Department of Computer and Electrical Engineering

North South University

Junior Design Project

Predicting the Power Output of a Combined Cycle Power Plant

Sakib Ahmed	ID#1712442042
Talha Yamin	ID#1712731642

Submitted to :

Dr. Tanzilur Rahman

Assistant Professor

Department of ECE

Spring 2020

ABSTRACT

We are at a stage in this world where the amount of primary natural resources being used to generate electricity, such as coal oil and gas, is slowly running out. Though there are alternatives these will still be the main sources of fuel for the future in Bangladesh along with most of the world. As such our current goal should be to find out ways to tackle this predicament. One such way that we discuss in this paper is to maximize the efficient electrical power output by using Combined Cycle Power Plants (CCPP) instead of Single Cycle ones as CCPP is able to effectively generate up to 50% more power from the same amount of fuel. It does so by converting waste exhaust heat into additional electrical power output. To make the best use of a CCPP's power output we propose exploring and discussing different prediction models using machine learning regression methods and finding out which model(s) are best capable of determining the effective electrical power output of the plant. The electrical output during operation under full load is affected by four main surrounding atmospheric parameters: Ambient Temperature (AT), Ambient Pressure (AP), Relative Humidity (RH) and Exhaust Vacuum (V). Our aim would be to use these parameters and find out how it influences the power output. The dataset used consists of this information which was gathered over a six-year time span. We analyze the dataset and investigate the best conditions using machine learning and data visualization techniques to estimate the electrical yield. We then compare our results with that of existing literature and see how far the accuracy we achieve has come along with discussing different methods other than machine learning.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Overview

In this chapter, we delve into the history of single cycle power plants and how they are quickly becoming antiquated pieces of technology in the face of newer and more useful alternatives such as the combined cycle power plant. We further give a brief explanation of machine learning and what role it plays in getting the most out of the power output of a combined cycle power plant.

1.2 Single Cycle Power Plants - End of an era

In today's world, almost everything runs on electricity and the demand for it could not be any lower. We know most power plants use natural resources, such as coal, oil, etc as fuel but these resources are finite and are running out every day. So it is our primary goal to save and utilize these resources to the best of our ability and that is where our proposed idea comes in. We propose using Combined Cycle Power Plants over traditional Single-cycle ones in Bangladesh. We discuss their benefits and how machine learning models can be used to predict the best conditions for the maximum power output. This knowledge would allow us to allocate resources accordingly to maximize efficiency as well as profit. If our models can make accurate predictions we will be able to increase the efficiency of the combined cycle power plant by up to 50%. Thus we would be getting a higher amount of electricity for the same amount of fuel burned. This paper deals with several regression models of machine learning that would be apt for such predictions.

1.3 Why use Combined Cycle Power Plants?

In Bangladesh, single cycle power plants are the norm and though they have done the job till now it is just not feasible anymore when thinking about their sustainability in the near future. Not only is it inefficient in power generation compared to a combined cycle power plant but it also emits enormous amounts of waste gases that are high in temperature, affecting the environment in general. [1] So both fuel usage and environment safety are at risk from continued usage of single cycle power plants, not to mention the fact that fuel usage is becoming more precious as time passes on and the air quality in Bangladesh is one of the worst. [3] Using single cycle power plants will only help to perpetuate that problem.

1.4 How Combined Cycle Power Plants work

A combined cycle power plant functionally works much like a single cycle power plant however it is structurally much more different and as a result, offers enhanced power generation for the same amount of fuel burned. Combined cycle power plants have two synchronous A/C generators as opposed to the single synchronous A/C generator in single cycle power plants as their names suggest. The second generator takes in the exhaust gases as a fuel input rather than discarding them as waste. It is able to do this using a Heat Recovery Steam Generator (HRSG). The HSRG captures excess heat that would have escaped as waste. The HRSG creates steam from the gas turbine exhaust heat and delivers it to the steam turbine. The steam turbine sends its energy to the generator drive shaft, where it is converted into additional electricity. As a result, extra electric power is produced and this extra bit is the reason why combined cycle power plants are preferred as overall electricity production is raised as high as 50%.[2]

Below, Fig.1.1 and Fig.1.2 illustrate the Chandpur 163MW Combined Cycle Power Plant (CCPP) project under the Bangladesh Power Development Board (BPDB). It is one of the priority projects of the Bangladesh Govt. The primary role of this undertaking was to provide sufficient electrical power to the country, namely in the Comilla zone. Other roles include, but are not limited to, managing the ever-increasing power demand of the country, decreasing transmission loss and increasing the stability of the national grid system. The Chandpur 163MW CCPP is situated beside the grid substation at Balurmath, New Truck Road on the Dakatia river bank of Chandpur town. The data for the following graphs were collected from the Chandpur power station. Graphs visualising monthly electricity generation for both Single Cycle(Fig.1.1) and Combined Cycle Power(Fig.1.2) mode are shown below to illustrate the differences in power generation. [1]

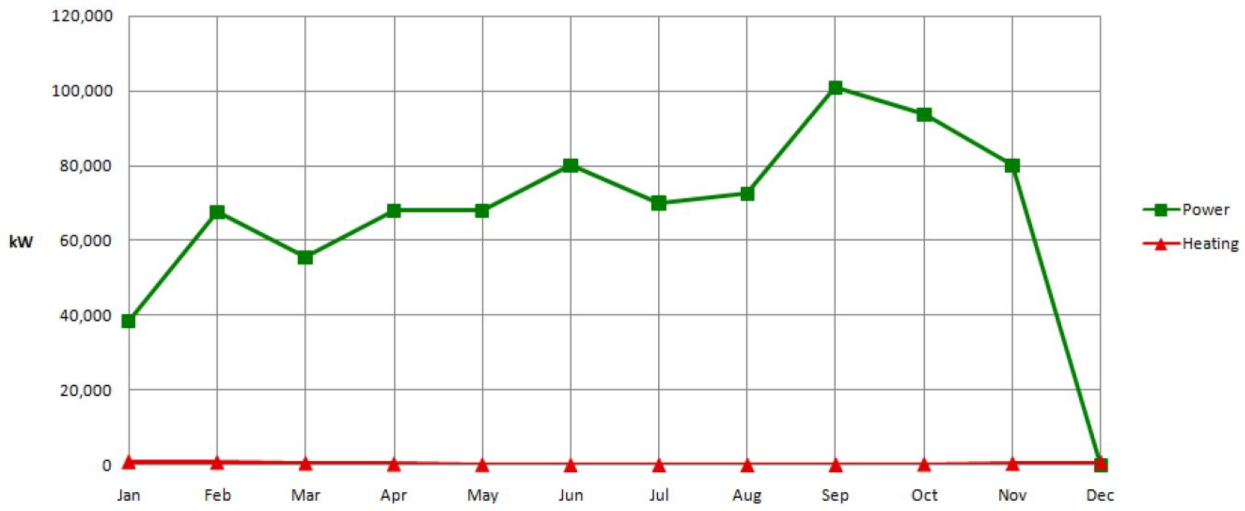


Fig.1.1: Electricity generation: single cycle mode [1]

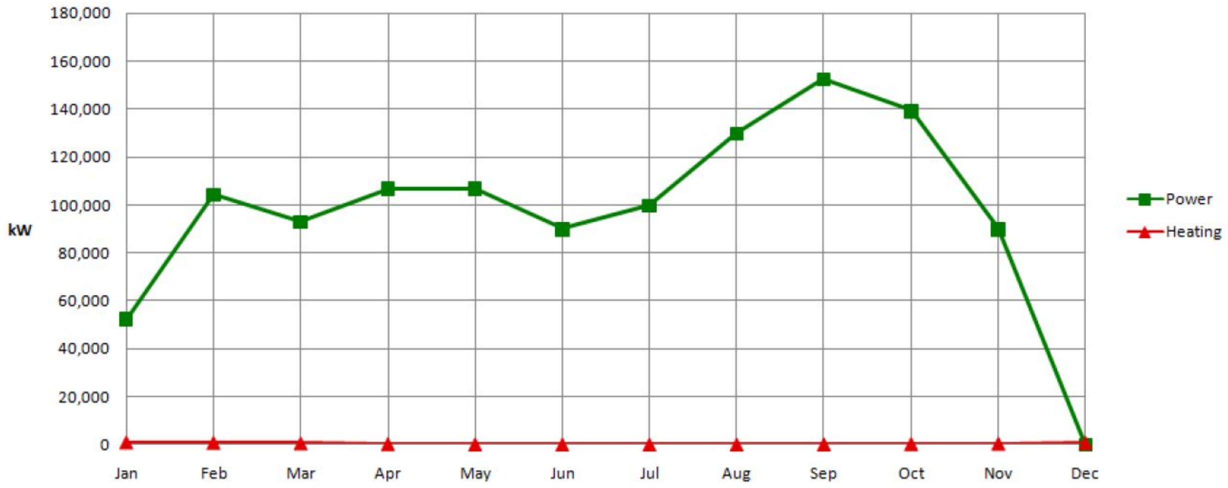


Figure 1.2: Electricity generation: single cycle mode [1]

Figure 1.1 represents the electricity generation data of a single cycle recorded in the year 2013. The net peak electricity generation of a single cycle is 100,933kW and net electrical energy is 584,350MWh. Whereas In the combined cycle, as shown in Figure 1.2, the net peak electricity generation is 152,572kW and net electrical energy is 856,714MWh. Comparing both the numerical and graphical data it can be observed that electricity production is tens of thousands more, which is a very significant amount, using combined cycle compared to the single cycle, all the while both are expending the same amount of gas as fuel.

1.5 Role and Objective of Project

1.5.1 What is machine learning

Computers using statistics and computer science to solve tasks, whether complex or simple, without being programmed to do so is known as machine learning. Similar to humans becoming better at a task through continued implementation of said task, and gaining experience, machine learning helps a computer or program to be better at a task through continued “training” using relevant information and “testing” to see if it can identify that same relevant information but in the same or even a different form. In this way, machine learning algorithms are used to complete and perfect tasks where predicting or identifying something is required. In the case of predicting, if we are predicting a single feature or variable it is known as regression in machine learning terminology. Hence machine learning regression models can be used.

1.5.2 How machine learning helps in power generation of a CCPP

The most prevalent issue investigated in machine learning is predicting a real value, known as regression. For this purpose, regression algorithms for machine learning were chosen to regulate a system's reaction to predict a numerical or real-valued target function. Many real-life problems, such as maintaining the proper power output of a power plant, can be solved and assessed as a regression issue by creating predictive models using machine learning methods. As such we have decided to use three such models, which are quite basic and simple, to enunciate our point.

CHAPTER 2: BACKGROUND STUDY

2.1 Overview

This chapter features various case studies and journals that were based on topics similar to our project. We picked the papers that were most similar to our workings so that we could gain ideas from them to help with our work.

2.2 Existing literature explanation

Some of the research that we have used for our project are given below:

- In his journal [5] Pınar Tüfekci examines and compares some machine learning regression methods to develop a predictive model, which can predict the full hourly load electrical power output of a combined cycle power plant. It uses the same dataset we use and therefore the features used are also the same. Among a few investigations, the best subset of the dataset is explored amidst all feature subsets in the experiments. The accuracy of the 3 best models out of the 15 different models they used was also compared
- In a case study [1], M. Hossain and colleagues have given a brief description of the prospects and efficiency of combined cycle and single cycle power plants in Bangladesh. They present a relevant case study to prove that the combined cycle power plants are more reasonable and economical to use in Bangladesh. For the longest time, single cycle power plants have been dominant when it came to power generation in Bangladesh. As we realize that, the most recent power age introduced limit in Bangladesh is more than 10,000 MW and it will be expanded if the single cycle power plants can be changed over into the join cycle power plants.
- In their journal [8] Bandic and colleagues present methods for predicting the power output of a combined cycle power plant using the same dataset and Random Forest Regression which we have also used. They compare error calculations from

5 different testing methods on their model as well as other methods different from ours.

- E. Elfaki and A. Ahmed[6], and B. Akdemir [7] both worked on predicting the power output of a combined cycle power plant using the same dataset however they used artificial neural networks. Since the dataset had characteristics of regression they were able to apply their neural network and use it for prediction. Much the same as our work they also performed statistical calculations on the errors between real values and estimated values to check the authenticity of the model.

CHAPTER 3: METHODOLOGY

3.1 Overview

This chapter discusses and shows - at length - the primary workflow diagram for the project along with the dataset and how it was cleaned. It then goes on to show visual representations of the dataset using heatmaps and scatterplots. Finally, it talks about the specific machine learning algorithms implemented and how they work.

3.2 Workflow Diagram

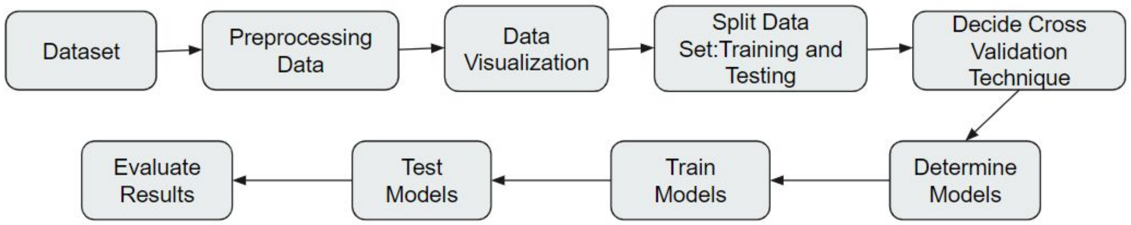


Figure 3.1: Workflow Diagram for the project

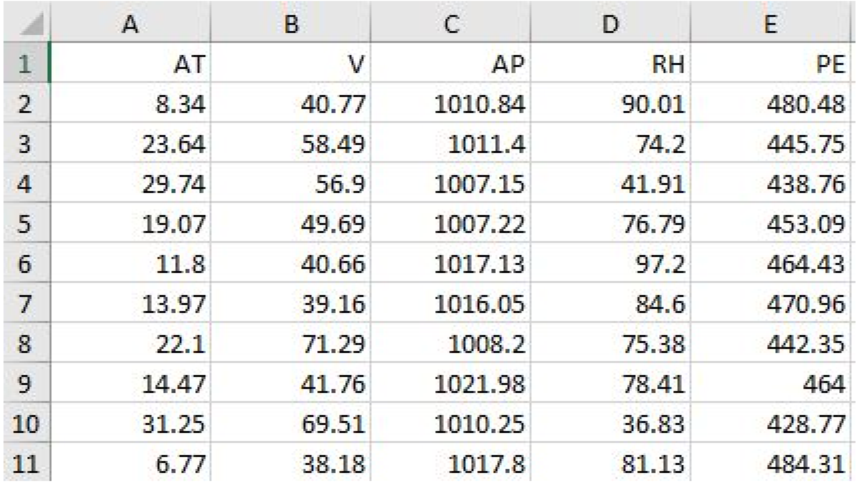
3.3 Dataset Description

The dataset that we have used contains 9568 data points which were recorded during operating hours of a combined cycle power plant over a 6 year time period, between 2006 to 2011. The readings come from a total of 674 different days of operation. It was acquired from the UCI Machine Learning repository. These data points were measured from 5 different features (4 inputs and 1 target). The variables that were considered and their ranges are given.

- Hourly ambient temperature (T) between 1.81 – 37.11 degrees Celsius,

- Atmospheric pressure (AP) in the range of 992.89–1033.30 mbar,
- Relative humidity (RH) from 25.56% to 100.16%
- Exhaust vacuum (V) within the range of 25.36–81.56 cm Hg.
- Net hourly power production (EP) within the range of 420.26–495.76 MW.

The first four features combine to some extent or another to predict the fifth feature. [4]



	A	B	C	D	E
1	AT	V	AP	RH	PE
2	8.34	40.77	1010.84	90.01	480.48
3	23.64	58.49	1011.4	74.2	445.75
4	29.74	56.9	1007.15	41.91	438.76
5	19.07	49.69	1007.22	76.79	453.09
6	11.8	40.66	1017.13	97.2	464.43
7	13.97	39.16	1016.05	84.6	470.96
8	22.1	71.29	1008.2	75.38	442.35
9	14.47	41.76	1021.98	78.41	464
10	31.25	69.51	1010.25	36.83	428.77
11	6.77	38.18	1017.8	81.13	484.31

Figure 3.2: Datasheet representing first 10 data points in the data set. [4]

3.4 Data Preprocessing

Data preprocessing is a vital process where we apply cleaning, integration, transformation, and reduction to the dataset for the machine learning algorithms. Many of the features may be unnecessary or repeated, so we apply feature selection and extraction giving us a minimum set of data allowing our algorithms to work quicker and more effectively. This will improve our results immensely. For our dataset, it was split into two different .xls files. The two files were then merged to remove duplicated data if any existed. One final integrated data set was then created.[4]

3.4.1 Handling Missing Values

With any dataset, it is important that the data must be refined to reduce the noise in the overall dataset to ensure the best prediction possible. With that under consideration, the missing values in data points were replaced with the mean value of the column thereby ensuring the data stays consistent.

3.4.2 Steps that were not considered

- Feature Selection - Since the dataset contained only just 5 features to predict the output, it was ideal not to use any feature selection to filter the features.
- Scaling - All the units on the features are adequately scaled. Hence, it was not ideal to perform scaling on any of the features.

3.5 Data Visualization

In order for us to deduce initial findings from the dataset, such as the relative relationships each independent feature had with the dependent feature, we decided to apply data visualization. This is an important method which would enable us to determine the importance the inputs had on the output of the given dataset from the first data point to the last. Below are some illustrations that helped us in visualizing the dataset.

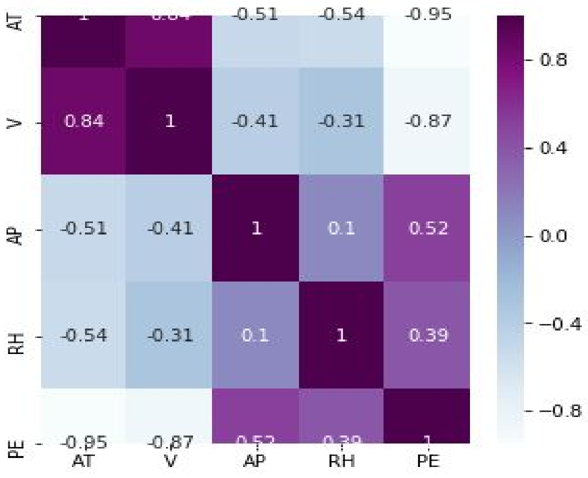


Figure 3.3: Heat Map of all 5 variables

Heat maps are a convenient method of understanding linear relations using colour intensity to illustrate how strong or weak the relationship between two variables maybe. Figure 3.3 shows a significant difference in colour between the spectrums of variables. Using the intensity scale on the right we can see that the variables ambient temperature and exhaust vacuum have the darkest regions, showing a high intensity. This shows us a strong correlation with the electrical power output, our dependent variable. Conversely, the other two variables ambient pressure and relative humidity have a low impact on the power output. This is shown by their light coloured regions which, as shown on the scale, have very low intensity.

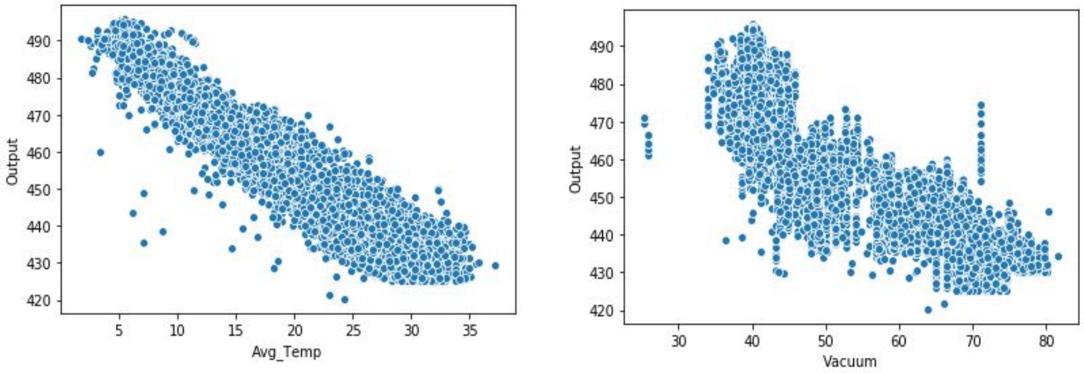


Figure 3.5: Scatter Plot Diagrams of Average Temperature and Vacuum

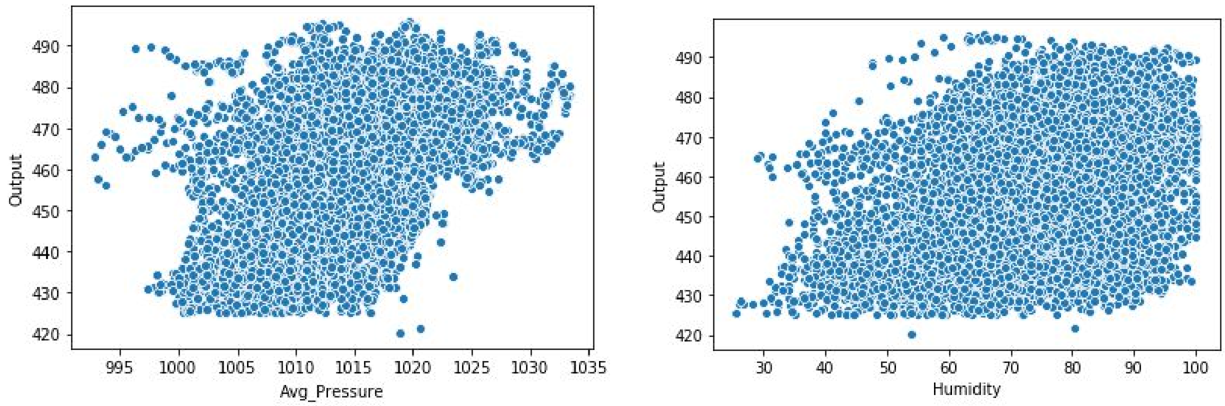


Figure 3.5: Scatter Plot Diagrams of Humidity and Average Pressure

The second method we used to visualize our dataset is a scatter plot diagram. We see that the features Average Temperature and Vacuum are likely to be the more important features from the dataset. The Relative Humidity and Atmospheric pressure graphs have points scattered all over the place, suggesting that these two features have little correlation with the output. On the other hand, the ambient temperature(AT) and exhaust vacuum(V) graphs show a linear relationship suggesting these have a higher correlation with the output. The observations from the scatter plots also confirm those from the heat map previously. Thus we have decided to use these two variables for our models.

3.6 Machine Learning Models and Application

3.6.1 Splitting the Dataset

The model needs to be built on one segment of the dataset, dubbed The Training Set. The rest of the dataset will be used for The Testing Set. The performance of the model should be similar in both sets proving that it can adapt to new sets and therefore new situations. 20% of the dataset was allocated for the test set. The rest, 80%, was used for the training set. Between these divisions, the dependent and independent variables were also divided. As a result, the training set had 7614 data points while the testing set had 1914 data points.

3.6.2 Machine Learning Models

The paper is a reflection of basic machine learning regression models being applied for prediction purposes and as such three basic regression algorithms were used. They are as follows.

- Multivariate Linear Regression

  Amongst the most simple and precise algorithms Multiple Linear Regression (MLR) predicts a continuous dependent variable (power output in our case), using two or more features. The dependent variable must be continuous whilst the independent variable may be categorical or continuous. The dependent variable (y) must be continuous. Here power output is continuous over a range of values. Independent variables(x1…xn) may be categorical or continuous. All of our independent variables are continuous. There need to be features with linear relationships. We’ve shown all the linear and non-linear features along with their combinations in the data visualisation section.

  The linear equation of the regression model is as follows:

  y = 454.6093 + (-1.9775)*Average_Temperature + (-0.2339)*Vacuum + 0.06214*Avg_Pressure + (-0.1581)*Humidity

- Support Vector Regression

Support Vector Machine can also be used as a regression method, keeping up all the fundamental highlights that portray the calculation (maximal edge). The Support Vector Regression (SVR) utilizes indistinguishable standards from the SVM for grouping, with just a couple of minor contrasts. Firstly, as the output is a real number it turns out to be extremely hard to anticipate the data at present, which has boundless conceivable outcomes. On account of regression, a margin of tolerance (epsilon) is set in estimate to the SVM which would have effectively mentioned the issue. Besides this fact, there is a greater confusing reason and the calculation is increasingly complicated. Thus it must be taken into consideration. However, the principle thought is consistently the same which is to limit errors, individualizing the hyperplane that amplifies the edge, remembering that part of the error is tolerated. Even though it is best suited for non-linear model functions, it is a basic and simple algorithm, which was one of our aims for using machine learning, and the contrasting results it provides are the reasons why it was chosen.

- Random Forest Regression

Random Forest Regression (RFR) is a version of ensemble learning so changes to the dataset may affect 1 or more trees but it would be harder to affect a “forest” of trees. We pick at random K data points(a subset) from the training set. We build n-numbered decision trees associated with each of these K data points. For a new data point, each of the n trees predicts the power output for that data point. So there are multiple predictions of power output from each tree for every data point. The new data point is assigned the average across all the predicted Y values from the multiple trees. Since the average of many predictions is taken it improves the accuracy of the predictions. Even though we aimed to use simple models, getting accurate predictions was still the primary goal so we chose RFR.

3.7 Cross-Validation

3.7.1 <u>What is Cross-Validation</u>

Cross-Validation allows us to compare different machine learning methods. It helps to decide which models were most accurate. Cross-Validation helps to give us a sense of how well these models may work in practice.

3.7.2 <u>k-Fold Cross-Validation</u>

For our project, we have used k-folds cross-validation. It is an exhaustive cross-validation technique. As an exhaustive technique, it tests the model in all possible ways. This method divides the data set into k training and validation/ testing sets. 1 subset is used for training while (k-1) subsets are used for testing. The error estimation is averaged for all k trials to get effective readiness of the model. Each k subset gets to be in the testing set at least once, and consequently, each (k-1) subset also gets to be in the training set at least once as well. k=2 Fold Cross Validation was used as the paper by paper by Pınar Tüfekci [5], which we compared our results with, did so as well. This form of cross-validation segregates the dataset into 2 portions for validation. In this approach, the dataset is split randomly 2 times into X train set, Y train set and X test set, Y test set.

3.7.3 <u>Why k-Fold Cross-Validation was used</u>

There is never enough data when it comes to training a model as training results can always be better. Reducing the dataset always runs the risk of overfitting. By reducing data we run the risk of reduced accuracy due to errors induced by bias. Dominant patterns may also not be recognized if enough data is not available for training. Therefore we need a way that provides ample data for training and keeps some for testing. In the end, every block of data is used for testing and we can compare models by seeing how well they performed.

CHAPTER 4: RESULTS AND ANALYSIS

4.1 Overview

In this chapter, we will analyze the results of our machine learning models. The accuracy of predicting each machine learning regression method is used to evaluate the overall comparison between actual and predicted values. We show these results numerically in the form of Mean Absolute Error and Root Mean Squared Error as well as visually in the form of scatterplots to support our results.

4.2 Model Results

The prediction accuracy of each machine learning regression method is used to evaluate the overall comparison between actual and predicted values. In this case, it was decided to evaluate the prediction accuracy by using the following performance criteria as the papers we compared our results with had these two error calculations in common.

- MAE

  The prediction error is the difference between the actual value and the predicted value for that instance. Given any test data-set, Mean Absolute Error of the model refers to the mean of the absolute values of each prediction error on all instances of the test data-set, giving us a broader understanding of the predictive accuracy.

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n} |y_j - \hat{y}_j|$$

Figure 4.1: Equation of MAE

- RMSE Score

  RMSE is a quadratic scoring rule which also measures the error's average size. It is the square root of the square average difference between forecast and real observation.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

Figure 4.2: Equation of RMSE

We obtain the following predictive accuracies after effectively operating the dataset through the aforementioned models.

Model	Mean Absolute Error(MAE)	Root Mean Squared Error (RMSE)
Support Vector Regression	10.29	13.00
Random Forest Regression (n=50 trees)	2.35	3.21
Multiple Linear Regression	3.57	4.44

Table 4.1: The three models with their predictive accuracies

To illustrate the accuracies visually, we have generated scatter plots to show how close the predicted values from each algorithm came close to the respective true values.

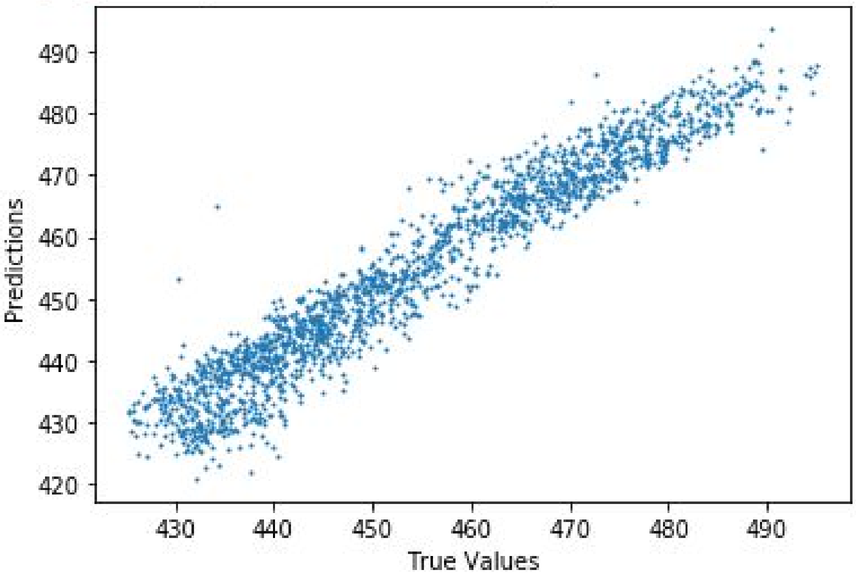


Figure 4.3: Multiple Linear Regression Predictions

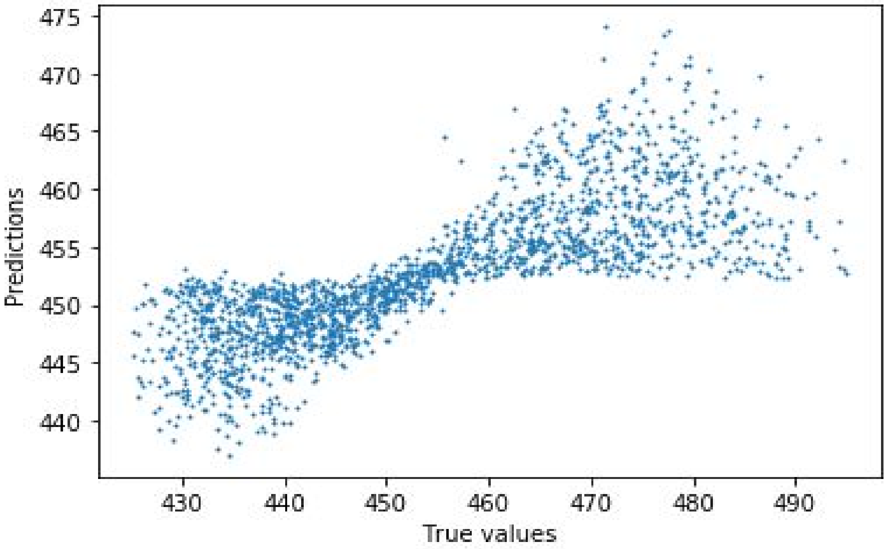


Figure 4.4: Support Vector Regression Predictions

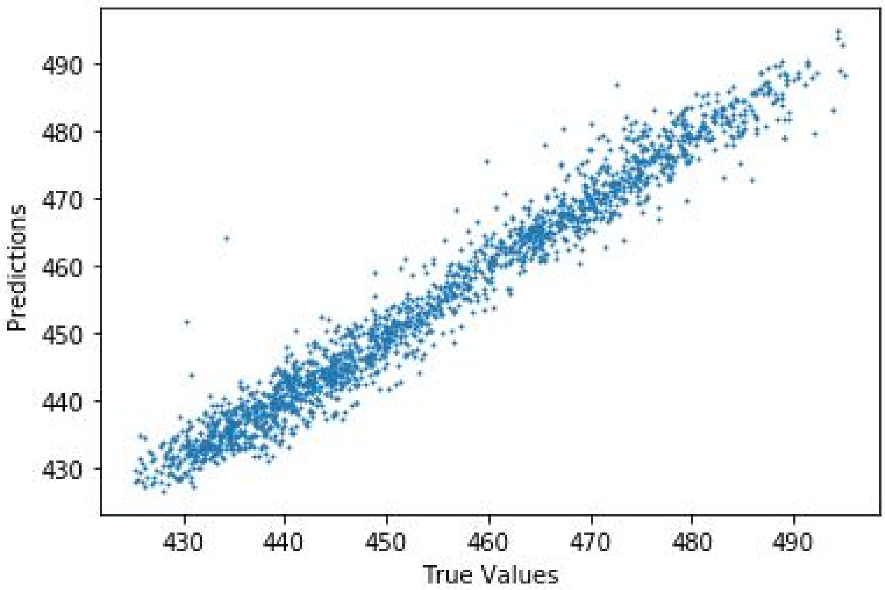


Figure 4.5: Random Forest Regression Predictions

4.3 Discussion and Comparison

From Table 4.1 we can see that Random Forest Regression has the lowest MAE and RMSE scores showing it had the highest accuracy amongst the three models. Multiple Linear regression also had scores trailing close to the Random Forest Regression model. But support vector regression had too high an error in its predictions as shown by the very high scores of MAE and RMSE showing it's inefficiency as a model in this particular case.

From the scatterplot results obtained, it can be seen that Support Vector Regression has the worst result showing little to no linearity having most points scattered everywhere[Figure 4.4] while Multiple Linear Regression comes in second showing better linearity[Figure 4.3] and lastly we see Random Forest Regression has the most linear graph with maximum points converging in the middle[Figure 4.5]. These results are further backed by our MAE and RMSE values from Table 4.1 where we saw that Random Forest Regression had the lowest error rate while MLR came in second and SVR had the worst error values thus further reinforcing our findings from the scatter plot diagrams.

Model	CV Accuracy with k=2, MAE	CV Accuracy with k=2, RMSE
Support Vector Regression	11.72	14.37
Random Forest Regression (n=50 trees)	2.69	3.70
Multiple Linear Regression	3.63	4.56
Bagging REP Tree	2.82	3.79
KStar	2.88	3.86
REP Trees	3.13	4.21

Table 4.2: Comparison of our results with that of the paper by Pınar Tüfekci [5] after applying 2-fold cross-validation

In Table 4.2 the first half of the table shows MAE and RMSE scores for our models after 2-fold cross-validation was applied. The second half of Table 4.2 shows the results of the paper by Pınar Tüfekci [5]. It shows the paper's best three models in terms of their MAE and RMSE scores after 2-fold cross-validation was applied. None of the models the paper used was the same as ours and as such, we picked the best three to compare with our three. Due to the paper having the MAE and RMSE scores we chose to compare it with ours. One thing to note is that the paper had shuffled the data set 5 times before applying 2-fold cross-validation. They did so to achieve maximum predictive accuracy on the dataset and therefore ended up doing 5 x 2-fold cross-validation as stated by them. [5] We, on the other hand, found the number of data points on the data set to be enough and as such did not shuffle the data set. This helped us to contrast our results with them as well.

As we can see from Table 4.2 the random forest regression MAE and RMSE score was the lowest yet again. It was lower than the best three models from the paper as well as our own models. This is due to the fact that random forest has several trees that help to improve accuracy, 50 in our case. There are multiple predictions of power output from each tree for every data point. Support vector regression had the highest MAE and RMSE, which in itself is quite high, showing again it is not suitable for this type of dataset. It is best suited for non-linear model functions which our data set isn't, as shown in previous scatterplots and heatmaps. Random Forest Regression shows that the graph is closest to the best fit line, whilst Multiple Linear Regression comes in second and Support Vector Regression shows the worst result.

Mode of testing	RMSE of Paper	Our RMSE
50% train, 50% Test	3.46	3.56
66% train, 34% Test	3.30	3.36
90% train, 10% Test	3.03	3.07
10-CV	3.25	3.45
20-CV	3.23	3.43

Table 4.3: Comparison of our RMSE scores with that of the paper by Bandic and colleagues [8] after trying 5 different methods of predictions on Random Forest Regression Model

Mode of testing	MAE of Paper	Our MAE
50% train, 50% Test	2.50	2.60
66% train, 34% Test	2.36	2.48
90% train, 10% Test	2.26	2.31
10-CV	2.29	2.44
20-CV	2.27	2.42

Table 4.4: Comparison of our MAE scores with that of the paper by Bandic and colleagues [8] after trying 5 different methods of predictions on Random Forest Regression Model

From both Table 4.4 and Table 4.3 we can see results of RMSE and MAE scores of our previous Random Forest Regression model and that of the paper by Bandic and colleagues [8] but this time it was after trying 5 different methods of predictions. Here the paper [8] first tried out 3 different test/train splits before applying k-fold cross validation twice where k=10 and 20 and we did the same to compare. In both tables our scores were slightly higher than that of the paper however we were able to maintain a consistency in the results from top to bottom with our scores just like they did with theirs. The small margin of difference between all our errors and their shows we were able to maintain a degree of accuracy through all 5 modes of testing.

4.4 Other approaches

Though machine learning has proved to be a successful method for predicting the power output for a combined cycle power plant artificial neural networks have also been popular among many works.

E. Elfaki and A. Ahmed[6] worked on predicting the power output of a combined cycle power plant as well using the same dataset however they used artificial neural networks. Since the dataset had characteristics of regression they were able to apply their neural network and use it for prediction. Similar to our work they also conducted similar statistical studies on the error between real values and estimated values to check the authenticity of the model.

B. Akdemir [7] used artificial neural networks on the same dataset to predict the power output of the combined cycle power plant and similar to us applied 2-fold cross-validation as well as find the mean squared error.

CHAPTER 5: CONCLUSION

In our paper, we have seen and discussed how much more efficient Combined Cycle Power Plants are than Single Cycle Power Plants. In today's world CCPP is of utmost importance as our environment is slowly deteriorating and fossil fuels are being used up. It is especially important in a country like Bangladesh, where getting up to 50% power for the same amount of fuel is always optimal. We have used machine learning and managed to predict the best outcomes for the given variables. The analysis of such a system takes a lot of computational power if done by a thermodynamic approach. Not only that but the results obtained may be disappointing and not dependable due to the assumptions taken. This is why we have used machine learning regression methods to analyse the prediction of the power output of the system and have shown results to be within minimum margins of error as compared with existing literature, and in some cases gotten better results.
For discovering the best way to predict the power output we applied 3 different machine learning regression methods. After applying all the models we found that Random Forest Regression proved to be the most effective model in predicting the outcomes. From the visualization of our variables we have seen average temperature to be the most influential variable.

With any machine learning model, the more the data we have the more accurate and precise predictions we get. In our case, the sample of data collected for our dataset was just 6 years. An ideal case would have been with a period of 10+ years helping us derive better predictions and giving greater accuracy. Another vital drawback of our dataset was that it consisted of only four different independent variables to work with. If we had more features we could have produced a more realistic result.

Although we have seen machine learning has proven to be fruitful in terms of predicting scenarios there remain more opportunities in the fields of artificial neural networks for further research as proved by more recent studies.

References

[1] M. Hossain, I. Zissan, M. Khan, Y. Tushar and T. Jamal, "Prospect of combined cycle power plants over conventional single cycle power plants in Bangladesh: A case study", 2014 International Conference on Electrical Engineering and Information & Communication Technology, 2014. Available: 10.1109/iceeict.2014.6919060 [Accessed 15 May 2020].

[2] COMBINED CYCLE POWER PLANT: HOW IT WORKS. [Online]. Available: https://www.ge.com/power/resources/knowledge-base/combined-cycle-power-plant-how-it-works [Accessed 15 May 2020]

[3] Air Pollution in Dhaka, Bangladesh. Real-time Air Quality monitoring project. [Online]. Available: https://aqicn.org/country/bangladesh/ [Accessed 15 May 2020]

[4]"makinarocks/awesome-industrial-machine-datasets", GitHub, 2019. [Online]. Available:https://github.com/makinarocks/awesome-industrial-machine-datasets/tree/master/data-explanation/Combined%20Cycle%20Power%20Plant?fbclid=IwAR2Gy8XaHHBNkBMIjAPjTYqsZMkPg5Us3_kGl_1pYGN3Zo_hg-Y3azG-47Y.
[Accessed 15 May 2020].

[5] Tüfekci, "Prediction of full load electrical power output of a baseload operated combined cycle power plant using machine learning methods", International Journal of Electrical Power & Energy Systems, vol. 60, pp. 126-140, 2014.
Available: 10.1109/iceeict.2014.6919060 [Accessed 15 May 2020].

[6] E. Elfaki and A. Ahmed, "Prediction of Electrical Output Power of Combined Cycle Power Plant Using Regression ANN Model", Journal of Power and Energy Engineering, vol. 06, no. 12, pp. 17-38, 2018.
Available: 10.4236/jpee.2018.612002. [Accessed 15 May 2020].

[7] Akdemir, Bayram. (2016). “Prediction of Hourly Generated Electric Power Using Artificial Neural Network for Combined Cycle Power Plant”, International Journal of Electrical Energy.
Available:4. 91-95. 10.18178/ijoee.4.2.91-95. [Accessed 15 May 2020].

[8] Bandic, Lejla & Hasicic, Mehrija & Kevric, Jasmin. (2020), “Prediction of Power Output for Combined Cycle Power Plant Using Random Decision Tree Algorithms and ANFIS.”
Available: 10.1007/978-3-030-24986-1_32. [Accessed 15 May 2020].