International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

IJISAE

ISSN:2147-6799 www.ijisae.org Original Research Paper

Comparative Analysis of Different Machine Learning Techniques Along with Hyper-Parameter Optimization for Prediction of Small Data Set Long Term Electricity Demand of Assam

Sarma Anurag*[1]; Nath Rupanjali[2]

Submitted: 08/12/2023 **Revised**: 13/01/2024 **Accepted**: 28/01/2024

Abstract: Electricity is a prime and compulsory source for development of any nation. Electricity demand of Assam is increasing at an alarming rate and to compensate this loss energy is generated from many non-renewable resources, as a result Assam fails to predict and generate the future electricity demand sustainably. This is due to many factors like rapidly rising population, literacy rate, industrialization, GDP and standard of living. Therefore, predicting the future electricity demand accurately will help in better decision making in implementing energy policies and provide information regarding future energy requirements and how to generate it sustainably. In this paper an attempt has been made to assess the effectiveness of different machine learning (ML) techniques for forecasting long term electricity demand of Assam for small dataset. The data collected is a small dataset consisting of several attributes that influence the energy consumption of Assam. A small dataset has been chosen as in many cases it is challenging to see that any product or system that has been in the market for a small time need to be predicted accurately for its future demand. So, the performance of the predicted results of the ML techniques will help in understanding the challenges of predicting small dataset. And to better understand the relationship between the attributes and the response Partial Dependence Plot has been used.

Keywords: *Energy, Forecasting, Machine learning, Performance, Small dataset*

1. Introduction

In today's era one of the biggest challenges that have been faced by Assam is to meet the growing energy demand sustainably. This is due to the rapidly rising population, Industrialization, GDP and the standard of living. Assam is mainly dependent on 4 types of energy sources, they are Gas, Coal, Hydro and renewable energy. Assam has 1,505 MW installed capacity of electricity generators as of FY 18[1]. The state has 1,027 MW of thermal generators (279 MW coal and 748.5 MW gas). Installed capacity of renewable sources wind, solar, small hydro plant (SHP), biomass power (BP), urban and industrial water power (U&I)) is 46.6 MW [1].

The electricity demand in Assam is increasing at an exponential rate and to compensate this need for electricity, Assam is producing high amounts of electricity from both Non-renewable and renewable energy sources, but the percentage of non-renewable energy sources is high as shown in figure 1.1. So, energy is a prime concern for sustainable development of Assam as the energy sector alone contributes more than 50% of Greenhouse Gas emission (GHG) in Assam **[2]** as shown in figure 1.2.

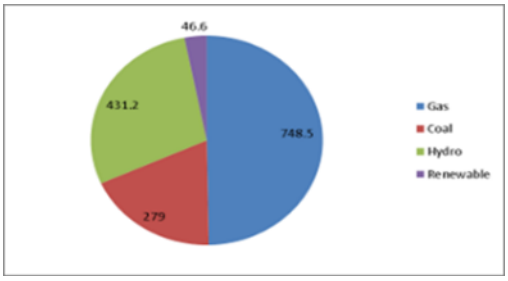


Fig-1.1 Energy generators installed capacity (2018)

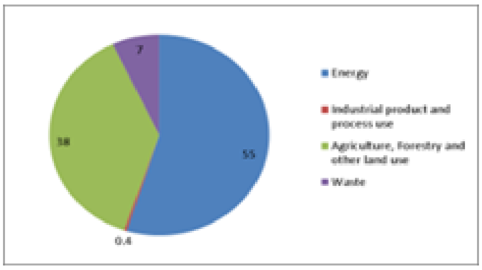


Fig-1.2 Sector wise contribution to economy wide GHG emission of Assam

Therefore, predicting the future electricity demand accurately will help in better decision making and provide information regarding future energy requirements and how to generate it sustainably. Forecasting is necessary in the energy sector for Assam because without forecast there will be no understanding between the production side and the consumption side which will lead to blackouts or excessive energy losses or large amount of GHG emissions. These surpluses and shortages significantly impact environment,

[1] Jorhat Engineering College, ASTU, Assam, India
ORCID ID : 0000-0002-8807-8050
[2] Jorhat Engineering College, ASTU, Assam, India
** Corresponding Author Email: sarmaanurag.90@gmail.com*

economic development, individual lives, and social stability. **[4]** So for Assam to achieve sustainable development goals (SDGs) it is of utmost important to predict the energy demand as Assam always suffer from blackouts during summer and surpluses during winter. Also, there is very limited research in the field of energy demand of Assam.

So, in this paper an attempt has been made to predict the future energy demand of Assam based on different ML techniques as these methods are computationally fast and reliable and accurate but based on no. of attributes and data used. Besides predicting the future energy demand a small dataset has been chosen as in many situations it is challenging to see that any product or system that has been in the market for a small time need to be predicted accurately for its future demand. So, with limited data, the performance of the predicted results of the ML techniques will help in understanding the challenges of predicting small dataset. For example, in many situations when a company tries to assess future demand, the available data contains only a small number of observations as either the product or system is new in the market or it is in the market only for a short time. In these situations, traditional modelling methods don't work well. So, researchers around the world are developing different ML techniques and deep learning techniques to tackle such type of challenges.

2. Methodology Applied

The electricity data has been collected from the archives of Statistical handbook of Assam and directorate of statistics and economics, Assam. After pre-processing of the data only selected attributes have been selected like Consumer price index (CPI), per capita GSDP at current prices, per capita income of Assam and Natural growth rate as these have large influence on the electricity demand of Assam. From the data set 70% data are used as training data set for ML algorithms and 30% data are used as test data set for checking the performance and prediction of the future energy demand. Further ML algorithms parameters like kernel function and root mean square error has been fined tuned with the help of hyper-parameter optimization. And finally, all the performance and predicted results are compared with each other for better decision making.

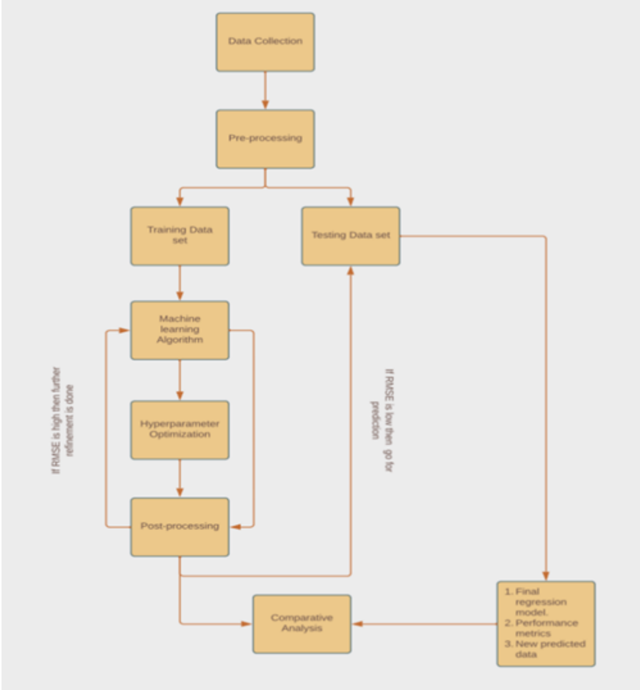


Fig.2.1: Flowchart of methodology applied

2.1. Machine learning Algorithms

Machine learning is an artificial intelligence technique where humans train the machines based on the available past data. As the number of data increases the performance of the ML algorithms increases provided suitable ML model is selected. i.e., MORE DATA > BETTER MODEL > HIGHER ACCURACY. Machine learning has emerged as a growing research field in prediction or forecasting as it can detect patterns in data set and adjust program actions accordingly and focuses on the development of these algorithms that can teach themselves to grow and change when exposed to new data. Further these algorithms parameters can be enhanced for better output which is called as Hyperparameter tuning or optimization and making the models either not suitable or accurate for future prediction.

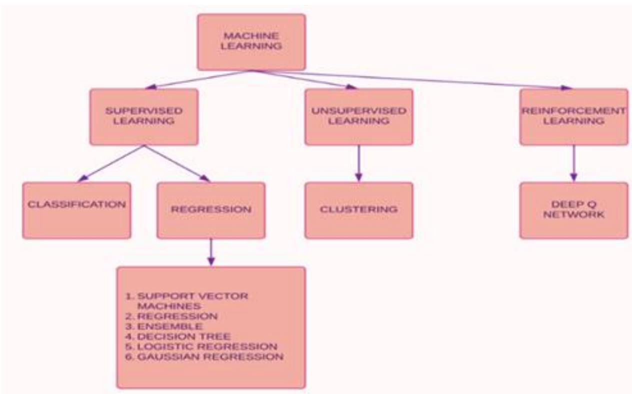


Fig.2.1.1: Classification of regression machine learning algorithms

2.1.1 Artificial Neural Network (ANN)

A computing model called a neural network is based on how the human brain works and is organized. It's a specific kind of machine learning algorithm made to spot patterns and then forecast or decide depending on the input data.

A neuron is the fundamental unit of a neural network. It is a mathematical function that accepts numerous inputs, gives each one a weight, adds them all up, and then passes the total through an activation function to create an output. Every input is multiplied by a corresponding weight, which establishes its significance or applicability. The model's activation function adds non-linearity and aids in simulating intricate connections between inputs and outputs.

In a neural network, neurons are arranged in layers. Input, hidden, and output layers are the three primary categories of layers. The initial input data is received by the input layer, and the final output or prediction is created by the output layer. As their name implies, the hidden layers are layers in-between the input and output layers. In order to discover and extract pertinent characteristics, they execute calculations and transformations on the incoming data. Weights serve as a representation of the connections between neurons in neighboring layers. These weights are modified during a neural network's training phase depending on the discrepancy between the expected output and the actual output. Backpropagation is an optimization technique that calculates the gradients of the error with respect to the weights and updates them appropriately. This correction procedure is often carried out using this approach. This iterative process keeps going until the neural network masters the ability to correctly anticipate or decide based on the input data.

The capacity of neural networks to generalize patterns from data and learn from it in order to generate predictions on fresh, unforeseen data is one of its strongest points. They may be used for a variety of tasks, including sentiment analysis, recommendation systems, natural language processing, picture and audio recognition, and many more. A neural network's design and configuration may change based on the particular issue and data characteristics, providing flexibility and adaptability across several domains.

2.1.2 Support Vector Machines (SVM)

For classification and regression problems, supervised machine learning algorithms called SVMs are utilized. SVMs are particularly useful when working with high-dimensional feature spaces or non-linearly separable data. The fundamental goal of SVMs is to identify the hyperplane that best divides the data into several groups. The hyperplane is a decision boundary in binary classification that maximizes the margin, or the separation between the hyperplane and the closest data points of each class. The closest data points to the hyperplane are referred to as support vectors. SVMs use a kernel function to translate the input data into a higher-dimensional feature space in order to locate the best hyperplane. Finding a linear decision boundary that matches a non-linear boundary in the initial feature space is made possible by this modification. The selection of the kernel function is critical and is based on the problem.

SVMs provide a number of benefits. Even when there are more features than samples, they are efficient at processing high-dimensional feature spaces. The margin maximization makes them less prone to overfitting. Furthermore, SVMs may detect intricate correlations in the data by using kernel functions. However, because the training time grows with the number of samples, SVMs may be computationally costly, especially with big datasets. SVMs could also depend on the choice of hyperparameters, including the kernel parameters and regularization parameter. For good performance, these hyperparameters must be tuned properly.

2.1.3 Gaussian Regression (GR)

Machine learning method for regression tasks is referred to as Gaussian regression, commonly known as Gaussian process regression (GPR) or Kriging. The link between the input variables and the related output values is modelled as a distribution over functions using a non-parametric Bayesian technique. The objective of Gaussian regression is to forecast a continuous output variable from a set of input variables. In contrast to linear regression, which estimates a single value as the forecast, Gaussian regression models the output as a distribution with a mean and covariance. This distribution illustrates the degree of uncertainty in the forecasts and offers a gauge of assurance or dependability.

Gaussian regression models learn from observed input-output pairs during the training phase by estimating the mean and covariance function parameters. The objective is to choose the distribution that best fits the training data while taking the uncertainty of the predictions into account.

There are various benefits of using gaussian regression. It can efficiently handle small datasets with a lot of noisy or few observations. It offers a logical framework for managing uncertainty in predictions, which makes it especially helpful in situations where accurate uncertainty estimates are essential, such active learning or decision-making in the context of uncertainty. Additionally, because various patterns and data structures may be captured depending on the kernel function selected, Gaussian regression enables flexible modelling of complicated interactions. Gaussian regression, however, may be computationally exhausting, particularly as the quantity of training samples rises. It becomes computationally costly and can be scalable when the covariance matrix is inverted to obtain the posterior distribution. To meet these issues, several approximation and optimization approaches have been created.

2.1.4 Multiple Linear Regression (MLR)

Multiple linear regression is a statistical method for modelling the connection between several independent variables and a dependent variable. In multiple linear regression, the objective is to identify the best-fitting linear equation that forecasts the value of the dependent variable based on the values of the independent variables.

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta z Xz + E$

where:

- Y is the dependent variable (the variable being predicted)
- β_0 is the y-intercept or constant term
- β_1, β_2, ..., βz are the coefficients associated with each independent variable (X_1, X_2, ..., Xz)
- X_1, X_2, ..., Xz are the independent variables (predictors)
- E represents the error term, which accounts for the variability not explained by the linear relationship

Keeping all other variables fixed, the coefficients β_1, β_2, ..., β_p show the change in the dependent variable corresponding to a one-unit change in the corresponding independent variable. Statistical techniques like maximum likelihood estimation and ordinary least squares (OLS) are used to estimate these coefficients from the provided data.

There are various advantages of using multiple linear regression. It enables the analysis of the unique contributions made by each independent variable while taking other variables' impacts into account. Additionally, it may be utilized to make predictions and comprehend the nature and significance of interactions between variables. Multiple linear regression can have certain drawbacks, though. Assuming a linear connection, it may not be the best regression approach if the relationship is not linear. It may also be sensitive to outliers and deviations from the fundamental presumptions. The model must be carefully interpreted and validated in order to guarantee its dependability.

2.1.5 Ensemble regression

A machine learning method called ensemble regression combines many different regression models to provide predictions that are more precise. It makes use of the notion that mixing the findings of many models rather than relying just on one can frequently produce superior outcomes. When dealing with complicated or noisy data, ensemble regression methods are very helpful since they may assist increase prediction accuracy, decrease overfitting, and handle multiple sources of uncertainty.

(a) Bagging (Bootstrap Aggregating): Bagging uses random sampling with replacement to create several subsets of the original training data. The next step is to train a different regression model using each subgroup. The ultimate result is produced during prediction by combining all of the separate models' predictions (for example, by taking the average). Bagging can increase prediction accuracy and aid to minimize variation.

(b) Random Forest: It is a bagging extension created especially for decision trees. Each decision tree in the ensemble is trained on a distinct sample of the data, and at each split, a random subset of characteristics is taken into account. The average of all the trees' forecasts yields the final prediction. Outliers, non-linear correlations, and high-dimensional data are not a problem for Random Forest.

(c) Boosting: Boosting is an iterative ensemble technique that combines weak learners, or straightforward models that perform marginally better than random guessing, to produce a robust regression model. Boosting trains models in a sequential manner using altered copies of the training data, with each succeeding model concentrating more on the cases that the preceding models incorrectly identified. Combining all of the model predictions, often using a weighted total, yields the final forecast. The boosting algorithms AdaBoost, Gradient Boosting, and XGBoost are all well-liked.

(d) Stacking: Stacking involves developing a variety of unique regression models on the same dataset and integrating the results with the help of a meta-model. Using the results of the individual models as inputs, the meta-model gains the ability to forecast the future. Stacking enables the ensemble to recognize numerous features and patterns in the data, possibly resulting in enhanced performance.

2.2 Hyperparameter Optimization/Tunning

Hyperparameter optimization plays an important role in the development of the model. It entails choosing the optimal collection of hyperparameters for a certain machine learning algorithm, which has a big influence on the model's performance and generalization capacity. The objective of hyperparameter optimization is to identify the best possible set of hyperparameter values that produces the greatest model performance. Usually, this is accomplished by using a cross-validation approach or assessing the model's performance on a validation set. Depending on the particular issue, a performance metric, such as accuracy, precision, recall, F1 score, or mean squared error, will be utilized for evaluation.

There are several methods for hyperparameter optimization, including manual search, grid search, random search, and more advanced techniques like Bayesian optimization and genetic algorithms.

2.2.1 Manual Search: In this method, hyperparameter values are manually chosen based on past knowledge and experience. While simple, it can take time and may not always produce the best outcomes.

2.2.2 Grid Search: In this method, each hyperparameter is given a predetermined set of values. The algorithm thoroughly assesses each conceivable combination of these variables before choosing the one that produces the best results. In vast hyperparameter spaces, grid search may be computationally costly while being easy to execute.

2.2.3 Random search: Hyperparameter values are chosen at random from predetermined ranges via random search. The highest performing collection of hyperparameters is selected after a certain number of iterations in which each random combination is evaluated. In cases when the hyperparameter space is vast and the number of iterations is constrained, random search is more effective than grid search.

2.2.4 Bayesian optimization: Bayesian optimization estimates the performance of various hyperparameter configurations using probabilistic models. It updates the probabilistic model, tests the model repeatedly, and chooses new hyperparameter values to test based on previous evaluations. This approach is very helpful when testing the model requires a lot of resources or is expensive.

Furthermore, it is crucial to keep in mind that the ideal hyperparameter values may change based on the dataset and the particular issue being handled, necessitating a re-optimization of the hyperparameters whenever new tasks or datasets are used.

In this study Bayesian hyperparameter optimization technique has been used as there are various benefits to Bayesian hyperparameter optimization. Compared to time-consuming search techniques like grid search and manual search, it is effective in searching the hyperparameter space and needs fewer evaluations. By taking uncertainty into account, it can deal with noisy or difficult-to-evaluate objective functions. The algorithm can effectively converge to the ideal configuration of the hyperparameters since it also offers a logical method to balance exploration and exploitation.

Optimizer used: Bayesian Optimization

Number of iterations: 100

Acquisition function: Expected improvement per second plus

Maximum training time: 300sec

3. Results and Discussions

A plot of per capita requirement of energy (kwh) for Assam along with different attributes that are having an effect on the electricity consumption is shown in figure-3 below. Assam is the most developed state in the Northeast area of India and its electricity consumption is increasing year by year. This rise of electricity consumption is possible only when the state is developing year after year. So, from the plot it can be seen that per capita income at current prices, GSDP at current prices, consumer price index for working class population is increasing year by year even though the natural growth rate is decreasing which leads to the conclusion that the electricity requirement is also increasing. As a result, forecasting of electricity demand along with these attributes will help in better prediction of electricity for future better decision making. Because with bad prediction will lead to loss of electricity or surplus of electricity demand in Assam.

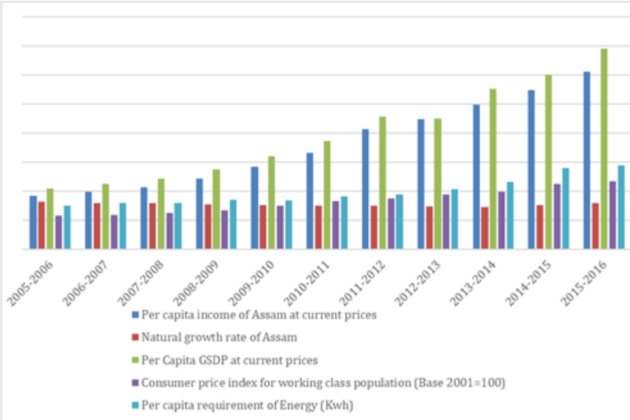


Fig.3- Year wise distribution of electricity requirement and different attributes affecting electricity consumption

3.1 Multiple Linear regression

MLR is the most widely used ML technique for prediction. In the present paper since the data collected are mostly linear either in upward or downward trend so MLR has been used for selected data set and it took 6.97 sec time to train the model. The results of both training and test validation of Stepwise MLR are given below:

Table3.1.1: - Training results of Stepwise MLR

Metrics	Value	Remarks
RMSE (Validation)	49.806	Satisfactory results as the RMSE, R-sqr, MSE and MAE showed that the model fits well with the given small data set
R-squared (Validation)	0.17	
MSE (Validation)	2480.7	
MAE (Validation)	37.269	

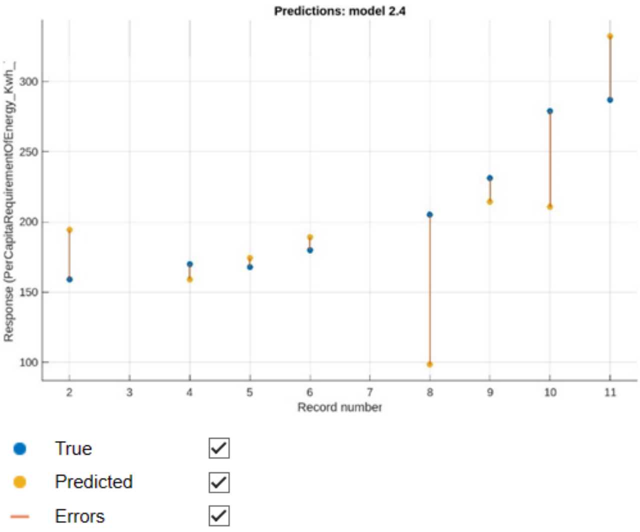


Fig.3.1.1- Response plot of MLR model

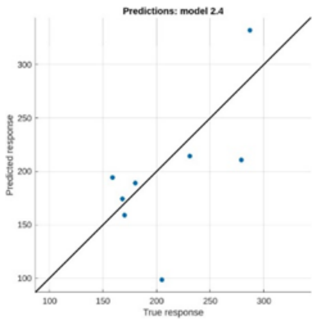


Fig. 3.1.2- Predicted plot of MLR

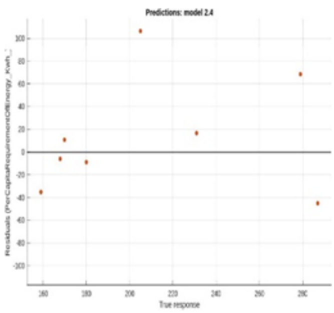


Fig. 3.1.3- Validation plot of MLR

Table 3.1.2:- Test results of Stepwise MLR

Metrics	Value	Remarks
RMSE (Test)	11.849	Test results are far better than training results which shows that the model can be used for small dataset forecasting
R-squared (Test)	0.48	
MSE (Test)	140.4	
MAE (Test)	9.2181	

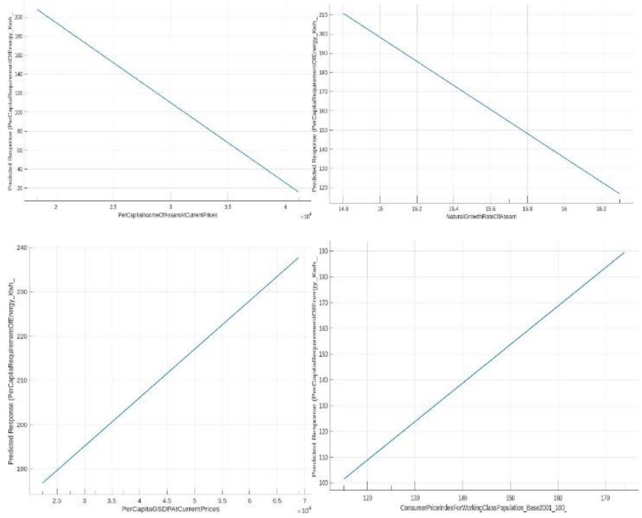


Figure 3.1.4:-

Plot-1 Partial dependence plot of Per capita requirement of energy vs Per capita income of Assam at current prices

Plot-2 Partial dependence plot of Per capita requirement of energy vs Natural growth rate of Assam

Plot-3 Partial dependence plot of Per capita requirement of energy vs Per capita GSDP at current prices

Plot-4 Partial dependence plot of Per capita requirement of energy vs Consumer price index for working class population based on year 2001

The Partial dependence plot (PDP) of MLR showed that both per capita income of Assam at current prices and natural growth rate of Assam showed negative impact on the per capita requirement of energy. But the features like GSDP at current prices and CPI for working class people showed positive impact on the per capita requirement of energy i.e., the probability of requirement of energy will increase in increase of GSDP at current prices and CPI for working class people. So, in case of MLR only two features or attributes or predictors are of interest for the target response.

3.2 SVM

SVM algorithm are best suited for those problems which are having small dataset and number of attributes and noises present in it. So, this model was used in the chosen small dataset and it took 8.059 sec time to train the model. The results of both training and test validation are given below:

Table 3.2.1:- Training results of SVM

Metrics	Value	Remarks
RMSE (Validation)	17.607	Satisfactory results as the RMSE, R-sqr, MSE and MAE showed that the model fits well with
R-squared (Validation)	0.90	
MSE (Validation)	310.01	

MAE (Validation)	13.863	the given small data set

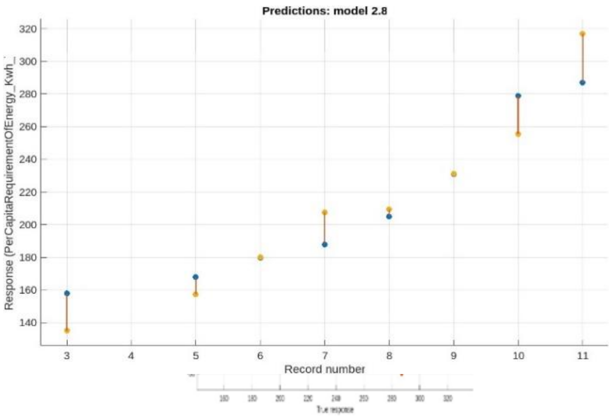


Fig. 3.2.2- Validation plot of SVM

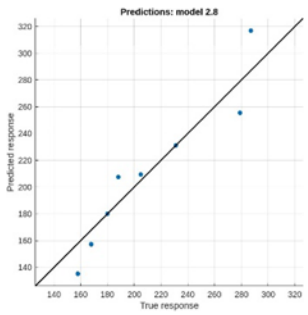


Fig. 3.2.3- Predicted plot of SVM

Table 3.2.2:- Test results of SVM

Metrics	Value	Remarks
RMSE (Test)	6.0747	Test results are far better than training results which shows that the model can be used for small dataset forecasting
R-squared (Test)	0.45	
MSE (Test)	36.901	
MAE (Test)	5.9501	

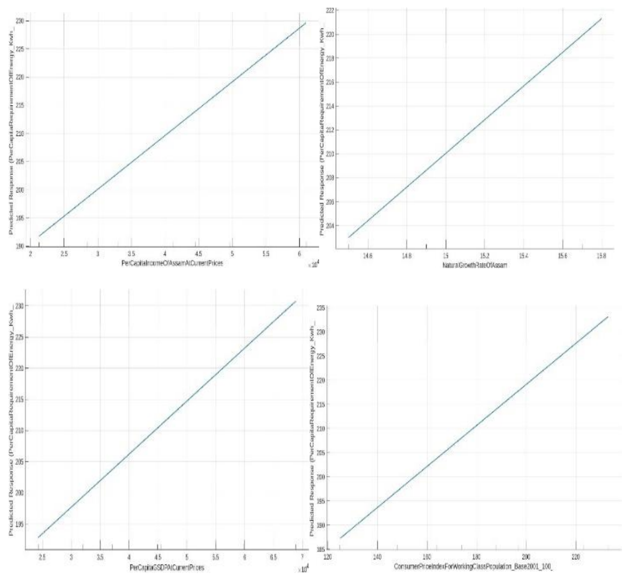


Figure 3.2.4:-

Plot-1 Partial dependence plot of Per capita requirement of energy vs Per capita income of Assam at current prices

Plot-2 Partial dependence plot of Per capita requirement of energy vs Natural growth rate of Assam

Plot-3 Partial dependence plot of Per capita requirement of energy vs Per capita GSDP at current prices

Plot-4 Partial dependence plot of Per capita requirement of energy vs Consumer price index for working class population based on year 2001

The PD (Partial dependency) plots of SVM model showed that with rise in all the attributes or predictors will increase the probability of Per capita requirement of energy of Assam.

The validation dataset used for training of the SVM model is used for hyperparameter tuning and it took 43.534 sec time to tune the model. Hyperparameters taken are: -

1. Kernal function: Quadratic
2. Box constraint: Auto
3. Kernal scale: Auto
4. Epsilon: Auto
5. Standardized data: Yes

The following table shows the results of the hyperparameter of SVM:

Table 3.2.3:- SVM model hyperparameter

Metrics	Value	Remarks
RMSE (Validation)	24.353	Hyperparameter of SVM validation model showed somewhat bad results w.r.t trained validation results. This shows that the parameters of SVM need to be further fine-tuned for low metric values
R-squared (Validation)	0.80	
MSE (Validation)	593.06	
MAE (Validation)	15.995	
RMSE (Test)	4.54	But test model is found acceptable as metric values fits best for the chosen dataset
R-squared (Test)	0.69	
MSE (Test)	20.614	
MAE (Test)	3.969	

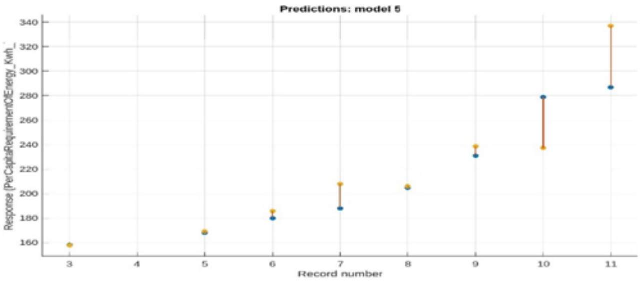
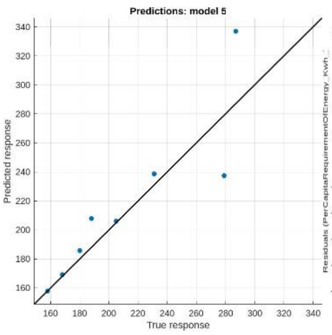


Fig. 3.2.5- Response plot of SVM hyperparameter model



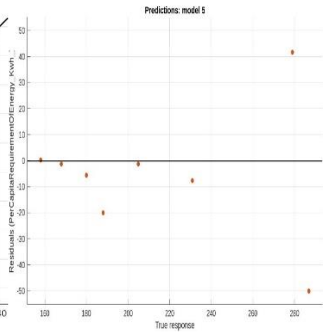


Fig. 3.2.6- SVM hyperparameter prediction plot

Fig. 3.2.7- SVM hyperparameter validation plot

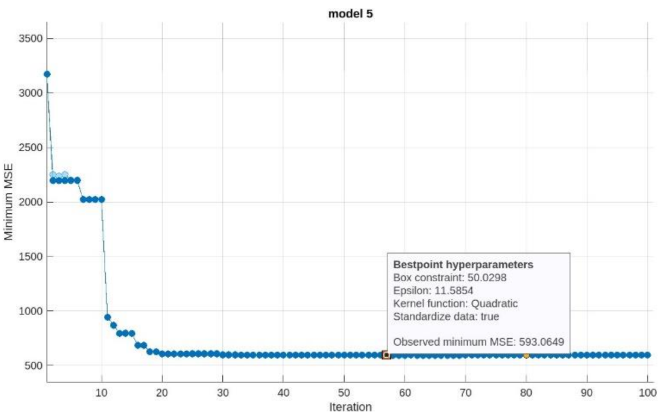


Fig. 3.2.8- Minimum MSE plot of SVM hyperparameter

3.3 GPR

Similar to SVM, GPR used in this model showed satisfactory results as these models can easily solve small data sets with many attributes. The model took 14.572 sec time to train the model. The results of both training and test validation are given below:

Table 3.3.1:- Training results of GPR

Metrics	**Value**	**Remarks**
RMSE (Validation)	19.249	Satisfactory results as the RMSE, R-sqr, MSE and MAE showed that the model fits well with the given small data set
R-squared (Validation)	0.88	
MSE (Validation)	370.51	
MAE (Validation)	14.628	

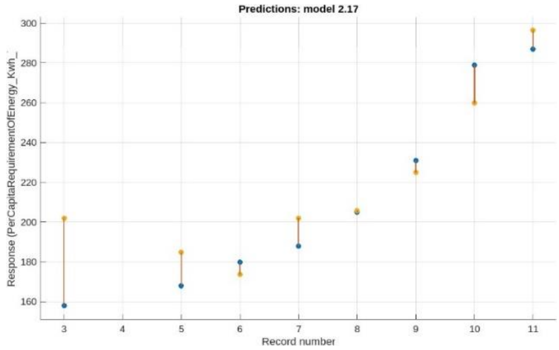
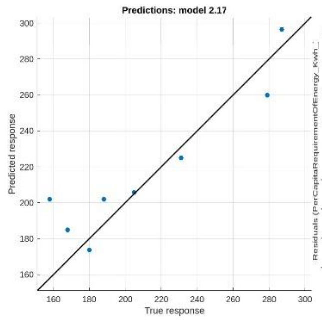


Fig. 3.3.1- Response plot of GPR model



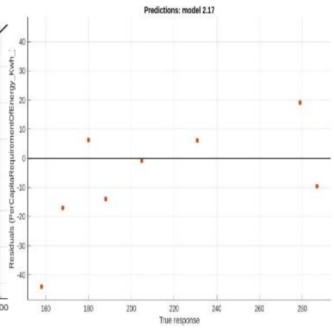
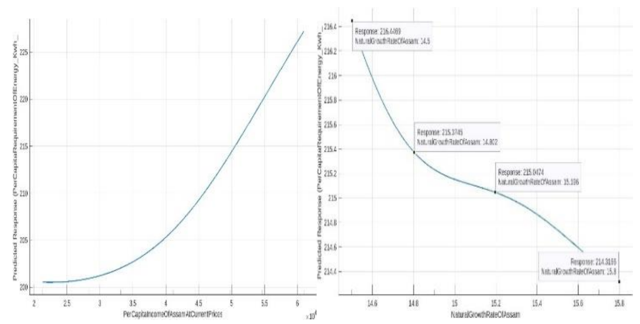


Fig. 3.3.2- Prediction plot of GPR **Fig. 3.3.3-** Validation plot of GPR

Table 3.3.2:- Test results of GPR

Metrics	**Value**	**Remarks**
RMSE (Test)	10.618	Test results are far better than training results which shows that the model can be used for small dataset forecasting
R-squared (Test)	-0.69	
MSE (Test)	112.75	
MAE (Test)	8.475	



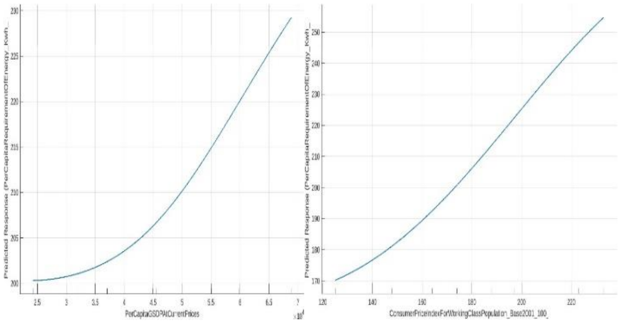


Fig. 3.3.4- Plot-1 Partial dependence plot of Per capita requirement of energy vs Per capita income of Assam at current prices

Plot-2 Partial dependence plot of Per capita requirement of energy vs Natural growth rate of Assam

Plot-3 Partial dependence plot of Per capita requirement of energy vs Per capita GSDP at current prices

Plot-4 Partial dependence plot of Per capita requirement of energy vs Consumer price index for working class population based on year 2001

The PD plots Per capita income of Assam at current prices, Per capita GSDP at current prices and Consumer price index for working class population based on year 2001of GPR model showed positive impact on the Per capita requirement of energy of Assam. But on the other hand, even though the natural growth rate of Assam is decreasing by some percentage per year without affecting the energy consumption but the PDP of Per capita requirement of energy vs Natural growth rate of Assam showed negative impact on the energy requirement i.e., with rise in population the probability of energy consumption will decrease. So, for this model only three attributes or features are of interest for the target response.

The validation dataset used for training of the GPR model is used for hyperparameter tuning and it took 40.781 sec time to tune the model. Hyperparameters taken are: -

1. Signal standard deviation: 34.822

2. Optimize numeric parameters: Yes

3. Kernal function: Isotropic matern 5/2

4. Kernal scale: Auto

5. Sigma: Auto

The following table shows the results of the hyperparameter of GPR:

Table 3.3.3:- GPR model hyperparameter

Metrics	**Value**	**Remarks**
RMSE (Validation)	10.593	Acceptable tuned hyperparameter GPR model far better than SVM model
R-squared (Validation)	0.96	
MSE (Validation)	112.22	
MAE (Validation)	9.7046	
RMSE (Test)	7.2865	Acceptable metrics value for concluding that GPR fits best for the chosen dataset
R-squared (Test)	0.21	
MSE (Test)	53.093	
MAE (Test)	6.8923	

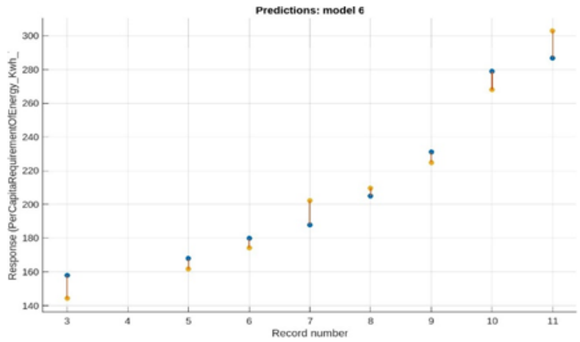


Fig. 3.3.5- GPR hyperparameter response plot

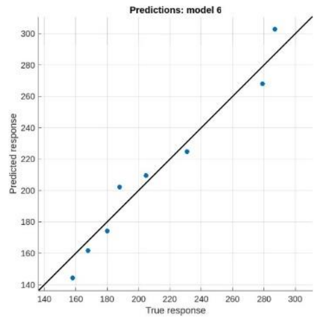


Fig. 3.3.6- GPR hyperparameter prediction plot

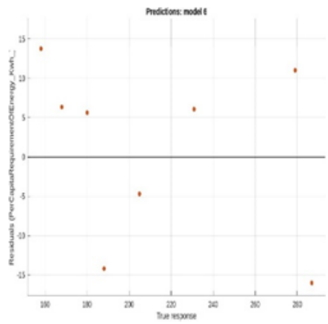


Fig. 3.3.7- GPR hyperparameter validation plot

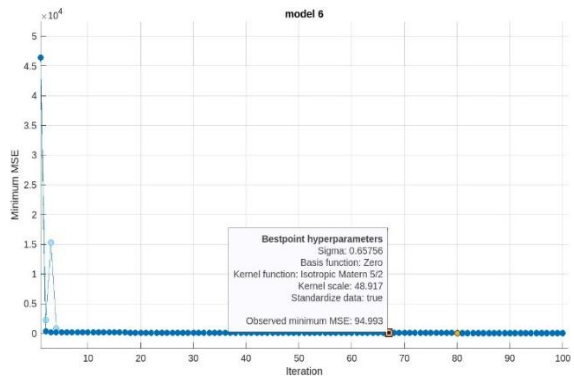


Fig. 3.3.8- GPR hyperparameter minimum MSE plot

3.4 ANN

The only deep learning method which gained importance within a short span of time and currently it is being used by almost all researchers for different purposes. ANN always gives better result w.r.t. forecasting even if the dataset is full of noises or small. Since ANN can solve small datasets with many attributes accurately so, it is being used in the current study. The model took 2.04 sec time to train the model. The results of both training and test validation are given below:

Table 3.4.1:- Training results of ANN

Metrics	Value	Remarks
RMSE (Validation)	5.833	Acceptable results of 3 layered NN showed that the model fits well with the given small data set and it is currently the best results of all ML techniques
R-squared (Validation)	0.96	
MSE (Validation)	34.023	
MAE (Validation)	4.812	

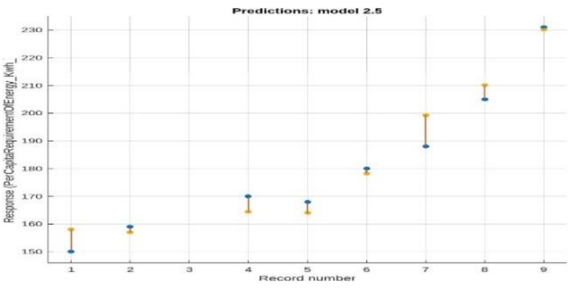


Fig. 3.4.1- ANN response plot

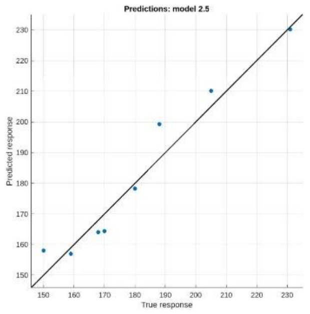


Fig. 3.4.2- ANN prediction plot

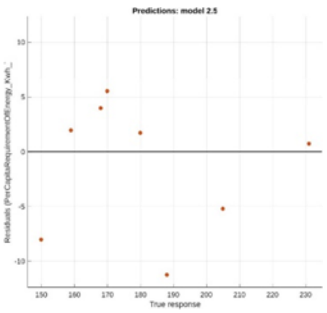


Fig. 3.4.3- ANN validation plot

Table 3.4.2:- Test results of ANN

Metrics	Value	Remarks
RMSE (Test)	32.229	RMSE of test data is greater than that of RMSE of training data. This might be because of overfitting of data or test dataset may have data that are unknown / not common and the test dataset used is small w.r.t. the training dataset.
R-squared (Test)	0.70	
MSE (Test)	1038.7	
MAE (Test)	28.082	

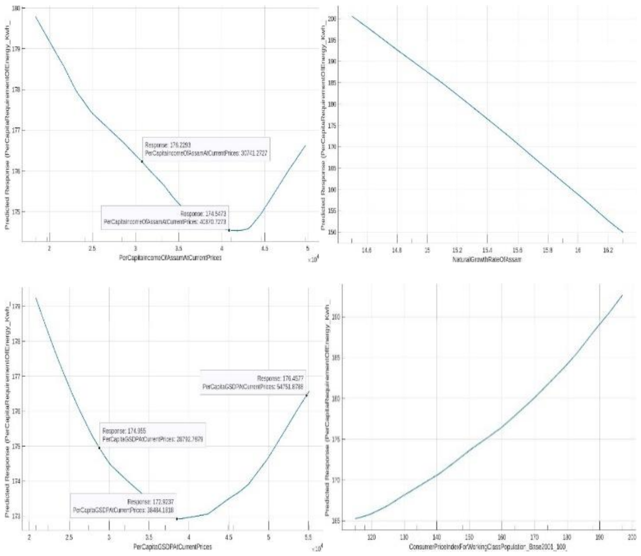


Figure 3.4.4:-

Plot-1 Partial dependence plot of Per capita requirement of energy vs Per capita income of Assam at current prices

Plot-2 Partial dependence plot of Per capita requirement of energy vs Natural growth rate of Assam

Plot-3 Partial dependence plot of Per capita requirement of energy vs Per capita GSDP at current prices

Plot-4 Partial dependence plot of Per capita requirement of energy vs Consumer price index for working class population based on year 2001

The PD plots of Per capita income of Assam at current prices and Per capita GSDP at current prices showed gradual decrease in the response as they are rising but after a certain break-even point both the predictors started to increase with increase in response. Also, the CPI for working class population based on year 2001 showed positive impact on the Per capita requirement of energy of Assam. In the case of ANN model three predictors or features are of interest except the Natural growth rate of Assam which is having negative impact on the selected response.

The validation dataset used for training of the NN model is used for hyperparameter tuning and it took 86.69 sec time to tune the model. Hyperparameters taken are:-

1. No. of connected layers: 1-3
2. Layer size range: 10-100
3. Iteration limit: 1000
4. Activation: ReLU, tanh, Sigmoid, None
5. Lamda range: 0-0.3
6. Standardized data: Yes

The following table shows the results of the hyperparameter of ANN:

Table 3.4.3:- ANN model hyperparameter

Metrics	Value	Remarks
RMSE (Validation)	26.199	1. The non-hyperparameter metrics of ANN is very much better than performance metrics of hyperparameter model. 2. ANN can have multiple different performance metrics as the hyperparameter tuning range is very high.
R-squared (Validation)	0.77	
MSE (Validation)	686.41	
MAE (Validation)	20.467	
RMSE (Test)	11.368	Acceptable metrics value for concluding that ANN test model fits best for the chosen dataset
R-squared (Test)	-0.93	
MSE (Test)	129.23	
MAE (Test)	9.4497	

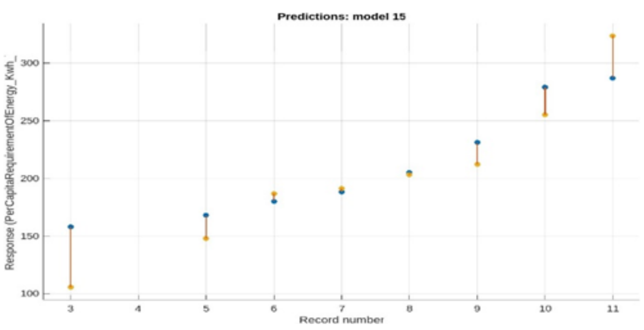
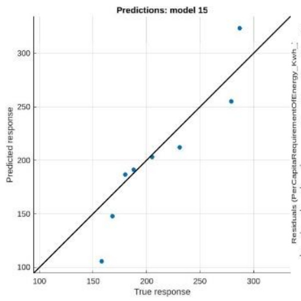


Fig. 3.4.5- ANN hyperparameter response plot



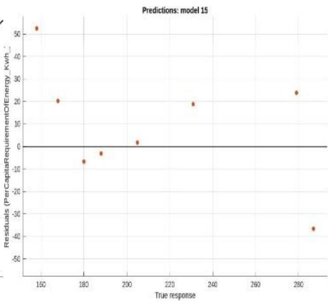


Fig. 3.4.6- ANN hyperparameter prediction plot

Fig. 3.4.7- ANN hyperparameter validation plot

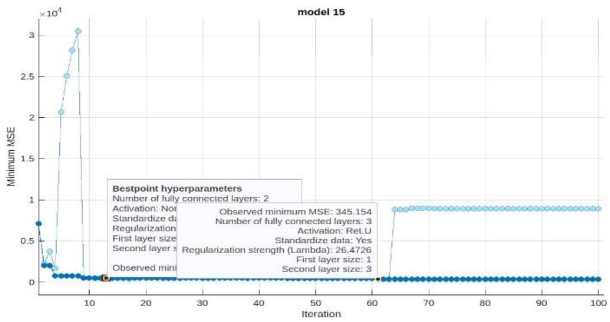


Fig. 3.4.8- ANN hyperparameter minimum MSE plot

3.5 Ensemble/Tree

Ensemble tree algorithm are best suited for those problems which are having number of attributes. So, this model was used in the chosen small dataset and it took 2.0611 sec time to train the model. The results of both training and test validation are given below:

Table 3.5.1:- Training results of Ensemble tree

Metrics	Value	Remarks
RMSE (Validation)	29.185	The performance metrics does not show any satisfactory results as the error difference is very high as shown in response plot also.
R-squared (Validation)	0	
MSE (Validation)	851.74	
MAE (Validation)	24.88	

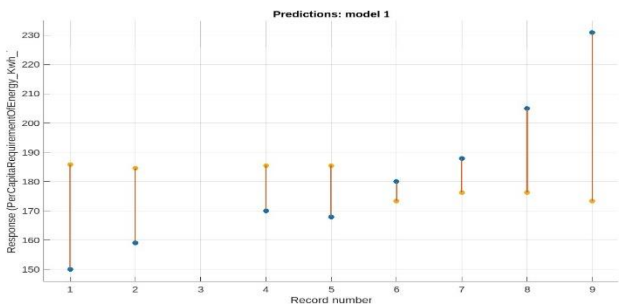
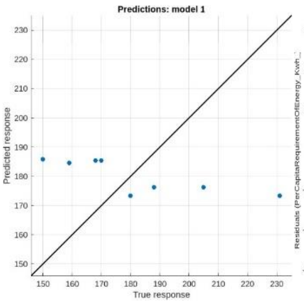


Fig. 3.5.1- Ensemble/Tree response plot



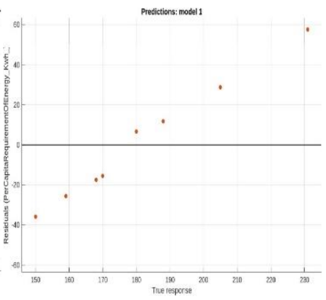


Fig. 3.5.2- Ensemble prediction plot **Fig. 3.5.3-** Ensemble validation plot

Table 3.5.2:- Test results of Ensemble tree

Metrics	**Value**	**Remarks**
RMSE (Test)	84.13	The metrics results proves that Ensemble is not suitable for small dataset as the results of test data is extremely high for prediction. Therefore ensemble is not recommended for small dataset
R-squared (Test)	-1.03	
MSE (Test)	7077.9	
MAE (Test)	75.542	

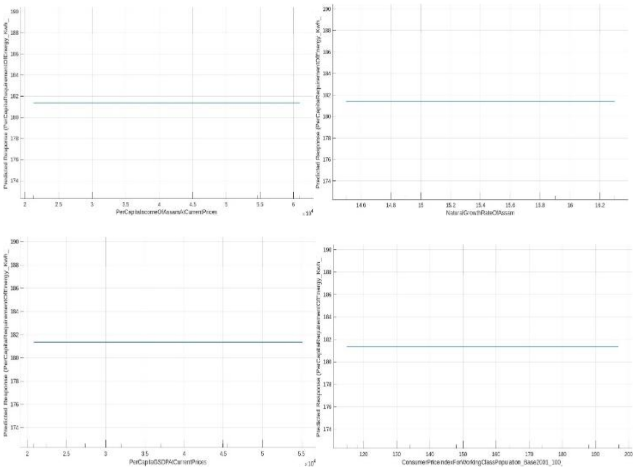


Figure 3.5.4:-

Plot-1 Partial dependence plot of Per capita requirement of energy vs Per capita income of Assam at current prices

Plot-2 Partial dependence plot of Per capita requirement of energy vs Natural growth rate of Assam

Plot-3 Partial dependence plot of Per capita requirement of energy vs Per capita GSDP at current prices

Plot-4 Partial dependence plot of Per capita requirement of energy vs Consumer price index for working class population based on year 2001

In case of Ensemble tree since all the predictors showed flat line thus it can be confirmed that this model shows no interaction between the predictors or features with the response. So, this model is not suitable for small datasets.

4. Conclusion

The challenges for analyzing small data sets can be substantial, and is often necessary in areas where it is infeasible to collect further data. Assam being the economic hub of North East India, prediction of energy demand became a useful task. But in the case of energy demand for Assam, the data collected or available is small. So, in such situation instead of using the traditional time series methods different ML and deep learning methods are used and the performance of these techniques are compared. And it is observed that these ML techniques are suitable for energy related prediction having small datasets. From the fig.4.1 and table 4.1 it can be seen that

(a) MSE for testing dataset is highest for Ensemble/tree and ANN whereas MSE for training dataset is highest for MLR and Ensemble/tree. Further from the Partial dependence plots it has been observed that Ensemble/tree model shows no interaction between the predictors with the response and MLR model shows interaction only with GSDP at current prices and CPI for working class people thereby accepting only two predictors for the response. Thus, from the above observations it can be concluded that Ensemble/tree and MLR is not suitable for the selected dataset with 4 attributes or predictors for prediction of per capita energy requirement of Assam.

(b) The partial dependence plot of GPR and ANN didn't accept Natural growth rate of Assam as a predictor and SVM is the only model accepting all the 4 selected attributes or predictors. But all the three models GPR, SVM and ANN showed better performance w.r.t. MLR and Ensemble.

(c) In case of training dataset performance metrics only ANN, GPR hyperparameter and SVM showed better results.

(d) But in case of testing dataset performance metrics SVM hyperparameter, GPR and ANN hyperparameter showed better results.

(e) Thus, from the point (c) and (d) it can be concluded that ANN, SVM, GPR and its hyperparameters can be used for small datasets except for ANN testing model because in the current study the RMSE of Trained model is very small than RMSE of Test model, this might be due to overfitting of data because of small dataset used.

(f) From PD plot point of view ANN and GPR can perform better with less predictors in comparison with SVM for prediction of Per capita requirement of energy of Assam.

(g) With reference to the results of the current paper it is recommended to use partial dependence plots for selecting the attributes i.e., in this case Per capita income of Assam at current prices, Per capita GSDP at current prices and Consumer price index for working class population based on year 2001 for forecasting Per capita requirement of energy of Assam.

5. Future Scope and Limitations

1. Hyperparameter optimization of SVM, GPR and ANN is not limited to the above parameters used. Its performance metrics can be further enhanced by changing these parameters.

2. Instead of using Bayesian optimization other hyperparameter optimization techniques can be used.

3. Comparative analysis can be done for Hybrid ML or traditional techniques for small datasets.

4. Number of predictors used will allow the ML and deep learning techniques to predict more accurately with better performance metrics.

Limitations for forecasting of small dataset is that finding the right forecasting model which is time consuming and with small data it is difficult to select the right attribute as too many attributes will lead to larger error.

Author contributions

Anurag Sarma: Data Collection, Methodology, Software, Investigation, Validation, Field study **Rupanjali Nath:** Conceptualization, Writing-Original draft preparation, Software, Writing-Reviewing and Editing

Conflicts of interest

The authors declare no conflicts of interest.

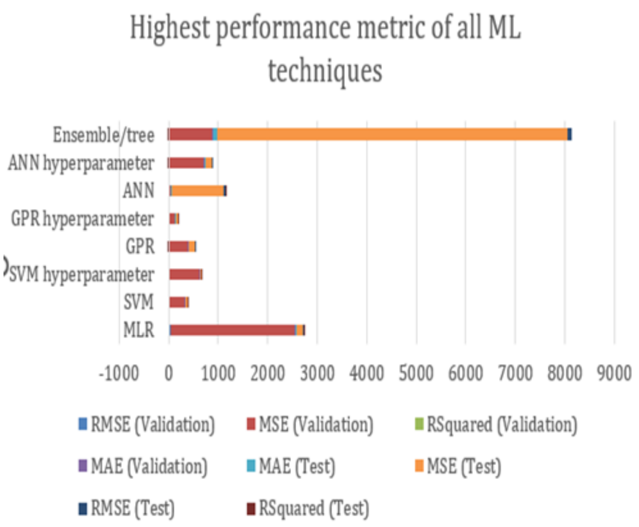


Fig.4.1- Highest performance metric of all ML techniques and its hyperparameters

Table 4.1- Performance metrics table of all ML techniques and its hyperparameters

Model Type	RMSE (Validation)	MSE (Validation)	RSquared (Validation)	MAE (Validation)	MAE (Test)	MSE (Test)	RMSE (Test)	RSquared (Test)
MLR	49.806	2480.68	0.17	37.269	9.22	140.4	11.85	0.475
SVM	17.607	310.01	0.9	13.86	5.95	36.901	6.075	0.45
SVM hyperparameter	24.353	593.05	0.8	15.995	3.97	20.614	4.54	0.69
GPR	19.248	370.51	0.88	14.628	8.475	112.75	10.62	-0.69
GPR hyperparameter	10.593	112.22	0.96	9.7046	6.892	53.09	7.28	0.21
ANN	5.833	34.023	0.96	4.812	28.082	1038.7	32.229	0.7
ANN hyperparameter	26.199	686.41	0.77	20.467	9.4497	129.23	11.368	-0.93
Ensemble/tree	29.185	851.74	0	24.88	75.542	7077.9	84.13	-1.03

References

[1] Anand A., & Suganthi L. (2018). Hybrid GA-PSO optimization of artificial neural network for forecasting electricity demand. *Energies, 11(4), 728*.

[2] Assam energy vision draft of APDCL (2018).

[3] Aylin C¸igˇdem Ko¨ne, Tayfun Bu¨ ke (2010) Forecasting of CO2 emissions from fuel combustion using trend analysis. Elsevier, Renewable and sustainable energy reviews 14 2906-2915

[4] Balachandran S, Velavan R and Rudramoorthy R. (2009) Energy Demand Forecasting for a District in Tamilnadu, India - An Analytical Approach. RERIC, International Energy Journal 10 187-194

[5] Baruah D. C., & Bora G. C. (2008). Energy demand forecast for mechanized agriculture in rural India. Energy Policy, 36(7), 2628-2636.

[6] Chaturvedi S., Rajasekar E., Natarajan S., & McCullen N. (2022). A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total

and peak monthly energy demand for India. Energy Policy, 168, 113097.

[7] Debnath Kumar Biswajit, Mourshed Monjur (2018) Forecasting methods in energy planning models. Elsevier, Renewable and sustainable energy reviews 88, 297-325

[8] Kale R. V., & Pohekar S. D. (2014). Electricity demand and supply scenarios for Maharashtra (India) for 2030: An application of long range energy alternatives planning. Energy Policy, 72, 1-13.

[9] Kumar A., Mittal A., & Pradhan A. K. (2022). Magnitude and determinants of energy intensity: evidence from Indian firms. Environmental Science and Pollution Research, 1-12.

[10] L. Suganthi, Anand A. Samuel (2012) Energy models for demand forecasting—A review. Elsevier, Renewable and sustainable energy reviews 16, 1223-1240

[11] Madeline Hui Li Lee, Yee Chee Ser et. al. (2022) A Comparative Study of Forecasting Electricity Consumption Using Machine Learning Models. Mathematics 2022, 10, 1329. https://doi.org/10.3390/math10081329

[12] Mubarok Hussain, Septyawan Ryan (2018) Analysis of PLN’s Electrical Energy Demand in the Area of Batam-Indonesia Using the Linear Regression Method. 2018 4th International Conference on Science and Technology (ICST), Yogyakarta, Indonesia

[13] Nandi Sumit and Basak Pijush (2014) Carbon Dioxide Emission in Assam, India- A Mathematical Study Based on Attributes. Res J. Chem. Environ. Sci. Vol 2 [4] August: 39-46

[14] N Sumit and B. Pijush (2014) Carbon Dioxide Emission in Assam, India- A Mathematical Study Based on Attributes. Res. J. Chem. Env. Sci. Vol 2 [4] August. 39-46

[15] Pokharel S. (2007) An econometric analysis of energy consumption in Nepal. Energy Policy 35 350–361

[16] Pradhan A. K., Rout S., & Khan I. A. (2021). Does market concentration affect wholesale electricity prices? An analysis of the Indian electricity sector in the COVID-19 pandemic context. Utilities Policy, 73, 101305.

[17] Qiang Wang, Shuyu Li, Rongrong Li (2018) Forecasting Energy Demand in China and India: Using Single-linear, Hybridlinear, and Non-linear Time Series Forecast Technique. Energy (2018), doi: 10.1016/j.energy.2018.07.168

[18] Sachan A., Sahu U. K., Pradhan A. K., & Thomas R. (2023). Examining the drivers of renewable energy consumption: Evidence from BRICS nations. *Renewable Energy, 202, 1402-1411*.

[19] Sang-Bing Tsai, Youzhi Xue et al. (2016) Models for forecasting growth trends in renewable energy. Elsevier, Renewable and sustainable energy reviews

[20] Saravanan S., Kannan S. and Thangaraj C. (2012) India’s electricity demand forecast using regression analysis and artificial neural networks based on principal components. ICTACT journal on soft computing, july, volume: 02, issue: 04

[21] Statistical handbook of Assam (2018)

[22] Trend analysis of GHG emissions in Assam (2013). GHG platform India

[23] Ullash K. Rout, Alfred Vo, Singh Anoop et al. (2011) Energy and emissions forecast of China over a long-time horizon. Elsevier, Energy 36 1-11

[24] Wang Q., Li S., & Li R. (2018). Forecasting energy demand in China and India: Using single-linear, hybrid-linear, and non-linear time series forecast techniques. Energy, 161, 821-831.

[25] Wen-jing Niu, Zhong-kai Feng et. al. (2021) Short-term electricity load time series prediction by machine learning model via feature selection and parameter optimization using hybrid cooperation search algorithm. IOP Publishing; Environmental Research Letters 16 055032