May 4, 2022

Executive summary

# The problem

This analysis attempted to construct a regression model to predict electrical energy output from a combined-cycle power plant (CCPP). Combined Cycle Power Plants (CCPP’s) are becoming more popular around the world for their greater efficiency and lower waste: Ramireddy (2012) found that CCPP’s can generate more than double the amount of electricity as conventional power plants. This study will use a multiple linear regression model to evaluate which, if any, variables can predict the plant’s electrical output.

## Hypotheses

### H0: A regression model that predicts electrical energy output cannot be constructed.

### H1: A regression model can be constructed to predict energy output.

# summary of the data-analysis process

The data for this analysis was made publicly available by the UC Irvine Machine Learning Repository and was downloaded as a Microsoft Excel (xlsx) file. The data set includes predictor variables of Ambient Temperature (AT), Ambient Pressure (AP), Relative Humidity (RH) and Exhaust Vacuum (V), and target variable Electrical Output (PE). The data was collected from a Combined Cycle Power Plant over 6 years (2006-2011) using sensors located around the plant that record these variables every second. For the analysis, Python programming language was used in a Jupyter notebook. An Ordinary Least Squares (OLS) regression model was created using the Statsmodel library.

# Findings and limitations

## The constructed model used the four independent variables to predict the target variable with an R-Squared value of 0.929; thus, the attempt to construct a regression model was reasonably successful. All variables had p-values of below 0.05. A partial snapshot of the model is below. This model asserts that temperature has the biggest effect on energy output (with higher temperatures creating less energy), while ambient pressure has the smallest effect (with higher pressure creating more energy). Exhaust vacuum and relative humidity have moderate effects, with higher vacuum and humidity correlated with lower energy output. Thus, maximum energy output is correlated with conditions of low temperature, exhaust vacuum and relative humidity, and high ambient pressure.

## Table Description automatically generated

## Linear regression is easy to implement in Python and has a very interpretable output, as shown above. However, there are limitations in any linear regression model, including the assumption of linearity between dependent and independent variables, as well as the assumption that multicollinearity between independent variables is not present. While both these assumptions present potential shortcomings in our model, neither is so large that the model should be discounted in any capacity.

# proposed actions

## Take similar measurements at one or more additional combined cycle power plants, ideally in various locations around the world.

### Further analyze new data.

### Realize geographic locations that are most ideal for maximizing energy output.

## Further analyze this data set using other techniques such as lasso regression or k-means clustering.

# Expected Benefits

## The goal of this analysis is to build a regression model that estimates electrical energy output based on ambient temperature and pressure, relative humidity and exhaust vacuum. Such a model would allow for accurate prediction of power output, which reduces issues such as power outages, as well economic and technical difficulties (Siddiqui et al., 2021). Furthermore, if optimal conditions for electrical output can be ascertained, it would be extremely beneficial to build or utilize CCPP’s in geographical locations where these conditions are most often met.

## References

Ramireddy, V. (2012, August 25). *An Overview of Combined Cycle Power Plant*. EEP - Electrical Engineering Portal. Retrieved May 1, 2022, from <https://electrical-engineering-portal.com/an-overview-of-combined-cycle-power-plant>

Siddiqui, R., Anwar, H., Ullah, F., Ullah, R., Rehman, M. A., Jan, N., & Zaman, F. (2021, December 23). *Power Prediction of Combined Cycle Power Plant (CCPP) Using Machine Learning Algorithm-Based Paradigm*. Wireless Communications and Mobile Computing. Retrieved May 1, 2022, from <https://www.hindawi.com/journals/wcmc/2021/9966395/>