishment of their daily basic activities.

1.1 Energy Management System

Energy management refers to the energy saving by analysis the energy usage in different ways such as supervising, controlling and sustaining the energy in the various field i.e. Residential field, Industrial field etc. Energy management is also done by the energy suppliers for meeting the demand of the dwellings and for generating their electricity. The energy management system are used by the energy supplier for maintaining the electricity supply in different fields without any hindrance. Energy Management is performed for minimizing the cost and helps in mitigating the effects on the climate due to increase in energy consumption. There are different benefits for energy management which are as follows [5]:

- Carbon Dioxide Reduction: The reduction in energy consumption helps in reducing the carbon emission and leads to decreasing its effect to the climate.
- Cost Reduction: The energy saving helps in the reduction of cost and reducing the electricity bills in any sector and giving rise to green environment.
- Risk Reduction: With energy saving helps in reducing its shortage in the peak periods and prevent to rise in the price of electricity.

1.1.1 Types of Energy Management

There are following ways to manage the energy such as home-side energy management (HEMS), demand-side energy management (DEMS) and supply-side energy management (SEM). They are explained below:

- **Demand-side Energy Management:** The focus of the DEMS to the save the energy by reducing the usage of electricity during the peak periods. it helps in modifying the demand of the end-users. By promoting the DEMS can leads to reduce the environmental impacts such as increase in the global warming. DEMS can be used by analysing the consumption of energy by individual electrical appliances in homes or in any other sector.

- **Supply-side Energy Management:** The main focus of SEM to the select between the different form of energy such as solar energy, wind energy, thermal energy etc. and than planning and configuring the energy in the most appropriate way such that it helps in the reduction of cost as well in the reducing its impact on the environment. SEM can be helpful in developing the buildings or an organisation which can leads to effective usage of the energy.

- **Home-side Energy Management:** The HEMS allows the users to optimize and manage the energy consumption which can be helpful in reducing the electricity bills effectively. It helps the consumer to know about the usage patterns of energy of their households by monitoring the consumption of whole dwelling so that, they can make use of energy consumption in an most efficient way.

1.2 Energy Audits and its Types

Energy Audit is a process to study about the energy consumption and wastage in an organisation or in any sector such as residential sector, industrial sector, transportation sector and agriculture sector. It aims at reduction of energy without affecting the comfort of the user and reducing the energy cost. They are carried on in a planned manner by energy suppliers or by the organization contributing to the energy providers.

1.2.1 Objectives of Energy Audits

There are various objectives of energy audit and are as follows:

- Energy audits helps to identify the loss of energy.
- It helps to establish the costs of various forms of energy purchased.
- It helps to identify the areas/sectors where the energy consumption is more and its wastage in that sector.
- It involves the high examination and analysis in the field of different form of energy in any area.

1.2.2 Types of Energy Audits

The energy audits are of two types:

- **Preliminary Energy Audit:** In this, the energy consumption analysis are done for a limited period of time such as for 10 days and highlights the energy cost and its wastage in any sector of the equipments used. It helps in a account in the demand of the energy supply.
- **Detailed Energy Audits:** In this, the energy consumption analysis for any duration is done for accounting the amount of energy used in any sector. The detailed study is done for reducing the energy consumption and its cost. For the detailed analysis of energy consumption it has been divided into following types:
  - **Utility Audit:** The analysis of energy consumption is done in terms of yearly,monthly energy consumption and suggestion about the energy saving are provided in any sector.
  - **Functional Audit:** The maintenance and operation of energy plant is suggested and the energy conservation measures are provided for its operations.

- **Overall system Audit:** The energy loss is analysed in different sector and the suggestions are made on the cause of the energy loss.
- **Modernization Audit:** The recommendation are made on the equipments usage for the energy saving process.

1.3 Methods of Dimensionality reduction with clustering

Dimensionality reduction is the the technique to extract or select the feature from high dimensionality data set which consist of high variability by procuring the set of different variables. It helps in the reduction of the variables in machine learning and statistics [1] and perpetuate the original data in the useful form. The importance of dimensionality reduction has increased because of the large existing data sets. In dimensionality reduction, the counsel prevalent in the high dimension data is extracted and exploited by certain dimensions depending on the technique used. It is cleaved in two categories:

- **Feature Selection:** It is used for converting the data sets consisting of high dimension data to a smaller subsets for creating a model. The methods used for selecting the features are filter, wrapper and embedded for building the model [1].
- **Feature Extraction:** It is used for transfigure the data sets consisting of many dimensions to a fewer dimensions. It includes the method of principal component analysis, linear discriminant analysis and generalized discriminant analysis.

1.3.1 Techniques of Dimensionality Reduction in Feature Extraction

There are different techniques of feature extraction in reducing the dimensions of data sets consisting of high dimension data. The techniques are used depending on the requirement of the machine learning model. They are categorized as supervised and unsupervised technique.

- **Supervised Technique:** Supervised technique refers to the findings of the data in the low-dimensional representation which can be used as the supervised learning task. It helps to prevent over-fitting of the data and provide the necessary interpretation and visualisation of the data. There are various methods for supervised dimensionality reduction such as linear discriminant analysis (LDA), Partial least squares discriminant analysis (PLS).

  - **Linear discriminant analysis (LDA):** It is used for the classification problems having more than two classes. LDA is the method used for pattern recognition and in statistics. It is also used for prediction in machine learning models for characterizing or separating the two or more events. It takes into account the scatter consideration within the classes as well as between the classes [44]. Firstly, it assumes the data in the linear space and secondly, the existence of linear separation between the classes. LDA is used for the separation between the multiple classes. The main constraint of LDA is that the number of feature should not exceeds the number of objects.

  - **Partial least squares discriminant analysis (PLS)**: It is used for finding the relation between dependent and independent variables in multivariate method. It helps to find the relation between the input variables and the different components of the input matrix. It is also used for correlation between the input matrix and the target value. It searches the different feature in the multivariate dimensions and developed the classes between them. Optimal number of feature are searched using the cross-validation method [11]. It is useful for the data set having less features.

- **Unsupervised Techniques:** It is based on the input data without having any fixed target set. It helps in generating the groups by matching the hidden patterns in the input data and grouping them into the clusters or in the classes. There are various methods for unsupervised dimensionality reduction such as principal component analysis, factor analysis.

  - **Principal Component Analysis:** It is the unsupervised technique for reducing the dimensions of the input data set. It matches the hidden patterns in the input data and group them accordingly. It describes the relationship between the quantitative variables and visualize the information contained in the data. Each feature is consider as the different dimension in principal component. Dimensions in the PC is equal to the number of feature the input data consist. The main goal of the principal component is to identify the variations in the data. It is also used to remove the redundancy and noise in the data and also in finding the relationship between correlated features. The basic working of principal component analysis is explained in the Figure 1.2. Firstly, the input data is preprocessed and the variables of the data are scaled the output matrix is obtained. The recognition of the components are found using the eigen values and eigen vector method. The data in PCA is determined by the usage of axis and help in the formation of principal components that amounts the variance of the inputs with each other.

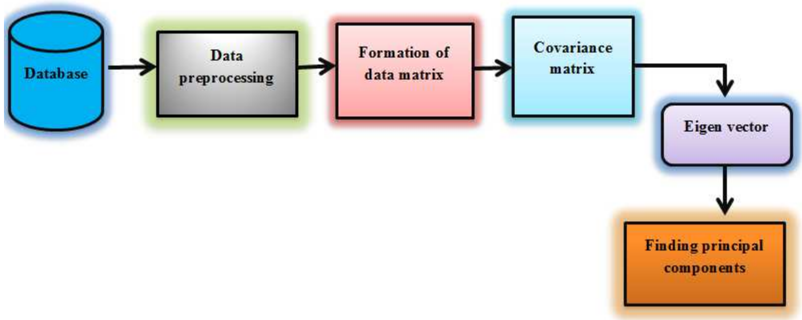


Figure 1.2: Principal component analysis working

- **Factor Analysis:** It is the procedure of unsupervised technique for reducing and summarizing the data. It doesn't make any distinction between the independent and dependent variables. The usage of factor analysis is to make the factors by identifying the correlation between the different features present in the data and to analyse the new set of uncorrelated variables with are present between the uncorrelated variable in the original data. The factors formed in this analysis are the combination of some common factors with the unique factors. The features are identified by the linear combination of factor. The model is represented by the equation (1.1):

$$Y_i = \alpha^{j1} F_1 + \alpha^{j2} F_2 + ... + \alpha^{jn} F_n + W_i U_i \tag{1.1}$$

where α_{ij} is the coefficient of variable on a common factor j, Y_i is the standard variable F_j is the common factor W_i is the coefficient of variable on a unique factor i U_i is the unique factor of variable i n total number of common factors.

1.3.2 Types of Clustering

Clustering refers to the grouping of data consisting of same patterns or objects. Clustering involves the analysis of multivariate data. It can be used in many fields such as biomedical, image reconstruction, energy patterns recognition , credit fraud detection etc. It is also used for finding the outliers in the data. There are different types of clustering and are explained below:

- **Hierarchical Method :** It creates the hierarchical decomposition of the data. For energy decomposition the hierarchy of the usage of appliances can be made for recognising its usage patterns. It has been further divided into two types:
  - **Agglomerative Approach :** In this, each feature has been into different group initially and than keep on merging till the feature with same patterns do not come under the one group or till the group group of feature has not been made. It has been known as bottom-up approach.
  - **Divisive Approach :** In this , each feature has been in one group initially and than keeps on separating according to their conditions until the condition is satisfied. This approach has been known as top-down approach.
- **Density-based Method:** This method of clustering has been based on threshold value. Each cluster in this grows when the density of its neighbour cluster exceeds by some threshold value.
- **Partitioning Method:** In this, the data is partitioned into different clus-

ters. If a data consist of 'n' features than 'k' patitions are made such that k$\leq$n. Each feature of data much contain in one cluster and each group must have one feature.

- **Grid-Based Method:** In this , the variables in the data form a grid based on the dimensions in the object space. A finite number of cells are quantized in the object space.

- **Model-Based Method:** This method of clustering automatically determines the number of clusters formed by finding the outlier or the noise in the data. It locates the cluster by using the density function.

1.4 Machine Learning Algorithms

Machine learning is an process that allows the system to work automatically with being explicitly programmed. It constructs the model according to the machine learning algorithm type. It divides the data into training, testing and validation according to the user requirement. It detects the patterns of the data and allow the computer to learn for future prediction. The historical data is used for the prediction of the applications. The process of machine learning has been depicted in the Figure 1.3. Machine learning has been used for weather forecasting, stock prediction, energy forecasting and for many other such types of applications. Machine learning algorithms are divided into three types:

- **Supervised Machine Learning Algorithms:** It can appertain to the known data and predict the output value by using the labelled data provided. It can be used for the analysis of known training data set. Model is provided with the target value on which the future prediction is to be made and rest attributes act as input to the model. Model is trained and then prediction is made by providing the input values.

- **Unsupervised Machine Learning Algorithms:**: It deals with the unlabelled data and explores the data and draws the inferences from it. It can be used to infer the hidden patterns from the data by using different techniques such as clustering, hidden markov models, principal component analysis etc.

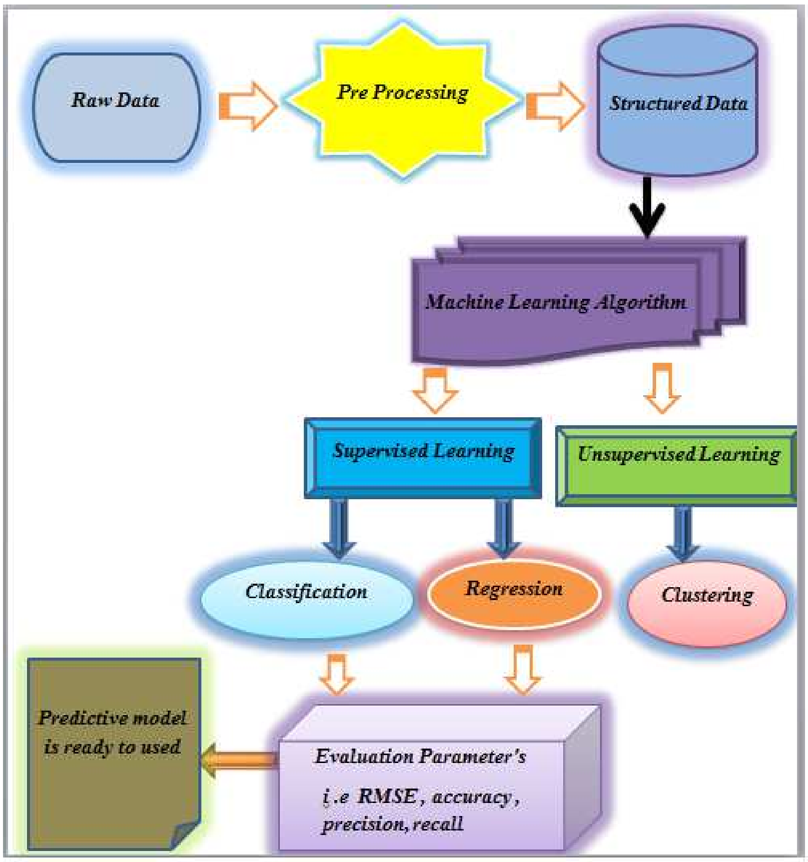


Figure 1.3: Machine learning process

- **Reinforcement Machine Learning Algorithms:** It can be cast-off for the learning method to interact with its environment by producing actions and discover error or records [4]. It determines the ideal behaviour within a context in order to maximize its performance. It gives reward to the most relevant characteristics.

This chapter was about the introduction of energy management techniques and the methods used for predicting the energy. It overviews about the need of energy prediction and management. In this, different machine learning types were also discussed along with the methods.

1.5 Thesis Organization

The thesis is organized into 6 chapters. A brief outline is given below:

- **Chapter 1**: This chapter has been about the basic introduction and machine learning process. It explains about the different energy management methods. It also give the details about the techniques/methods used before the energy prediction models.

- **Chapter 2**: Chapter 2 describes the literature review of the machine learning prediction methods and optimization techniques.

- **Chapter 3**: In this, the problem was discussed which was seen in the analysis done in Chapter 2. The research gaps were discussed and the objectives were formulated to overcome the research gaps and the problems faced.

- **Chapter 4**: This chapter describe the proposed methodology to meet the required objectives formulated in Chapter 3. The steps and the techniques used for the energy prediction and optimization are discussed in detailed and the algorithms are formulated.

- **Chapter 5**: This chapter study the performance of technique and the prediction models. The results of the models were compared with different evaluation parameters. The hybrid optimization was implemented and results are compared with the best model.

- **Chapter 6**: The last Chapter gave the conclusion from the implementation and the result formed for achieving the desired objectives for this thesis. The future scope of the thesis have been discussed.

Chapter 2

Literature Review

In this section, the work done by various researchers have been discussed on energy prediction and optimization. With the advent of digitalization, the occupants depend on the electrical equipment for their basic activities which lead to the increase in the energy consumption in dwellings so the dire need for energy prediction came into existence. The main aim for energy prediction and optimization is to provide the information to energy supplier about the amount of power to be supplied in the households so that the occupants are able to their households chores with ease.

2.1 Energy Management: Existing Methods used in Residential Sector

There are various methods used in the residential sector for energy management. The different methods are focused on different types of management in households. Some of the methods has been discussed below along with there approach used for managing the energy in households:

The Table 2.1 describes the previous work done on machine learning techniques. It describes the work of authors with their different approaches for energy management such as HEMS, DEMS etc. M. Aiad et al. [7] has worked on the non-intrusive load monitoring method for energy saving. The main focus was on the the appliances with coincidence power consumption. For this, author proposed a cluster splitting approached based on cluster cohesion. L. Yao et al. [54] develops an energy management system for home with an aim to solve the problem of energy grids and the load scheduling problems for occupants with the mixed integer linear programming algorithm. The cost minimization was achieved by the authors with the energy management system by the controlled in the peak hours. P. Tsai et al.[49] used a genetic algorithm for optimizing the energy efficiency of the appliances. J. Flora et al. [31] initiated adaptive k-means algorithm by considering the load shapes for measuring the power consumption of appliances based on hourly basis. The author in [6] propound a demand-side energy management for a grid connected households with a locally generated photovoltaic energy. W. Kleiminger et al. [27] investigated the occupancy detection by undertaking the ground truth occupancy data. Z. Wang et al. [52] proposed the mean-shift clustering and multidimensional linear discriminate method.

J. Han et al. [21] developed a home energy management system by comparing the energy of the same household appliances. H.Murata et al. [38] used the non-intrusive monitoring system and the regression methods for calculating the accuracy of the appliances. However, previous studies have not took all the household electrical appliances together.

Table 2.1: Existing Methods for Energy Management

Author	**Technique Used**	**Application**	**Approach**	**Challenges**
M.Aiad et al. [7]	Cluster Splitting	Non-intrusive load monitoring	HEMS	• To estimate the power consumption of appliances that switch there state frequently. • To measure the power of continuously varying loads appliances.
L.Yoa et al. [54]	MILP	Optimal energy management	HEMS	• Scheduling of all household appliances in all the environmental conditions with different load characteristics.• Integrate the energy saving method with carbon reduction framework.
P.Tsai et al. [49]	Genetic algorithm	Scheduling	HEMS	• Calculate the energy consumption for all household appliances.• Diversify the energy saving resources.
J.Flora et al. [31]	Adaptive K-mean algorithm, Hierarchical clustering	Segmentation	DEMS	• Integrating the energy saving.

to be cont'd on next page

Table 2.1: Existing Methods for Energy Management (Cont.)

Author	**Technique Used**	**Application**	**Approach**	**Challenges**
L.Wang et al. [6]	Load clustering, Load Scheduling Optimization	Autonomous appliance scheduling, Prosumer concept	DEMS	• Develop an energy efficiency method with a carbon reduction framework.
W.Kleiminger et al.[27]	Classification Algorithm	Occupancy Detection	HEMS	• To detect the home appliances during sleep hours.• Investigation of sensor fusion methods.
Z.Wang et al. [52]	Mean Shift Clustering, MLDM	Non-intrusive	DEMS	• To find an innovative and more of identification of appliances.
J.Han et al. [21]	GHEMS	Energy comparison of appliances	HEMS	• To diversify energy efficiency of home appliances.Increasing the magnitude of saving the energy instead of replacing them.
H.Murata et al. [38]	Regression Methods	Estimating the states of household appliances	HEMS	• To estimate the power consumption of all electrical appliances.

The different appliances has been used by different authors for energy management in residential sector using different methods. The Table 2.2 elaborates the appliances used by the authors using different methods for energy management. There are some appliances such as Refrigerator, TV, Air Conditioner have been implemented by many techniques such as Load Clustering, Multidimensional linear discriminate method(MLDM), Cluster Splitting etc. So, there is a dire need for energy management by considering all of the possible home appliances.

Table 2.2: Appliances used in Different Methods

Techniques Used	**CA**	**RM**	**CS**	**MSC**	**MLDM**	**AKMA**	**HC**	**LC**	**LSO**	**MILP**	**GA**
Refrigerator	✓	✓	✓	✓	✓	×	×	✓	✓	×	×
TV	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×

Table 2.2: Appliances used in Different Methods (Cont.)

Techniques Used	**CA**	**RM**	**CS**	**MSC**	**MLDM**	**AKMA**	**HC**	**LC**	**LSO**	**MILP**	**GA**
Washing Machine	✓	×	×	✓	✓	×	×	✓	✓	×	✓
AC's	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	×
Dishwasher	×	×	✓	×	×	×	×	×	✓	✓	✓
Chimney	×	×	✓	×	×	×	×	×	×	×	×
Lights	×	✓	✓	×	×	✓	✓	✓	✓	×	×
Fan	×	✓	✓	✓	✓	×	×	✓	✓	×	×
Water Filters	×	×	✓	×	×	×	×	×	×	×	×
Grinder	×	×	✓	×	×	×	×	×	×	×	×
Mixer	×	×	✓	✓	✓	×	×	×	×	×	×
Toaster	×	×	✓	×	×	×	×	×	×	×	×
Kettle	✓	✓	✓	×	×	×	×	×	×	×	×
Iron	×	×	✓	×	×	×	×	×	×	×	×
Routers	✓	×	×	×	×	×	×	×	×	×	×
Printer	×	×	×	×	×	×	×	×	×	×	×
Geyser	×	×	✓	×	×	×	×	×	×	×	×
Heaters	×	×	✓	✓	✓	✓	✓	✓	✓	×	×
CCTV's	×	×	×	×	×	×	×	×	×	×	×
Hair Dryer	×	×	×	✓	✓	×	×	×	×	✓	×
Bulbs	×	×	✓	×	×	×	×	×	×	×	×
Cooler	×	×	×	×	×	×	×	×	×	×	✓
Microwave Oven	✓	×	✓	✓	✓	×	×	✓	✓	×	×
Straightener	×	×	×	×	×	×	×	×	×	×	×
Smoke Alarms	×	×	×	×	×	×	×	×	×	×	×
Tumble Dryer	✓	×	✓	×	×	×	×	✓	✓	×	×
Induction Cooker	×	×	×	✓	✓	×	×	×	×	×	×
Vacuum Cleaner	×	×	×	×	×	×	×	×	×	×	✓

2.2 Energy Management: Machine learning algorithms

Energy prediction play a crucial role in every sector i.e. transport sector, residential sector, agriculture sector, industrial sector etc. The residential sector energy demand has been increasing with each passing year as the usage of electronic appliances has been increased in the dwellings. The residential sector prediction can be made on a particular house, residential buildings consisting of different apartment. The usage of electricity depends on the various factors such as number of occupants and their behaviour, climatic conditions, demographic location etc.

The pattern in which the appliances have been used daily plays a crucial role in the prediction of energy.
Table 2.3 describes the different models used by authors for predicting the energy consumption in the residential sector. Multi-Layer Preceptron and Random Forest has been used by the authors [50] for the prediction of power consumption in residential buildings. The accuracy of multi-layer preceptron achieved was 95 % and that of random forest was 90.83 %. Basu et al. [10] worked on to forecast of the appliance usage at a particular hour. The approach is restricted to the prediction of the appliance usage by taking only its consumption and time the appliance was used. The prediction was based on whether the appliance will be turned on at particular interval of time.
Zhoa et al. [55] explored on the different methods used for the prediction of energy consumption for building and elaborates the work on widely used different methods such as Artificial Neural Networks (ANNs) and Support Vector Machine (SVMs). Kwac et al. [31] proposed the segmentation method for energy consumption using hourly data of households by taking the load shapes of the energy consumed in different hours. The authors used the two clustering techniques i.e. Adaptive K-mean based on threshold and hierarchical clustering and concluded the lifestyle pattern during different periods of time in a day. Various kinds of approaches have been used by the authors for predicting the energy consumption. Tian et al. [48] applied the ordinary least squares(OLS) and spatial regression analysis in domestic sector for urban areas. The regression analysis and Lagrange multiplier statistical test were performed on electricity and gas used based on the council tax and domestic energy consumption. The simulation was performed to indicate the regression analysis performance.

The authors [28] have used the different time series models to forecast energy consumption based on different methods to utilize the energy. The prediction done was on renewable energy such as coal, oil and natural gas and concluded the accuracy of 99.5%, 92.8% and 98.6% respectively. Diao et al. [16] proposed the energy consumption in residential building by taking the occupancy factor and followed the statistical approaches to identify the behavior of occupants. Reinhardt et al. [43] has abstracted the method for predicting the power consumption by time

series pattern for the electrical appliances based on the different signatures and concluded that 90 minutes advanced predictions can be achieved for the energy consumption. Mishra et al. [37] proposed the method to reduce the electricity bill by cutting the power from the main grid and using the power stored in the battery during low-peak demand. This has been achieved by predicting the next day energy consumption using machine learning algorithm. The proposed methodology leads to the 10-15 % of energy saving.
Hernandez et al. [22] critique about the relationship between the weather variable and the energy consumption. The correlation between different weather variables and the total energy consumption was observed. Electricity forecasting for hourly electricity loads for next 24 hours by taking the activity pattern in the households [19]. Different machine learning algorithm were used to predict the power consumption in residential sector at appliance level i.e. Random Forest (100%), KNN (27.17%), Rpart (35.16%), NNET (42.05%), SVR (34.67%). The prediction of solar power was performed by considering weather forecast by using machine learning algorithm by Sharma et al. [45]. The different regression algorithms were compared and the results was obtained and compared based on the accuracy. It was observed that SVM-based prediction achieved the highest accuracy among all the models i.e. it was 27% more accurate than other models.

From the above review, it can be clearly seen that very few authors have worked on the appliance level prediction of energy in the residential sector. So, there is a need to predict the power consumption at the appliance level which could be useful for the energy suppliers as well for the occupants.

Table 2.3: Machine Learning Models

Author	**Models**
F. Wahid et al. [50]	MLP, Random Forest
N. Sharma et al. [45]	Support vector machine
K. Gajowniczek et al. [19]	Random Forest,SVR, knnr, NNET, Rpart
L. Hernandez et al. [22]	PCA, Correlation
A. Mishra et al. [37]	EWMA, LR, SVMs

to be cont'd on next page

Table 2.3: Machine Learning Models (Cont.)

Author	**Models**
W. Tian et al. [48]	OLS, spatial regression analysis, Lagrange multiplier statistical test
J.F. Kwac et al. [31]	K-Means, hierarchical clustering
H.X. Zhoa et al. [55]	ANN, SVM
K. Basu et al. [10]	Decision tree, Decision table, BN
U. Kumar et al. [28]	Grey-Markov, Grey-Model
L. Diao et al. [16]	Pattern Clustering
A. Reinhardt et al. [43]	Time series analysis

2.3 Energy Management: Optimization Techniques

Energy Optimization has been performed for minimizing the cost, error rate in residential sector. Several optimization techniques has been used by considering different electronic appliances in different sectors.

Table 2.4 describes the different optimization methods used for minimizing the energy consumption used in residential sector. Rasheed et al. [42] applied the binary multiple knapsack for optimization technique to reduce the electricity bills without affecting the user comfort. The optimization was carried out on three types of appliances in the response to the behaviour, weather conditions and electricity prices. The results were verified by simulation on the optimization results. The behaviour of occupants for usage of different appliances in homes was carried out for finding the energy consumption patterns by Singh et al. [46]. An unsupervised approach has been implemented for incremental data mining by frequent pattern mining applied to the energy consumption data. Cottone et al. [15] proposed the method for optimizing the energy of appliances by recognizing the user activities. The activities of the users were extracted by information theory approach. Further, the knapsack optimization problem was defined for optimizing the energy need of for households.

Li et al. [33] developed the optimization method and compared it with different existing multi-objective algorithms. The GenOpt and artificial neural network (ANN) was performed and the results was compared with the non-dominated

sorting genetic algorithm (NSGA-II), multi-objective particle swarm optimization (MOPSO), the multi-objective genetic algorithm (MOGA) and multi-objective differential evolution (MODE) and MODE achieved the optimized result. The realistic scheduling mechanism was purposed by Mahmood et al. [34] for classifying the appliances with their time of use(TOU) and by different constraints. The Binary Particle Swarm Optimization (BPSO) was implemented for appliance utility and cost effectiveness. Wang et al. [51] propound the electricity forecasting with adaptive particle swarm optimization and compared the result with the existing time series models i.e. seasonal exponential smoothing (S-ESM), weighted support vector machines (W-SVM).
Subbiah et al. [47] build the energy demand model for appliance usage with the activity performed by the occupant and calculated the energy consumption based on the various constraints such as appliances rating , the duration of the appliance used and the type of activity done in the home. The method used was based on individual modelling approach and different data set was used to achieve the goals. Ardakani et al. [8] forecast the electricity energy consumption with optimized and artificial neural network(ANN). The forecasting model used was multi-variable regression and ANN and the results was optimized by using Particle Swarm Optimization (PSO) and Improved PSO (IPSO). The final results indicated that IPSO- ANN performed better as compared with PSO and achieved the MAPE 1.94 and 1.51 respectively for both the data sets used for the forecasting. The authors [40] had used ant colony optimization with support vector machine model for removing the redundant information for predicting the short-term forecasting of electricity.

As per the authors knowledge, very few authors have considered the hybrid approach of optimization with the best machine learning algorithm by taking the climatic conditions. There are few authors who have worked on the appliance level forecasting with different climatic conditions. So, there is a need to implement the machine learning algorithm for energy forecasting and then perform the optimization technique with a hybrid approach.

Table 2.4: Optimization Techniques

Author	**Optimization Technique**
M.B. Rasheed et al. [42]	Binary Multiple Knapsack
D.Mahmood et al. [34]	Binary Particle Swarm Optimization
D. Niu et al. [40]	Ant Colony Optimization
F.J. Ardakani et al. [8]	Particle Swarm Optimization
K. Li et al. [33]	NSGA-II, MOPSO, MODE, MOGA
J. Wang et al. [51]	Adaptive Particle Swarm Optimization
P. Cottone et al. [15]	Knapsack Optimization

2.4 Energy Management: Hybrid Optimization Techniques

Some of the authors have proposed hybrid optimization techniques for energy minimization for improving the accuracy and reducing the error rate. Different hybrid techniques used by authors are discussed below:

Table 2.5 discusses about the different hybrid optimization techniques used by the author for energy minimization. It elaborates about the hybrid methods used for energy consumption in the households. Li et al. [32] worked on the annual power forecasting by using the generalized regression neural network (GRNN) along with fruit fly optimization algorithm (FOA). The basic purpose of using the FOA with GRNN was to select the value of parameters for load forecasting. The FOAGRNN proposed model was compared with the existing models i.e. GRNN, PSOGRNN, SALSSVM and OLS_LR. The proposed work focuses on the annual power forecasting not on the individual factors that effects the usage of power. Hong [23] developed a monthly electricity hybrid forecast model SRSVRCABC. The model was hybrid with support vector regression model and compared with different existing models. The proposed model improves the forecasting accuracy. The main focus of the author was on the monthly total electricity consumption rather than to focus on the individual factors responsible for electricity consumption. Multi-Objective genetic algorithm (NSGA-II) was proposed by Eskander et al. [18] to maximize the energy saving annually by minimizing the investment. The proposed work focuses on a particular area in Portugal without considering the users comfort. The main drawback of the work was the replacement of the appliances for minimizing the energy.

The author [30] proposed the ensemble multi-linear preceptron approach and choose the 4 appliances to model i.e. chiller, pump, fan, and reheat device and then integrated the four models to optimize the energy value by using PSO. The optimization was performed on total energy consumption of HVAC system and the consumption was reduced by 7%. Wang et al. [20] minimized the customer cost by estimating the electricity consumption in households for Demand-side management. The management of the resources was implemented by using the particle swarm optimization. The results were obtained and it was observed that the optimized resource management reduces the cost upto 16%-17% and optimized resource management with optimized appliance selection leads to 19-21% cost savings. The author [13] proposed a hybrid fruit fly optimization with support vector regression model for predicting the monthly electricity consumption for improving the accuracy of the electricity forecasting. The results were compared with different techniques and the proposed model performed better than the previous research.

Table 2.5: Hybrid Optimization Techniques

Author	**Hybrid Optimization Technique**
Hong-ze Li et al. [32]	FOAGRNN
Wei-Chiang Hong [23]	SRSVRCABC
Monica Eskander et al. [18]	Multi-Objective genetic algorithm (NSGA-II)
Andrew Kusiak et al. [30]	ensemble MLP with PSO
Lingfeng Wang et al. [20]	BPSO
Guohua Cao et al. [13]	SFOASVR

This chapter was about the different survey done on various techniques and methods. The section 2.2 and 2.1 gave the details about the various machine learning algorithms and methods used for predicting the energy consumption in different sectors and the evaluation parameters to compare the various models. The section 2.4 depicted about the different optimization technique applied to minimize the energy consumption and the section 2.5 revealed about the different optimization technique integrated with machine learning model for energy minimization in different fields. From the above, literature survey it concludes that very few authors have worked upon the energy prediction based on the appliance level along with the optimization. So, there is a need to implement such a method that works on the appliance level prediction and optimization.

Chapter 3

Problem Statement

This chapter is concentrated about the problem statement and research gaps in the existing literature survey and the objectives to be achieved while making the energy optimization model.

3.1 Problem Statement

Energy Prediction and optimization has become a critical issue nowadays as there has been increase in the usage of electrical appliances in the households and due to which the emission of carbon dioxide has increased. As the energy suppliers are not able to fulfil the requirement of the electricity for the households which is a basic requirement for accomplishing their daily activities without any hurdle. So,there is need for energy management for home appliances.

3.2 Research Gaps

Most of the work has been done on the energy prediction and optimization using different factors relating to energy. There were some gaps found during the literature survey about the energy prediction and optimization. The research gaps have been discussed below:

- The existing energy prediction models are focused on the accumulated energy consumption of the dwellings. But the proposed model focuses on the individual home appliances consumption as different appliances consumes different amount of energy at a particular time interval ([34], [15], [40], [7]).
- As per the research, different factors are considered for the energy prediction such as demographic factor, time at which the appliances has been used, climatic condition etc. Very less work has been done in the field of the climatic condition by focusing on individual appliances. The proposed method has taken all the climatic as well as weather conditions for the energy prediction based on the appliances consumption ([19], [48],[37]).

- Different methods and machine learning models has been used for the energy prediction. In the proposed work the method and the models used improved the accuracy and the energy optimized value based on the climatic factors as compared with the pervious work ([50] , [55], [19], [10]).
- Different hybrid models has been used in the existing work such as ([32], [23], [18], [30],[20])for optimizing the energy but climate was not the factor used for energy optimization. The proposed model focuses on the hybrid optimization model such as PSO and ABC with climatic variables for energy minimization which improved the minimized energy value.

3.3 Objectives

The research gaps perceived in the literature review has been effectuated by the following objectives:

- To explore the various energy prediction machine learning algorithms and the energy optimization techniques for minimising the energy of the dwellings.
- To propose a hybrid energy optimized technique for energy suppliers based on energy usage for home appliances.
- To validate the contemplate approach using the climatic conditions such as wind speed, humidity, precipitation, temperature etc.

Chapter 4

Research Methodology

This chapter is about the methodology proposed to compensate the research gaps discussed in the section 3.2. The proposed methodology is divided into two phases for energy prediction and optimization as shown in Figure 4.1.

The phase 1 of the methodology is about the energy prediction of home appliances by using different method and machine learning models. The prediction is based on different climatic variables such as wind speed, visibility, temperature etc. The phase 2 is about the energy optimization based on the inputs to the model. The optimization performed in the proposed methodology is hybrid optimization by finding the best model in the phase 1. The optimization of the energy will be useful to the end users to perform their daily activities without any hurdles.

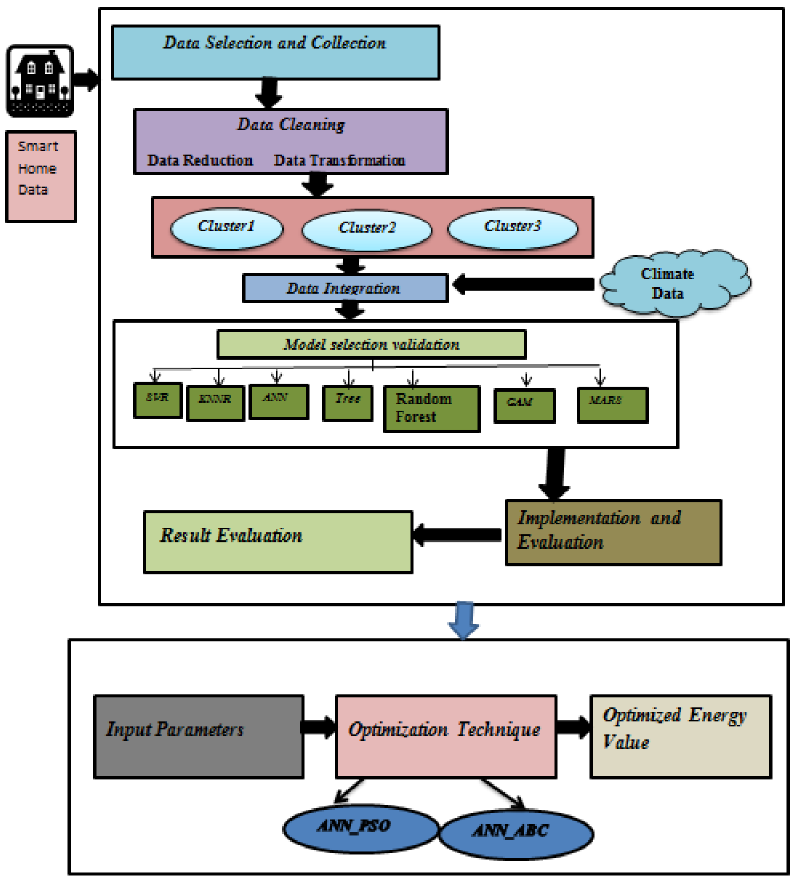


Figure 4.1: Hybrid Energy Management Model

4.1 Data Description

Smart Home data consisting of different appliance power consumption was taken into consideration along with its climatic conditions. The data set used in this paper for the prediction of power is taken from Almanac of Minutely Power data set (AMPds) [36]. The data set is a record of one home around the time period of 2 years (from April 1,2012 to March 31,2013) consisting of 21 sub-meters with the time interval of 1 minute[35]. The Climate data set consisted of the same time period but with the interval of one hour. There was no missing value in the home data set but climate data set consist of many missing values which were removed by KNN method by using the Euclidean distance. For the prediction purposes the appliances data set was converted into the same time interval as that of climate by taking the mean values for the particular day.

Figure 4.2 summarizes the power consumed by some of the appliances used in the data set according to the wind speed. The graph below shows the power consumed on the April 2, 2012 according to the temperature. The secondary axis shown is the temperature axis in degree Celsius. Figure 4.3 encapsulates the monthly energy consumed by the total household in the year 2013 according to the average outside temperature. The energy consumed in the month of July and September when the outside temperature was 18 °C was least. From this we can interprets that below this temperature the energy consumed by household was more as compared with the month of July and September. The Table 4.1 indicates the appliance data set with Appliances ID used for indicating the names of appliances in the figures.

Table 4.1: Appliance Data set [36]

ID	**Appliances**
B1E	North Bedroom
B2E	Master/South Bedroom
BME	Basement Plugs & Lights
CDE	Clothes Dryer
CWE	Clothes Washer

to be cont'd on next page

Table 4.1: Appliance Data set (Cont.)

ID	Appliances
DNE	Dining Room Plugs
DWE	Dishwasher
EBE	Electronics Workbench
EQE	Security/Network
FGE	Kitchen Fridge
FRE	HVAC/Furnace
GRE	Garage
HPE	Heat Pump
HTE	Instant Hot Water Unit
OFE	Home Office
OUE	Outside Plug
TVE	Ent Tv/PVR/AMP
UTE	Utility Room Plug
WOE	Wall Oven
RSE	Rental Home

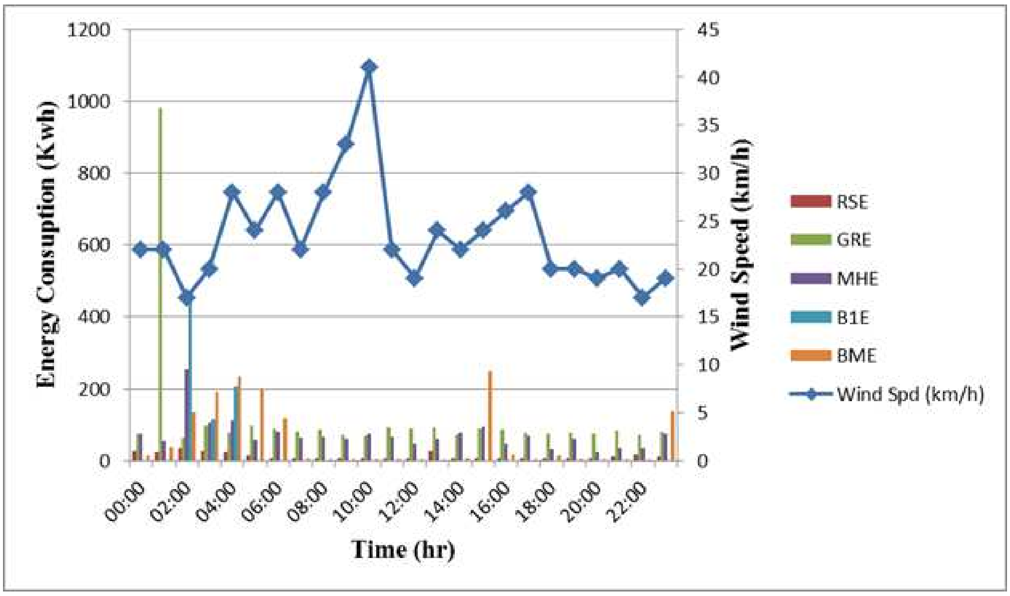


Figure 4.2: Power Consumed according to the Wind Speed

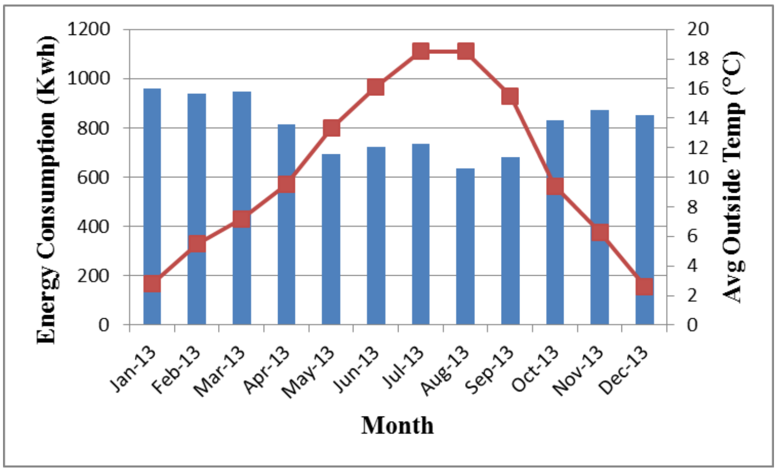


Figure 4.3: Monthly Energy Consumption in the year 2013

4.2 Principal Component Analysis with Clustering

Power analysis was essential for the prediction of the power consumed by household appliances. It helps in analyzing the patterns followed by households occupants for the usage of the appliance at any time period. The statistical approach used for reducing the observations followed by different appliances [29]. It was used to transform the smart home data of appliance to form a new system by measuring the variance according to the dimensionality of the data set and forming the clusters by using the k-means clustering algorithm according to the pattern followed for the usage of the appliance.The scaling of the variables is done in PCA to find the variances in the input data set. The scaling is performed by using equation (4.1)

$$S = \frac{x_i - \sum_{i=1}^{n} x}{Std(x)} \tag{4.1}$$

where, x stands for input data, n is for number of variables and Std indicates the standard deviation for the x values. Each observation was considered as different dimension and the relation was found by using the covariance using the eigen values. The clusters formed were the linear combinations of the original data. The main motive to perform the PCA was to identify the hidden patterns for the usage of the appliances, reducing the dimensionality of the data set [26] by using clustering and identifying the highly correlated variables [41]. It is an unsupervised approach which involves various set of features Z_1 , Z_2 , Z_P and no associated response. PCA works with the dimensions used in the data. The dimensions are the number of variables but every dimension has its own importance. Each dimensions originated by PCA is the linear combinations of p features.

$$X^1 = \phi^{11} Z^1 + \phi^{21} Z^2 + \phi^{31} Z^3 + ... + \phi^{P1} Z^P \tag{4.2}$$

$$X^2 = \phi^{12} Z^1 + \phi^{22} Z^2 + \phi^{33} Z^3 + ... + \phi^{P1} Z^P \tag{4.3}$$

$$X^{i1} = \phi_{11} Z_{i1} + \phi_{21} Z_{i2} + \phi_{31} Z_{i3} + ... + \phi_{P1} Z_{iP} \tag{4.4}$$

$$maximize_{\phi_{11},......,\phi_{P1}} \{1 \setminus \{n\} \sum_{i=1}^{n} (\sum_{j=1}^{P} \phi_{j1} Z_{ij})^2 \} subjectto \sum_{j=1}^{P} \phi_{j1}^2 = 1 \tag{4.5}$$

The equation (4.2) was used for finding the first principal component analysis with Z_1 ,Z_2, ., Z_P as features where X^1 is first principal component, ϕ^{P1} is the loading vector comprising of first principal component, Z^1,........,Z^p are normalized vector. The first PCA give the result of the highest variation features. The second PCA gives the result of the variables with low variation with each other. The equation (4.3) was used to calculate the second principal component analysis.

The equation (4.4) and equation (4.5) was for finding the variance and eigen decomposition value for optimizing the variables. The equation ($\sum_{j=1}^{P} \phi_{j1} Z_{ij}$) is used for normalizing vector.

PCA with clustering have been performed in two phases. In the algorithm 4.1 describes the working of PCA in phase 1 for reducing the dimension by finding the hidden patterns and in algorithm 4.2 describes the phase 2 working with the clustering algorithm [39]. The pseudo code proposed is as follows:
Input Vector: A = A_1,A_2,........,A_n
Output: set of k clusters

Algorithm 4.1 Phase 1: PCA to reduce the Dimension of Appliance data set

1: Organise the dataset in the form of matrix A
2: Scale the variable (x) using equation (4.1)
3: The eigen values and the eigen vectors are calculated of the matrix A
4: The correlation between the variables and Principal Component found
5: The quality of variables are indicated by square cosine and square coordinates
6: The significant variables are found in the principal components
7: Principal component (PC's) having the largest eigen values with largest variances are selected
8: The new matrix M is formed with PC's
9: The reduced data set D is projected to new axis by applying M to A

Algorithm 4.2 Phase 2: K-Means Clustering with reduced data set (D)

1: Input vectors are selected randomly for initialization of the cluster
2: The cluster center is found which is closet to the input vector
3: The input vector is assigned to the corresponding cluster
4: By using the mean the cluster centers are updated of the assigned input vector
5: The step 2 & 3 are repeated until no change in the mean value

4.3 Machine Learning Models

The prediction problem for consumption of energy with climate conditions is very crucial as energy consumption in the dwellings have various socio- economic factors. The prediction of energy has been calculated by using the different climate variables such as temperature, wind speed, humidity etc. at appliance level. Clustering was performed by using PCA and according to clusters formed the total power of those appliances was calculated and data was integrated with the climate variables.The different regression models were implemented during training and testing such as Support Vector Regression, K-Nearest Neighbour Regression, Random Forest, Tree, Artificial Neural Network (ANN), and General Additive Model using Splines (GAMs), Multivariate Adaptive Regression Splines (MARS).

$$P = \sum_{i=1}^{n} A_i \tag{4.6}$$

The equation (4.6) was used for the sum of power of different appliances according to clusters formed based on their hidden patterns where A_i is the power of an appliance and P is the total power calculated of the particular cluster formed.

4.3.1 Support Vector Regression (SVR)

Support vector regression are used for maintaining the features that are used as predictors for model training and testing. The basic working of SVR is based on data set. It maps the non-linear data into feature dimension space of the original data [14]. The main purpose of SVR is mapping of data from training data set to the feature space. The feature space optimized hyperplane by formulating the non-linear relationship between response variable and the predictor variable. The basic SVR function is formulated in equation (4.7).

$$f(y) = \gamma * \Theta(y) + \beta \tag{4.7}$$

where Θ(y) is the feature that are mapped to input data y and γ and β are the coefficient.

4.3.2 K-Nearest Neighbour Regression

It was started in 1970s for pattern recognition as a non-parametric technique. It predicts the target based on some distance of the k nearest neighbours. The function used for predicting the energy consumption using knn.reg( ). This function was used for regression algorithm. The parameters are calculated by the function for the prediction. The function was rum on R platform with different parameters tuning for achieving the good accuracy.

4.3.3 Random Forest

Random Forest is categorized as ensemble-learning models. It combines the different regression tree and it root node represents the different path a variable with highest importance choose to predict the response variable value. It was used for finding the variable importance. It selects the different sample and then successively draws the relationship between the response and predictor variables.

4.3.4 Regression Trees

They are the variants of the decision tree.. It splits the branch by predicting the predictor variables which is contributing the most for the response variable by repeating partitioning the node that splits the data into various smaller nodes. It finds the different regions of blocks for the response variable prediction. In, Regression trees the regions are dividing into the number of blocks R1 , R2 , ... , Ri. The selection of the prediction based on the cut point and the root node split into smaller regions of the form $\{Xi/Xi < p\}$ and X/ Xi$\geq$p where (Xi , Yi) i.e. X is the predictors and Y is the response variable and i = 1, 2,...... ,n and p is the value for the splitting of the node predicted. In this paper, rpart and tree package was used for the training of the regression tree. The pruning of the regression tree was done using the tree package, The different values was obtained for both the packages of the regression tree.

4.3.5 Artificial Neural Network

Artificial Neural Network is based on the collection of the connected nodes having three different layers i.e. input, hidden, output layers. The output of the training data depends on the weights of different input variables [12]. A Figure 4.4 represents the overall block diagram used in this paper which dependent on the inputs i.e. are different climate variables Input1, Input2,.......,Inputn which provides the neurons to the second layer which is the hidden layer and give rise to the patterns according to input variables. The output is the prediction of the power consumption.

For electricity forecasting, ANN has been widely used because of its capability to deal with the non- linear data. It deals with non- linear data based on the climatic condition and the usage of appliance. For climatic indicators as input data, electricity forecasting is defined as function of different climatic variables such as temperature, humidity, visibility, wind speed etc. as defined by the equation (4.8).

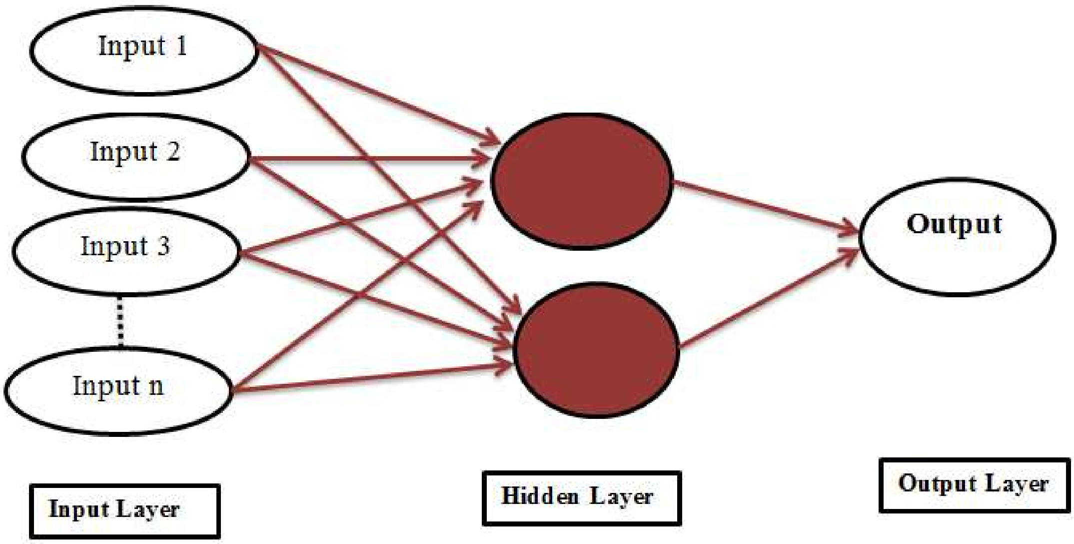


Figure 4.4: Block Diagram for Neural Network

$$E(t) = f(C_{temperature}(t), C_{Humidity}(t), C_{Visibilty}(t), C_{WindSpeed}(t), C_{DewPoint}(t), ..) \tag{4.8}$$

The mean square error calculated during the model training and testing phase depends on the pattern consumption of the usage of the appliances, the applied mode and the input data. The mean square error will be used as the fitness function during the hybrid optimization of PSO and ABC. The equation (4.9) is used as the fitness function/objective function.

$$MSE = \frac{\sum_{j=1}^{n}(Actual_i) - Predicted_i^2}{n} \tag{4.9}$$

4.3.6 General Additive Model using Splines (GAMs)

GAM is a procedure which apprehends the non-linearitys in the data and helps to fit nonlinear models. Gam are the generalized version of linear models in which the predictors rely linearly or non-linearly on smooth nonlinear functions like splines, polynomial or step functions etc. In this paper, we have used splines as they fit smooth linear function on the wreath of predictors (x1, x2,......, xp).In the equation (4.10) $\frac{y_i}{f(x)}$ is the regression function on different predictors. ϵ_i is the some noise in the different nonlinear functions and f$_j$ x$_{ij}$ are the different nonlinear functions on the predictors (xp) where p is the number of different variables on which the regression function was calculated [53] where $\sum_{j=1}^{p} f_j(x_{ij})$ = $f_1(x_{i1})+f_2(x_{i2})+ ......+f_p(x_{ip})$ i.e. the sum of different functions used on different predictors. We have used gam() function in R to fit GAM using splines with an approach back-fitting.

$$y = F(x) = \sum_{j=1}^{p} f_j(x_{ij}) + \epsilon_i = f_1(x_{i1}) + f_2(x_{i2}) + ... + f_p(x_{ip}) \tag{4.10}$$

4.3.7 Multivariate Adaptive Regression Splines (MARS)

It was introduced by Jerome H. Friedman in 1991.MARS is a generalization of stepwise linear regression which takes the form of an expansion in splines basis functions. The basis functions as well as the variables associated with each functions are automatically determined by the data [9]. In the equation (4.11) $\beta_i(x)$ is the weighted sum of basis functions in which one basis function is multiplied by its coefficient.

$$F(x) = \sum_{j=0}^{k} \Theta_i \beta_i(x) \tag{4.11}$$

where $\beta_i(x)$ is basis function which can be of three types and Θ_i is constant coefficient.

4.4 Model Evaluation Parameters

Before going to any conclusion regarding the regression models, it is important to know the parameters that are essential to evaluate any model. With the comparison of indicators only it can be found which model best suites to the data set. There are different evaluations parameters from which the performance of the model can be judged. They are described in the following sub sections:

4.4.1 Analysis of Residual

Residual are the difference between the actual value and the predicted value of the target variable. They are represented by the scatter plots known as residual plots. If the difference between predicted and actual value was less than the model is said to be fit for the dataset. Residual plots are the unpredictable random component of each observation for the target variable.

$$r_i = x_i - \hat{x}_i \tag{4.12}$$

In equation (4.12), r_i is residual, x_i are the actual values and $\hat{x}_i$ are the predicted values for i = 1,2,3,......,N. The one way to check the residual graph are to check the sum and the mean of the graph as the sum and the mean is always be equal to 0.

4.4.2 Coefficient of Determination (R^2)

It is the method for validating the result of the model. It is variance that is calculated of the dependent variable which is predictable from the independent variables. It indicates the measure of the observed variable based on the variation of the independent variables with the help of different models. Its value ranges from 0 to1.

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{4.13}$$

where,
Sum of Squares Total: SST $= \sum (y - \bar{y})^2$
Sum of Squares Regression: SSR $= \sum (\acute{y} - \bar{\acute{y}})^2$
Sum of Squares Error: SSE $= \sum (y - \bar{\acute{y}})^2$
If the value of R^2 is 1 it indicates the regression line perfectly fits the data. It is used to measure the goodness of fit for the data.

4.4.3 Root mean square error

It is used to measure the difference between the actual value and the predicted value by the model. The differences between the values are known as residuals. If the difference between the actual and predicted value is less than the model formed is said to be good and if the difference is large than the predicted model is said to be unfit for the data. In the equation (4.14) y^i are the predicted values by the model and y_i are the actual values randomly taken by the model during the prediction for i=1, 2,......, n where n belongs to number of variables used in the dataset.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i) - y_i^2}{n}} \tag{4.14}$$

4.4.4 Accuracy

Accuracy is the measure of the percentage of the predicted and the actual values. It indicates the "how accurate our model have performed the testing of the input data". It is calculated by the equation (4.15).

$$Accuracy = \frac{\sum_{i=1}^{n}(abs(Actual_i - predicted_i) * 100 \leq err}{n} \tag{4.15}$$

where, Actual$_i$ is actual target, predicted$_i$ is predicted target, err is the acceptable error and n is the total number of instances.

4.5 Energy Management: Optimization Techniques

Energy optimization is necessary after the prediction of energy for the energy suppliers sa that they could provide the minimum amount of energy to the dwellings to accomplished their basic work of the households. Energy optimization leads to the minimizing the energy value according to climate variables which could be useful for the households as well for the environment. This section describes about the optimization techniques which could be helpful for the future energy prediction methods:

4.5.1 Particle Swarm Optimization (PSO)

Particle Swarm Optimization was inspired from the swarm intelligence to find the shortest route for their activities and introduced by Kennedy and Eberhart in 1995 [17]. It consist of particle update (p$_i$), velocity update (v$_i$) and comparing of their best values. The PSO algorithm updates the velocity and the position vector according to the input data. The equation (4.16) and equation (4.17) describes the velocity update and position update.

$$V_{id} = w * v_{id} + const_1 * ran_1 * (pbest_{id} - x_{id}) + const_2 * ran_2 * (gbest_{id} - x_{id}) \tag{4.16}$$

where w is the interia weight,d are the number of parameters to be optimized, $const_1$ and $const_2$ are the acceleration constants and ran_1 and ran_2 are the random numbers between the range 0-1. The $const_1$ and $const_2$ are used to find the optimal path by moving the each particle towards the pbest and gbest.

$$X_{id} = x_{id} + V_{id} \tag{4.17}$$

4.5.2 Artificial Bee Colony Optimization (ABC)

Artificial bee colony algorithm was inspired from the honey bees nature. It is a meta-heuristic approach proposed by Karaboga [24] and further was developed by Karaboga and Bastruck [25] in 2005. It consist of three phases i.e. employed bee phase, onlooker bee phase and scout bee phase. The main goal in this optimization is to produce the best solution for each individual bee. The employed bee locates a new food source f_i in the region of the current source y_i. The equation (4.18) used by the employed bee for locating the food source:

$$f_{jk} = y_{jk} + \Phi_{jk}(y_{jk} - y_{ik}) \tag{4.18}$$

where i$\varepsilon(1, 2, ..., SN)$ and k$\varepsilon(1, 2, ..., D)$ are the indexes which are randomly chosen and j and i needs to be different from each other.The Φ_{jk} is between -1 & 1. The greedy selection mechanism is used by the employed bee's for memorizing the better solution.

The second phase in ABC is known as onlooker bee phase which chooses the food source by using the probability according to the fitness function/objective function used by the employed bees. The equation (4.19) was used for calculating the probability.

$$Prob_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_i} \tag{4.19}$$

where fit refers to the fitness function in the equation (4.9). The third and the final phase is scout bee phase in which the food source which cannot be improved by number of different cycles performed is removed from the entire population and the employed bee of that source become scout. The new random source position

f_i is found by scout bee using equation (4.20).

$$y_j^k = y_{min}^k + ran[0,1](y_{max}^k - y_{min}^k) \tag{4.20}$$

The y_{min}^k and y_{max}^k are the upper and lower bounds of parameter k.

The three phases of the algorithm are repeated in cycles known as maximum cycle number (MCN) until a termination condition is satisfied. If a termination conditions are met the algorithm passes the value to the ANN network created in the initial phase of the algorithm for calculating the optimal weights and for mean square error . If the updated weight satisfy the condition the model is fully trained and ready of the testing otherwise the loop goes on till the condition was not satisfied.

4.5.3 Training of neural network by PSO (PSO _ ANN)

The proposed hybrid algorithm PSO_ANN is described in algorithm 4.3. The evaluation of the fitness function used in the equation (4.9) was calculated and the values of pbest and gbest was updated until the termination conditions are satisfied. If the termination conditions are satisfied the loop is passed to the neural network created at the initial phase for calculating the optimal weight and the mean square error by equation (4.9). Finally, the weights are updated and model is ready to use for testing if the termination conditions are met otherwise the step is repeated until the conditions are not satisfied.

Algorithm 4.3 PSO_ANN algorithm

```
Initialize the ANN network with input data
for
   all particles do
   Initialize the parameters of PSO
end for
loop
   for  do
      all particles
      Calculate the new velocity using equation (4.16)
      Calculate the new position using equation (4.17)
      Calculate the fitness value ate new position
   end for
```

13: Find the pbest value and set gbest value
14: **end loop**
15: If the termination condition is not satisfied go to Step 2
16: Bring the pbest and gbest value found in step 12 to the neural network initialize in Step 1
17: **loop**
18: **for do**
19: Calculate the optimal weight of the input data
20: **end for**
21: Calculate the MSE using equation (4.9)
22: **end loop**
23: If the termination condition is not satisfied go to Step 17

4.5.4 Training of neural network by ABC (ABC_ANN)

The proposed hybrid algorithm ABC_ANN is described in algorithm 4.4. The evaluation of the fitness function used in the equation (4.9) was calculated and the values of pbest and gbest was updated until the termination conditions are satisfied. If the termination conditions are satisfied the loop is passed to the neural network created at the initial phase for calculating the optimal weight and the mean square error by equation (4.9). Finally, the weights are updated and model is ready to use for testing if the termination conditions are met otherwise the step is repeated until the conditions are not satisfied.

Algorithm 4.4 ABC_ANN algorithm

1: Initialize the ANN network with input data
2: cycle=1
3: Initialize the food source position f_i
4: Fitness function evaluation of food source using equation (4.9)
5: Evaluate the best source food and the value in gbest
6: **repeat**
7: **for do**
8: Each component y
9: Employed Bee Phase
10: **for do**
11: Each employed bee i
12: y component of gbest is replaced by using the y component of bee i
13: Calculate the value of pbest
14: **if** pbest > gbest **then**
15: gbest is replaced by pbest
16: **end if**
17: New food source position is evaluated using equation (4.18)

```
18:         Fitness function evaluation of food source using equation (4.9)
19:       end for
20:       Calculate the probability P_i using equation (4.19)
21:       Onlooker Bee Phase
22:       for  do
23:         Each onlooker bee i
24:         Select a f_i depending on P_i
25:         y component of gbest is replaced by using the y component of bee i
26:         Calculate the value of pbest
27:         if pbest > gbest then
28:           gbest is replaced by pbest
29:         end if
30:         New food source f_i position is evaluated
31:         Fitness function evaluation of food source using equation (4.9)
32:       end for
33:     end for
34:     Scout Bee Phase
35:     if employed bee = Scout then
36:       Replace it with new random source position
37:     end if
38:     Best solution is stored in some value
39:     Compare the best solution with pbest
40:     Store the best value
41:     cycle =cycle +1
42: until cycle = MCN
43: If the termination condition is not satisfied go to Step 2
44: loop
45:     for  do
46:       Calculate the optimal weight of the input data
47:     end for
48:     Calculate the MSE using equation (4.9)
49: end loop
50: If the termination condition is not satisfied go to Step 44
```

Chapter 5

Implementation and Results

The two phases described in the methodology in the chapter 4 have been implemented and the comparison have been done of various learning algorithms in the following section. Further, the hybrid optimization was performed by selecting the best machine learning algorithm. The implementation have been done step by step and the results are verified below:

5.1 Results of Phase 1 of Energy Management Model

Many experiments were performed for predicting the power consumption according to weather conditions. Different methods and learning models were implemented to get the desired output. Initially, the data set described in the section 4.1 was divided into two parts i.e. appliance data set and the climate data set. The step 1 of the EMM was performed on the appliance data set. The PCA with K-Means clustering was implemented after data pre-processing step. PCA was performed in R by using the PCA( ) function and FactoMineR package with the help of Eigen values for retaining the variance by each principal component analysis and then applying the K-means clustering to merge the different appliance according to their usage patterns by the occupants in their homes. Each Eigen value obtained for calculating the variance by each principal component analysis when divided by 10 leads to the variance percentage. The cumulative variance percentage obtained by adding the successive proportions of variance percentage. Figure 5.1 depicts the results of deciding the number of principal component. The scree plot indicated the eigen values in the order from largest to smaller for determining the number of dimensions needed. In the Figure 5.2 shows the relationship between all the variables present in the data. The variables are grouped together according to their relationship formed. The Quality of variables was measured by square cosine (cos2). The Figure 5.3 represents the quality of each variable by cos2 using dimension1-2 of principal component. The appliance MHE has highest quality as compared with others. There was very less difference between

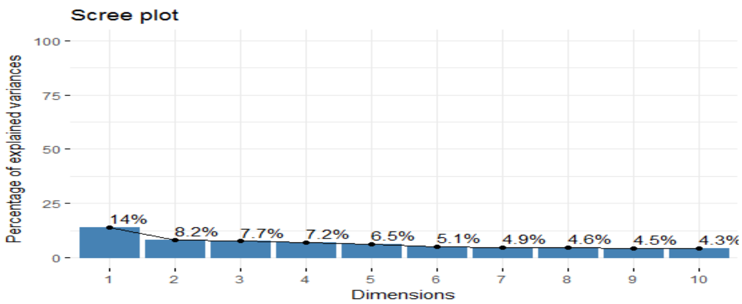


Figure 5.1: Eigen values with the variance percentage

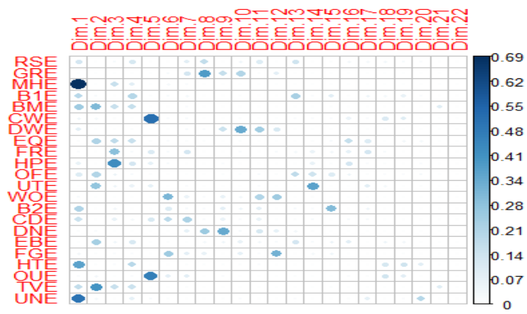


Figure 5.2: Variable correlation

the quality of appliance TVE, BME and UNE. From the Figure 5.3 it interprets that the quality of appliance EQE, B1E and B2E was same. A high cos2 value indicated that a variable representation is good on principal component and low cos2 value indicated that the representation of that appliance is not at all good on principal component.

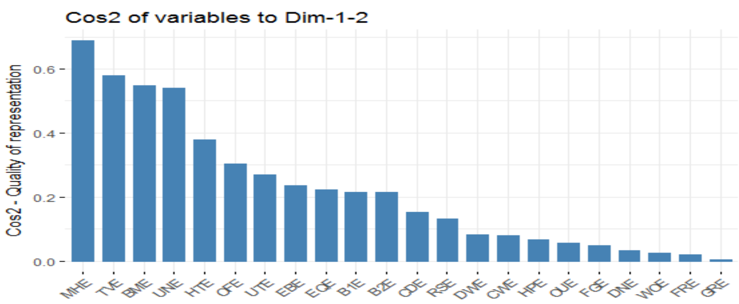


Figure 5.3: Quality of variable by Cos2

The contribution of the variables with principal component was also analysed. The Figure 5.4 elaborates that the appliance that contributed with dimensions 1-2 of principal component. There were many appliances that had less contribution on both the dimensions 1-2. The highest contribution was of appliances MHE, TVE, BME, UNE and rest had less contributed. Contribution with respect to other dimensions was also performed and the results obtained were different in all the cases. So, only the results obtained by dimension1-2 are represented as they were having highest percentage compared to others. The clustering by K-means was

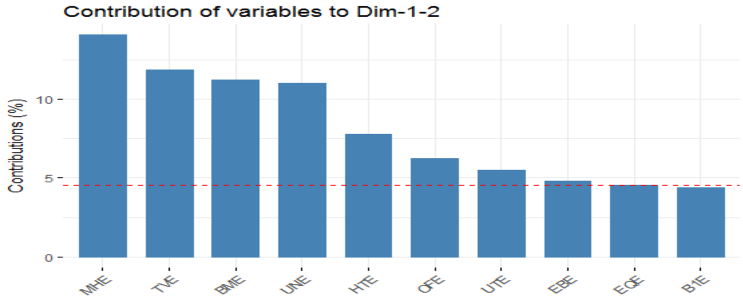


Figure 5.4: Variable contribution by Dim 1-2

performed according to the pattern followed by the households for the usage of the particular appliance. The Figure 5.5 elaborates on three different clusters formed with the usage of principal component. The cluster 1 appliances were UTE, EBE, OFE, EQE , cluster 2 consisted of appliances B1E,UNE, HTE, DWE, B2E, RSE, BME, TVE, DNE,GRE and lastly, cluster 3 consisted of FRE, HPE, CDE, CWE, OUE as shown in table 5.1.

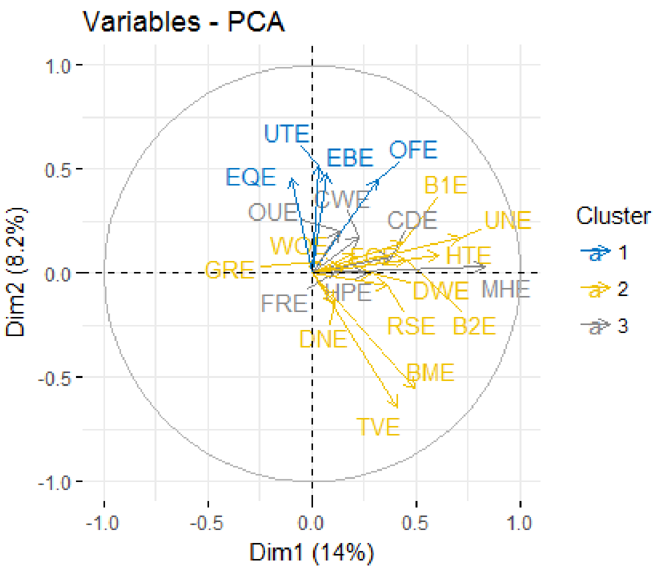


Figure 5.5: Cluster of appliances

Table 5.1: Clustered Appliances

Cluster	**Appliances**
Cluster 1	UTE, EBE, OFE, EQE
Cluster 2	B1E,UNE, HTE, DWE, B2E, RSE, BME, TVE, DNE,GRE
Cluster 3	FRE, HPE, CDE, CWE, OUE

The cluster 1 consisted of the appliances that consumed continuous energy and cluster 2 consisted of the appliances that have moderate type of energy consumption while that of cluster 3 have high energy consumed appliances. The Figure 5.6 depicts the two days energy consumed by the appliances. The cluster 1 shown in blue consisted of the appliances that have continuous or low energy consumption and the red color depicts the cluster 2 appliances that consume the moderate or the mid-peak energy consumption while that of cluster 3 shows the high energy consumption appliances.

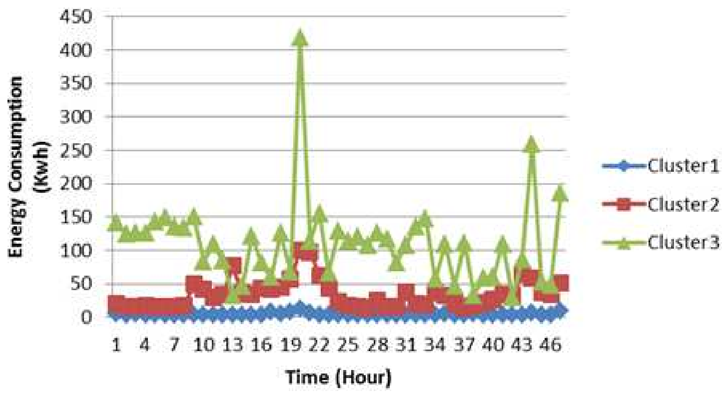
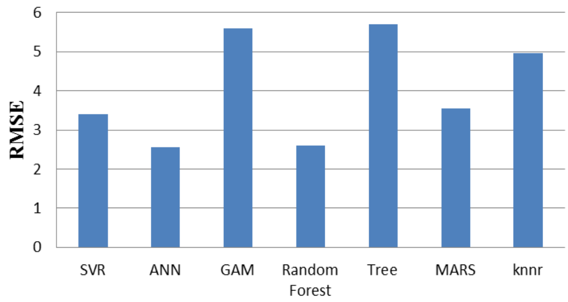
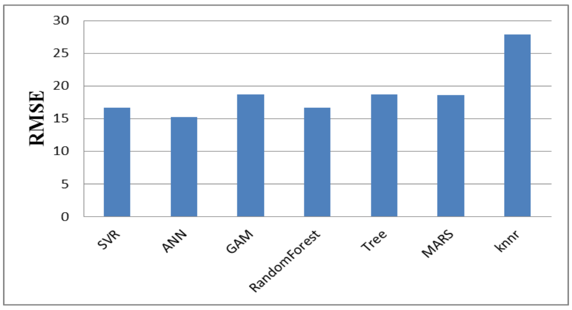


Figure 5.6: Energy consumption by cluster of appliances

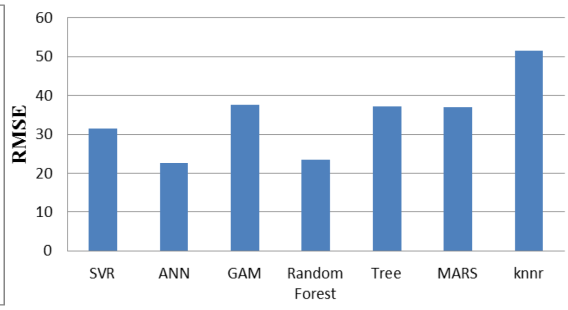
After the cluster formation by PCA, the cluster of appliances was integrated with the climate variables such as temperature, wind speed, visibility, humidity etc. Many machine models were implemented for prediction for energy as explained in section 4.3. The models were compared with there evaluation parameters and the best model was used for the optimization purposes. The Figure 5.7 depicts the RMSE values for all the cluster after performing the models in R. The RMSE value in Figure 5.7a depicts of cluster 1 where ANN (2.55) has the lowest value while that of regression tree (5.59) has the maximum RMSE value. The cluster 2 shown in the Figure 5.7b has RMSE least of ANN (15.23) while maximum value of RMSE achieved was of K-nearest neighbour Regression (27.89) while that of cluster 3 Figure 5.7c has the maximum value of RMSE is 51.58 in K-nearest neighbour Regression while the minimum value of ANN (22.59). It has been observed that all the three clusters consist of the minimum value was of ANN while the maximum RMSE value varies according to the clusters. The Figure 5.8 unveils about the average RMSE value of all the models for energy prediction. The overall average of all the clusters in the households for ANN (13.45) was minimum while that of knnr (28.14) was maximum for one-day energy consumption. It concludes the ANN has achieved the minimum RMSE value in case of clusters energy consumption as well as in average energy consumption.



(a) RMSE of cluster 1



(b) RMSE of cluster 2



(c) RMSE of cluster 3

Figure 5.7: RMSE Values

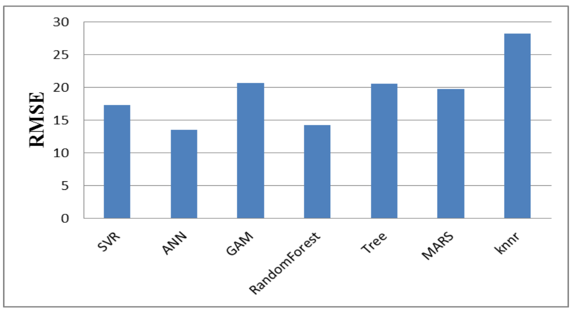
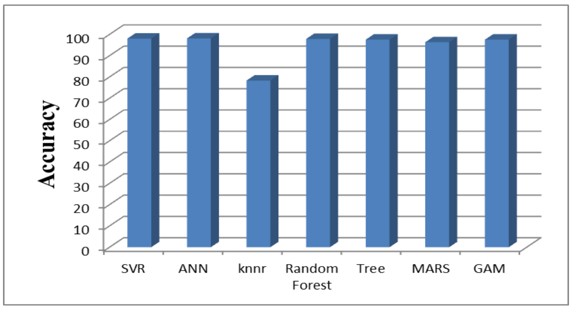


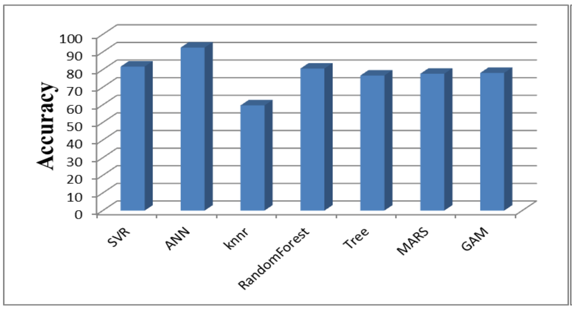
Figure 5.8: Average RMSE value for energy consumption

The Figure 5.9 indicates the accuracy comparison of all the three clusters obtained after model testing. The accuracy obtained in the cluster 1 as shown in the Figure 5.9a of ANN (98.11%) is maximum as compared to all other. The accuracy obtained of other models in cluster 1 has not much difference with ANN i.e. SVR (98.04%), GAM (97.46%) while the minimum accuracy obtained was of knnr (78.34%). The accuracy results of cluster 2 are depicted in Figure 5.9b. The

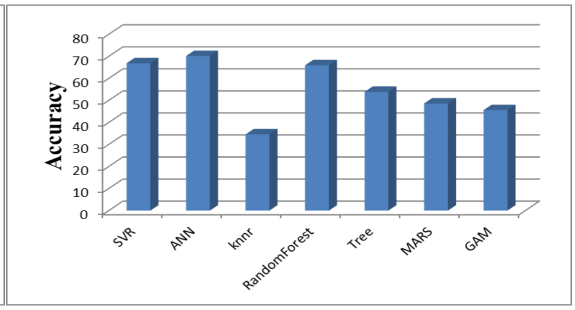
ANN (92.50%) obtained the maximum accuracy in cluster 2 while the minimum accuracy was of knnr (59.69%). Similarly, the cluster 3 accuracy of ANN (70.17%) was maximum while of knnr (34.6%) was minimum as depicts in the Figure 5.9c. The average accuracy for energy consumption in ANN (86.92%) while that of knnr (57.54%) as shown in the Figure 5.10. The ANN has achieved the maximum accuracy in case of average energy prediction as well in the form of cluster energy prediction. In this section, PCA with K-Means clustering was performed



(a) Accuracy of cluster 1



(b) Accuracy of cluster 2



(c) Accuracy of cluster 3

Figure 5.9: Accuracy Values for Clusters

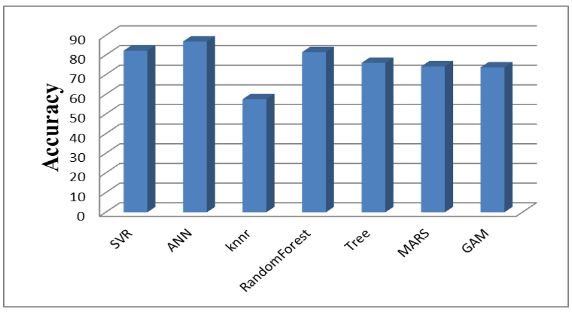


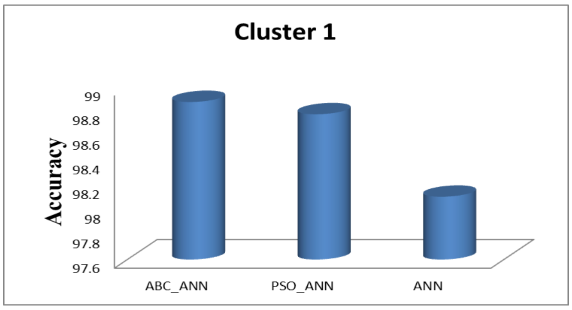
Figure 5.10: Average Accuracy value for energy consumption

and different clusters were implemented by various machine learning models. The models were evaluated based on their evaluation parameters and ANN performed the better among all the models.

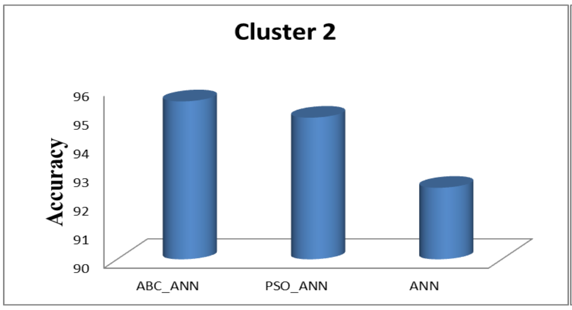
5.2 Results of Phase 2 of Energy Management Model

From the section 5.1, it can be clearly concluded that ANN performed better in terms of accuracy as well as in terms of RMSE. The hybrid optimization was performed with ANN. The two optimization was implemented i.e. PSO_ANN and ABC_ANN for minimizing the energy consumption in all the three clusters. The results of hybrid optimization were compared with the ANN model in the subsequent section. The Figure 5.11 compares the accuracy achieved after the optimization of energy with the ANN model. In the cluster 1 the accuracy achieved by ANN (98.11%),PSO_ANN (98.78%) and ABC_ANN (98.88%) are shown in the Figure 5.11a. From this, it can be clearly seen that the ABC_ANN and PSO_ANN performed better than ANN which helps in minimizing the energy value. The cluster 2 depicts that the accuracy achieved by ANN (92.50%),PSO_ANN (94.95%) and ABC_ANN (95.52%) are shown in the Figure 5.11b. Similarly, the accuracy of cluster 3 in Figure 5.11c unveils the accuracy of ANN (70.17), PSO_ANN (79.95%) and ABC_ANN (80.67%). From all the three clusters , it can be perspicuously seen that the ABC_ANN performed better than PSO_ANN and ANN. The Figure 5.12 elaborates about the average accuracy of PSO_ANN (91.22%), ABC_ANN (91.69%) and ANN (86.92%) for one-day energy consumption. The average accuracy achieved by ABC_ANN (91.69%), PSO_ANN (91.22%), ANN (86.92%) as depicts in the Figure 5.12. The ABC_ANN achieves more accuracy as compared with the other two models.

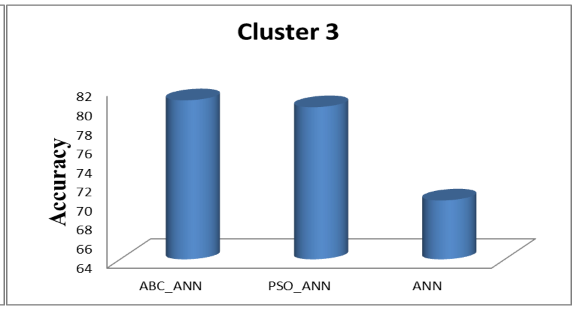
The error curves of all the three clusters were obtained in case of ANN, PSO_ANN, ABC_ANN. The red color in the figures indicated the error rate of ABC_ANN while the green color gives the error rate of PSO_ANN and the light blue indicates the ANN model. The Figure 5.13c indicates the RMSE value of cluster 3 i.e. ANN (31.58), PSO_ANN (22.41), ABC_ANN (21.82). Similarly, the RMSE values of cluster 2 in Figure 5.13b unveils that the ABC_ANN (13.09) has less error than the PSO_ANN (13.09) and ANN (15.23). The Figure 5.14 shows the average RMSE value for ABC_ANN (13.22),PSO_ANN (13.88), ANN (17.98). It clearly indicates that ABC_ANN has less RMSE as compared with other PSO_ANN and



(a) Accuracy Comparison of cluster 1



(b) Accuracy Comparison of cluster 2



(c) Accuracy Comparison of cluster 3

Figure 5.11: Accuracy Comparison

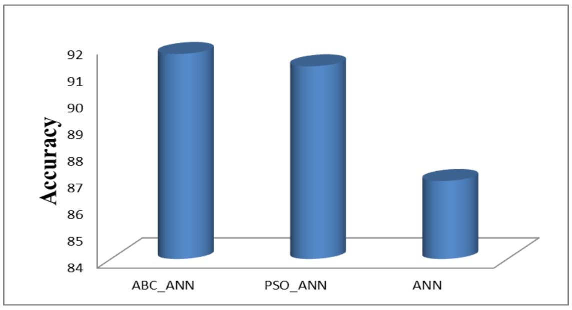
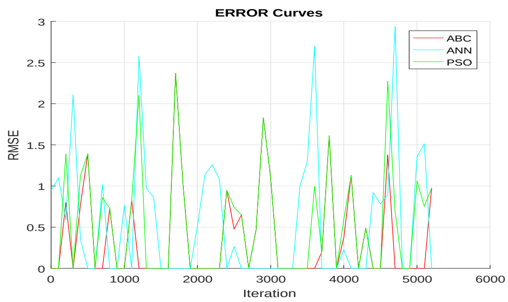
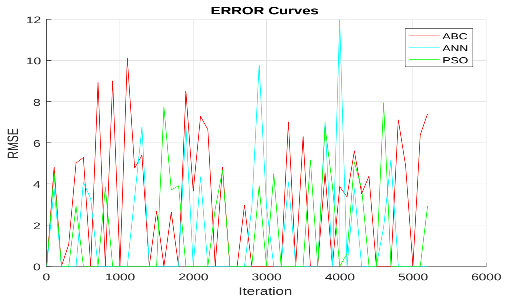


Figure 5.12: Average Accuracy Comparison for energy consumption

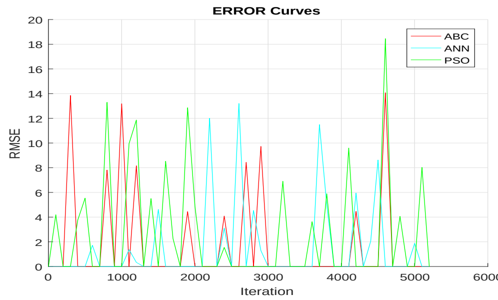
ANN. Therefore, ABC_ANN performed better than PSO_ANN. The Figure 5.15 and 5.16 shows the comparison of Actual,ANN predicted, Optimized predicted values. In the cluster 1 Figure 5.15a and 5.16c shows the average value of actual, ANN predicted, PSO-optimized and ABC-optimized value for one day are 4.83, 8.09, 6.09 and 5.09 respectively. The optimized values are near to the actual values but the ANN predicted values are very far from the actual values. The values for cluster 2 Figure 5.15b and 5.16b for one day are 39.32, 38.40, 37.13 and 38.29 respectively. Finally, the values of Actual, ANN-predicted, PSO-Predicted and ABC-Predicted are 117.19, 82.54, 115.34, 114.94 respectively in Figure 5.15c and 5.16c. Therefore, the PSO_ANN and ABC_ANN performed better than the ANN model.



(a) Error Curve of cluster 1



(b) Error Curve of cluster 2



(c) Error Curve of cluster 3

Figure 5.13: Error Curve

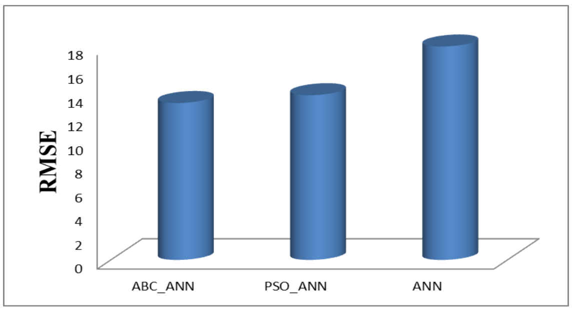
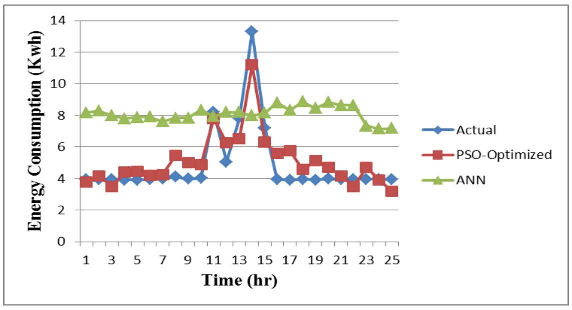
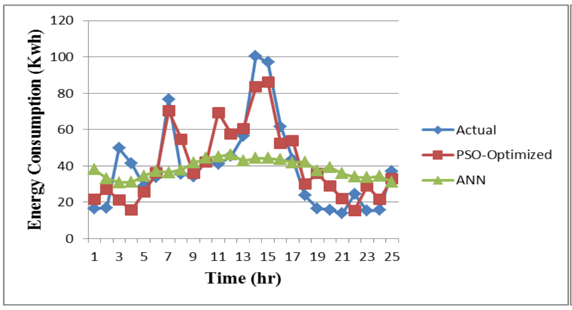


Figure 5.14: Average RMSE Comparison for Energy Consumption

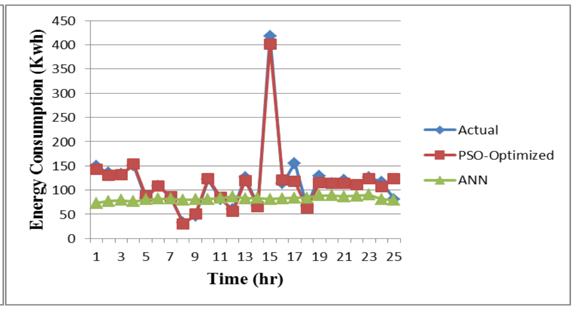
Convergence graph is a plot for comparing the different optimization algorithms with their objective value vs. Iterations. It tells about the performance of the optimization algorithms. The Figure 5.17 indicates the convergence graph of all the three clusters and compared the optimization performance of ABC_ANN with PSO_ANN according to there minimum objective value obtained. The convergence rate for cluster 3 shown in the Figure 5.17c elaborates that ABC_ANN (3.4645e+03) performs better than PSO_ANN (8.8415e+03). The cluster 2 in the Figure 5.17b indicates that ABC_ANN (681.83) obtains the less value from PSO_ANN (2140.4).Therefore, from all the three clusters it can be clearly seen that the convergence rate of ABC_ANN is better than PSO_ANN in all the clusters.



(a) PSO_ANN values of cluster 1

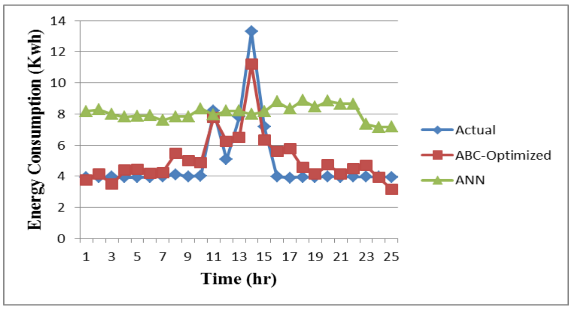


(b) PSO_ANN values of cluster 2

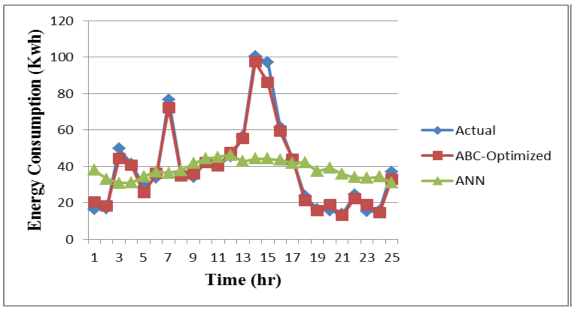


(c) PSO_ANN values of cluster 3

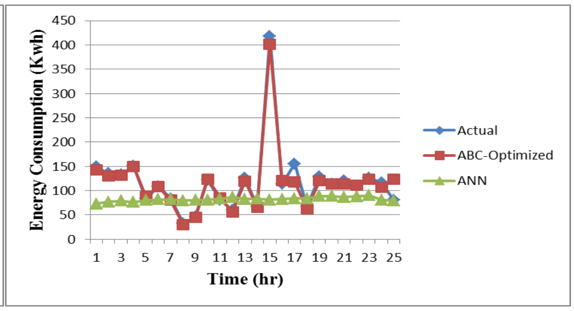
Figure 5.15: PSO_ANN Optimized Values



(a) ABC_ANN values of cluster 1



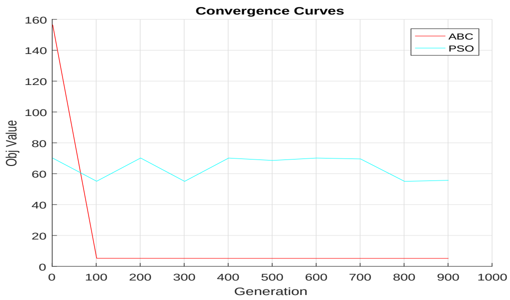
(b) ABC_ANN values of cluster 2



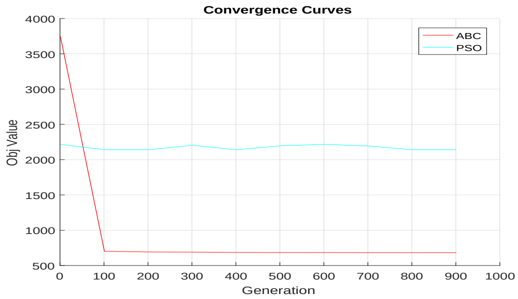
(c) ABC_ANN values of cluster 3

Figure 5.16: ABC_ANN Optimized Values

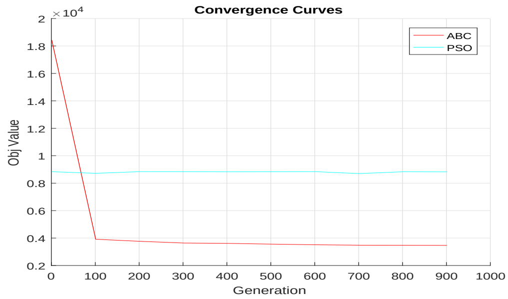
In this section, hybrid optimization was implemented and the results were compared with the ANN models. The comparison based on accuracy and RMSE was made with ANN, PSO_ANN and ABC_ANN. The ABC_ANN performed better in all the cases based on accuracy and RMSE values.



(a) Convergence Curve of cluster 1



(b) Convergence Curve of cluster 2



(c) Convergence Curve of cluster 3

Figure 5.17: Convergence Curve

5.3 Comparison with existing model

The comparison of some of the machine learning models with the same data set was done with the existing models in the [19]. The optimization of energy was

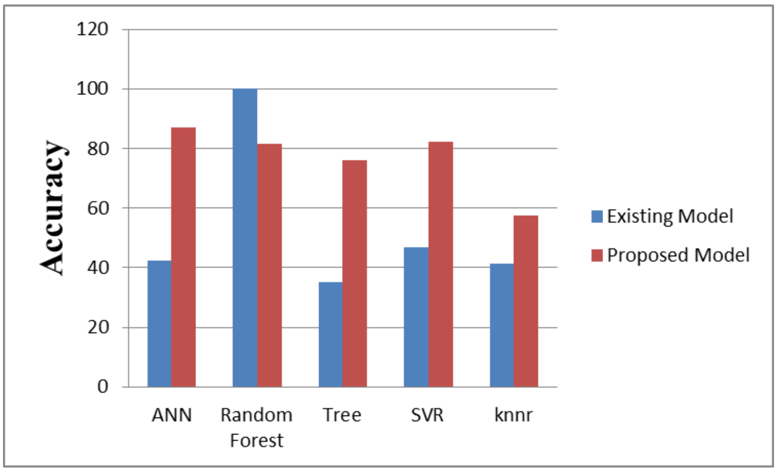


Figure 5.18: Comparison of Proposed Model with Existing Model

not performed only the prediction was performed. The prediction was not based on the cluster formation but in this prediction was performed based on the cluster with climatic conditions. The average value of all the cluster was compared with the existing model. The Figure 5.18 unveils the accuracy results of the existing model with the proposed model. The ANN in the existing model achieved very less accuracy i.e 42.26% while in this the ANN in the proposed model achieved the maximum average accuracy i.e. 86.92% . The random forest (100%) achieved the maximum accuracy in the existing model while in the proposed model random forest (81.48%) achieved average accuracy. The other existing models such as SVR (47.02%), Tree (35.12%), knnr (41.37%) has less accuracy as compared with proposed model i.e. SVR (82.22%), Tree (76.03%), knnr (41.37%)

This chapter was about the results and implementation of both the phases of proposed energy management model. In comparison with machine learning models, ANN performed better from all other models whereas in the hybrid optimization ABC_ANN performed better than ANN model as well as from PSO_ANN. Lastly, the result of machine learning models was compared with the existing model.

Chapter 6

Conclusion & Future Scope

This chapter is about the final inferences from all the above methods and techniques used for implementing the energy management models and to further discussion about the future scope of the proposed model.

6.1 Conclusion

In this pragmatic study of electricity forecasting and optimization of households with different climatic conditions various methods and techniques were implemented and the results were verified with the existing models. Firstly, PCA with K-Means clustering was performed on the different electrical appliances used in the dwellings for their work. Further, after the cluster formation the appliances were integrated with different climatic conditions and the prediction models were implemented and different inferences were made based on evaluation parameters. Lastly, the best model was obtained and was hybrid with the optimization techniques for minimizing the energy value. The following inferences were made from the above study:

- The three clusters were formed with matching hidden patterns using the method two algorithms in the section 4.2. The clusters consisted of the appliances having high energy consumption, mid-peak energy consumption and continuous energy consumption. Further, the three clusters were merged with different climatic variables such as temperature, wind speed, visibility etc.

- Seven different machine learning models were implemented for energy prediction i.e. ANN, knnr, SVR, Random Forest, Tree, GAM's, MARS. All the models were compared with their accuracy achieved and the RMSE values. The maximum accuracy was achieved by ANN (86.92%).

- The best model was hybrid with ABC and PSO optimization technique and the hybrid optimization was performed i.e. PSO_ANN, ABC_ANN. The hybrid techniques was compared with the ANN model in terms of accuracy and error curves were obtained. The accuracy achieved by ABC_ANN (91.69%) was maximum when compared with PSO_ANN (91.22%) and ANN (86.92%) model.

- Both the hybrid techniques were compared with the help of convergence graph and it was found that ABC_ANN performed better than PSO_ANN in case of energy minimization.

This analysis proved that applying PCA with clustering before prediction models help in achieving the better results from the existing work. The hybrid technique help in minimizing the energy consumption and minimization helps the energy suppliers for providing the demand of electricity in dwellings to accomplish their daily activities without any hindrance.

6.2 Future Scope

From the above study,the future scope of this research can be extended in various fields such as:

- The feedback system will be deployed in residential sector for the betterment of energy management.
- It can be expanded for the different sectors i.e. industrial sector, transportation sector and agriculture sector.
- The system can be scaled by using deep learning and artificial intelligence methods.

References

[1] Dimensionailty reduction. [Online]. Available: https://www.techopedia.com/definition/30392/dimensionality-reduction

[2] Electricity consumption in indian homes. [Online]. Available: http://cprindia.org/news/6512

[3] Energy use in homes. [Online]. Available: https://www.iea.org/statistics

[4] Machine learning. [Online]. Available: http://www.expertsystem.com/machine-learning-definition

[5] The what, why, and how of energy management. [Online]. Available: http://www.energylens.com/articles/energy-management

[6] C. O. Adika and L. Wang, "Autonomous appliance scheduling for household energy management," *IEEE transactions on smart grid*, vol. 5, no. 2, pp. 673–682, 2014.

[7] M. Aiad and P. H. Lee, "Non-intrusive monitoring of overlapping home appliances using smart meter measurements," in *Power and Energy Conference at Illinois (PECI), 2017 IEEE*. IEEE, 2017, pp. 1–5.

[8] F. Ardakani and M. Ardehali, "Long-term electrical energy consumption forecasting for developing and developed economies based on different optimized models and historical data types," *Energy*, vol. 65, pp. 452–461, 2014.

[9] M. S. Balshi, A. D. Mcguire, P. Duffy, M. Flannigan, J. Walsh, and J. Melillo, "Assessing the response of area burned to changing climate in western boreal north america using a multivariate adaptive regression splines (mars) approach," *Global Change Biology*, vol. 15, no. 3, pp. 578–600, 2009.

[10] K. Basu, L. Hawarah, N. Arghira, H. Joumaa, and S. Ploix, "A prediction system for home appliance usage," *Energy and Buildings*, vol. 67, pp. 668–679, 2013.

[11] L. A. Berrueta, R. M. Alonso-Salces, and K. Héberger, "Supervised pattern recognition in food analysis," *Journal of chromatography A*, vol. 1158, no. 1-2, pp. 196–214, 2007.

[12] M. R. Biswas, M. D. Robinson, and N. Fumo, "Prediction of residential building energy consumption: a neural network approach," *Energy*, vol. 117, pp. 84–92, 2016.

[13] G. Cao and L. Wu, "Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting," *Energy*, vol. 115, pp. 734–745, 2016.

[14] Y. Chen, P. Xu, Y. Chu, W. Li, Y. Wu, L. Ni, Y. Bao, and K. Wang, "Short-term electrical load forecasting using the support vector regression (svr) model to calculate the demand response baseline for office buildings," *Applied Energy*, vol. 195, pp. 659–670, 2017.

[15] P. Cottone, S. Gaglio, G. L. Re, and M. Ortolani, "User activity recognition for energy saving in smart homes," *Pervasive and Mobile Computing*, vol. 16, pp. 156–170, 2015.

[16] L. Diao, Y. Sun, Z. Chen, and J. Chen, "Modeling energy consumption in residential buildings: A bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation," *Energy and Buildings*, vol. 147, pp. 47–66, 2017.

[17] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on.* IEEE, 1995, pp. 39–43.

[18] M. M. Eskander, M. Sandoval-Reyes, C. A. Silva, S. Vieira, and J. M. Sousa, "Assessment of energy efficiency measures using multi-objective optimization in portuguese households," *Sustainable Cities and Society*, vol. 35, pp. 764–773, 2017.

[19] K. Gajowniczek and T. Zḱabkowski, "Electricity forecasting on the individual

household level enhanced based on activity patterns," *PloS one*, vol. 12, no. 4, p. e0174098, 2017.

[20] N. Gudi, L. Wang, and V. Devabhaktuni, "A demand side management based simulation platform incorporating heuristic optimization for management of household appliances," *International Journal of Electrical Power & Energy Systems*, vol. 43, no. 1, pp. 185–193, 2012.

[21] J. Han, C.-S. Choi, W.-K. Park, and I. Lee, "Green home energy management system through comparison of energy usage between the same kinds of home appliances," in *Consumer Electronics (ISCE), 2011 IEEE 15th International Symposium on*. IEEE, 2011, pp. 1–4.

[22] L. Hernández, C. Baladrón, J. M. Aguiar, L. Calavia, B. Carro, A. Sánchez-Esguevillas, D. J. Cook, D. Chinarro, and J. Gómez, "A study of the relationship between weather variables and electric power demand inside a smart grid/smart world framework," *Sensors*, vol. 12, no. 9, pp. 11 571–11 591, 2012.

[23] W.-C. Hong, "Electric load forecasting by seasonal recurrent svr (support vector regression) with chaotic artificial bee colony algorithm," *Energy*, vol. 36, no. 9, pp. 5568–5578, 2011.

[24] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Technical report-tr06, Erciyes university, engineering faculty, computer engineering department, Tech. Rep., 2005.

[25] D. Karaboga and B. Akay, "Artificial bee colony (abc) algorithm on training artificial neural networks," in *Signal Processing and Communications Applications, 2007. SIU 2007. IEEE 15th*. IEEE, 2007, pp. 1–4.

[26] S. Kirbiz, Y. Ulker, and B. Gunsel, "A pattern recognition framework to blind audio watermark decoding," *AEU-International Journal of Electronics and Communications*, vol. 63, no. 2, pp. 92–102, 2009.

[27] W. Kleiminger, C. Beckel, T. Staake, and S. Santini, "Occupancy detection from electricity consumption data," in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. ACM, 2013, pp. 1–8.

[28] U. Kumar and V. Jain, "Time series models (grey-markov, grey model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in india," *Energy*, vol. 35, no. 4, pp. 1709–1716, 2010.

[29] T. Kurita, "Principal component analysis (pca)," in *Computer Vision*. Springer, 2014, pp. 636–639.

[30] A. Kusiak, M. Li, and F. Tang, "Modeling and optimization of hvac energy consumption," *Applied Energy*, vol. 87, no. 10, pp. 3092–3102, 2010.

[31] J. Kwac, J. Flora, and R. Rajagopal, "Household energy consumption segmentation using hourly data," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 420–430, 2014.

[32] H.-Z. Li, S. Guo, C.-J. Li, and J.-Q. Sun, "A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm," *Knowledge-Based Systems*, vol. 37, pp. 378–387, 2013.

[33] K. Li, L. Pan, W. Xue, H. Jiang, and H. Mao, "Multi-objective optimization for energy performance improvement of residential buildings: A comparative study," *Energies*, vol. 10, no. 2, p. 245, 2017.

[34] D. Mahmood, N. Javaid, N. Alrajeh, Z. A. Khan, U. Qasim, I. Ahmed, and M. Ilahi, "Realistic scheduling mechanism for smart homes," *Energies*, vol. 9, no. 3, p. 202, 2016.

[35] S. Makonin, "Ampds2: The almanac of minutely power dataset (version 2)," 2016. [Online]. Available: http://dx.doi.org/10.7910/DVN/FIE0S4

[36] S. Makonin, F. Popowich, L. Bartram, B. Gill, and I. V. Bajic, "Ampds: A public dataset for load disaggregation and eco-feedback research," in *Electrical Power & Energy Conference (EPEC), 2013 IEEE*. IEEE, 2013, pp. 1–6.

[37] A. Mishra, D. Irwin, P. Shenoy, and J. Kurose, "Smartcharge: Cutting the electricity bill in smart homes with energy storage," in *Future Energy Systems:*

Where Energy, Computing and Communication Meet (e-Energy), 2012 Third International Conference on. IEEE, 2012, pp. 1–10.

[38] H. Murata and T. Onoda, "Estimation of power consumption for household electric appliances," in *Neural Information Processing, 2002. ICONIP'02. Proceedings of the 9th International Conference on*, vol. 5. IEEE, 2002, pp. 2299–2303.

[39] D. Napoleon and S. Pavalakodi, "A new method for dimensionality reduction using k-means clustering algorithm for high dimensional data set," *International Journal of Computer Applications*, vol. 13, no. 7, pp. 41–46, 2011.

[40] D. Niu, Y. Wang, and D. D. Wu, "Power load forecasting using support vector machine and ant colony optimization," *Expert Systems with Applications*, vol. 37, no. 3, pp. 2531–2539, 2010.

[41] R. Platon, V. R. Dehkordi, and J. Martel, "Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis," *Energy and Buildings*, vol. 92, pp. 10–18, 2015.

[42] M. B. Rasheed, N. Javaid, A. Ahmad, M. Jamil, Z. A. Khan, U. Qasim, and N. Alrajeh, "Energy optimization in smart homes using customer preference and dynamic pricing," *Energies*, vol. 9, no. 8, p. 593, 2016.

[43] A. Reinhardt, D. Christin, and S. S. Kanhere, "Predicting the power consumption of electric appliances through time series pattern matching," in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings.* ACM, 2013, pp. 1–2.

[44] S. Sharavanan and M. Azath, "Lda based face recognition by using hidden markov model in current trends," *International Journal of Engineering and Technology*, vol. 1, no. 2, pp. 77–85, 2009.

[45] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, "Predicting solar generation from weather forecasts using machine learning," in *Smart Grid Communica-*

tions (SmartGridComm), 2011 IEEE International Conference on. IEEE, 2011, pp. 528–533.

[46] S. Singh and A. Yassine, "Mining energy consumption behavior patterns for households in smart grid," *IEEE Transactions on Emerging Topics in Computing*, 2017.

[47] R. Subbiah, A. Pal, E. K. Nordberg, A. Marathe, and M. V. Marathe, "Energy demand model for residential sector: a first principles approach," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 3, pp. 1215–1224, 2017.

[48] W. Tian, J. Song, and Z. Li, "Spatial regression analysis of domestic energy in urban areas," *Energy*, vol. 76, pp. 629–640, 2014.

[49] P. Tsai and B. Pan, "Intelligent electricity bill saving system for smart home," in *Fuzzy Theory and Its Applications (iFUZZY), 2015 International Conference on.* IEEE, 2015, pp. 105–108.

[50] F. Wahid, R. Ghazali, A. S. Shah, and M. Fayaz, "Prediction of energy consumption in the buildings using multi-layer perceptron and random forest," *IJAST*, vol. 101, pp. 13–22, 2017.

[51] J. Wang, S. Zhu, W. Zhang, and H. Lu, "Combined modeling for electric load forecasting with adaptive particle swarm optimization," *Energy*, vol. 35, no. 4, pp. 1671–1678, 2010.

[52] Z. Wang and G. Zheng, "Residential appliances identification and monitoring by a nonintrusive method," *IEEE transactions on Smart Grid*, vol. 3, no. 1, pp. 80–92, 2012.

[53] S. N. Wood, "mgcv: Gams and generalized ridge regression for r," *R news*, vol. 1, no. 2, pp. 20–25, 2001.

[54] L. Yao, J.-Y. Shen, and W. H. Lim, "Real-time energy management optimization for smart household," in *Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2016 IEEE International Conference on.* IEEE, 2016, pp. 20–26.

[55] H.-x. Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3586–3592, 2012.

List of Publications

International Journal

1. Jasmeet Kaur, Anju Bala *Predicting power of Home Appliances Based on Climatic Conditions*, International Journal of Energy Sector Management, Emerald. [Under Minor Revision, ESCI]

2. Jasmeet Kaur, Anju Bala *A Hybrid Energy Management Approach for Home Appliances using Climatic Forecasting*, An International Journal of Building Simulation, Springer. [Under Major Revision, SCI Expanded]

3. Jasmeet Kaur, Anju Bala *A Comparative Analysis: Energy Optimization Techniques for Residential Sector*, Mitigation and Adaptation Strategies for Global Change, Springer. [Under Review, SCI Expanded]

International Conference

1. Jasmeet Kaur, Anju Bala *Review of Machine Learning Techniques for Optimizing Energy of Home Appliances* , Information and Communication Technology for Competitive Strategies, Springer. [Accepted, In Press]