PREDICTION OF FULL LOAD ELECTRICAL POWER OUTPUT OF A BASE LOAD OPERATED COMBINED CYCLE POWER PLANT USING IBM WATSON

A MINI PROJECT REPORT

Submitted to

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, HYDERABAD

In partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted by

THOTA PRAVEEN KUMAR	**19UK5A0531**
SAMBARAJU DINESH	**19UK5A0512**
LINGAM REVENTH	**17UK1A0582**

Under the esteemed guidance of

Ms. ANUSHA

(Assistant Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE

(Affiliated to JNTU Hyderabad & Approved by AICTE,
New Delhi) Bollikunta , Warangal – 506005

2018-2022

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE

(Affiliated to JNTU Hyderabad & Approved by AICTE, NewDelhi)

Bollikunta , Warangal – 506005

2018 – 2022

CERTIFICATE

This is to certify that the Mini Project Report entitled **" PREDICTION OF FULL LOAD ELECTRICAL POWER OUTPUT OF A BASE LOAD OPEARETED COMBINED CYCLE POWER PLANT USING IBM WATSON"** is being submitted by ***T.PRAVEENKUMAR(H.NO:19UK5A0531),S.DINESH(H.NO:19UK5A0512) , L.REVENTH (H.NO:17UK1A0582)*** in partial fulfillment of the requirements for the award of the degreeof Bachelor of Technology in Computer Science and Engineering to Jawaharlal Nehru Technological University Hyderabad during the academic year 2021-22, is a record of workcarried out by them under the guidance and supervision.

Ms.ANUSHA KUMAR — *Dr. R. NAVEEN*

Project Guide — *Head of the Department*

EXTERNAL EXAMINAR

ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr. P. PRASAD RAO,** Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this mini project in the institute.

We extend our heartfelt thanks to **Dr. R. NAVEEN KUMAR**, Head of the Departmentof CSE, Vaagdevi Engineering College for providing us necessary infrastructure andthereby giving us freedom to carry out the mini project.

We express heartfelt thanks to **Ms.SHIVANI KAPOOR**, Program Manager, SmartBridgeEducational Services Private Limited, for their constant supervision aswell as for providing necessary information regarding the mini project and for theirsupport in completing the mini project, mini project and internship.

We express heartfelt thanks to the guide, **Ms. ANUSHA** Assistant professor, Department of CSE for her constant support and giving necessaryguidance for completion of this mini project.

Finally, we express our sincere thanks and gratitude to my family members, friendsfor their encouragement and outpouring their knowledge and experience throughoutthe this.

INDEX

ABSTRACT

The utilization of renewable energy to lessen climate change and global warming has become an expanding pattern. To further develop the prediction capacity of renewable energy, different prediction techniques have been created. Predicting the full load electrical power output of a baseburden power plant is significant to amplify the benefit from the accessible megawatt-hours. Thispaper looks at and analyzes some machine learning relapse strategies to foster a prescient model, which can foresee the full hourly burden electrical power output of a combined cycle powerplant. The base burden activity of a power plant is affected by four primary boundaries, whichare utilized as info variables in the dataset, like ambient temperature, atmospheric pressure, relative humidity, and exhaust steam pressure. These boundaries influence electrical power output, which is considered the objective variable. The dataset, which comprises this informationand target variables, was gathered over six years. In light of these variables, the best subset of thedataset is explored among all component subsets in the examinations.

INTRODUCTION

1.1 Introduction

Producing electricity removed from the fuel goes through a few phases. This can be accomplished in a combined cycle power plant. This kind of innovation will create two sorts of energy electricity and steam.Consolidating the cycles will produce half more than the single cycle innovation[1]. In the first place, the gas will consume in Gas Turbine and blended in with the air that comes from the air filter. Predicting a genuine worth, known as relapse, is the most widely recognized issue investigated in machine learning. Therefore, machine learning algorithms are utilized to control the reaction of a framework for predicting anumeric or genuine esteemed objective element[2]. Some genuine issues can be tackled as relapse issues and assessed utilizing machine learning methods to foster predictive models. That blend will turn the generator and by its turn will produce the electricity. The heat lost from the gas turbine will be caught in the Heat Recovery Steam Generator (HRSG).

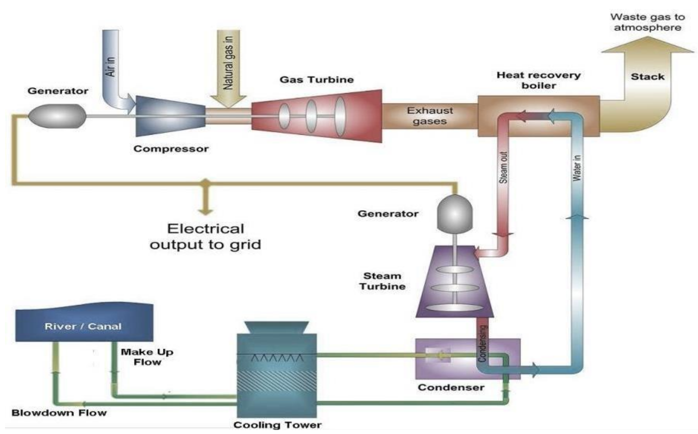

Figure 1: The combined cycle power plant

1.3 Motivation

Generating electricity extracted from the fuel goes throughseveral phases. This can be achieved in a combined cycle power plant. This type of technology will produce two types of energy electricity and steam. Combining the cycles will generate 50% more than the single cycle technology [1]. First,the gas will burn in Gas Turbine and mixed with the air that comes from the air filter. That combination will spin the generator and by its turn will generate the electricity. The heatlost from the gas turbine will be captured in the Heat RecoverySteam Generator (HRSG). An HRSG creates steam using boiled water that will spine an additional turbine generator toproduce electricity. Finally, the stem will become liquid to recital it by using a condenser [2]. Figure 1 shows the combined cycle power plant processes.

Many countries establish this technology to generate moreelectricity due to the high demand to consider the environmentand economic aspects. The main part of this plant is a gas turbine that will generate power.

The data set presented consists of four inputs temperature(AT), ambient pressure (AP), relative humidity (RH) and exhaust vacuum (V) that will be used to predict the total electric power (EP) hourly which is the output. The dataset consists of 9,568 observations captured every hour by somesensors located around the combined cycle power plant. All ofthese observations were collected over six years from 2006 to2011 [3][4].

The aim of this paper is to examine and compare four models which are Multiple linear regression, Multilayer perceptron, k-nearest neighbor, and random forest algorithm.

The remaining of the paper is structured as follows: SectionII describes the technical background. Section three discussesthe dataset and features selection. Section four presents the related work. Section five explains the model design, followed by results, and discussion in Section six. Finally, theconclusion and future work section are presented in Section seven.

2. COMBINED PINCH AND ENERGY ANALYSIS

Pinch analysis is a typical strategy for the plan of thermal cycles and is identified with the utilities. While the strength of pinch analysis has been accentuated as far as addressing the principle highlights of a framework graphically and showing focusing on data for measure alteration, the benefits of energy analysis are likewise applied in this work, in that the utilization of energy analysis permits thought of any energy frameworks[4]. The composite curve (CC) and the grand composite curve (GCC) are significant apparatuses in pinch innovation. They are exhibited utilizing temperature versus enthalpy tomahawks. The energy targets change by the CC and GCC as far as heat loads[5]. For an expanded application for heat and power frameworks, composite curves and great composite curves are created. As shown in Fig.2, the composite curves for heat exchangers and boilers can be changed over to exergy composite curves and elegant composite curves[6]. The concealed regions decide the exergy annihilation identified with the heat move measure. By consolidating pinch and energy analysis, assessment of the shaft work for both power age and refrigeration frameworks can be acquired precisely[7].

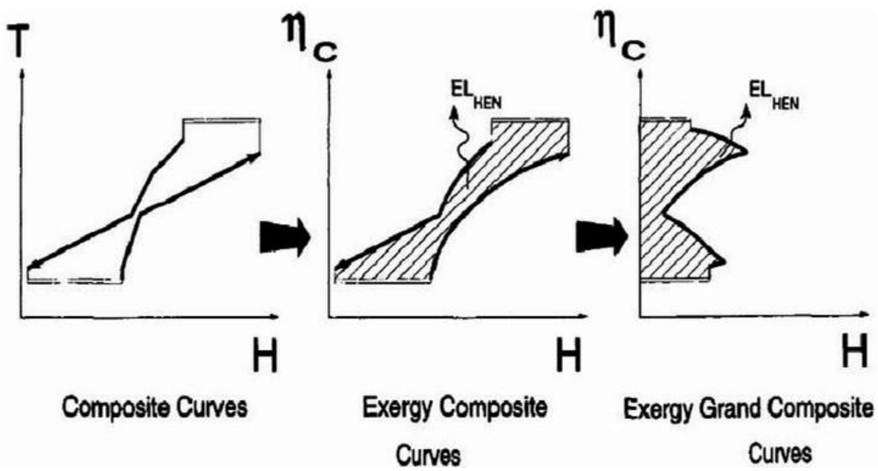

Figure : Energy transformation from CC to ECC and EGCC

$$n = 1 - \frac{T0}{T}$$

To conquer the constraint of the ηc-H graph, the energy level chart was characterized by Feng in 1997,where Ω signifies the energy level and H the measure of energy.

$$\Omega = \frac{Ene\cancel{rgy}}{\text{Energy}}$$

Both energy and energy models for an entire framework can accordingly be addressed at the same time onthis outline[8]. The energy level representation (ELR) in light of the combined pinch and energy analyses develop on the previous system of the thermodynamic way to deal with heat integration

$$\Omega = 1 - \frac{T_0}{T}$$

3. LITERATURE SURVEY

Srivastava and Yadav have examined the presence of a combined cycle utilizing a fume pressure refrigeration framework. The outcome shows that the explicit plant work of the combined cycle increments by 4 % and plant effectiveness by 0.39 %.

Najjar has examined water smelling salts assimilation chiller. More detailed improvement in straightforward cycle proficiency and power by cooling the gulf air utilizing an assimilation framework.

Mohanty and Paloso have utilized lithium-bromide, a double-effect absorption chiller, which delivered as much as 11 %additional electricity from a similar gas turbine power plant.

Bies et al. contemplated utilizing a lithium bromide double-effect absorption chiller to cool warm ambientair entering a gas turbine blower.

Dawoud et al. determined a normal increase of 19.7% in the power output of a GE Frame6B CT situated in Oman if an ingestion chiller was utilized to cool the delta air to the C

Yokoyama and Ito researched how gas turbine cogeneration plants' unit size and cost are influenced by bay air cooling with ice stockpiling. They analyzed limits with and without air cooling and decided on yearly expenses.

4. TECHNICAL BACKGROUNG

4.1 Multiple linear Regression.

Multiple linear regression is a simple linear regression with more than one input. It is an algorithm that represents the prediction of linear relationship between multiple inputs and one output to find the best fit structure of the data based on thenumber of the inputs [5]–[8]. For example, finding the best plane if the inputs is two independent variable [9]. The formula of multiple linear regression is:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + e_i, \quad i = 1, 2, \ldots, n \qquad (1)$$

y = output or dependent variable, xi = inputs or predictor variables, β_0 =

constant value (intercept) .

βn = coefficient value for each input in the equation.

ei = residuals or error of the model.

4.2 Multilayer Perceptron.

Multilayer perceptron is a type of neural networks that usesmore than one perceptron to predict the output or the signals from a given input [10]–[12]. Each node represents a functionit could be a step function or any other. Taking the four layersexample the first layer is the inputs the last layer is the outputsand the inner layers are the hidden layers as shown in figure 2[13]. In the perceptron algorithm, the important parameters areweight and bias. The weight changes whenever there is an error. For the training model, there are three steps. First, the model will pass the input throw the layers and that called Forward pass. Then, it will calculate the error between the desired output and the model output. Finally, the modelupdates the weight to match the desired one and that called backward pass [14]. Whenever there are more hidden layers the training model will be faster [13].

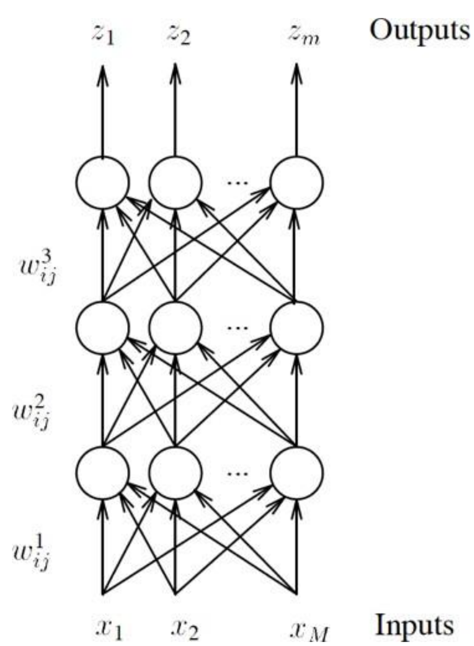

Figure : Multilayer Perceptron .

4.3 K-Nearest Neighbors

K-NN is a simple machine learning technique commonly used in classification problem. KNN does not build the model from the training data, that is why it is classified under lazy category [15]. KNN determines the input's class based on the number of neighbors given which is K [4]. For example, if the k = 3 and we have two classes A and B and we want to define the class of input x. First, KNN algorithm will defined the region of 3 nearest neighbors of x. the most repeated class will identify the class of x. Figure 3 shows how increasing the value of K can change the class of the input [16].

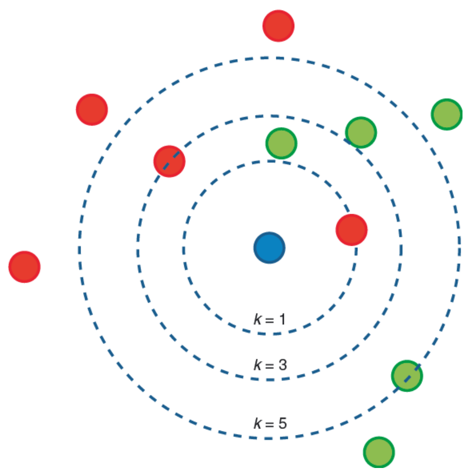

Figure 3 K-NN example.

4.3 Random Forest Algorithm

This algorithm commonly used for classification but there is an application for it in the regression filed. The random forest algorithm is a learning algorithm consists of large numbers of independent decision trees [17]–[19]. For each tree, there is a prediction for the class or the output. In each intermediate node there is a condition to specify which leaf node. The most predicted class in the decision is the output model [20]. See the figure below (a), (b), (c) and (e) trees predict the red class while the (d) tree predicts the green class. Thus, the final class is red. This prediction process goes for every single input [16].

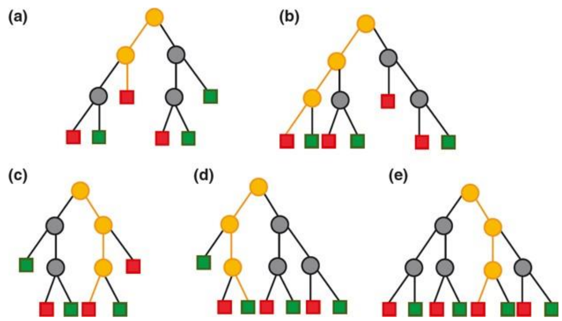

Figure : Random forest example..

5. DATASET PREPROCESSING AND FEATURE SELECTIONS

5.1 Dataset Preprocessing.

The Dataset discusses the electrical power that generates from the gas turbine which will affect the economy of the country. It is consisting of five attributes ambient temperature(AT), ambient pressure (AP), relative humidity (RH), exhaust vacuum (V) and electric power (EP). All these attributes related to the gas turbine situation. ambient temperature, ambient pressure, relative humidity and exhaust vacuum are inputs, and the electric power is output. There are no missing values or rows in the dataset. Each attribute shows its effects on the output so we can say all attributes or features are important.

The description for each column in the dataset:

- Ambient temperature: the temperate of the gas turbine in Celsius.
- Ambient pressure: the pressure of the gas in millibar.
- Relative humidity: ratio between the highest humidity and the current humidity in percentage.
- Exhaust vacuum: the tall of the exhaust vacuum in cm.
- Electric power: the total energy created by gas turbine in MW.

5.2 Outliers.

There are no outliers in the dependant value of the model.

Thus, there is no need to remove any value in the dataset.

5.3 Class of Attributes.

For the regression problems we have to check the type ofeach feature or attribute and we have to convert the nominal variable to dummy variable for regression problem. The typeis numeric for all the dependent variable and the independentvariables Also, there is no need to convert any nominal valueto dummy variable and there is no data matrix for the data.

5.4 Descriptive Summary and Normality.

According to Table 1, The mean of the dependent variableis 454.365. Median equals 451.55. Mode equals 468.8. The minimum value in the dependent variable is 420.26 and the maximum is 495.76. The skewness is 0.3065 which means thedependent variable distribution is approximately symmetric because it located between 0.5 and -0.5. In addition, the

Table 1 Descriptive summary of the dependent variable.

Parameters	**Values**
Mean	454.365
Median	451.550
Mode	468.800
Minimum	420.260
Maximum	495.760

kurtosis value is -1.049 which means the distribution is not normally distributed it should be 0 to be normally distributed.Q1 is 439.8, Q3 is 468.4 and the interquartile range (IQR) is 28.6 from:

$$IQR= Q3 - Q1 \qquad (2)$$

For the right outliers they should be greater than 551.3. The left outliers should be less than 396.9. Applying the following equations:

$$\text{Right outlier} > Q3 +1.5(IQR) \qquad (3)$$

$$\text{Left outlier} < Q1 - 1.5(IQR) \quad (4)$$

From the above results, that is why we do not have any outliers. Moreover, the variance is 291.3 that means the data not spread out from the mean. The standard deviation is 17.1this means there is no problem in the range of the data and thesize of sample is good.

Applying Anderson-Darling normality test shows us that the dependent variable not normally distributed because the p value less than 0.05 thus, we reject the null hypothesis which is the normality. Figure 5 shows the histogram which shows that data is not normally distributed.

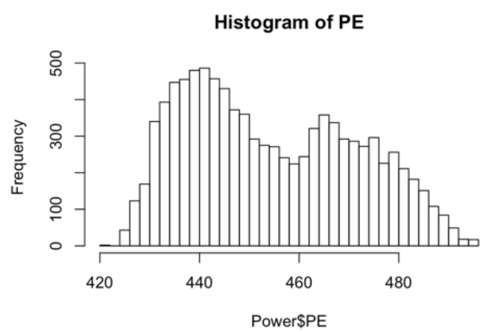

Figure : Histogram of the electric power variable .

5.4 Feature Selections

Applying the stepwise selection function shows that all variables are important for the machine learning prosses because there was no discard or deletion for any variable. Moreover, the ols_ste_all_possible function shows that all variables are important for the prediction process. This function calculates the mean absolute error, R square, and more for each regression subset. The smallest MAE is 156880.41 for the subset that has all the variables. Also, it hasthe largest value for R square which is 0.9298 which means using this subset can predict more accurate results. Other Methods in Weka shows that deleted any variables increases mean absolute error even if the evaluator shows that one of the variables could be removed. Thus, there is no need to delete any variable.

6. RELATED WORK

There are several previous studies about predicting the power of combined cycle power plant using machine learning methods. One study done by Tüfekci [22] examined and compared fifteen regression methods including simple linear regression, multiple linear regression and REPTree. The comparison based on the mean absolute error and a RootMean-Squared Error. The best method was the Bagging method with REPTree algorithm with a mean absolute error of 2.818 and a Root Mean Squared Error of 3.787.

Moreover, Tüfekci represented a study predict the power of combined gas by comparing machine learning methods. conventional multivariate regression, additive regression, k- NN, feedforward ANN, and K-Means clustering were used to generate local and general forms. KNN best for a fine-tuned dataset. ANN yielded good results in speed and memory tests[4].

Islikaye and Cetin [23] used seven machine language methods to predict the total power of a combined cycle power plan. The best results were for K-NN, Linear Regression and RANSAC regressions.

Another study used Only ANN with random subset and different hidden layers to enhance the prediction of net power[2].

7. MODEL DESIGN

For each algorithm that I will use in this comparison, the data will be split into 80% training and 20% testing using cross-validation technique with K=10 means the training will tack 9 datasets and one dataset for the testing. Thus, there will be 9568 for the training and 1914 for the testing beads on Weka. The variance inflation factor is under 10 for all variable that means the is no multicollinearity problem exists. temperature variable has 5.92 vif, Exhaust Vacuum variable has 3.88 vif, Ambient Pressure has 1.46 vif and, Relative Humidity has 1.70 vif.

The R square of the model is 0.9298 which means that only 90% of model can be predicted. Moreover, the p-value of the model less than 0.05 as figure 13 shown. Thus, we reject the null hypothesis. That means the power is affecting by all the other attributes and it is statically significant. All the information above from the MLR model.

8. COMPUTATIONAL APPROACH

A computer program has been created for energy, combined pinch-energy, energy-economic, energy destruction, and energy destruction–level examinations of the 423-MW combined cycle and 315- MWsteam gas-terminated power plants considered here. With this info information, the adiabatic flame temperature, the number of moles of burning items and the stream paces of enthalpy (MW) and entropy (MW=K) are determined[9]. Then, at that point, the net stream paces of different energies and entropies, the energy efficiencies of the hardware, and the energy deficit rate for every segment are assessed. ELR and ECDL are built dependent on the after-effects of energy and pinch examinations. Then, at that point, the plan and execution assessment dependent on heat and pressure should be possible, permitting the presentation of hardware with pressure impacts like the turbine to be assessed all the more precisely.

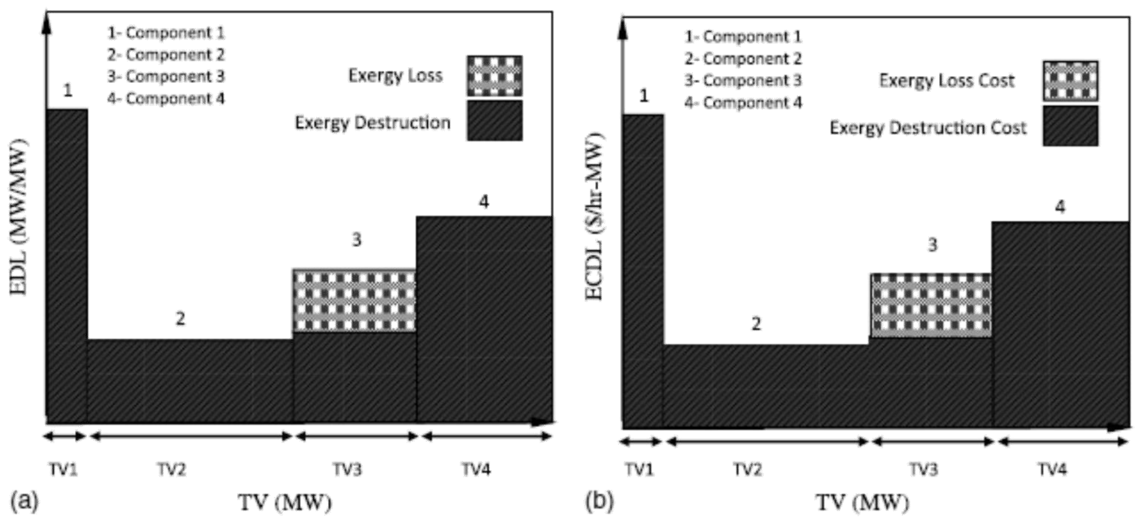

Figure 4: EDL/ECDL Representation :(a) EDL; (b) ECDL

The heat move rate to or from every hardware is resolved to fulfil the energy and energy balances for the part. Once energy balances for the parts, intersections and the plant limit are set up, the unit cost of different energies and items is determined by settling the expense balance conditions[10]. Also, the program can plot combined pinch-energy representations for every segment dependent on energy level boundaries, just as EDL and ECDL[11].

9. RESULTS AND FORECASTING PERFORMANCE

Altogether, 41 kinds of estimations of anticipating exactness were assembled in this investigation—table 1 records estimations of gauging exactness utilized by more than ten examinations. Mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) are three estimations most habitually utilized[12]. Renewable energy can be addressed in enhanced units, and the upsides of renewable energy vary a ton for various examinations[13]. To stay away from the impacts of units and upsides of renewable energy, MAPE is determined to portray gauging precision.

Sources of Energy / Measure-ments	Solar	Wind	Hydro-Power	Biomass	Geothermal	Wave	Tidal	Total
MAE	17	26	1	3	1	2	0	50
MAPE	12	23	0	1	0	0	3	39
RMSE	36	31	1	4	0	2	3	77
R^2	9	8	1	3	1	1	0	23
NRMSE	9	4	0	0	0	0	0	13
MSE	2	8	0	1	1	0	0	12

Table: 1 Measurement of forecasting accuracy used by more than ten studies

By and large, numerous models were presented in each investigation. Subsequently, Table 2 shows the best presentation bring about by each study.MAPE esteems under 10% are exceptionally precise predictions[14]. Subsequently, most estimating correctness's of gathered investigations are high as far as MAPE.

Sources of Energy	MAPE (%)/References	Average MAPE (%)
Solar	17.72,2.5,54,7.43,0.22,2.78	9.01867
Wind	8.1082, 1.66, 3.38,6.530,3.871, 17.1076	5.75465
Tidal	0.9743, 2.048	3.784
Biomass	3.783	1.5842

Table 2: Mape Values In Energy Prediction

Moreover, the coefficient of assurance (R2) is another estimation determined in this investigation for analysis[15]. The assurance coefficient addresses the extent of the fluctuation in the reliant variable that islogical by the independent variables T. It tends to be seen that most upsides of R2 are bigger than 0.8. Moreover, eight articles utilized R2 and MAPE at the same time as gauging execution estimations.

10. CONCLUSION

Because of concerns brought about by climate change and global warming as of late, renewable energy is blasting. Hence, precise prediction of renewable energy power is significant, and many related approacheshave been directed. Energy analysis also regularly includes the assurance of proportions of execution: energy destruction proportions, energy misfortune proportions, and energy efficiencies. In this, such proportions of execution are thought of. The CCPP, where the dataset is provided for this examination,has begun to utilize this created predictive model for the following day's hourly energy output. First, the applications of machine-learning techniques to renewable energy have been expanding, and the employments of artificial intelligence techniques and mixed-race models in solar-energy and wind-energy predictions are the larger part. The Combined Cycle Power Prediction, where the dataset is provided for this approach, has utilized this created predictive model for the following day's hourly energy output.

11. FUTURE WORK

The aim of this paper was to determine the best machine learning model to predict the total electric power that could begenerated by gas turbine. The random forest algorithm showsthat it is the best model by having the least error.

For the future work, other algorithm could be used for thecomparison such as Support Vector Machines (SVM), K starand Decision stump.

12. REFERENCES

[1]. H. J. Koo, S. Lim, J. Choe, S.-H. Choi, H. Sung, and K.-H. Do, “Radiographic and CT features of viral pneumonia,” Radio- Graphics, vol. 38, no. 3, pp. 719–739, 2018.

[2]. T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, “Automated detection of COVID- 19 cases using deep neural networks with X-ray images,” Computers in Biology and Medicine, vol. 121, Article ID 103792, 2020.

[3]. Vishal Dineshkumar Soni. (2018). IOT BASED PARKING LOT. International Engineering Journal For Research & Development, 3(1), 9. https://doi.org/10.17605/OSF.IO/9GSAR

[4]. I. Ahmad and K. Pothuganti, "Smart Field Monitoring using ToxTrac: A Cyber-Physical System Approach in Agriculture," 2020 International Conference on Smart Electronics and Communication (ICOSEC), 2020, pp. 723-727, doi: 10.1109/ICOSEC49089.2020.9215282.

[5]. sridevi Balne, Anupriya Elumalai, Machine learning and deep learning algorithms used to diagnosis of Alzheimer’s: Review, Materials Today: Proceedings, 2021, https://doi.org/10.1016/j.matpr.2021.05.499.

[6]. B. Sekeroglu and I. Ozsahin, “Detection of COVID-19 from chest X-ray images using convolutional neural networks,” SLAS TECHNOLOGY: Translating Life Sciences Innovation, vol. 25, no. 6, pp. 553–565, 2020.

[7]. V. D. Soni and A. N. Soni , “Cervical cancer diagnosis using convolution neural network with conditional random field, ” 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021,pp 1746-1751.

[8]. T. Gao, “Chest X-ray image analysis and classification for COVID- 19 pneumonia detection using Deep CNN,” 2020, https://www.medrxiv.org/content/10.1101/2020.08.20.20178913v2.

[9]. Ankit Narendrakumar Soni (2019). Crack Detection in buildings using convolutional neural Network. JOURNAL FOR INNOVATIVE DEVELOPMENT IN PHARMACEUTICAL AND TECHNICAL SCIENCE, 2(6), 54-59.

[10]. T. Zebin and S. Rezvy, “COVID-19 detection and disease progression visualization: deep learning on chestX-rays for classification and coarse localization,” Applied Intelligence, 2020.

[11]. Ankit Narendrakumar Soni (2018). Application and Analysis of Transfer Learning-

Survey. InternationalJournal of Scientific Research and Engineering Development, 1(2), 272-278.

[12]. R Alugubelli. (2016). Exploratory Study of Art ificial Intelligence in Healthcare. Internat ional Journal ofInnovat ions in Engineering Research and Technology, 3(1), 1–10.

[13]. A. Sedik, M. Hammad, F. E. Abd El-Samie et al., “Efficient deep learning approach for augmenteddetection of Coronavirus disease,” Neural Computing & Applications, 2021.

[14]. Roopha Shree Kollolu Srinivasa (2020), INFRASTRUCTURAL CONSTRAINTS OF CLOUD COMPUTING, International Journal of Management, Technology and Engineering, Volume X, Issue XII, DECEMBER/2020, Page No: 255-260.

[15]. Ankit Narendrakumar Soni (2019). Spatical Context Based Satellite Image Classification-Review.

International Journal of Scientific Research and Engineering Development, 2(6), 861-868.