

**Reg no:39110505**

**Name: KODATI NAGA SATYA SAI MANIKANTA**

# **Prediction Of Full Load Electrical Power Output Of A Base Load Operated Combined Cycle Power Plant Using Machine Learning**

## **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 base burden power plant is significant to amplify the benefit from the accessible megawatt-hours. This paper 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 power plant. The base burden activity of a power plant is affected by four primary boundaries, which are 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 information and target variables, was gathered over six years. In light of these variables, the best subset of the dataset is explored among all component subsets in the examinations.

## **1.introduction**

### **1.1 overview**

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 a numeric 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).

## **1.2 Purpose**

Aim of this work is to apply and experiment various options.effects on feed-forward artificial neural network (ANN) which used to obtain regression model that predicts electrical output power (EP) of combined cycle power plant based on 4 inputs. More specifically, this work uses MATLAB neural networks toolbox to study stochastic behavior of the regression neural, effect of number of neurons of the hidden layers, effect of data subset size for training, effect of number of variables as input, different training functions results, data preprocessing, and statistical errors.

## **2.Lirerature survey**

### **2.1 Existing problem**

Machine learning is the process of equipping computers with the ability to learn by using data and experience like a human brain. The main aim of machine learning is to create models, which can train themselves to improve, perceive complex patterns, and find solutions/predictions to new problems by using the previous data .

Machine learning, more specifically the field of predictive modeling, is primarily concerned with minimizing the error of a model or making the most accurate predictions possible, at the expense of explainability. As such, linear

regression was developed in the field of statistics and is studied as a model for understanding the relationship between input and output numerical variables but has been borrowed by machine learning. It is both a statistical algorithm and a machine learning algorithm .

The concept of linear regression was first proposed in 1894. Linear regression is a statistical test applied to a data set to define and quantify the relation between the considered variables. It is a modeling technique where a dependent variable is predicted based on one or more independent variables. Linear regression analysis is the most widely used of all statistical techniques .

Dimensionality reduction is used in machine learning to avoid the curse of dimensionality and to convert the system from high to low dimension without sacrificing the important information in features. Ideally, the reduced representation should have a dimensionality that corresponds to the intrinsic dimensionality of the data . The dimensionality reduction can be performed either by manually selecting the required features or by using specific techniques that reduce the system's dimension . One of these techniques is Principal Component Analysis (PCA). PCA is a mathematical method that uses algorithms to reduce dimensions in a high-dimensionality system to a low dimension while keeping the maximum number of variations in the resulted features . PCA works by finding directions with the highest variation of data; called principal components where working with these reduced features is much more efficient than modeling with thousands of numbers for each sample .

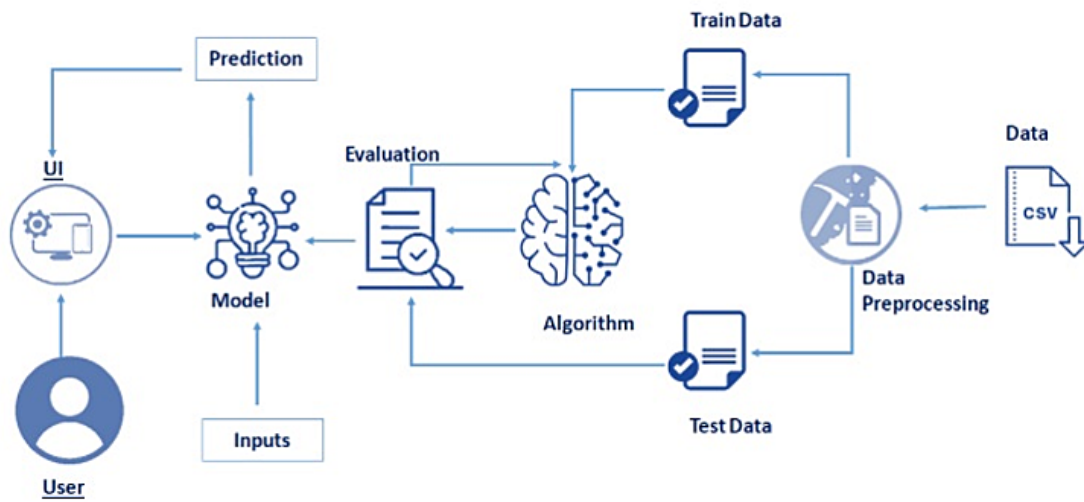
## **2.2 Proposed Solution**

In order to find accurate and efficient ways of predicting hourly electrical energy output, the researchers in utilized a dataset collected over 6 years whose data points corresponded to average hourly sensor measurements when the plant was set to work with full load. The input features were ambient temperature, relative humidity and ambient pressure, which are known to be major factors in gas turbines, as well as exhaust vacuum measured from the steam turbine. They utilized conventional multivariate regression, additive regression, k-NN, feed-forward ANN and K-Means clustering to form local and global predictive models. They found that even with simple regression tools such as k-NN smoother, it is

possible to predict net yield with less than 1 % relative error on average. Using more sophisticated tools and proper preprocessing it is possible to significantly increase the performance. The research in explained the used methodology by: first, based on the input variables, the best subset of the dataset is explored among all feature subsets in the experiments. Then, the most successful machine learning regression method is sought for predicting full load electrical power output. Thus, the best performance of the best subset, which contains a complete set of input variables, has been observed using the most successful method with the best mean absolute error and root-mean-squared error.

### 3.Theoretical Analysis

#### 3.1 Block diagram



#### 3.2 Hardware / Software Designing

- Processor – [Intel Xeon E2630 v4](#) – 10 core processor, 2.2 GHz with Turboboost upto 3.1 GHz. 25 MB Cache
- Motherboard – ASRock [EPC612D8A](#)
- RAM – 128 GB DDR4 2133 MHz
- 2 TB Hard Disk (7200 RPM) + 512 GB SSD

- GPU – [NVidia TitanX Pascal](#) (12 GB VRAM)
- Intel Heatsink to keep temperature under control Storm Trooper Cabinet

## 4.Experimental Investigations

Exploratory data analysis is an approach to analyzing data sets to summarize their main characteristics, often with visual methods and used for determine how best to manipulate data sources to get the answers you need, making it easier for data scientists to discover patterns, spot anomalies, test a hypothesis, or check assumptions.

**head():**To check the first five rows of the dataset,we have a function called **head()**.

Head() method is used to return top n (5 by default) rows of a DataFrame or series.

**Tail():** To check the last five rows of the dataset, we have a function called **tail()**.

## Understanding Data Type and Summary of features

How the information is stored in a DataFrame or Python object affects what we can do with it and the outputs of calculations as well. There are two main types of data those are numeric and text data types.

- Numeric data types include integers and floats.
- Text data type is known as Strings in Python, or Objects in Pandas. Strings can contain numbers and / or characters.
- or example, a string might be a word, a sentence, or several sentences.
- There is no categorical data present in our dataset. But it is not necessary that all the continuous data which we are seeing has to be continuous in nature. There may be a case that some categorical

data is in the form of numbers but when we perform `info()` operation we will get numerical output. So, we need to take care of those type of data also.

**describe():** functions are used to compute values like count, mean, standard deviation and IQR(Inter Quantile Ranges) and give a summary of numeric type data.

## Checking For Null Values

1. After loading it is important to check the complete information of data as it can indication many of the hidden information such as null values in a column or a row

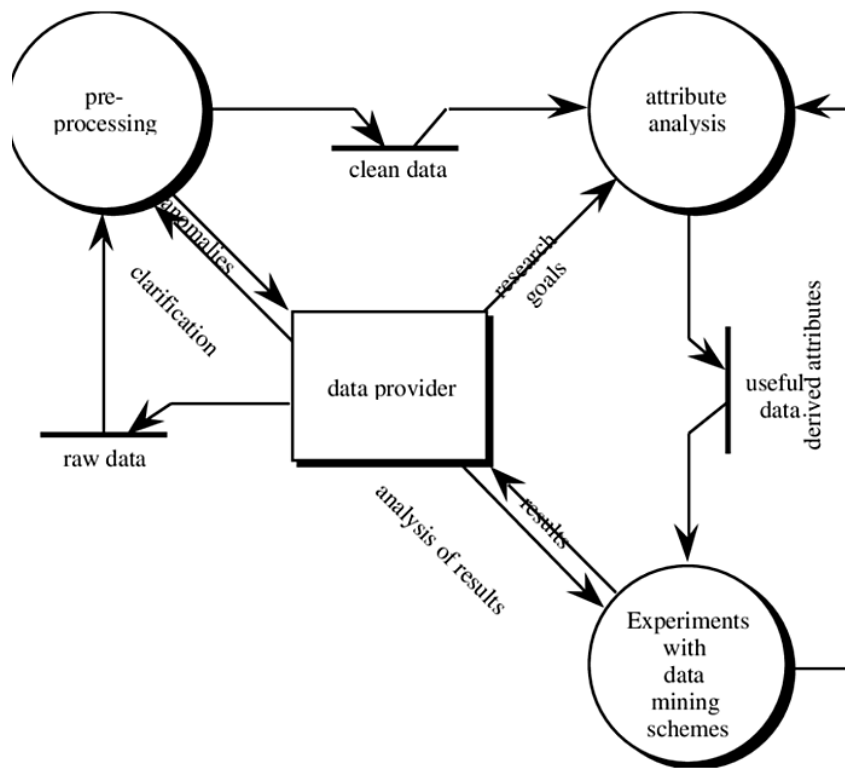
2. Check whether any null values are there or not. if it is present then following can be done,

a. Imputing data using Imputation method in sklearn

b. Filling NaN values with mean, median and mode using `fillna()` method.

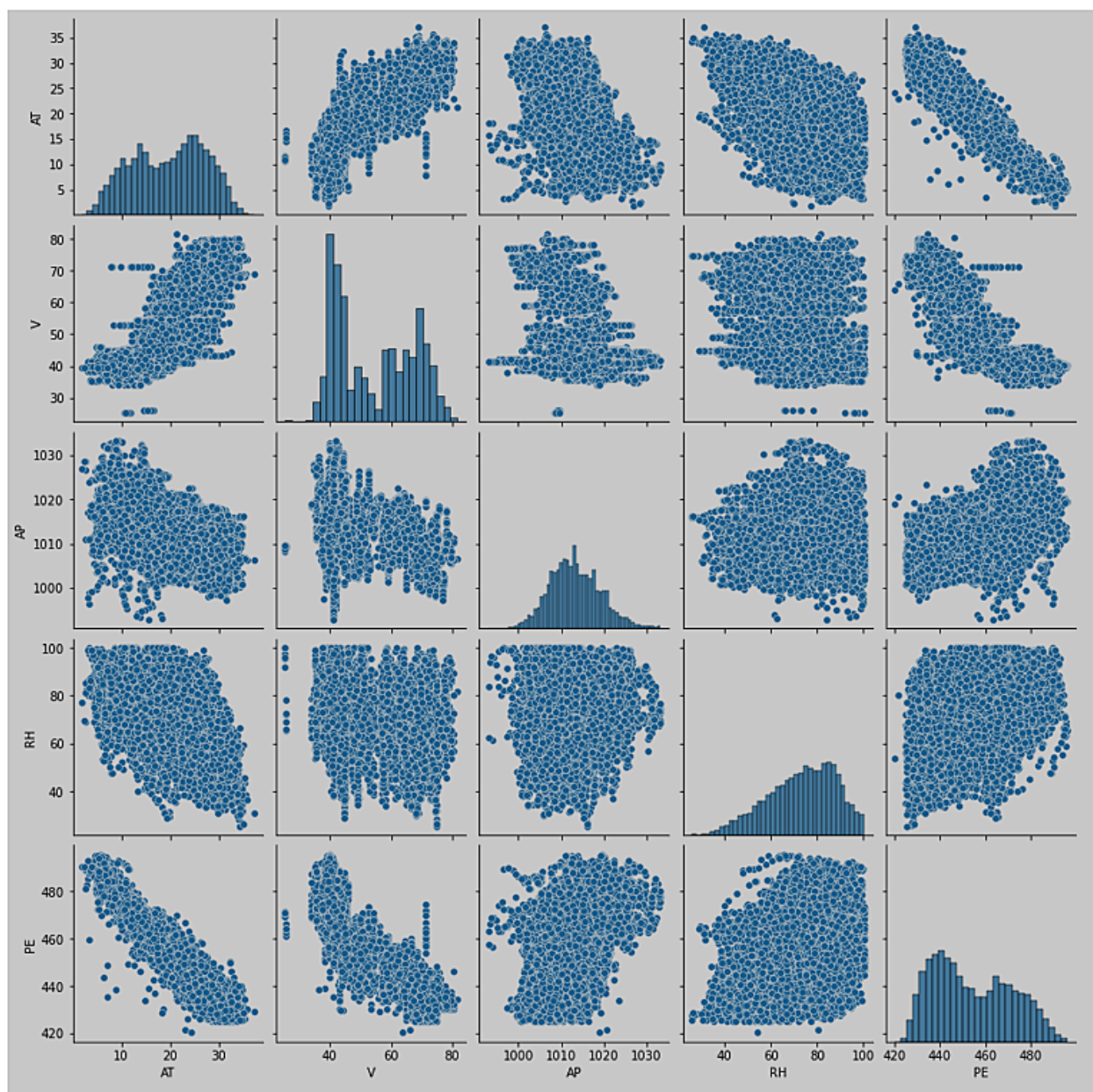
We will be using `isnull().sum()` method to see which column has missing values by counting the total sum of null in each column.

## 5. Control Flow



## 6.Result

Exploratory data analysis (EDA) Exploratory data analysis resulted in clean data by the means of no missing values, outliers or duplications in our dataset according to the output of our code. This pathed the way for the next objective of modeling. The scatter matrix in an obvious linear relationship between AT and PE. The PE vs AT relationship is enlarged for a better visuality , a direct and linear relationship exists between PE and AT.





### Introduction

The Combined Cycle Power Plant or combined cycle gas turbine, a gas turbine generator generates electricity and waste heat is used to make steam to generate additional electricity via a steam turbine. The gas turbine is one of the most efficient one for the conversion of gas fuels to mechanical power or electricity. The use of distillate liquid fuels, usually diesel, is also common as alternate fuels.

More recently, as simple cycle efficiencies have improved and as natural gas prices have fallen, gas turbines have been more widely adopted for base load power generation, especially in combined cycle mode, where waste heat is recovered in waste heat boilers, and the steam used to produce additional electricity.

The basic principle of the Combined Cycle is simple: burning gas in a gas turbine (GT) produces not only power, which can be converted to electric power by a coupled generator, but also fairly hot exhaust gases. Routing these gases through a water-cooled heat exchanger produces steam, which can be turned into electric power with a coupled steam turbine and generator.

## PREDICTION OF ELECTRICAL OUTPUT POWER OF COMBINED CYCLE POWER PLANT

Ambient Temperature(AT):

Exhaust Vacuum(V):

Ambient Pressure(AP):

Relative Humidity(RH):

## PREDICTION OF ELECTRICAL OUTPUT POWER OF COMBINED CYCLE POWER PLANT

Prediction of Electrical Output is: 485.4654999999997

## 7.Advantages And Disadvantages

### Advantages:

1. Easily identifies trends and patterns

Machine Learning can review large volumes of data and discover specific trends and patterns that would not be apparent to humans. For instance, for an e-commerce website like Amazon, it serves to understand the browsing behaviors and purchase histories of its users to help cater to the right products, deals, and reminders relevant to them. It uses the results to reveal relevant advertisements to them.

## **2. No human intervention needed (automation)**

With ML, you don't need to babysit your project every step of the way. Since it means giving machines the ability to learn, it lets them make predictions and also improve the algorithms on their own. A common example of this is anti-virus softwares; they learn to filter new threats as they are recognized. ML is also good at recognizing spam.

## **3. Continuous Improvement**

## **4. Handling multi-dimensional and multi-variety data**

## **5. Wide Applications**

## **Disadvantages**

### **1. Data Acquisition**

Machine Learning requires massive data sets to train on, and these should be inclusive/unbiased, and of good quality. There can also be times where they must wait for new data to be generated.

### **2. Time and Resources**

ML needs enough time to let the algorithms learn and develop enough to fulfill their purpose with a considerable amount of accuracy and relevancy. It also needs massive resources to function.

### **3. Interpretation of Results**

## 4. High error-susceptibility

## 8.Applications

### 1. Image Recognition:

Image recognition is one of the most common applications of machine learning. It is used to identify objects, persons, places, digital images, etc. The popular use case of image recognition and face detection is, **Automatic friend tagging suggestion**:

Facebook provides us a feature of auto friend tagging suggestion. Whenever we upload a photo with our Facebook friends, then we automatically get a tagging suggestion with name, and the technology behind this is machine learning's **face detection** and **recognition algorithm**.

It is based on the Facebook project named "**Deep Face**," which is responsible for face recognition and person identification in the picture.

### 2. Speech Recognition

While using Google, we get an option of "**Search by voice**," it comes under speech recognition, and it's a popular application of machine learning.

Speech recognition is a process of converting voice instructions into text, and it is also known as "**Speech to text**", or "**Computer speech recognition**." At present, machine learning algorithms are widely used by various applications of speech recognition. **Google assistant**, **Siri**, **Cortana**, and **Alexa** are using speech recognition technology to follow the voice instructions.

### 3. Traffic prediction:

If we want to visit a new place, we take help of Google Maps, which shows us the correct path with the shortest route and predicts the traffic conditions.

It predicts the traffic conditions such as whether traffic is cleared, slow-moving, or heavily congested with the help of two ways:

- **Real Time location** of the vehicle from Google Map app and sensors
- **Average time has taken** on past days at the same time.

Everyone who is using Google Map is helping this app to make it better. It takes information from the user and sends back to its database to improve the performance.

#### **4. Self-driving cars:**

One of the most exciting applications of machine learning is self-driving cars. Machine learning plays a significant role in self-driving cars. Tesla, the most popular car manufacturing company is working on self-driving car. It is using unsupervised learning method to train the car models to detect people and objects while driving.

#### **5. Medical Diagnosis:**

In medical science, machine learning is used for diseases diagnoses. With this, medical technology is growing very fast and able to build 3D models that can predict the exact position of lesions in the brain.

It helps in finding brain tumors and other brain-related diseases easily.

## **9. Conclusion**

At the beginning of this article, we set out to develop a predictive model for full-load output power (PE) based on the dataset provided. We explored the dataset to find out if we had missing values or other problems, then played around with 4 features subset selections on 3 different machine learning regression algorithms.

We were able to discover that using a complete set of parameters or features on the Random Forest Regression algorithm yielded the best results.

## **10.Future Scope**

AT and V have a strong negative linear relations to the output PE, AP and RH have weak positive linear relations to the output PE. This can be further shown with the correlation coefficients between inputs and the output (PE) .

The linear relations strength between the variables are of correlation R and correlation strength  $R^2$ . It can be seen that AT and V are strongly linearly related to each other and to the output PE, while AP and RH have weak linear relations to all other variables and output.

Although it is obvious that the governing variables are AT and V, the effect of absence and presence of each of variables is studied.

## **11.Bibilography**

1. Hoang, T.-D., Pawluskiewicz, D. K. (2016). The efficiency analysis of different combined cycle power plants based on the impact of selected parameters. International Journal of Smart Grid and Clean Energy, 5 (2), 77–85. doi: <https://doi.org/10.12720/sgce.5.2.77-85>.

2. Combined cycle power plant: how it works. Available at: <https://www.ge.com/gas-power/resources/education/combined-cyclepower-plants>.

3. Tüfekci, P. (2014). Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. International Journal of Electrical Power & Energy Systems, 60, 126–140. doi: <https://doi.org/10.1016/j.ijepes.2014.02.027>

4. Moayedi, H., Mosavi, A. (2021). Electrical Power Prediction through a Combination of Multilayer Perceptron with Water Cycle Ant Lion and Satin Bowerbird Searching Optimizers. Sustainability, 13 (4), 2336. doi: <https://doi.org/10.3390/su13042336>
5. Sholahudin, S., Han, H. (2015). Heating Load Predictions using The Static Neural Networks Method. International Journal of Technology, 6 (6), 946. doi: <https://doi.org/10.14716/ijtech.v6i6.1902>
6. Dehghani Samani, A. (2018). Combined cycle power plant with indirect dry cooling tower forecasting using artificial neural network. Decision Science Letters, 7, 131–142. doi: <https://doi.org/10.5267/j.dsl.2017.6.004>
7. Çelik, Ö. (2018). A Research on Machine Learning Methods and Its Applications. Journal of Educational Technology and Online Learning, 1 (3), 25–40. doi: <https://doi.org/10.31681/jetol.457046>
8. Brownlee, J. (2016). Linear Regression for Machine Learning. Machine Learning Algorithms. Available at: [https:// machinelearningmastery.com/linear-regression-for-machine-learning/](https://machinelearningmastery.com/linear-regression-for-machine-learning/)
9. Kumari, K., Yadav, S. (2018). Linear regression analysis study. Journal of the Practice of Cardiovascular Sciences, 4 (1), 33. doi: [https://doi.org/10.4103/jpcs.jpcs\\_8\\_18](https://doi.org/10.4103/jpcs.jpcs_8_18)

## Appendix

```
from flask import Flask, render_template, request # Flask is a application
# used to run/serve our application
# request is used to access the file which is uploaded by the user in our application
# render_template is used for rendering the html pages
import pickle # pickle is used for serializing and de-serializing Python object
structures

app=Flask(__name__) # our flask app

@app.route('/') # rendering the html template
def home():
    return render_template('home.html')

@app.route('/predict') # rendering the html template
def index() :
    return render_template("index.html")

@app.route('/data_predict', methods=['POST']) # route for our prediction
def predict():
    at = request.form['at'] # requesting for at data
    v = request.form['v'] # requesting for v data
    ap = request.form['ap'] # requesting for ap data
    rh = request.form['rh'] # requesting for rh data

    # converting data into float format
```

```
data = [[float(at), float(v), float(ap), float(rh)]]
```

```
# loading model which we saved
```

```
model = pickle.load(open(r'D:/Flask app/Flask app/CCPP.pkl', 'rb'))
```

```
prediction= model.predict(data)[0]
```

```
return render_template('predict.html', prediction=prediction)
```

```
if __name__ == '__main__':
```

```
    app.run(port=5050,debug=False)
```









