Analysis of Deep Learning and predictive models for energy consumption forecasting of Buildings

By

Muhammad Usman

00000119268

Supervisor

Dr. Rizwan Ahmad

A thesis submitted in conformity with the requirements for the degree of *Master of Science* in Information Technology

School of Electrical Engineering and Computer Science (SEECS)

National University of Sciences and Technology (NUST)

Islamabad, Pakistan

(August 2019)

Approval

It is certified that the contents and form of the thesis entitled **"Analysis of Deep Learning and predictive models for energy consumption forecasting of Buildings"** submitted by **Muhammad Usman** have been found satisfactory for the requirement of the degree

Advisor : Dr. Rizwan Ahmad
Signature :____________________
Date :____________________

Committee Member 1: Dr. Arsalan Ahmad
Signature :____________________
Date :____________________

Committee Member 2: Dr. Salman Abdul Ghafoor
Signature :____________________
Date :____________________

Committee Member 3: Dr. Safdar Abbas Khan
Signature :____________________
Date :____________________

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by Mr/Ms Muhammad Usman, (Registration No. 00000119268), of School of Electrical Engineering & Computer Science (SEECS) (School/College/Institute) has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/M Phil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature:____________________

Name of Supervisor:Dr. Rizwan Ahmad

Date :____________________

Signature (HOD): ____________________

Date :____________________

Signature (Dean/Principal):____________________

Date :____________________

Certificate of Originality

I, *Muhammad Usman* declare that this thesis titled "Analysis of Deep Learning and predictive models for energy consumption forecasting of Buildings" and the work presented in it are my own and has been generated by me as a result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a Master of Science degree at NUST
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at NUST or any other institution, this has been clearly stated
3. Where I have consulted the published work of others, this is always clearly attributed
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work
5. I have acknowledged all main sources of help
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself

Muhammad Usman,

00000119268

Copyright Notice

I am dedicating this thesis to *my beloved wife and children*

Abstract

The modern world is living in a digital age which demands more energy for continuous operation of devices, gadgets, machines, household, etc. Consequently, there must be some system for energy management in terms of harvesting, generation, reservation, and balanced use. Which entails research for efficient machine learning forecasting models to better predict future consumption. Since most of the modern machine learning models are data-driven rather Engineering. Therefore for this study, we are focusing to predict the building energy consumption using deep learning and other contemporary models trained on real data captured for four years which was consumed by an educational building situated in London, United Kingdom. These six machine learning classifiers are, Bagging, Boosting, Random Forest, Deep Neural Network (DNN),Support Vector Regression (SVR), Artificial Neural Network (ANN). The prediction models are fed the same data for four years on various criteria such as outer temperature, solar radiation, wind speed, humidity, and working day indicator. The last year data was used for testing the predictive values of all the models. Results have shown that the last month in test data seems to be an outlier as dropping it improves performance by 2 %. Furthermore, the comparison also made for office day known as working day and non-working day known as non office day using weekday indicator. The trained models are used to predict electricity consumption units and all classifiers are compared with actual utilization units of electricity for last year. Results reveal that ANN proves itself to the best of all five approaches achieving a Mean Absolute Percentage Error (MAPE) of 6.41% where DNN, SVR, Bagging, Boosting, Random Forest has achieved MAPE of 11.15%, 9%, 7.46%, 8.46% and 9.84% respectively. This work can be extended to other building energy-related problems with respect to management, conservation, mitigation, and proper utilization.

Keywords: *Artificial Neural Network (ANN), Deep Neural Network (DNN), Bagging,* *Boosting, Random Forest, Prediction, Energy Forecasting, Support Vector Regression (SVR)*

Acknowledgments

I pay my gratitude to Allah almighty for blessing me a lot and without His guidance I couldn't complete this task. I am thankful to Dr.Rizwan Ahmad for guiding and encouraging me to complete my thesis. His timely and efficient contributions helped me to shape the thesis into its final form and I express my gratefulness for his sincere supervision all the way. I am also thankful to Dr. Khuram Pervez Amber, Department of Mechanical Engineering, Mirpur University of Science and Technology (MUST), Mirpur AJK, who provided the data set and timely guidance for successful completion of this research.

I am also thankful to my GEC members (Dr Salman Abdul Ghafoor, Dr Safdar Abbas Khan and Dr Arsalan Ahmad) and Department of Computing for providing me with an academic base, which enabled me to complete this thesis. I'd like to thank my parents who have provided me every comfort of life and their utmost affection and support made me reach where I am today.

Contents

List of Figures

List of Tables

List of Algorithms

List of Main Abbreviations

Abbreviations

ANN	Artificial Neural Network
DNN	Deep Neural Network
kWh	Kilo Watt per Hour
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MRE	Mean Relative Error
MSE	Mean Square Error
NRMSE	Normalized Root Mean Square Error
NWD	Non-Working Day
RMSE	Root Mean Square Error
WD	Working Day

Chapter 1

Introduction

1.1 Motivation

Since we are living in a digital age in which we are totally driven by digitally equipped devices. There is mushroom growth of these computing devices known as smart devices, gadgets, Wearable Sensor Body Area Network Devices, Computing Devices, and other Electric and Electronic Devices. Each device has some sensors in it, populating a huge amount of data continuously. Even in underdeveloped countries, a remote and far-flung area man is equipped with 2 devices on average. According to Statista [1] each person on earth will have 6 devices on average in 2020. These devices may contain smartphones, tabs, PCs, Laptops, Servers, Wearable Sensor Devices, RFID Smart Tags, etc. All these devices require Energy for their continuous and smooth working. Furthermore, the concept of Electric Cars is prevailing to have less hazardous and ecofriendly [2]. In future Electrical Cars demand efficient energy management [3]. The idea of Green homes which require managed usage of electricity to mitigate its consumption and designing of houses in such a way to gain maximum benefit from environment factors e.g. light, air ventilation, use of insulating material to avoid buildings from outside temperature, etc so that they become ecofriendly. All the above discussion forces researchers to work for all facets of energy whether it is harvesting, conservation, storage, management, prediction for future use, etc. It has been proved by researchers that 40 % of overall Energy is consumed by buildings and it is increasing with a ratio of 1.8 % every year [4].

In Buildings major demand of energy consumption is from Academic and administration buildings as they require more electricity. Due to the availability of big data and plenty of storage on clouds, it is possible to have history data related to energy consumption along with all those parameters which may affect its demand [5, 6]. Nowadays machine learning is playing a key role in eliciting meaningful information after training over a large number of data sets. This information is further fed to some trained model for predicting some target variable, the likelihood, auto text generation, speech generation, text, character and speech recognition, clustering and grouping of data, visualization of multidimensional data, etc. The country like Pakistan is in dire need of saving energy as it is coping up with energy crises. So it is better to predict energy consumption by employing different machine learning models and analyzing their performance. Deep learning is a new area of research which is becoming popular due to its performance that surpasses all other non-deep learning-based models. Albeit deep learning is evolved from the old model of neural network which was not in use due to non-availability of enormous computing machines. It became popular as it performs very well in the field of computer vision after the invention of Convolutional Neural Network [7]. We want to study and analyze different deep learning architectural models to forecast the energy consumption for higher education institute.

1.2 Overview

Energy has an inevitable role in the modern world to properly run machinery for human benefit . It is the driving force in any field of a nation. Many researchers are focusing on diverse applications of energy utilization, conservation, mitigation, prediction, etc. Researches have been carried out in each sector which revolves around energy, viz. energy harvesting, solar energy adaptation, eco-friendly energy, energy reservation, green buildings, etc. Prediction of energy beforehand is of utmost need in all domain of energy consumption. In recent research, it is an admitted fact that about 40 % of the energy consumption in these days is due to the building sector. Another research reveals the fact that there is an annual 3 % increase in energy in the UK and USA, while about 1.5 % in the rest of the world. China alone is the second-largest energy consumer in world [8]. In this research, our main focus is to forecast the energy consumed by a higher education institution by applying different well-known machine learning models.

Since machine learning prediction models are playing an important role both in supervised as well as in unsupervised learning. In supervised learning, we have some dependent variable which can be computed using independent variables. We have to train our model using training data to correctly predict the dependent variable as well as avoid over and underfitting. While in unsupervised learning we use clusters, fragments, and groups of unknown data so that we achieve such clusters or groups which will better reflect the meaning of data points. In this study, we have real-time 5 years of data of higher education institute situated in London. In recent past, there was also a study carried out on the above-mentioned data. But only two prediction model one is statistically based named multiple regression model [9]. It is a modification of linear regression model which is used to predict the ground truth when we are able to make line closer to training data points while in multiple regression model we still intended to draw a virtual line in hyperplane and then predict the test data point by applying different learned constants.

These constants play the role of coefficients applied to independent variables to predict the dependent variable value. Overfitting and underfitting are two terms involved in machine learning, to avoid overfitting in our trained model we apply different approaches e.g. m-fold method. The second approach which was used in previous research was the genetic programming approach. In this approach, we make learned generations multiple times and our goal is to find that generation which is closest to the training data. In this research, our main focus is to utilize certain well-known prediction models on our data set which has outperformed in different domains e.g. image recognition, speech and text recognition, machine health prognosis, intrusion detection, etc.

The models which are achieving unmatched results, known as neural networks. We employed two models of the neural network one is ANN while other is DNN. The main difference in ANN and DNN is that in ANN we used sigmoid function and gradient descent for activation and backpropagation. While in DNN we use the different number of weights as happened in Convolutional Neural Network (CNN), different activation functions e.g. Relu, sigmoid, tanh, etc. Similarly, Pooling layers (Max Pool, Average Pool) are used to downsize the problem. While more and more hidden layers, as well as neurons, are introduced for self-feature extraction. In the final layer of the DNN, softmax function is applied to get predicted output. We have also employed a novel and well-known approach which perform outclass in non-separable data points. This model is Support Vector Regression, which finds out those data points who can serve as support vector. Support vectors are those points which can place a maximum separable line between two groups or clusters. Then in regression, we place the unknown data point at in that cluster which best fits.

For performance evaluation and assessment of the different machine learning algorithms on the real data, we use six different error evaluation parameters viz. Round Mean Square Error (RMSE), Mean Absolute Error (MAE), Normalized Round Mean Square Error (NRMSE), Mean Absolute Percentage Error (MAPE), Mean Relative Error(MRE). These different parameters reveal different performance factors of the prediction model.

For this study, the data we have gathered is from an institutional building situated in UK, known as South Bank Techno Park. The five parameters of observation/measurements of the above-stated building viz.

- Mean of Daily weather Temperature measured in Kelvin
- Mean of daily global radiance (W/m2)
- Mean of Daily Humidity (%)
- Mean of Daily wind Velocity (m/s)
- Last parameter is Week Day Indicator

- Daily Consumption of Electricity (Wh/m^2)

This study is based on data gathered over a span of 5 years i.e. from Jan, 2007 to Dec, 2011. We split the data into two halves. For training purpose, we will use 4 years data and for testing purpose, we will use one-year data for prediction and match it to the original data to evaluate and assess the performance.

First part of the thesis, There is a brief introduction. In second chapter our main focus will be on literature review while in the third chapter we will briefly tell the methodology in which we describe how mathematically these models will be applied to the given data and how they will be trained to forecast the future energy utilization.

1.3 Applications of Machine Learning

The most common application in supervised learning are business, economics, manufacturing, pharmaceutics and pharmaceutical, wired and wireless telecommunication, bioinformatics, data mining, etc.

1.3.1 Association Analysis

The field of association analysis, we employ association rule mining an approach which works on co-relation to find an association among different things. These type of elicitation is used in a retail store, customer likelihood, etc.

1.3.2 Supervised Learning

Supervised learning is a subfield under umbrella of machine learning. In supervised learning, target variable, which have some ground-truth value or dependent variable is used to map all independent variables to work for it. We train our model or classifier to predict to infer about this dependent value. Then after training, we use some test data to assess how well our classifier is predicting. Overfitting and underfitting are two terms related to model in machine learning.

Over fitting

While training the model, we train the model in a way that for each data point fed to model for training purpose. If after training the model predicts the exact output as expected then our classifier will be over fitted and it will predict the specific value which may not be the Case in an unknown data point. To avoid overfitting, we apply some regularization using hyperparameters to cost the model if during training it does predict some deviated value.

Under fitting

In machine learning, it is the case when we train our model with a limited number of data set. So our model is well trained to properly predict the dependent variable on some test data point.

In supervised learning we do face recognition, regression type problems, knowledge extraction, trends in social and other media applied on big data, outlier detection, intrusion detection.

1.3.3 Unsupervised Learning

Unsupervised learning is use when we have huge data and in beginning, we dont know much about data to elicit information, we have to make clusters or chunks of data in such a way that after clustering we may infer something interesting and informative. So many clustering algorithms like K Nearest Neighbor, mean based algorithm, etc. are applied to visualize that data. Mostly unsupervised learning is applied in the census, city and town planning, customer segmentation, in compression, bioinformatics, etc.

1.3.4 Reinforcement learning

This subfield of machine learning can also be placed in unsupervised learning. In this type of learning, there is no ground truth value. Rather a reward-based strategy is applied. An action is taken after the model predicts some action. After that model will be rewarded or penalized according. The goal of the model is to maximize rewards. Game playing, Robot, etc. are common examples of Reinforcement learning.

1.4 Contribution

Pakistan is in energy crisis for the last two decades [10, 11]. If we are able to predict the energy consumption accurately then we will be aware of the demand supply gap and take apt steps to address the problem. If we are able to identify some deviation from the predicted energy then we can further proceed to stop the theft in energy as happening in some area of our country. This work will lead to analyze other data sets to be employed using deep learning model and can be utilized for their own specific field prediction.

Chapter 2

Literature Review

Recent researches proved that Buildings are the most expensive sector of energy consumption as compared to other economic sectors. In research conducted by S.Gul, it is proved that about 30-45 % energy demand is only from building sector [12]. They in their research focus on detail energy consumption of an educational institute. According to their research two approaches are required to mitigate the energy consumption. First, is long term planning in building design as well as properly maintaining the central Heating, Ventilation and Air Conditioning (HVAC) may result in lesser consumption. While in research conducted by Berardi tells about how different countries are consuming electricity and found that a major part of consumption is due to building energy [13]. They also demand some good future prediction model for effective use of electricity. Another research did by [14] on different papers published on energy and its different domain from 1976-2016. They demand in their research that energy efficiency approaches are highly demanding. Energy consumption forecasting is done by many researchers using different approaches. Chengdong Li [15] combines Stake Autoencoders (SAE) and Extreme Learning Machine (ELM). In this research, they adopted many other prediction models. Another research for energy consumption in the manufacturing sector of China is carried out by [16]. In a review paper by AS Ahmad and MY Hassan [17] on comparison of ANN and SVR and other engineering approaches taken by other researchers for prediction. In nutshell, we can conclude that there are three different approaches for prediction Statistical, Artificial Intelligence (AI), and Engineering. Engineering approaches for energy forecasting involves architectural and design approaches including the material used, climate factor, weather condition, thermal properties of the building material, etc. While in statistical approach main focus remains on previous data and finding out most dominating parameters affecting the prediction. Some most famous statistical models are Gaussian Mixture model, conditional Demand Analysis, Regression models and autoregressive moving average [18–22]. AI is playing the main role in the field of prediction, prognosis, and forecasting. Machine Learning comes under the umbrella of AI, has a grip in the field of prediction. The models in AI mainly focuses on historical data upon which they trained their model and then the prediction is done using these trained models. Machine Learning provides a variety of supervised and unsupervised learning models. But the most surpassing models are neural network-based. Because they have the ability to learn a complex model by means of weights and back-propagation. Researchers have tried to overcome some problems faced by neural network e.g. diminishing gradients by introducing Long Short Term Memory (LSTM) cell. Some examples of AI prediction models used for forecasting are SVR, Genetic Programming and Genetic Algorithm, ANN , Fuzzy Logic, are used by different researchers [23–25]. Due to increased demand for energy in industry and economy, management and prediction of energy is now more demanding and it becomes an integral part of the energy industry. Kovacic and Sarler [26] used two statistical and AI model for natural gas utilization forecasting in Slovenia for Company of Steel Store which alone uses 1.1 % Gas of the country. They achieved the performance of GP over MR. In another work, K. P. Amber et al. [27] extended the earlier work to compare with ANN, DNN and SVR. In another study , Silva et al. [28], uses GP, find out different impacting Factors like social, political and economic, and after investigating they proposed a forecasting model which predicts long term energy. The proposed the solution till 2050. The sector was industrial sector in Brazil. Residential buildings forecasting for Energy Consumption was also studied by Biswas et al. [29] in a research. TxAIRE Research Houses were used for data collection. They used coefficients of determination R^2 for performance of the ANN. The values of R^2 were remain between 0.87 and 0.91 in their research. In Oil Industry ANN has also performed significant tasks [30][31]. In the study conducted by Zeng et al. [31] for electricity consumption forecasting for oil driven pumps, they employ 3 distinct machine learning approaches to forecast electricity. The main use of electricity was for pumps which were used to extract oil in China. They trained their model for optimization by employing step by step tweaking after each error process and for performance evaluation of results they uses MAPE and RE on SVR, ANN and MR.

Astonishing thing is in this study ANN performs well and achieved an error of ±5%.

A study was conducted in Poland [32] to predict the values of diurnal and seasonal gas consumed using several environmental and calendar variables. The conclusion of the study was in the favor of ANN. They proposed the consumption for hourly basis and for daily gas utilization. The MAPE value of this study calculated to 9.0 %.

Support Vector Regressions are very good predictors and performing in various diversified applications. They are used for prediction of electricity load as studied by [33][34] and also used electricity consumption as assessed by [35],hydropower consumption [36] and heating demand [37]. In another study conducted by [38] in 2017 for forecasting electricity consumption for educational building some independent variable found to be less effective for statistical approach of multiple regression.

Chapter 3

Proposed Methodology

AI, superset of machine learning plays an imperative role in all domains where data is involved. Since it is apparent that in this age we are having data everywhere. Therefore, applying different machine learning algorithm to find out some meaning full information which is worthwhile for future management is indispensable nowadays. For energy consumption forecasting we have applied six different prediction models which are SVR, ANN, DNN, Bagging, Boosting and Random Forest. Now let us take a look into all of these models, their pros and cons, etc.

3.1 Support Vector Regression

SVR is supposed to be very good machine learning classifier before the neural net and in certain cases still, it is. They are applicable to both separable also known as linear as well as non-separable also known as nonlinear or complex structural problems. Support Vector Machine is basically an update on generalized portrait algorithm which was developed in Russia in 60-70 by Vapnik Lerner and Chervonenkis [39]. Now again in AT & T lab, they have been designed by Vapnik and co-workers for industrial utilization. In Industry there were at that time focused to work for Optical Character Recognition (OCR). But it also proves its capability for object recognition at that time and becomes the best machine learning classifier for prediction, classification, regression, etc. However with the advent of high-performance computing at the end of 20th century it was made possible to simultaneously compute tensors so again an era of research is opened for neural networks and after recurrent neural network and convolution neural network

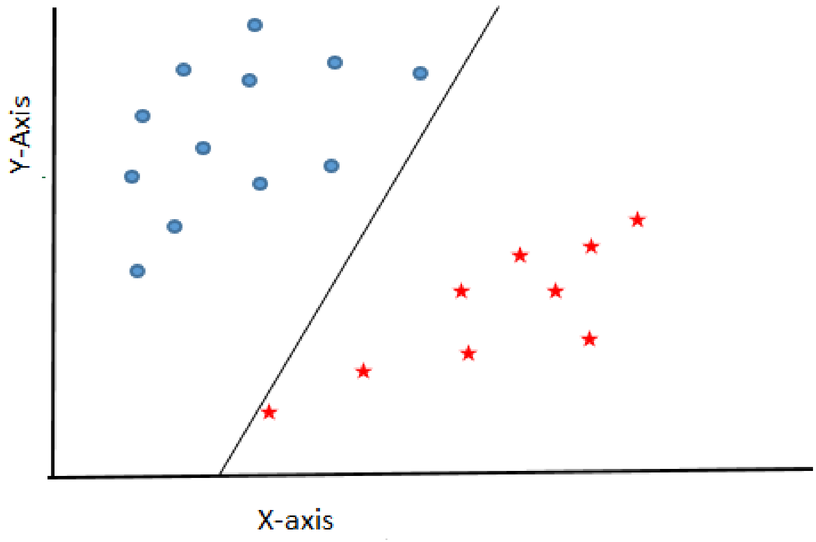

Figure 3.1: A separation line between two data points.

Support Vector Regression regression was considered to be a bit behind but still it is playing roles in many fields.

3.1.1 Working of Support Vector Regression for Linear Problem

Suppose we have been given data set comprised of data points in the form of (x,y). We want to separate all the data points in such a way that we place a line between different classes in such a way that this line is at a maximum margin from the closest point of each cluster or group. These closest points to each cluster are known as support vectors.

Diagrammatically it can be represented in Fig.3.1.

Some two dimensional data is represented in Fig.3.1 with a separable line. Then another line is placed between same data in Fig.3.2. In Fig.3.3 so many separable lines have been shown. Finally in Fig.3.4 it is apparent that we want to separate both circled and stars data points into two groups. We can place a separable line as mentioned in Fig.3.1, and also a line can be placed as in Fig.3.2. In Fig.3.3 so many lines have been shown which are separating both the classes of data points. But our intention is to classify in such a way that it remains at a maximum distance from the support vectors (farthest points from the marginal line) as mentioned in Fig.3.4 the circled points of both groups are representing the support vectors.

Mathematically we can formulate it as a problem of finding some w and b in such a way that $\langle \mathrm{w}^T, x_i \rangle + b \geq 1$ subject to $y_i = 1$; $\langle \mathrm{w}^T, x_i \rangle + b \leq -1$ if $y_i = -1$ is maximized for all data points(x_i, y_i).
A better approach is to attain those random values of w and b such that $\phi(w) = \frac{1}{2}w^T w$ acquire minimum value subject to for all $(\mathrm{x}_i, y_i), providedthat$ $\mathrm{y}_i(\langle \mathrm{w}^T, x_i \rangle + b) \leq 1$.

If the training data is in such a way that it is difficult or even impossible to separate the data then we introduce slake variables, and leave some of errors.

After introducing slake variables new mathematical form will be attain some weights w and biases b such that $\phi(w) = \frac{1}{2}w_w^T + C \sum \zeta i$ have its minimum value for all (x_i, y_i) $(\langle \mathrm{w}^T, x_i \rangle + b) \geq 1 - \zeta i$ and $\zeta i \geq 0$ for all i where C is overhead or variable used for regularization.

The above-specified algorithm will work for linearly separable data points. But if the data set is in such a way that it is noisy or quite hard to draw a separate line between data points of two different groups then we apply the idea of moving the data points from original dimension to some new dimension. We apply some function known a kernel function as dot kernel is used in this research. Such that it will transform the data points dimension.

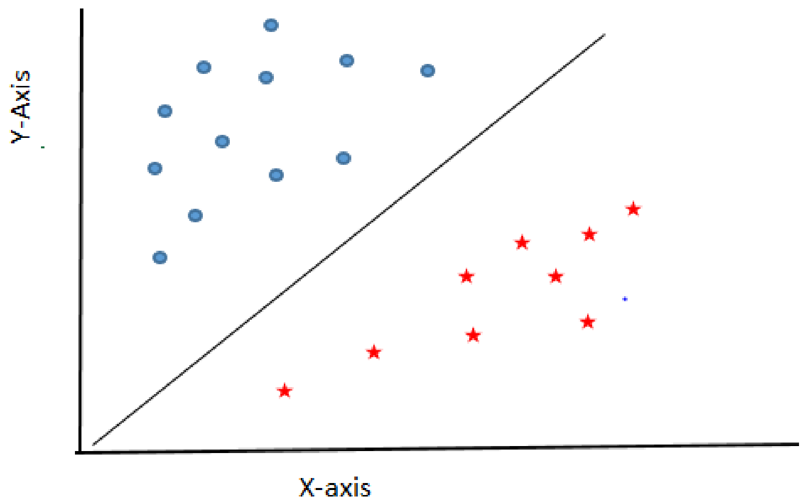

Figure 3.2: Another separation line between same data points

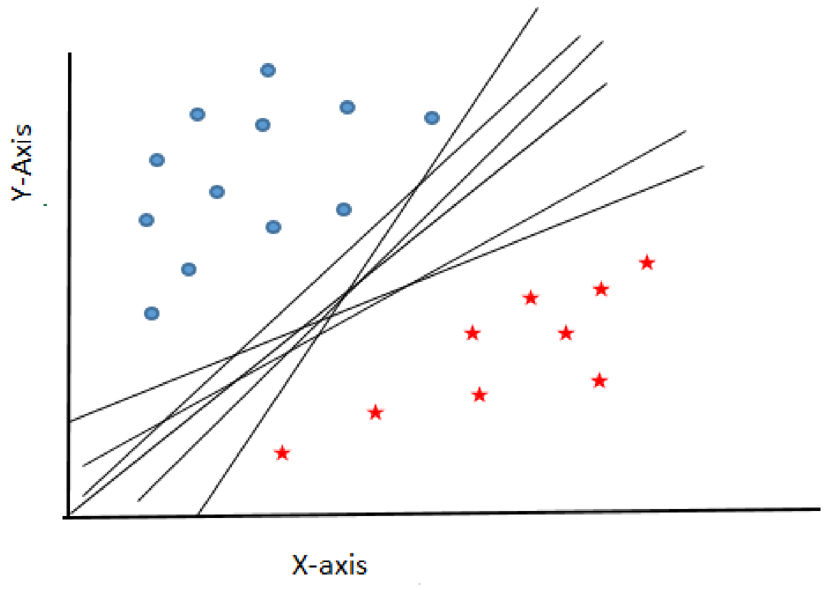

Figure 3.3: So many separation lines between same data

Diagrammatically the concept is shown in Fig.3.5 to Fig.3.6 transforming one line data to plane data. While the same concept is applied in Fig.3.7 and in Fig.3.8 transforming the data from 2 dimension to 3 dimension to make it separable.

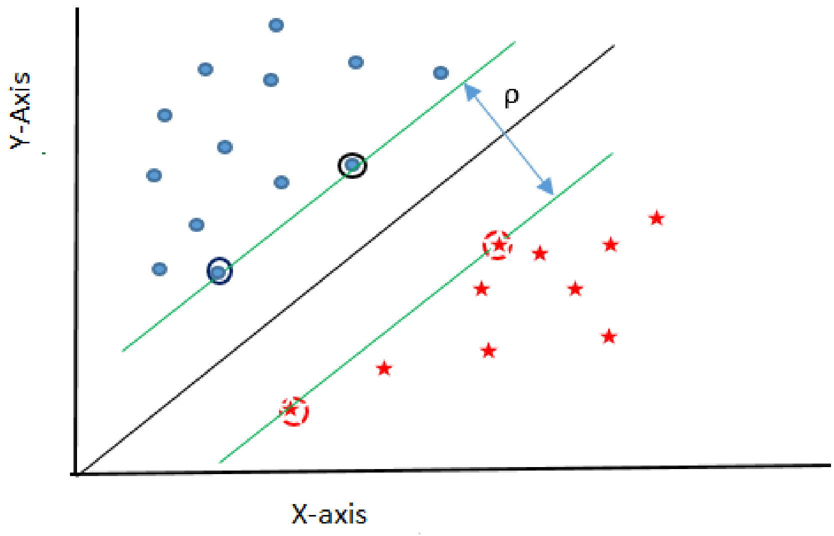

Figure 3.4: Required separation line between data points at maximum margin

Area of applications of Support Vector Regressions are used in web data and alphanumeric data classification, Image Recognition, and Classification, protein and DNA strand classification, malignant cell Classification, optical character recognition, handwriting recognition, etc.

3.2 Artificial Neural Network

Our next prediction model is an ANN which has a very good impact on energy consumption forecasting as is evident from the recent research [9, 40, 41]. Talking about the evolution of Neural Network, first thinking about the neural network is made in about 1958 by McCulloch and Pitts and it was named perceptron. However major contribution was made by Hopefield energy approach. ANN is basically a copy or mimicry of the human brain. Since it almost maps on the human brain, therefore, the terms involved in an ANN are somehow a human-made version of terms involved in a biological neural

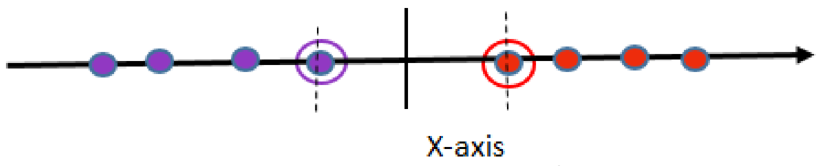

Figure 3.5: When data is not linearly separable

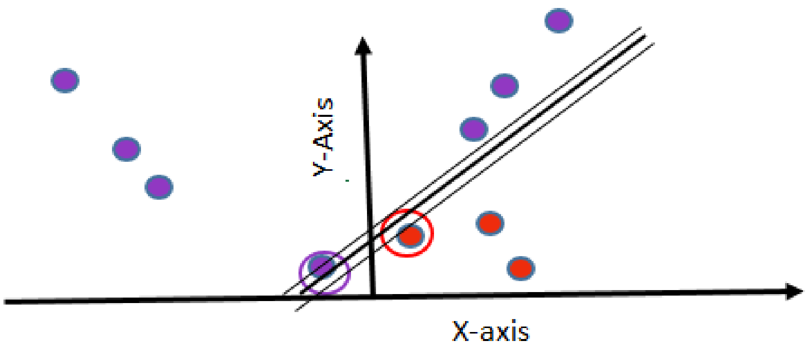

Figure 3.6: Transforming data from one dimension to another

network. A human brain neuron consists of the following four elements.

- Dendrite: These are branches which are responsible for taking input
- Soma: Almost all of the processing done here in soma
- Axon: It is the output generated from neurons and send via connection to other neuron
- Synapses: It works as connection between axon and dendrite

A Human brain neuron is also shown in figure 3.9. Now all above-mentioned parts of a human neuron are maps to artificial neuron in the table 3.1.

Depiction of a simple ANN can be seen in Fig. 3.10.

The simplest ANN is an architectural composition of at minimum three layers connected to each other, the very first layer also known as input layer used for feeding independent variables, the central layer also known as hidden layer have some activation function and the third layer which is called output layer. As depicted in the figure 3.11.

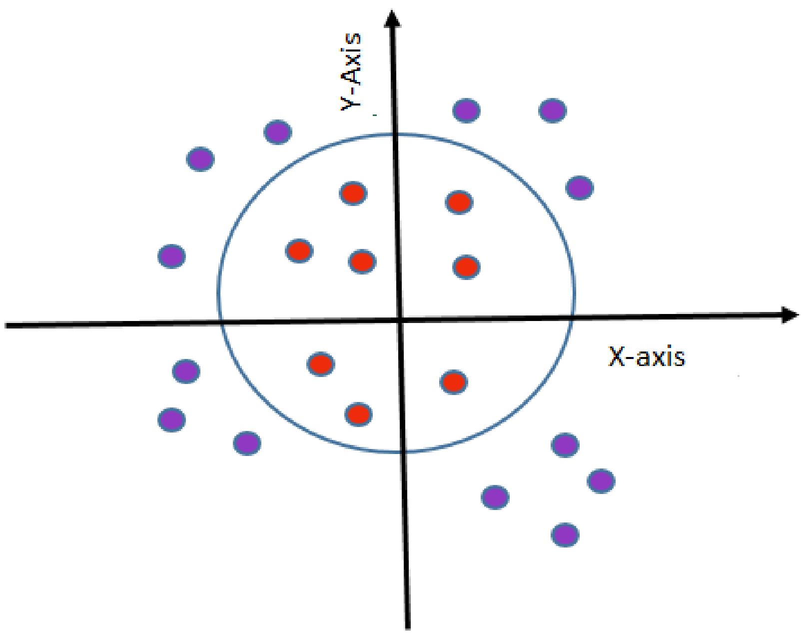

Figure 3.7: Another view of data which is not separable

Table 3.1: An analogy between human brain neuron and artificial neuron

Sr.No.	Human Neuron	Artificial Neuron
1	Dendrite	Input
2	Soma	Node
3	Axon	Output
4	Synapses	Weight

There are two phases of any classifier for prediction

- Training
- Testing

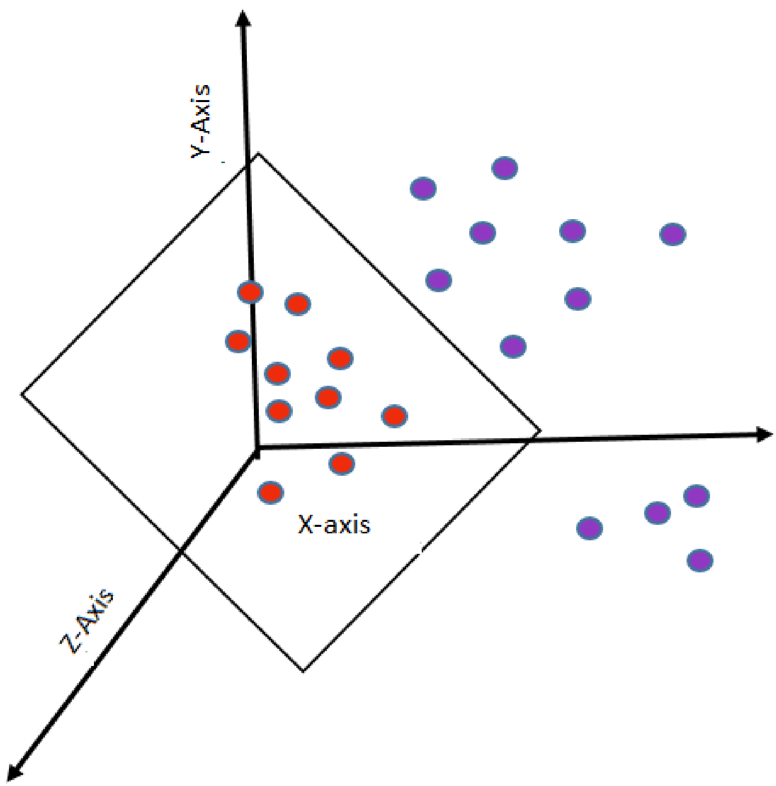

Figure 3.8: Transformation of Data from one dimension to other

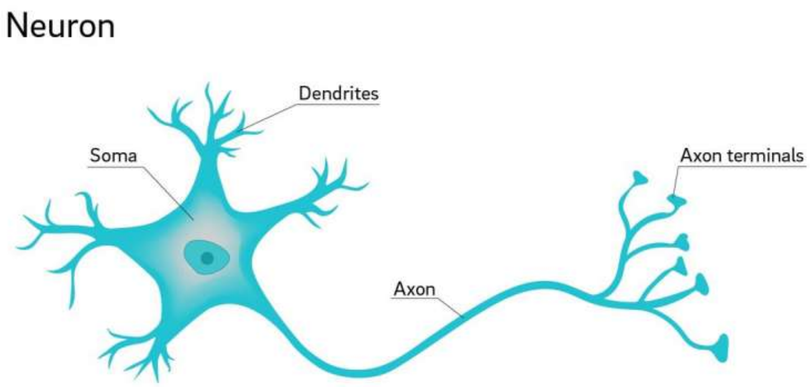

Figure 3.9: A Human Brain Neuron [42].

3.2.1 Training

During the training phase, we provide all the data points are sample data in the form of tensor fed to the first layer serves as input layer and take dot product with weights initially weights are kept random. A bias is added to each single neuron input and passes next This output is then fed into some activation function and the output is fed to the next layer as input. In this fashion, the output of each neuron in layer is passed

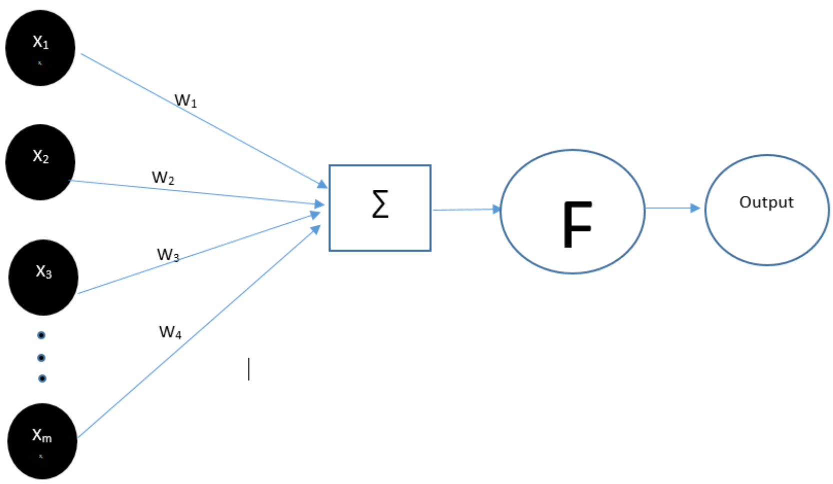

Figure 3.10: Basic Neural Network Diagram

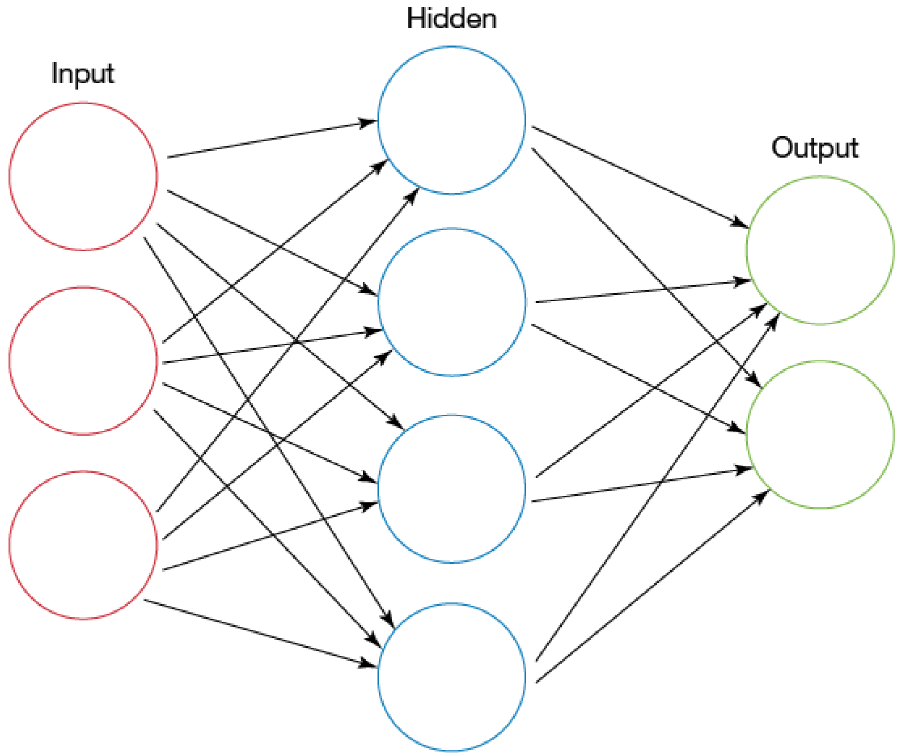

Figure 3.11: Diagram of Simplest Artificial Neural Network with three layers [43].

through some function and passed to next layer. During this Research, we have applied 10 layers and each layer composed of 100 neurons. In the last layer, we apply some mapping function which gives output and this output is passed to some loss function. The loss function basically finds out how much this output has deviated from the ground truth or the real value. Again taking this loss function we compute and backpropagate throughout the network. Normally we use the derivative of the loss to backpropagate in the network. In this way when one data point passed through the network, random weights and biases are adjusted. Again we pass another data point and repeatedly do this work until all of the data points are passed on and we got final weights. The common output from each layer can be mathematically expressed as

$$O_i x = \frac{1}{1 + e^{-\sum(w_i x_i + b)}} \tag{3.2.1}$$

3.2.2 Testing

After the training phase, we pass some other data point to assess and evaluate how well our network is trained. During the testing phase, we only our network work in a fashion of feed-forward network. While during training it works in feedforward and backpropagation network.

Activation Functions

Most famous and popular activation functions are

- Logistic or Sigmoid Function
- Tanh
- ReLu(Rectified Linear Unit)
- leaky Relu
- Randomized Leaky ReLu

Logistic or Sigmoid Function

It is a mathematical function which is real, bounded, differentiate-able and it has a nonnegative derivative. The sigmoid function is a very famous function used for activation. It squashes the value of input between 0 and 1 as shown in the figure below. Mathematically we can express the sigmoid function as

$$e^x = \frac{1}{1 + e^{-}x} \tag{3.2.2}$$

Tanh

Tanh is also another popular activation function. It squashes value between -1 and 1. It is monotonic. Its derivative is not monotonic. It is mainly used for classification. Its graph is shown in the figure below while mathematically it can be represented as

$$tanh(x) = \frac{sinh(x)}{cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{3.2.3}$$

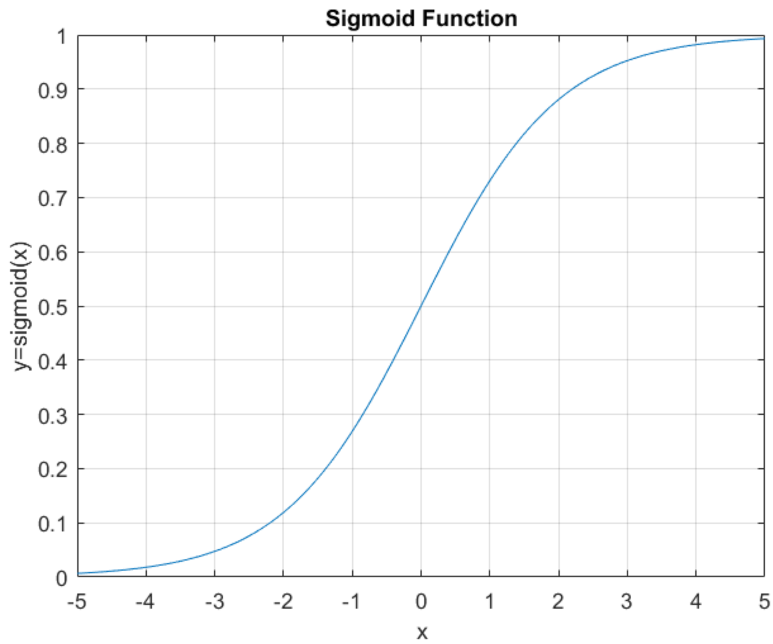

Figure 3.12: Sigmoid Activation Function.

ReLu Rectified Linear Unit Leaky Relu Randomized Leaky ReLu

In Rectified Linear Unit the negative values are squashed to zero while all positive values remain unchanged. So sometime negative values may not properly train the

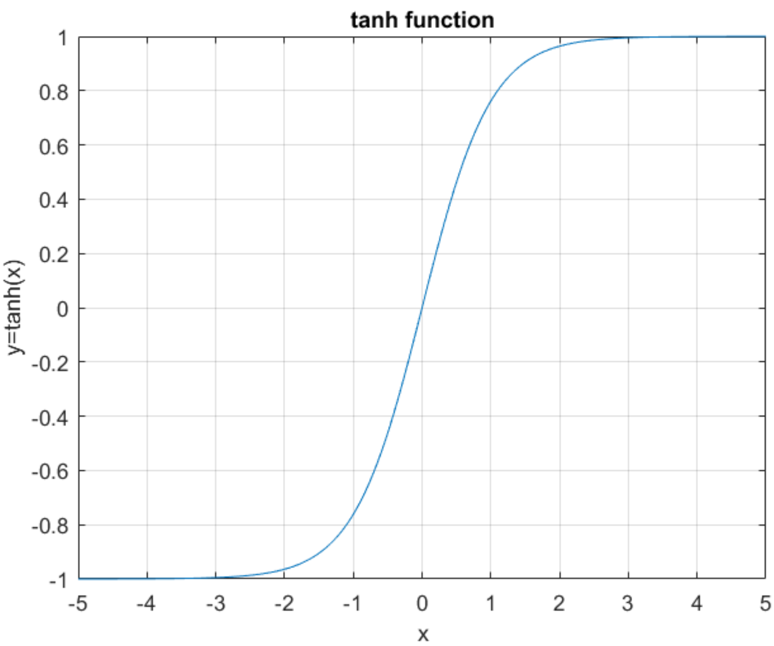

Figure 3.13: Tanh activation Function.

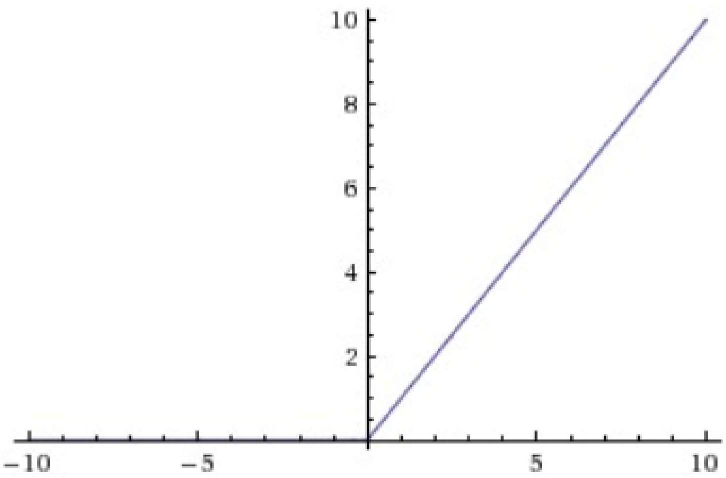

Figure 3.14: Rectified Linear Unit (ReLu) [44].

model. Since they have the same effect after passing through Relu. To solve the dying problem of Relu we use a modification in Leaky Relu which converts the input factor to be multiplied by 0.01 when the value of the input is not 1 then it is called randomized Leaky Relu.

Maxout

Maxout activation is performing some good learning of model over sigmoid and tanh as it take large steps. In maxout activation the output will be maximum of all the inputs. Mathematically maxout activation is given below

$$O_i(x) = max(w_1x + b1, w_2x + b2, ...) \tag{3.2.4}$$

3.3 Deep Learning

Deep Learning can be expressed a neural network with several layers of neurons between input, hidden and output. Nowadays the most precise and powerful statistical classifier for inferring many common user patterns without human intervention is built on algorithms from deep learning. In deep learning, we may use its variants like Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) etc.

Uses of Deep Learning

Recognition of faces, objects, behavior, words, speech, etc.

Image Analysis

- Detection (e.g. disease)
- Recognition (e.g. objects)
- Identification (e.g. persons)

Data Mining

- Classification
- Change and Deviation Detection
- Knowledge Discovery

Prognosis

- Ozone prognosis
- Weather Forecast
- Stock market prediction

3.3.1 Deep Neural Network

Our Next classifier or prediction model is a DNN. The prefix in the name "deep " is due to the fact in these networks we mostly use multiple and deeper layers. The layers and nodes are increased due to the availability of enormous computing as a common man have access to the cloud server. The most popular DNN which fascinates and make DNN famous in this age are CNN. They became popular when they surpass all classifiers in ImageNet Competition. Famous Layers in CNN are ConvNet, Fully Connected (FC), Pooling and Relu.

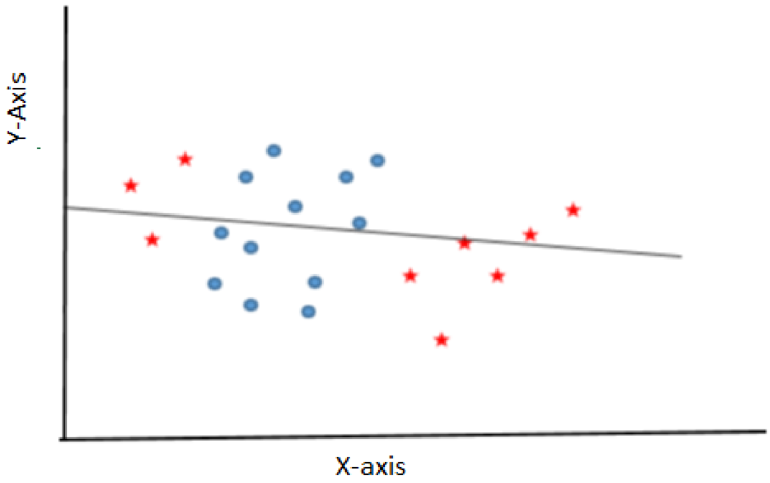

Figure 3.15: A single line is applied to non linear data.

In the below Diagrams the working of DNN is shown

Now let us discuss each layer of DNN in detail so that we understand their working.

The very first layer known as input layer is multiplied by some random weights by applying a kernel also known as a filter to our input tensor(matrix) and do convolution

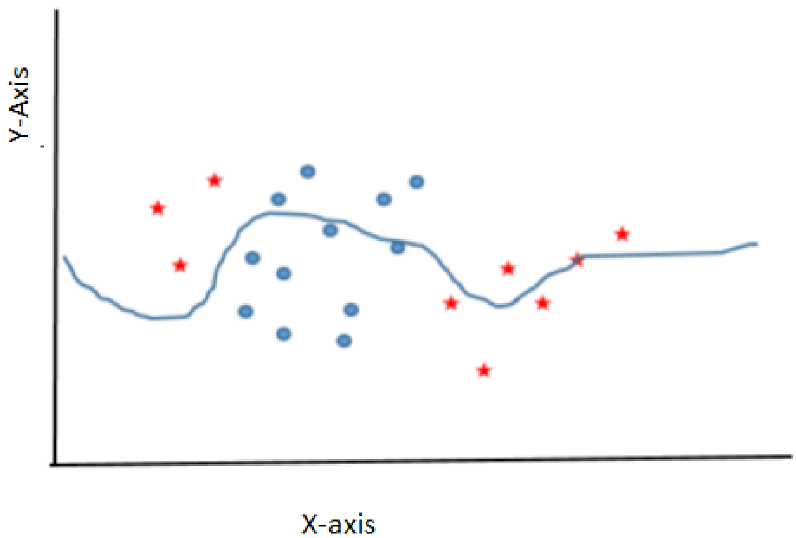

Figure 3.16: Phases of learning of DNN.

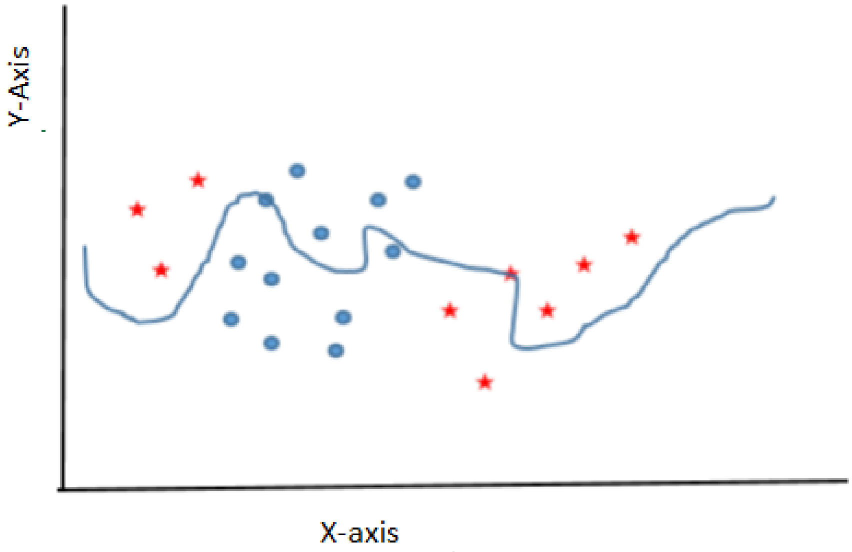

Figure 3.17: Phases of learning of DNN.

process. We may take a lot number of filters. Because each filter will extract different feature from the input. This is how in the DNN there is no need for human intervention

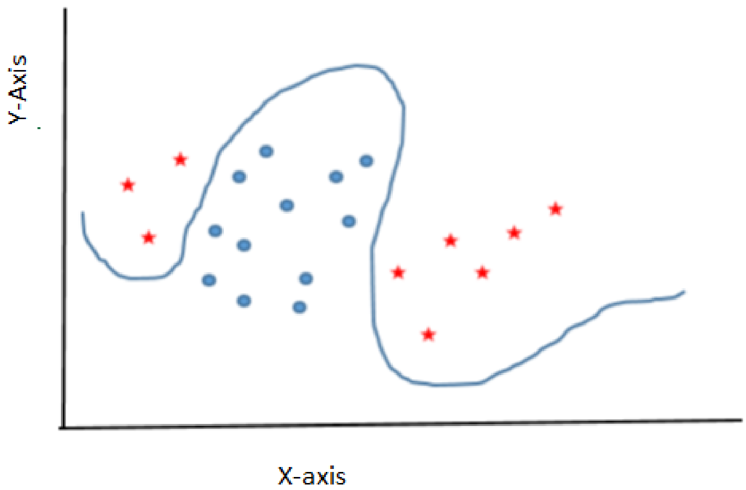

Figure 3.18: Finally trained DNN with non linear separable line.

for feature extraction. Sometime we apply zero padding to avoid border values missing when applying filters. If we are giving the size of W1 * H1* D1 as input tensor and suppose we are applying K number of filters with F size of filter and depth with D1 and stride S and padding P. Normally we apply Zero padding so that no noise input will take part. The Depth of the Filter will remain the same. The depth of the output will become the number of filters so the next output will be W2*H2*K with K number of biases added to it.

Now, this output will have processed by some activation function. In CNN we apply Rectified Linear Unit (Relu) which basically squashed the out to zero. It means that if some input is negative then the output will be zero otherwise unchanged. For image classification, these two layers are multiple time applied and in between them A pooling layer is also introduced. Pooling layer is used to downsize the input. Two Pooling approaches are the most common. Average pooling and Max pooling are two famous techniques. In Max pooling, we basically pick the maximum value of 2* 2 matrix.

In the final layer, we apply fully connected layer which maps our input to certain discrete or continuous values. The same Loss function is applied here and the same back-propagation is applied using SGD (Stochastic Gradient Descent) or some other optimization approach. Along with a derivative of the loss function, we also apply some regularization to avoid overfitting. For this study, we use Maxout as our activation function and softmax for our fully connected layer [45]. Mathematically

$$O_i(x) = max(w_1x + b, w_2x + b, w_3x + b) \tag{3.3.1}$$

The softmax function for mapping the values in final layer to get the predicted values can be expressed as

$$f_O = \frac{e^O}{\sum e^O} \tag{3.3.2}$$

where

$$O = wx + b \tag{3.3.3}$$

3.4 Ensemble Learning

our two new classifiers are basically based ensemble learning. In ensemble learning, we combine and compile prediction on the basis of different models or classifiers or learners. Two most popular ensemble learners are applied for this study to assess their performance in comparison with the deep learning model. While in machine learning predicting some statistically-based approach, some factors affect our results. These factors may be noise or bias or maybe variance. Ensemble models are used to overcome these factors. The only noise is not covered by ensemble models as they are created by some bad data. A very famous example of ensemble learning is the story of elephant seeing by blind men. In that story, each blind man tells only that part which is only guessed by him. So no one was able to predict that elephant properly. But if we combine all of their guesses about elephants than we can easily understand an elephant. This analogy works in ensemble learning. We take many weak learners and predict using them. After their prediction, we combine them.

Before understanding about bagging and boosting. Lets discuss another term known as bootstrapping.

Bootstrapping

Bootstrapping is an approach in which we divide our data set in a random number of chunks or sample with replacing every sample. Since these chunks are sample may have different data points from data sets so randomly replacing them produce varying variance and bias. Since probability is equal for selection of the data samples. Therefore, we may have better mean and standard deviation. Now let us see step by step how can we calculate mean

1. Take random samples from data set using replacement.
2. Calculate the mean of each sample.
3. Now take the overall mean by averaging the means.

3.4.1 Bagging

Bagging is made by combining two terms, first is bootstrap and second is aggregation. As in above we have discussed bootstrapping. Now in bagging, we basically train our classifiers on each of bootstrap. After that, we combine their prediction depending on the task. For the classification problem, we use the technique of voting in which we count the maximum number of votes. While in regression problems as in our case we use averaging for regression. In algorithmic way we may define bagging as

Algorithm 1 Bagging Algorithm

Training Phase:

1. *Initialize all variables(parameters)*
2. *Take C as ensemble Learner*
3. *N, number of classifiers to be trained*
4. *M number of attributes of each data set* :

0: **for** $i = 1$ to N **do**

Take a bootstrap sample $S_i fromDataset$

Generate a classifier $C_i astrainingset$

Add this classifier to ensemble C

0: **end for**

return C

Regression Phase:

1. *Take the predicted value of each classifier*
2. *Average the value* =0

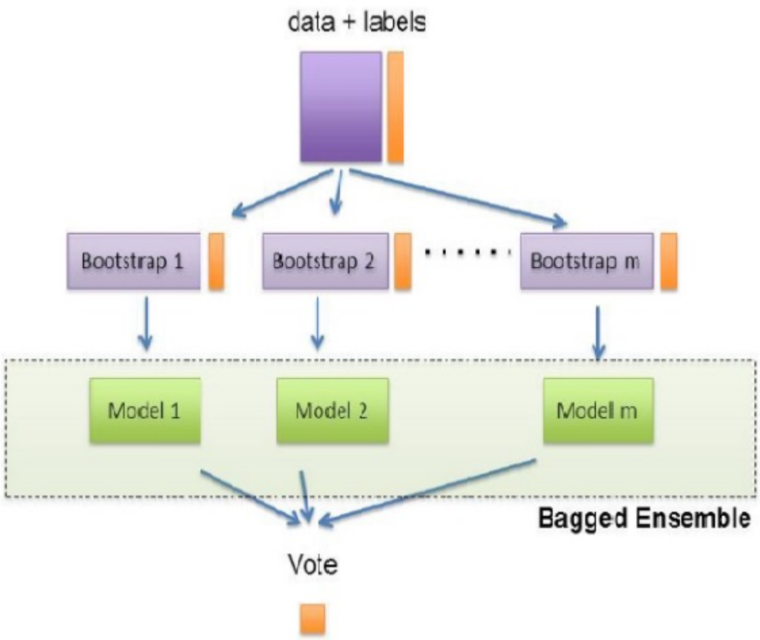

Figure 3.19: How Bagging splits in bootstrap and combine for voting [46].

3.4.2 Boosting

Boosting is also an approach in which we combine multiple base classifiers in a way that boost or make better accuracy which is collectively better than all of those. The main difference between bagging and boosting is that in bagging we train a classifier in parallel then we combine while in boosting the classifiers are trained sequentially.

In boosting we use sample space to predict certain value and then we redistribute the weights for improvement. Correct strategy got more weight while wrong prediction strategy is reduced further. In algorithmic way we may define boosting as

---

Algorithm 2 Boosting

Boosting

1. Take a random subset of training Data Sets as t1 and avoid replacing from the sample C to train a weak learner C1

2. Pick another random sample Data Set t2 still without replacing from the training set but adding half of the data points that was wrongly predicted in previous iteration to train a learner C2

3. Find the training examples t3 in the training set C on which C1 and C2 both were not agree to train a third weak learner C3

4.Sum up all the weak learners via majority voting or average the value for regression

: =0

3.4.3 Random Forest

A random forest also falls in the category of ensemble learning for prediction. It combines multiple classifiers result. Normally in the random forest we use trees and combine their result to get the final result. Interestingly Random forest is also a bagging approach. Then the question arises why we are specifying it as an extra case. The answer is in Random Forest regression we are using decision trees as models while in the above-specified bagging approach we are using different neural networks and taking their average to predict the electricity consumption. For further understanding of decision trees, we have to understand how decision tree works.

Decision Tree

A decision tree is a graph which is used to make a decision upon certain criteria on a node inside a graph and a final decision is made upon reaching a leaf node. Node presents the test while branch present outcome of the test. Each leaf node presents a label for a classification problem. Decision trees are also called CART (Classification and Regression Tree). For the decision, we use entropy or information gain. Some common terms involved in the decision tree are

1. Leaf Node
2. Splitting
3. Decision Node
4. Pruning
5. Sub Tree
6. Parent and Child Node
7. Root Node

Chapter 4

Results and Performance Analysis

The data for this study is acquired from an educational Institute known as South Bank Techno Park London. The image of the building from which the data is captured is shown in figure 4.1. The data was measured on daily basis for a span of 5 years. The data spanning over year 2007 to 2010 is fed for training while the data of year 2011 is given for testing the performance of each classifier.

1. Daily Consumption of Electricity (Wh/m^2)
2. Daily Temperature Mean taken in Kelvin
3. Daily global radiance Mean (W/m^2)

Figure 4.1: South Bank Techno Park London [9].

Table 4.1: Description of the Building

Description	**Building**
Building Type	Offices
Total Covered Area	7811
Construction Year	2003
Building Location	South West
Office Hours	0800 - 1800
office Close Time	2200
Cooling Mechanism	Natural ventilation and few Split Air Conditioner
Heating Method	2 boiler
Lighting type	CFL
Total Lifs	2
Total LV supplies	2
Total floor	3

4. Daily Humidity Mean (%)

5. Daily Wind Velocity Mean (m/s)

6. Week Day Indicator (1 for working 0 for non-working).

Before discussing further lets take a look on building detail.

For data analysis phase all related data to electricity consumption e.g. weather data, Electricity Consumption data, building occupied data, wind speed data is selected. All the descriptive information such as building type, Total Covered Area, Construction Year, Building location, Operation Hours, Heating System, etc. are mentioned in 4.1.

It is apparent from the chart that the requirement of heating in winter is raised consequently the consumption is also raised in winter while in summer since heating is not as much required therefore consumption of electricity is mitigated.
For our study, we gather the real values of daily mean actual measured values of temperature, radiance, wind speed and humidity. The wind speed and humidity is taken for London from the Kings College Environment Research Group.It was accessed from their website [47]. Weekday indicator is also introduced as during weekday the consumption raises as compared to non-week days [48].

4.1 Performance Evaluation

All six classifiers have been trained on the same data and tested on the same data i.e. the year 2007 to year 2010 was used for training that model and the year 2011 was used to test the model for prediction. A total number of Six Classifiers are used to predict electricity consumption.

The assessment on six different predictive models have been analyzed using Mean Absolute Error, Mean Relative Error, Mean Absolute Percentage Error, Root mean Square Error, and Normalized Root Mean Square Error for all six performance metrics. The formulas for these metrics are given in Equations.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_{real,i} - Y_{Predict,i})^2}{n}} \tag{4.1.1}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_{real} - Y_{predict,i}}{Yreal}\right| \tag{4.1.2}$$

$$MRE = \left|\frac{Y_{real} - Y_{predict}}{Yreal}\right| \tag{4.1.3}$$

$$MAPE = \frac{100}{n}\sum_{i}^{100}\left|\frac{Y_{real,i} - Y_{predict,i}}{Yreal}\right| \tag{4.1.4}$$

$$NRMSE = \frac{RMSE}{Y_{predict,max} - Y_{predict,min}} \tag{4.1.5}$$

4.1.1 Artificial Neural Network

Our first classifier which outperforms all other classifier is placed first and the statistical data related to ANN is presented in the below table. It is quite interesting that if last month is dropped from testing then we get a drastic change of 2 % in error. The Respective MRE, MAPE, and NRMSE values are given by 5.95%, 6.41% and 9.68%. Getting prediction of 11 months may affect the results that Respective MRE, MAPE and NRMSE values are given by 4.32%, 3.89% and 6.28% The Result shows that ANN outperforms in all of the prediction models.

Table 4.2: Predicted data of (a) 12 months (b) 11 months using ANN

(a)

	Real values	Forecast values
		ANN
Min	426	408
Max	155	194
Mean	311	320
Median	347	356
N	271	214

(b)

	Real values	Forecast values
		ANN
Min	426	408
Max	176	194
Mean	317	321
Median	351	357
N	250	214

The graph in figure 4.2 and in figure 4.3 represents the real and forecasted electricity consumption units for whole year & 11 months respectively using ANN.

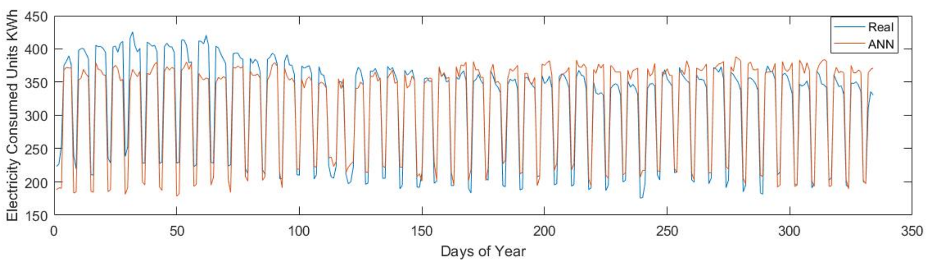

Figure 4.2: Prediction graph of 11 months utilization using ANN.

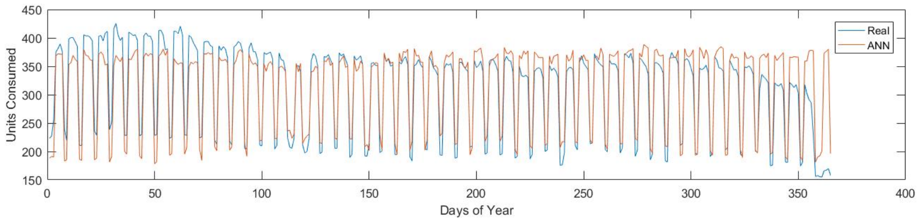

Figure 4.3: Prediction graph of 12 months utilization using ANN.

4.1.2 Support Vector Regression

The traditional Classifier before deep learning or neural networks which performs very good in different domains is Support Vector Regression. All the statistical data related to Support Vector Regression is presented in the below table. It is also relevant to talk that if last month is dropped from testing then we get a drastic change of about 2 % in error also in SVR. The Respective MRE, MAPE, and NRMSE values are given by 8.41%, 9.02% and 13.07%. Getting prediction of 11 months may affect the results that The Respective MRE, MAPE, and NRMSE values are given by 6.67%, 6.31% and11.15%

The Result shows that ANN outperforms in all of the prediction models.

Table 4.4: Predicted data of (a) 12 months (b) 11 months using SVR.

(a)

	Real values	Forecast values
		SVR
Min	426	398
Max	155	160
Mean	311	322
Median	347	359
N	271	238

(b)

	Real values	Forecast values
		SVR
Min	426	398
Max	176	160
Mean	317	323
Median	351	359
N	250	238

The graph in figure 4.4 and in figure 4.5 represents the actual and predicted electricity consumption units for whole year and 11 months respectively using SVR.

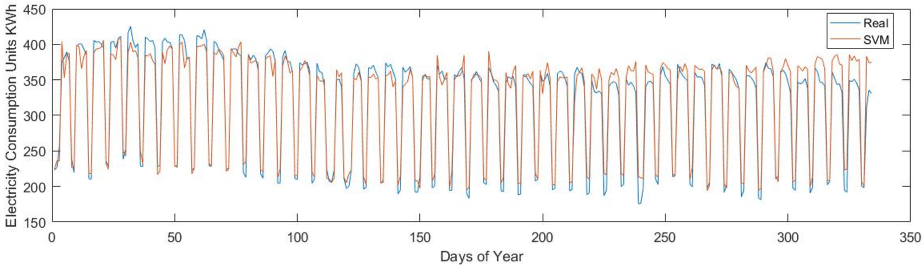

Figure 4.4: Prediction graph of 11 months utilization using SVR.

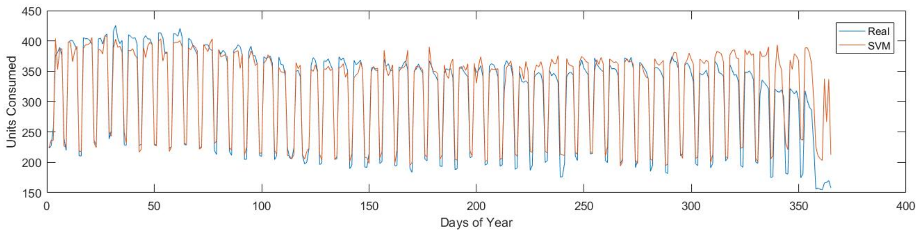

Figure 4.5: Prediction graph of 12 months utilization using SVR.

4.1.3 Deep Neural Network

Although DNN is supposed to perform excellently due to data deficiency for training, it seems it produces some bad results. However if in future we may have larger data for training. All the statistical data related to DNN is presented in the below table. Still, it is worth considering that in the case of DNN, last month of the testing year proves to be an outlier. Dropping the last month we get a still change of about 2 % in error also in DNN. The Respective MRE, MAPE, and NRMSE values are given by9.56%, 11.15%

Table 4.6: Predicted data of (a) 12 months (b) 11 months using DNN.

(a)

	Real values	Forecast values
		DNN
Min	426	402
Max	155	233
Mean	311	331
Median	347	356
N	271	169

(b)

	Real values	Forecast values
		DNN
Min	426	402
Max	176	233
Mean	317	331
Median	351	356
N	250	169

and 13.02%. Getting prediction of 11 months may affect the results that The Respective MRE, MAPE, and NRMSE values are given by 7.61%, 8.08% and 11.03%.

The graph in figure 4.6 and in figure 4.8 represents the actual and predicted electricity consumption units for whole year and 11 months respectively using DNN.

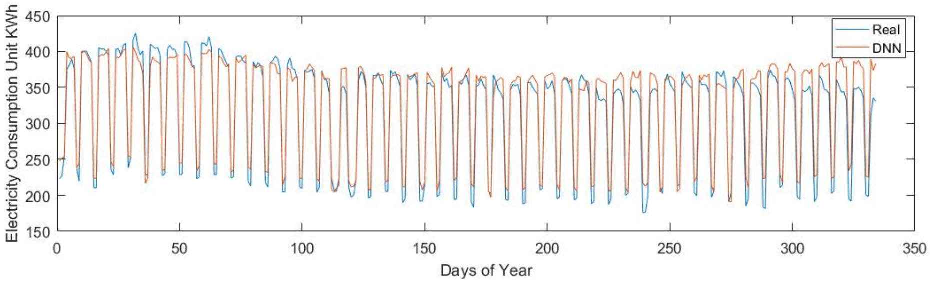

Figure 4.6: Prediction graph of 11 months utilization using DNN.

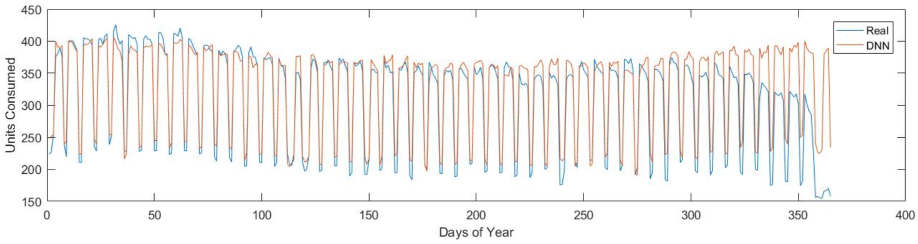

Figure 4.7: Prediction graph of 12 months utilization using DNN.

4.1.4 Bagging

Bagging is an ensemble technique used for prediction by the averaging result of different weak classifiers while predicting the value in a parallel fashion. It is worth notable that Bagging performs as a second-best classifier for this data set after ANN. All the statistical data related to Bagging prediction is presented in the below table. We have presented both 11 month and 12-month graphs. Dropping the last month we get a still change of about 2 % in error also in Bagging. The Respective MRE, MAPE, and NRMSE values are given by 6.95%, 7.46% and 10.74% . Getting prediction of 11 months may affect the results that The Respective MRE, MAPE, and NRMSE values are given by 5.12%, 4.66% and 7.33%

The graphs shown in figures 4.8 and in 4.9 represents the real and forecasted electricity consumption units for whole year and 11 months respectively using Bagging.

Table 4.8: Predicted data of (a) 12 months (b) 11 months using Bagging

(a)

	Real values	Forecast values
		Bagging
Min	426	393
Max	155	198
Mean	311	321
Median	347	359
N	271	195

(b)

	Real values	Forecast values
		Bagging
Min	426	393
Max	176	198
Mean	317	322
Median	351	359
N	250	195

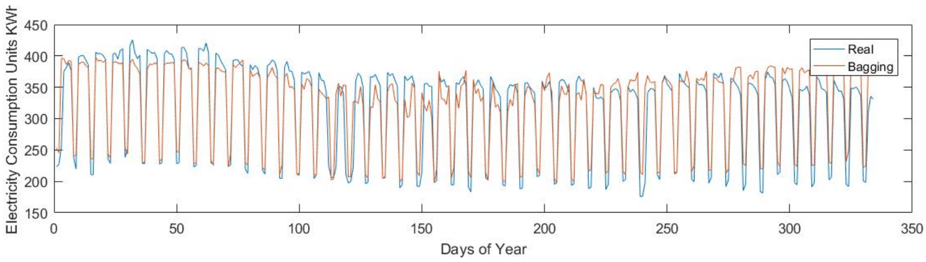

Figure 4.8: Prediction graph of 11 months utilization using Bagging.

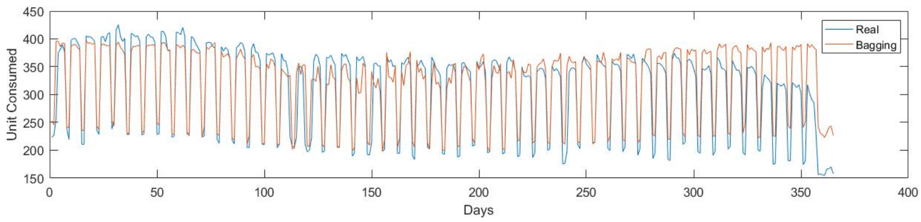

Figure 4.9: Prediction graph of 12 months utilization using Bagging.

4.1.5 Boosting

Boosting is also an ensemble model used for prediction by averaging the result of different weak classifiers while predicting the value in a sequential fashion. It is worth notable that Boosting performs as a third-best classifier for this data set after ANN and Bagging. All the statistical data related to Boosting prediction is presented in the below table. we have presented both 11 month and 12-month graphs. Dropping the last month we get a still change of about 2 % in error has happened in all other. The Respective MRE, MAPE, and NRMSE values are given by7.49%, 8.46% and 12.74%. Getting prediction of 11 months may affect the results that The Respective MRE, MAPE, and NRMSE

Table 4.10: Predicted data of (a) 12 months (b) 11 months using Boosting.

(a)

	Real values	Forecast values
		Boosting
Min	426	406
Max	155	191
Mean	311	327
Median	347	364
N	271	216

(b)

	Real values	Forecast values
		Boosting
Min	426	406
Max	176	191
Mean	317	326
Median	351	364
N	250	216

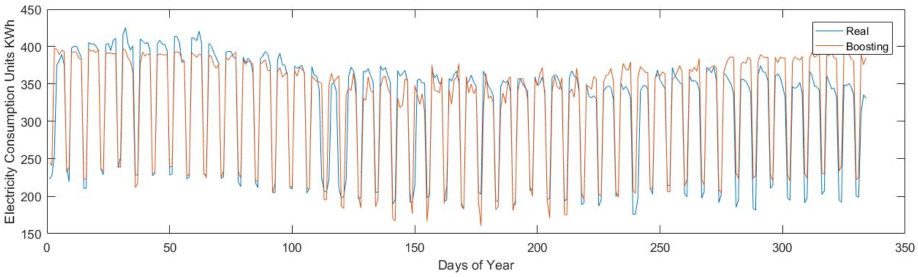

Figure 4.10: Prediction graph of 11 months utilization using Boosting.

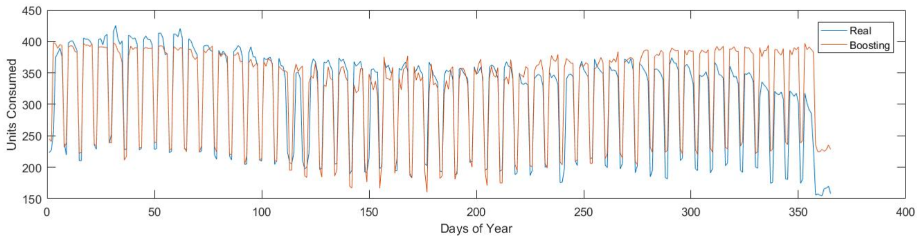

Figure 4.11: Prediction graph of 12 months utilization using Boosting.

values are given by 5.10%, 4.84% and 7.17%. The graph depicted in figures 4.10 and in 4.11 represents the actual and predicted electricity consumption units for whole year and 11 months respectively using Boosting.

4.1.6 Random Forest

Random Forest is our last model which grows the different level of decision trees. Results showed that it does not perform well for prediction. All the statistical data related to Random Forest prediction is presented in the below table. we have presented both 11

Table 4.12: Predicted data of (a) 12 months (b) 11 months using Random Forest

(a)

	Real values	Forecast values
		Random Forest
Min	426	420
Max	155	206
Mean	311	332
Median	347	361
N	271	214

(b)

	Real values	Forecast values
		Random Forest
Min	426	420
Max	176	206
Mean	317	332
Median	351	361
N	250	214

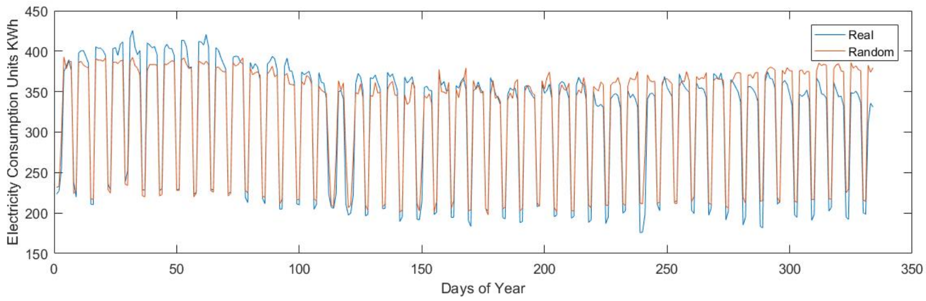

Figure 4.12: Prediction graph of 11 months utilization using Random Forest.

month and 12-month graphs. Dropping the last month we get a still change of about 2 % in error has happened in all other. The Respective MRE, MAPE, and NRMSE values are given by8.89%, 9.84% and 14.17% Getting prediction of 11 months may affect the results that The Respective MRE, MAPE, and NRMSE values are given by6.84%, 6.80% and 11.69%. The graph which is shown in figures 4.12 and in 4.13 represents the actual and predicted electricity consumption units for whole year and 11 months respectively using Random Forest.

Below are both of the graphs.

4.1.7 Error Analysis of all the models

All the results are shown in below table using all the above-specified metrics. It is obvious that ANN is still proving to be the best of all contemporary classifiers in terms of mean absolute error. It hits the NRMSE of 9.68% which is best over all parallel models. The achievement of SVR, Random Forest, and DNN is the same Where the efficiency of SVR and DNN remains close enough in terms of NRMSE. Similarly, if we consider the MAPE metric, it is quite vivid that ANN performs better in comparison with all other approaches. MRE also advocates the efficacy of ANN over all other classifiers. However, Bagging reached very near to ANN.

Table 4.14: Performance metrics of different Models for 1 year.

Metric Type	ANN	DNN	SVR	Bag	Boost	Random Forest
RMSE (W/m^2)	26	35	35	29	34	38
MAE (W/m^2)	17	27	24	20	21	25
MRE%	5.95%	9.56%	8.41%	6.95%	7.49%	8.89%
MAPE%	6.41%	11.15%	9.02%	7.46%	8.46%	9.84%
NRMSE%	9.68%	13.02%	13.07%	10.74%	12.74%	14.17%

Now as presented in previous results of each classifier. It is astonishing that if we consider 11 month's data then we achieve very good performance that reveals that the last month of our test data is an outlier. Below is the result for 11 month's prediction.

Table 4.15: Performance metrics of different Models for 11 months.

Metric Type	ANN	DNN	SVR	Bag	Boost	Random
RMSE (W/m^2)	16	28	28	18	18	29
MAE (W/m^2)	13	22	20	15	15	20
MRE%	4.32%	7.61%	6.67%	5.12%	5.10%	6.84%
MAPE%	3.89%	8.08%	6.31%	4.66%	4.84%	6.80%
NRMSE%	6.28%	11.03%	11.15%	7.33%	7.17%	11.69%

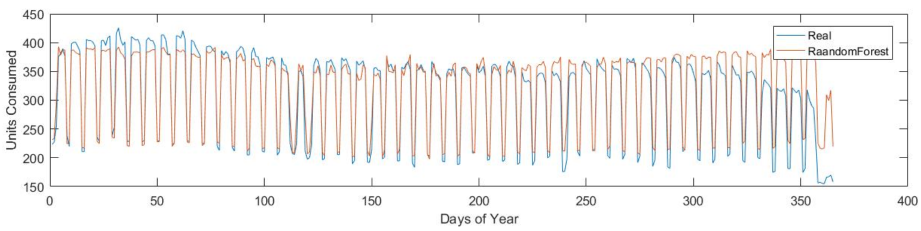

Figure 4.13: Prediction graph of 12 months utilization using Random Forest.

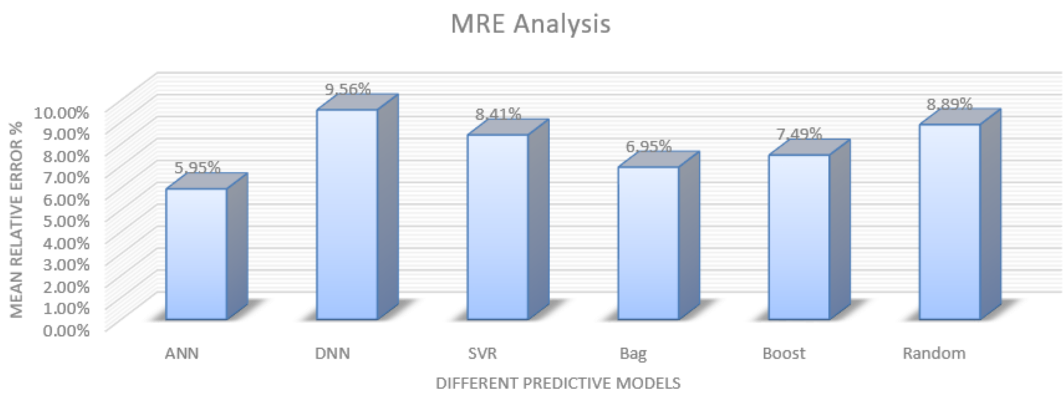

Figure 4.14: Error Analysis bar Chart for different models.

Below is the boxplot of all the predictive models for the whole year in figure 4.15.

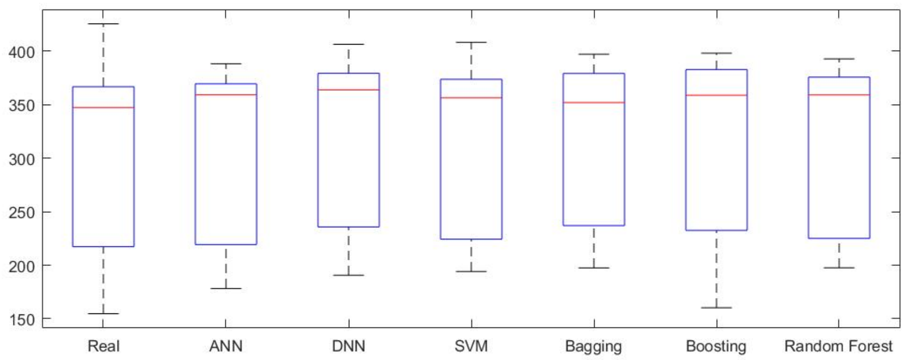

Figure 4.15: Boxplot of different Model for complete year.

4.1.8 Month wise electricity consumption

Month wise electricity consumption is shown in below charts. In figure 4.16 electricity consumption is shown for entire year. All predicted and actual. While in figure 4.17 electricity consumption for only working week days is presented. Similarly in figure 4.18 electricity consumption for only non working week days is presented.

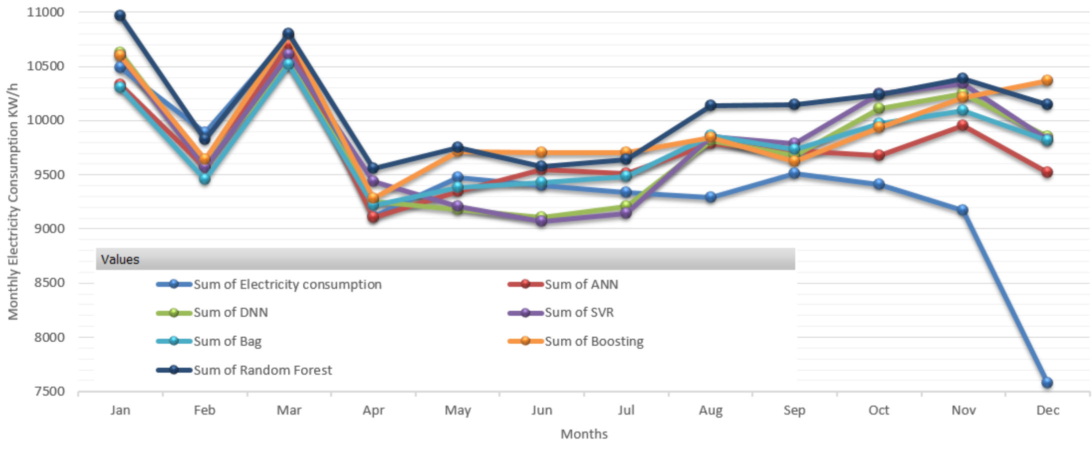

Figure 4.16: Month wise electricity consumption line comparison chart.

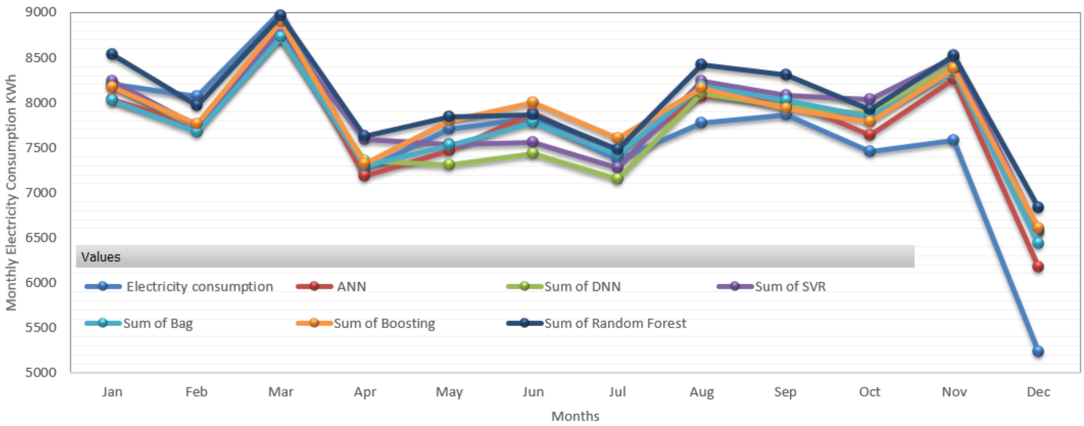

Figure 4.17: Month wise electricity consumption line comparison chart for working days.

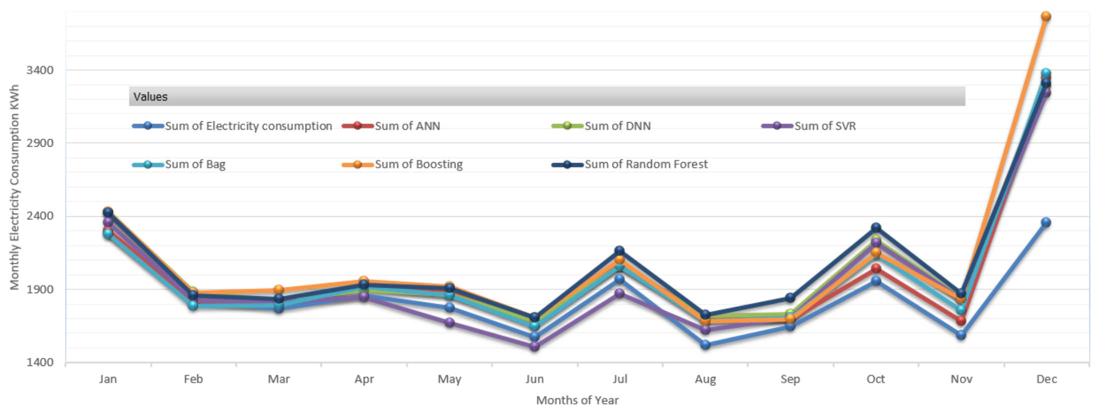

Figure 4.18: Month wise electricity consumption line comparison chart for Non working days.

Chapter 5

Conclusion & Future Work

5.1 Conclusion

Real-time data of an educational building for the energy utilization of five years is fed to different Modern machine learning data-driven models and analyzed for different Energy consumption prediction approaches including Bagging, Boosting, ANN, DNN, SVR, and Random Forest. The predicted utilization units for the working weekdays, nonworking days and for all the months were analyzed with the actual utilization of energy units. It was assessed that under consideration predictive models forecasted electricity consumption of office days of all months. The range with in which all falls is about 4.0%. It seems that december in testing year was an outlier as dropping it improves performance 2%. In terms of monthly consumption, the first three months, except the DNN model, all other five models predicted electricity consumption quiet well. Inclusively we can say that all of the classifiers predicted one year of electricity consumption of building with an mean relative error of less than 10.0%. Summing it all together, we reach a conclusion that ANN, when applied to given building data, has achieved a quiet high level of accuracy as compared to all other five models. If we analyze the boxplot we can observe that data distribution of ANN and Bagging for 2nd and 3rd quartile and median are quite close to the real values of these two methods. Both methods seem to achieve higher performance than others in overall achievements. In terms of MRE, ANN outperforms all other methods with 5.95% error. Second to that is Bagging. Where Boosting, SVR, Random Forest DNN falls after respectively. The accuracy in prediction makes ANN better than all other classifiers despite performance metrics. ANN proves to do well with the same complexity as compared to other exhaustive models such as DNN and SVR. Researches have shown that DNN performs well when it has a lot of data because it can automatically extract feature as well as there is a need for proper variable and activation function selection. But data was limited in terms of DNN. Hence performance of DNN seems not as good as it considered to be since availability of data was limited. This study can further extended to analyze the performances of different predictive models for different type of buildings resided in Pakistan.

5.2 Future Work

Since there is a shortage of energy especially in Pakistan. Consequently, it demands more care and properly managed the use of electricity in its Residential Buildings including administration buildings as well as an educational and residential building. Since this study is done using the dataset of London where temperature varies with Pakistan Temperature a lot. Therefore, we need more HVAC services in Pakistan. So it is need of the time to collect data of different buildings on a daily basis. So that we may perform different machine learning tasks subject to the availability of data.

It is worthy to note that the President of Pakistan has recently launched an initiative for the application of AI on Data. So this kind of work will only be possible if we have real-time data of our buildings.

References

[1] Number of network connected devices per person around the world from 2003 to 2020. URL https://www.statista.com/statistics/678739/forecast-on-connected-devices-per-person/.

[2] Christopher Reuß. Production of electric cars in the smart factory. *ATZproduction worldwide*, 6(2):58–58, May 2019. ISSN 2524-8812. doi: 10.1007/s38312-019-0018-7. URL https://doi.org/10.1007/s38312-019-0018-7.

[3] Mike Seddon. Cars and clouds: New drivers for efficient power management. *Additional Conferences (Device Packaging, HiTEC, HiTEN, & CICMT)*, 2019(DPC):000671–000705, 2019. doi: 10.4071/2380-4491-2019-DPC-GBC1KEYNOTE_ONSemiconductor. URL https://doi.org/10.4071/2380-4491-2019-DPC-GBC1KEYNOTE_ONSemiconductor.

[4] Payam Nejat, Fatemeh Jomehzadeh, Mohammad Mahdi Taheri, Mohammad Gohari, and Muhd Zaimi Abd. Majid. A global review of energy consumption, co2 emissions and policy in the residential sector (with an overview of the top ten co2 emitting countries). *Renewable and Sustainable Energy Reviews*, 43:843 – 862, 2015. ISSN 1364-0321. doi: https://doi.org/10.1016/j.rser.2014.11.066. URL http://www.sciencedirect.com/science/article/pii/S1364032114010053.

[5] A. GonzÃąlez-Vidal, A. P. Ramallo-GonzÃąlez, F. Terroso-SÃąenz, and A. Skarmeta. Data driven modeling for energy consumption prediction in smart buildings. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 4562–4569, Dec 2017. doi: 10.1109/BigData.2017.8258499.

[6] Giovanni Tardioli, Ruth Kerrigan, Mike Oates, James OâĂŸDonnell, and Donal Finn. Data driven approaches for prediction of building energy consumption at

urban level. *Energy Procedia*, 78:3378 – 3383, 2015. ISSN 1876-6102. doi: https://doi.org/10.1016/j.egypro.2015.11.754. URL http://www.sciencedirect.com/science/article/pii/S1876610215024868. 6th International Building Physics Conference, IBPC 2015.

[7] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436 EP –, May 2015. URL https://doi.org/10.1038/nature14539.

[8] Xiaodong Cao, Dai Xilei, and Junjie Liu. Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade. *Energy and Buildings*, 128, 06 2016. doi: 10.1016/j.enbuild.2016.06.089.

[9] KP Amber, MW Aslam, and SK Hussain. Electricity consumption forecasting models for administration buildings of the uk higher education sector. *Energy and Buildings*, 90:127–136, 2015.

[10] Hafiz Bilal Khalil and Syed Jawad Hussain Zaidi. Energy crisis and potential of solar energy in pakistan. *Renewable and Sustainable Energy Reviews*, 31:194 – 201, 2014. ISSN 1364-0321. doi: https://doi.org/10.1016/j.rser.2013.11.023. URL http://www.sciencedirect.com/science/article/pii/S1364032113007727.

[11] Ahmad Bilal Awan and Zeeshan Ali Khan. Recent progress in renewable energy âĂŞ remedy of energy crisis in pakistan. *Renewable and Sustainable Energy Reviews*, 33:236 – 253, 2014. ISSN 1364-0321. doi: https://doi.org/10.1016/j.rser.2014.01.089. URL http://www.sciencedirect.com/science/article/pii/S1364032114001154.

[12] Mehreen S Gul and Sandhya Patidar. Understanding the energy consumption and occupancy of a multi-purpose academic building. *Energy and Buildings*, 87:155–165, 2015.

[13] Umberto Berardi. A cross-country comparison of the building energy consumptions and their trends. *Resources, Conservation and Recycling*, 123:230–241, 2017.

[14] Carmen de la Cruz-Lovera, Alberto-Jesús Perea-Moreno, José-Luis de la Cruz-Fernández, José Alvarez-Bermejo, and Francisco Manzano-Agugliaro. Worldwide

research on energy efficiency and sustainability in public buildings. *Sustainability*, 9(8):1294, 2017.

[15] Chengdong Li, Zixiang Ding, Dongbin Zhao, Jianqiang Yi, and Guiqing Zhang. Building energy consumption prediction: An extreme deep learning approach. *Energies*, 10(10):1525, 2017.

[16] Bo Zeng, Meng Zhou, and Jun Zhang. Forecasting the energy consumption of chinaâĂŹs manufacturing using a homologous grey prediction model. *Sustainability*, 9(11):1975, 2017.

[17] AS Ahmad, MY Hassan, MP Abdullah, HA Rahman, F Hussin, H Abdullah, and R Saidur. A review on applications of ann and svm for building electrical energy consumption forecasting. *Renewable and Sustainable Energy Reviews*, 33:102–109, 2014.

[18] Xiaoshu Lü, Tao Lu, Charles J Kibert, and Martti Viljanen. Modeling and forecasting energy consumption for heterogeneous buildings using a physical–statistical approach. *Applied Energy*, 144:261–275, 2015.

[19] Aurore Lomet, Frédéric Suard, and David Chèze. Statistical modeling for real domestic hot water consumption forecasting. *Energy Procedia*, 70:379–387, 2015.

[20] Hengfang Deng, David Fannon, and Matthew J Eckelman. Predictive modeling for us commercial building energy use: A comparison of existing statistical and machine learning algorithms using cbecs microdata. *Energy and Buildings*, 163: 34–43, 2018.

[21] Zhanyu Ma, Hailong Li, Qie Sun, Chao Wang, Aibin Yan, and Fredrik Starfelt. Statistical analysis of energy consumption patterns on the heat demand of buildings in district heating systems. *Energy and Buildings*, 85:464–472, 2014.

[22] Nelson Fumo and MA Rafe Biswas. Regression analysis for prediction of residential energy consumption. *Renewable and sustainable energy reviews*, 47:332–343, 2015.

[23] Xiaoyu Wang, Dongkun Luo, Xu Zhao, and Zhu Sun. Estimates of energy consumption in china using a self-adaptive multi-verse optimizer-based support vector machine with rolling cross-validation. *Energy*, 152:539–548, 2018.

[24] Yu-Rong Zeng, Yi Zeng, Beomjin Choi, and Lin Wang. Multifactor-influenced energy consumption forecasting using enhanced back-propagation neural network. *Energy*, 127:381–396, 2017.

[25] S Hr Aghay Kaboli, A Fallahpour, J Selvaraj, and NA Rahim. Long-term electrical energy consumption formulating and forecasting via optimized gene expression programming. *Energy*, 126:144–164, 2017.

[26] Miha KovaÄDiÄD and BoÅ¿idar Åăarler. Genetic programming prediction of the natural gas consumption in a steel plant. *Energy*, 66:273 – 284, 2014. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2014.02.001. URL http://www.sciencedirect.com/science/article/pii/S0360544214001340.

[27] K.P. Amber, R. Ahmad, M.W. Aslam, A. Kousar, M. Usman, and M.S. Khan. Intelligent techniques for forecasting electricity consumption of buildings. *Energy*, 157:886 – 893, 2018. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2018.05.155. URL http://www.sciencedirect.com/science/article/pii/S036054421830999X.

[28] Felipe L.C. Silva, Reinaldo C. Souza, Fernando L. Cyrino Oliveira, Plutarcho M. Lourenco, and Rodrigo F. Calili. A bottom-up methodology for long term electricity consumption forecasting of an industrial sector - application to pulp and paper sector in brazil, 2018. ISSN 0360-5442. URL http://www.sciencedirect.com/science/article/pii/S0360544217321217.

[29] M.A. Rafe Biswas, Melvin D. Robinson, and Nelson Fumo. Prediction of residential building energy consumption: A neural network approach. *Energy*, 117:84 – 92, 2016. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2016.10.066. URL http://www.sciencedirect.com/science/article/pii/S0360544216315006.

[30] Jie Wang and Jun Wang. Forecasting energy market indices with recurrent neural networks: Case study of crude oil price fluctuations. *Energy*, 102:365 – 374, 2016. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2016.02.098. URL http://www.sciencedirect.com/science/article/pii/S0360544216301608.

[31] Chunlei Zeng, Changchun Wu, Lili Zuo, Bin Zhang, and Xingqiao Hu. Predicting energy consumption of multiproduct pipeline using artificial neural networks. *Energy*, 66:791 – 798, 2014. ISSN 0360-5442. doi: https://doi.org/10.1016/

j.energy.2014.01.062. URL http://www.sciencedirect.com/science/article/pii/S036054421400084X.

[32] Jolanta Szoplik. Forecasting of natural gas consumption with artificial neural networks. *Energy*, 85:208 – 220, 2015. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2015.03.084. URL http://www.sciencedirect.com/science/article/pii/S036054421500393X.

[33] YouLong Yang, JinXing Che, YanYing Li, YanJun Zhao, and SuLing Zhu. An incremental electric load forecasting model based on support vector regression. *Energy*, 113:796 – 808, 2016. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2016.07.092. URL http://www.sciencedirect.com/science/article/pii/S0360544216310167.

[34] Mayur Barman, N.B. Dev Choudhury, and Suman Sutradhar. A regional hybrid goa-svm model based on similar day approach for short-term load forecasting in assam, india. *Energy*, 145:710 – 720, 2018. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2017.12.156. URL http://www.sciencedirect.com/science/article/pii/S0360544217321990.

[35] Guohua Cao and Lijuan Wu. Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting. *Energy*, 115:734 – 745, 2016. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2016.09.065. URL http://www.sciencedirect.com/science/article/pii/S0360544216313068.

[36] Shuai Wang, Lean Yu, Ling Tang, and Shouyang Wang. A novel seasonal decomposition based least squares support vector regression ensemble learning approach for hydropower consumption forecasting in china. *Energy*, 36(11):6542 – 6554, 2011. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2011.09.010. URL http://www.sciencedirect.com/science/article/pii/S0360544211006074.

[37] Nima Izadyar, Hossein Ghadamian, Hwai Chyuan Ong, Zeinab moghadam, Chong Wen Tong, and Shahaboddin Shamshirband. Appraisal of the support vector machine to forecast residential heating demand for the district heating system based on the monthly overall natural gas consumption. *Energy*, 93:1558 – 1567, 2015. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2015.10.015. URL http://www.sciencedirect.com/science/article/pii/S0360544215013791.

[38] Khuram Pervez Amber, Muhammad Waqar Aslam, Anzar Mahmood, Anila Kousar, Muhammad Yamin Younis, Bilal Akbar, Ghulam Qadar Chaudhary, and Syed Kashif Hussain. Energy consumption forecasting for university sector buildings. *Energies*, 10(10), 2017. ISSN 1996-1073. doi: 10.3390/en10101579. URL http://www.mdpi.com/1996-1073/10/10/1579.

[39] Alexey Ya. Chervonenkis. *Early History of Support Vector Machines*, pages 13–20. 10 2013. ISBN 978-3-642-41135-9. doi: 10.1007/978-3-642-41136-6_3.

[40] Chirag Deb, Lee Siew Eang, Junjing Yang, and Mattheos Santamouris. Forecasting energy consumption of institutional buildings in singapore. *Procedia Engineering*, 121:1734–1740, 2015.

[41] Chirag Deb, Lee Siew Eang, Junjing Yang, and Mattheos Santamouris. Forecasting diurnal cooling energy load for institutional buildings using artificial neural networks. *Energy and Buildings*, 121:284–297, 2016.

[42] Introduction to neural networks, . URL https://medium.com/datadriveninvestor/introduction-to-neural-networks-a0fe9ec0a947.

[43] Convolutional neural networks (cnns / convnets), . URL https://cs231n.github.io/convolutional-networks//.

[44] Modeling one neuron), . URL https://http://cs231n.github.io/neural-networks-1///.

[45] Kangji Li, Hongye Su, and Jian Chu. Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: A comparative study. *Energy and Buildings*, 43(10):2893–2899, 2011.

[46] Boosting, bagging, and stacking âĂŤ ensemble methods with sklearn and mlens), . URL https:https://medium.com/@rrfd/boosting-bagging-and-stacking-ensemble-methods-with-sklearn-and-mlens-a455c0c982d /.

[47] Erg (eastern research group, inc.) 2013, 2013. URL https://www.erg.com/.

[48] Bank holidays dates 2013, 2014 and 2015 (uk public holidays), . URL http://www.projectbritain.com/bankholidays.html.