

# Wind Power Forecasting Model Through Data Analysis of Causes and Impacts of Global Warming

Jun Seok Hwang<sup>1</sup>, Seoyeon Kim<sup>2</sup>, Seunghyun Lee<sup>3</sup>, Shiwoo Lee<sup>4,\*</sup>, Kenneth Chisoon Park<sup>5</sup>

<sup>1</sup>Williston Northampton School, Easthampton, United States

<sup>2</sup>Branksome Hall Asia, Seogwipo-si, South Korea

<sup>3</sup>Korean Minjok Leadership Academy (KMLA), Gangwon-do, South Korea

<sup>4</sup>Seoul Foreign School, Seoul, South Korea

<sup>5</sup>North London Collegiate School Jeju, Seogwipo-si, South Korea

## Email address:

shiwoolee0@gmail.com (Shiwoo L.)

\*Corresponding author

## To cite this article:

Jun Seok Hwang, Seoyeon Kim, Seunghyun Lee, Shiwoo Lee, Kenneth Chisoon Park. Wind Power Forecasting Model Through Data Analysis of Causes and Impacts of Global Warming. *International Journal of Energy and Environmental Science*.

Vol. 6, No. 5, 2021, pp. 134-142. doi: 10.11648/j.ijees.20210605.14

**Received:** September 23, 2021; **Accepted:** October 8, 2021; **Published:** October 19, 2021

---

**Abstract:** In this article, we aim to spread the need for change by informing the reader about global warming and its consequences. Global warming is a phenomenon in which average global temperatures rise due to the increase in greenhouse gases (GHG) emitted and trapped in the atmosphere. These specific gases are special in that they contribute to the flow of heat within the atmosphere as they trap more heat from the sun within the atmosphere. The most influential of these gases is CO<sub>2</sub>, which most of the data analysis in this article will be focusing on. There are many issues and problems that are attributed to global warming, along with a wide array of data that relates to the topic. By analyzing such data, one can find out the global trend in climate change and even come up with potential solutions for this increasingly threatening event. The lack of action taken by world leaders is due to general negligence and a lack of an effective and coherent solution to tackle climate change. A probable solution would be to find a surrogate to the main cause of GHG emissions, which is the substantial exploitation of conventional (nonrenewable) energy sources. Renewable energy sources are excellent substitutes to conventional energy sources and one of the most effective types is wind. This energy source is currently implemented in many areas around the globe and is producing better results than in the past when conventional energy sources were in use. Although wind turbines are more efficient than solar panels in producing electricity because they are not affected by the presence of sunlight, the amount of electricity generated can fluctuate due to various factors. To adjust to these varying conditions, machine learning regression algorithms can be managed to create a wind power forecasting model that predicts how much electricity can be produced on that day according to the varying factors; it will allow people to be less dependent on those factors. This article explores the trends for climate change and the avenues for change. It must be noted that averting one's gaze away from this urgent issue resolves nothing.

**Keywords:** Global Warming, Greenhouse Gas, CO<sub>2</sub> Emissions, Renewable Energy, Wind Energy, Machine Learning Regression Algorithm

---

## 1. Introduction

Global Warming is a phenomenon in which the average annual global temperature increases. Throughout earth's history, there have been eras, such as the late Mesozoic era, where the earth's temperature was higher than now at 35.5°C

[1]. However, the global warming we face today differs in that it is not natural - rather, it is caused by human activity [2]. Even a small increase in the average annual global temperature can have detrimental environmental consequences that affect both humanity and mother nature. In terms of humanity, global warming will devastate many low-lying population centers and



## 2. The Severity of the Current Global Situation

### 2.1. The Problems

The figure above shows significant keywords of global warming. Hundreds of articles per topic were drawn from Google News and imported as data. These topics include global warming, climate change, and carbon dioxide emission. Keywords were selected from 100 articles of each respective topic, combined, and placed in a cloud-shaped picture. The font size denotes its significance. The most conspicuous keyword in the provided picture is 'climate change'. It implies that the Earth is experiencing a serious transition. Additionally, China and Europe are notable keywords in the picture. This sheds light on the importance of China and Europe within the current global situation.

### 2.2. Lack of Attention

As shown in Figure 2, 'Global Warming' and 'Greenhouse Gas' are receiving significantly less attention compared to other keywords in the United States. The use of 'pytrends' in python allows access to the global trend, which is the of googling the keyword received in a month, of a certain region; in this particular graph, the region is the United States and the keywords are 'COVID-19', 'Black Lives Matter', 'BTS', 'Global Warming', and 'Greenhouse Gas'. 'COVID-19' had the highest search result among the five keywords with 'Black Lives Matter', 'BTS', 'Global Warming', and 'Greenhouse Gas' following after, respectively. Approximately from the mid of February of 2020,

'COVID-19' started to be researched and eventually culminated at about August of 2020. 'BTS', meanwhile, remained constant, for the whole period shown in the graph above. Although it was searched significantly less, meaning that it received less attention than BTS, it was given more concern before COVID-19 was dispersed from its origin. The little fluctuation it had was zeroed after the immediate catastrophe, COVID-19 became known to the public. Here, the inherent lack of awareness of 'Global Warming' and 'Greenhouse Gas', which is closely related to the previous keyword, is manifestly visible. This is a clear indicator of the lack of public attention regarding 'Global Warming' despite it being of a different magnitude in terms of its severity and significance compared to other issues that are receiving more recognition.

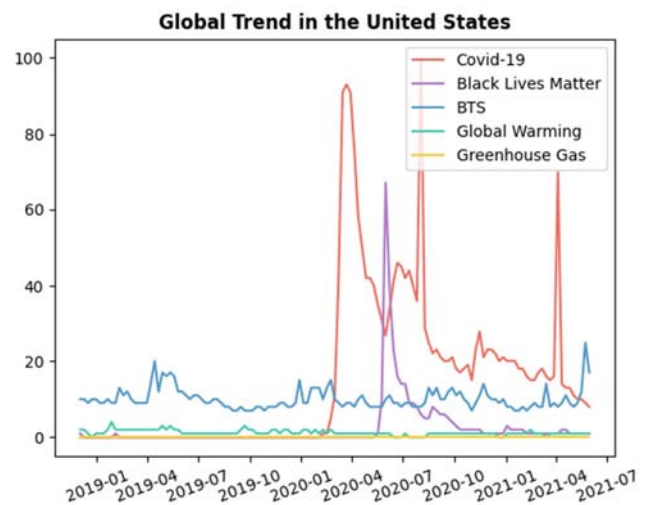


Figure 2. GHG Emissions by sector.

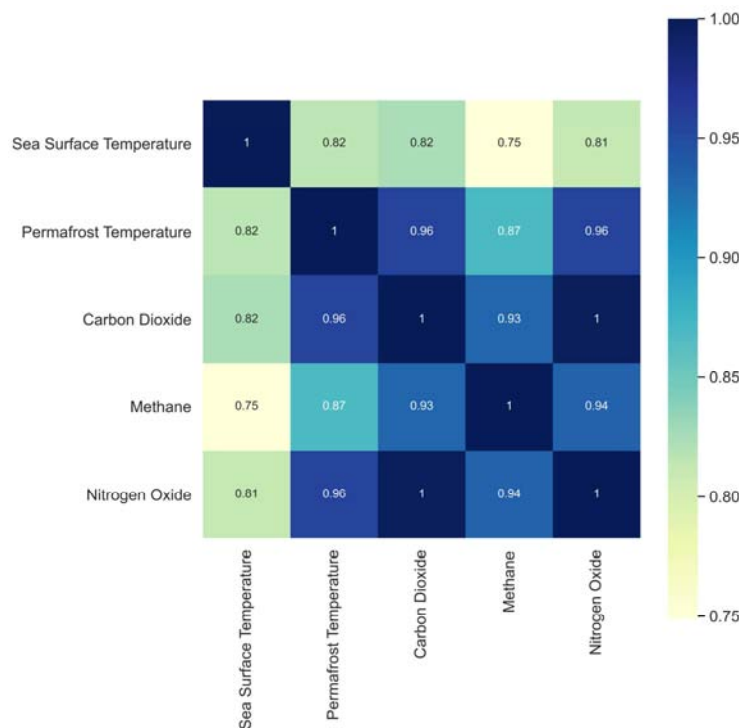


Figure 3. Correlation between greenhouse gas emissions and temperatures.

### 2.3. Severeness

Figure 3 shows the strength of correlation between the words between the respective variables from the x and y axes. 1 being the strongest relationship possible and 0 meaning no correlation whatsoever. It can be seen that all the boxes that have the same x and y variables aligned are all at a correlation strength of 1, since they are the same thing. However, there is also one combination with a correlation strength of 1. This correlation is the correlation between carbon dioxide and nitrogen oxide. It also seems to be that the further the representation heads to the bottom right, the stronger the correlation strength is.

As depicted in Figure 4, the average number of heatwaves per year and intensity expresses the average temperature above the local threshold during heatwaves are both

increasing. In general, both graphs demonstrate an increase throughout the decades (the 1960s to 2010s). In the 'Frequency by Decade' graph, a dramatic rise from the lowest mark, 2.04 (2dp.) heat waves to the highest point, 6.02 (2dp.) Heatwaves is visible. This means that the average number of heatwaves per year increased rapidly from the 1960s to 2010s. On the other hand, in the 'Intensity by Decade' graph, a ceiling trend from the lowest point, 1.99°C (2dp.) in the 1960s to a peak of 2.46°C (2dp.) in the 2010s is visible. Whether it be dramatic or gradual, both the average number of heatwaves per year and the average temperature above the local threshold during heat waves have risen over the past decades and the global warming problem is aggravating as time passes.

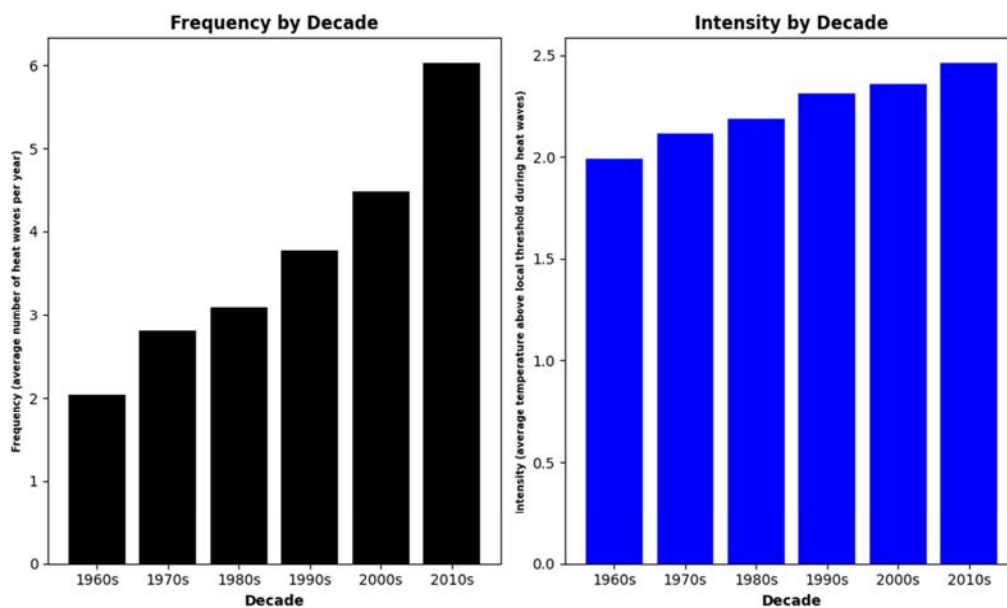


Figure 4. Frequency and Intensity by Decade.

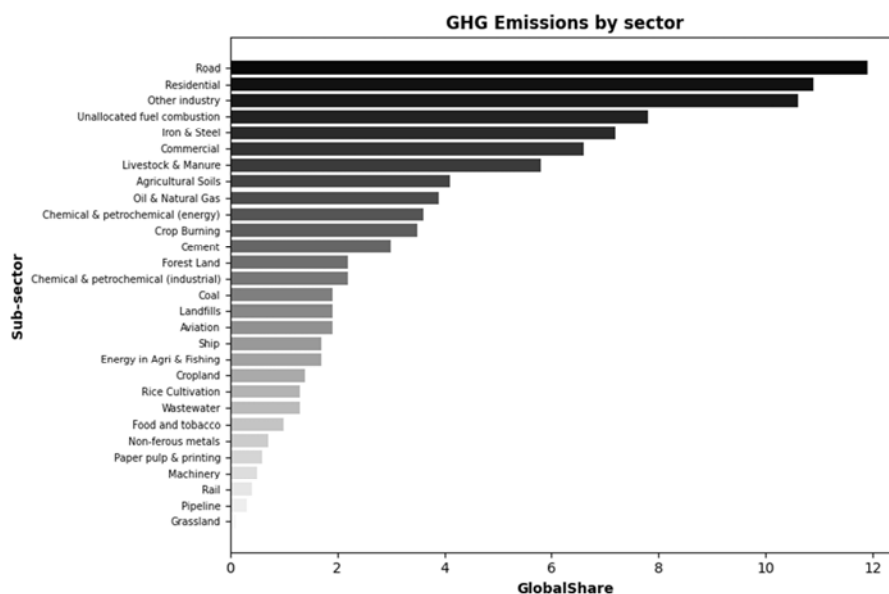


Figure 5. GHG Emissions by sector.

## 2.4. Emission Rates

As presented in Figure 5, various sectors contribute to form the whole GHG Emissions. Figure 5 shows GHG Emissions by sector with each of the sub-sectors (exhibited on the y-axis) global share presented on the x-axis in percentage, where the term, 'sub-sector', are factors that cause GHG Emissions. As presented, there are profuse sub-sectors that are part of the combustion of fossil fuels. Of the Top 5, three sub-sectors, 'Road', 'Residential', and 'Unallocated fuel combustion'. Directly relate to the burning of fossil fuels. The majority of automobiles on the road, domestic machinery, and the factors of 'Unallocated fuel combustion' use fossil fuels to be managed. A dramatic difference in the percentage range of the GHG emission can be seen from the highest sub-sector being 'Road' by contributing 11.9% and the lowest being 'Grassland' by taking 0.1% of the whole GHG emission.

Figure 6 shows the average annual greenhouse gas emission rate of countries like China, the United States, South Korea, the United Kingdom, and the average of 193 countries. China and the United States show massive differences in emission rates compared to other countries. Also notice the abrupt increase in the emission rate in China; the emission rate in 2018 is approximately 4 times greater than the emission rate in 1991. Other countries such as South Korea and the United Kingdom show relatively static and low emission rates. Despite this, the average emission rate is even lower than both South Korea and the United Kingdom, indicating that the majority of the countries have relatively low emission rates.

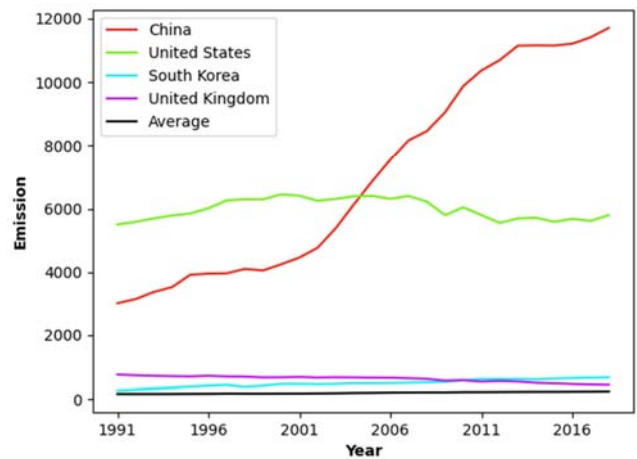


Figure 6. Average Annual Greenhouse Gas Emission Rate by Country.

The average annual greenhouse gas emission rate of all countries was 244.874715 MTCO<sub>2</sub>E IN 2018, indicating that the total greenhouse gas emission per day was

$$244.874715 \times 193 = 47260.8 \text{ (MTCO}_2\text{E)}$$

since the data is based on 193 countries. According to the Greenhouse Gas Equivalencies Calculator made by the EPA, the total emission is equal to the greenhouse gas emission by 10,278 passenger vehicles driven from one year, and the sequestered CO<sub>2</sub> by 781,467 tree seedlings grown for 10 years.

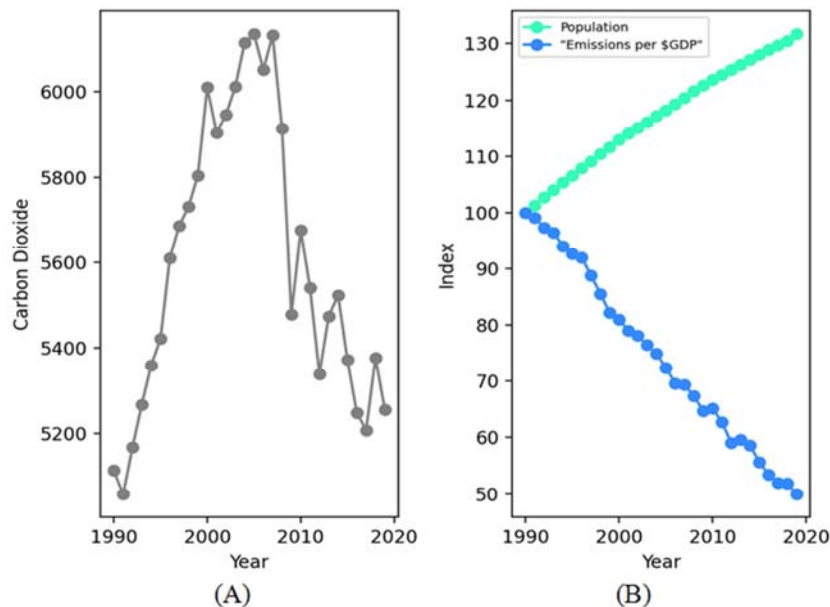


Figure 7. (A). Annual Carbon Dioxide Emissions; (B). Relative Ratio between Carbon Dioxide emissions per GDP and Population by year.

As exhibited in Figures 7(A) and 7(B), Figure 7(A) shows the emission of Carbon dioxide, and Figure 7(B) indicates the relative ratio between population and CO<sub>2</sub> emissions per GDP. According to Figure 7(B), the emissions per GDP gradually

diminished by 50 percent compared to 30 years ago. In contrast, the human population has been increasing for the last 30 years and currently has a 30 percent bigger population than in 1990. Despite the increase in population, the CO<sub>2</sub>



emission culminated in 2005 but decreased approximately 14 percent in 2019. The decrease in carbon dioxide emissions could be predicted to some extent by using calculus. Figure 7(A) has a similar shape to the Logistic Growth Curve which finds applications in a range of fields, including biology, economics, and geoscience. The inflection point which is 1997 implies Carbon dioxide emissions would not increase indefinitely because they would soon reach an asymptote. The asymptote actually formed from 2003 to 2007. After the graph reached the asymptote, the emissions gradually decreased.

### 2.5. The Effect of Time

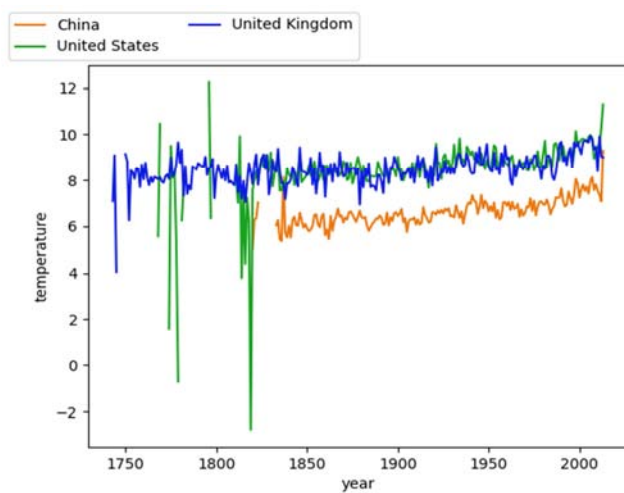


Figure 8. Global Land Temperatures by Country.

As displayed in Figure 8, China, the United States, and the United Kingdom's average land temperature increased and the fluctuation of the global land temperature lessened as time passed. The line graph above shows the varying land temperatures in countries throughout time from the 1750s to the 2000s in the United States, United Kingdom, and China. It is important to note that all 3 countries made more or less an inclination in average land temperatures, peaking in the 2000s. United States' global land temperatures show how human settlement causes changes in the natural environment as it slowly develops more and more and the industrial revolution molds the world. This is also interesting as it shows the increase in stability as time flows, as the land temperatures used to fluctuate up 800% or even down 1000%, but as the time gap from the present decreases, the less extreme the increases and decreases become, only amounting to a 150% increase at most. However, it must be noted that the average land temperature has increased as a whole in the more recent decades.

### 2.6. Consequences

As demonstrated in Figure 9, the average percentage of glaciers each year is gradually decreasing relative to that in 1950. Figure 9 shows the average volume of four glaciers (South Cascade Glacier, Gulkana Glacier, Wolverine Glacier, and Lemon Creek Glacier) with respect to that in the year

1950. There is a visible downwards trend and in fact an acceleration in the rate at which the glaciers are melting due to increasing energy demand and consequently, increasing greenhouse gas emissions. There are specific time periods - that being 1997 and 2008-where the graph exhibits an anomalous behaviour where the rate of decrease falters. This roughly coincides with two major financial crises (1997 Asian Financial crisis and 2008 Global financial crisis) where human activity was significantly impaired. This is an excellent piece of data that shows how human activity such as manufacturing and the gas emission caused by it results in direct consequences.

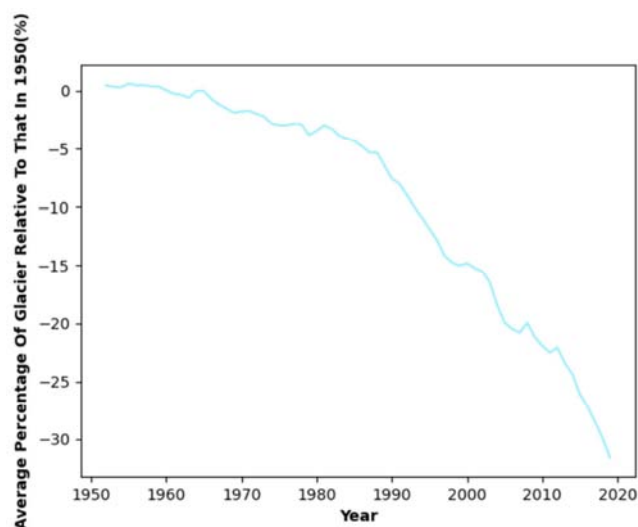


Figure 9. Average Percentage of Glaciers Each Year In Respect to that in 1950.

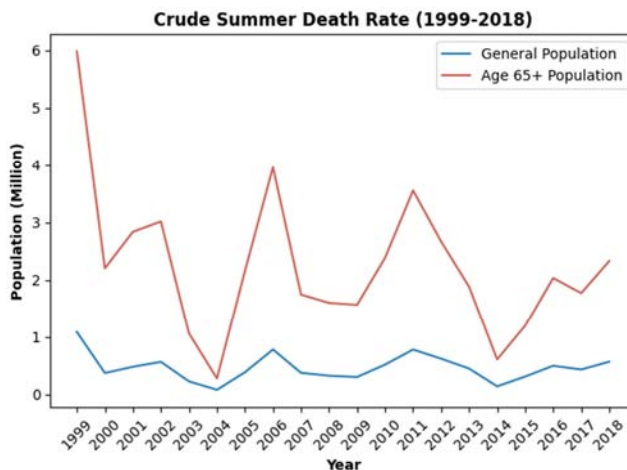


Figure 10. Crude Summer Death Rate (1999-2018).

As shown in Figure 10, the crude summer death rate is immense and this, acknowledging Figure 8, is due to the increase in global land temperature. Figure 10 presents a crude summer death rate per million of both the general population and the age 65+ population due to the high temperature over the years of 1999 to 2018. As shown, the death rate for 'Age 65+ Population' is higher than 'General Population'. These deaths were caused by 'Heat-

related Cardiovascular Disease' and more than a third of all heat deaths relate and are aftermaths of climate change, which is due to global warming [11]. While the 'General Population' has a moderate change of summer death rate over the period, the 'Age 65+ Population' has a rapid fluctuation. The summer death rate of the 'General Population' has a rather small difference between the climax, 1.08% (2dp.) in 1999, and the bottommost mark, 0.08%(2dp.) in 2004. On the other hand, the summer death rate of the 'Age 65+ Population' reached a peak amounting 5.98% (2dp.) in 1999 and again touched the lowest point amounting to only 0.28% (2dp.) in 2004. A conspicuous similarity between the two populations is that the highest and the lowest marks were, respectively, in the years 1999 and 2004.

### 3. Solution: Prediction of Wind Energy

#### 3.1. Possible Solution for Reducing CO<sub>2</sub> Emission

Renewable energy is energy that is defined as being resources through natural means. Renewable energy includes solar energy, wind energy, geothermal energy, and more [12]. The most prominent of these types being solar and wind. This technology is already being used today, and it is showing great results. Although solar panels are more popular, it is without a doubt that wind energy is more efficient overall. Not only do wind turbines generate more electricity than solar panels, but they also work day and night unlike solar panels, and they release less CO<sub>2</sub> into the air [13].

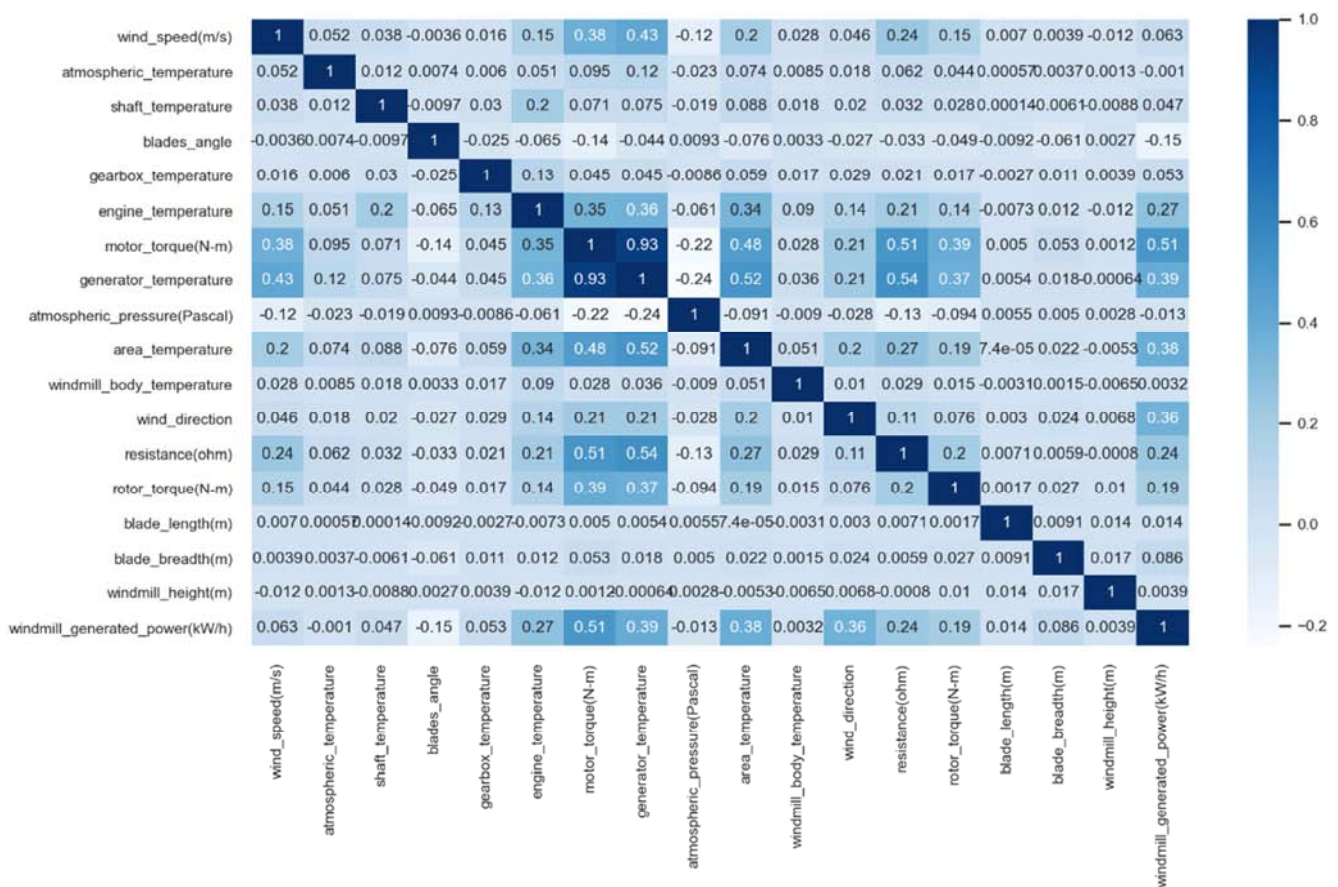


Figure 11. Correlation of Values Between Wind Datas.

#### 3.2. Theoretical and Actual Analysis of Wind Energy

The above figure represents the correlation between the theoretical data of wind and the actual data of wind. The closer the value of each grid is to 1, the closer the two datas are. The color of the grid is blunter as the value is further away from 1. The main diagonal of the graph is constituted only of 1s, showing that identical factors of the two data have identical values. Other factors show non-identical relations, but notice that the values are close to 1 near the center region of the graph. Since the order that

the factors are in does not have specific criteria and is in a random order, the pattern shown above is a mere coincidence.

#### 3.3. Machine Learning Regression Algorithm

In this section, the machine learning prediction model was implemented to predict the power generated by the features in Figure 11 using three algorithms: K-Neighbor regression, Linear regression, and XGB regression.

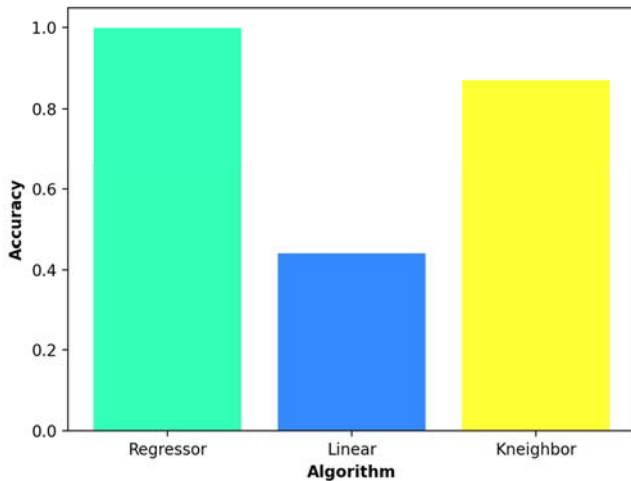


Figure 12. Accuracy by Algorithm.

K's nearest Neighbors regression, often called the KNN regression, is an algorithm used in Python to approximate a certain outcome from an independent variable by assessing other data sets in its neighborhood. It does this by averaging all the data points in its neighborhood and applying it to the independent variable given [14].

An example of an algorithm with a similar function as KNN is linear regression. It also predicts an outcome, the dependent variable when it is given the independent variable. The main difference is that it is designed to find the linear average of a set of data. Its advantages are that it performs remarkably well when provided with linearly separable data and that it is uncomplicated to train. Nevertheless, it has deficiencies and they are that it only estimates the linearity between dependent and independent variables and that it is susceptible to high signal-to-noise ratio (SNR) [15].

Another type of predicting algorithm is the XGB Regression, which uses gradient boosting decision trees to calculate the most likely dependent variable from an independent variable when given a data set [16]. This means that decisions are made through splitting decisions which narrow the dependent variable down into a more specific area, eventually calculating an answer. This type of prediction algorithm is good for classification and is better suited for less extreme numbers than the other types of prediction algorithms [17].

A wind power forecasting model built by the utilization of these three machine learning regression algorithms allows for easing the complexity of wind fluctuation. This wind power forecasting model predicts how much wind can be generated according to the different factors shown in Figure 11. Wind, despite its superiority among the renewable energy sources in steadiness, also fluctuates in the amount that it generates due to various factors presented in Figure 11. This wind power forecasting model will allow people to rely more on wind energy than the conventional energies that contribute greatly to global warming. The process of teaching an algorithm to predict is largely constructed of two factors: training and testing. The algorithm divides the given dataset into the 'train

dataset' and the 'test dataset' to enact these two sections. The 'train dataset' is "used to fit the machine learning models" and the 'test dataset' is "used to evaluate" how the machine learning model performs in reality [18]. K-neighbors regression scored 86.9 (3sf.)% in training and 57.7 (3sf.)% in testing for accuracy while linear regression resulted in 44.1 (3sf.)% in training and 43.8 (3sf.)% in testing. On the other hand, XGB regression excelled by totaling 99.8 (3sf.)% in training and 95.3 (3sf.)% in testing. All K-neighbors regression, linear regression, and XGB regression were trained and tested and the result was that for the prediction of wind energy generation, XGB regression had the highest accuracy among the three machine learning regression algorithms in both the training and testing dataset.

## 4. Conclusion

Global warming is a fast-approaching problem. It is undeniable that global warming is real and threatening, proven by all data across the globe. Because of this, it is crucial to create and initiate possible solutions or at least inhibit the continuation of global warming. The most probable solution to this is most likely using renewable energy to replace fossil fuels. From the collected data, it seems that this will be the most helpful. Additionally, it is possible to know how effective wind energy will be against global warming through coding and data analysis. For future work, we are going to implement the prediction model that can predict the total power from various renewable resources.

## References

- [1] Goethe-Universität Frankfurt am Main. (2015, May 28). Extreme global warming of the Cretaceous period punctuated with significant global cooling. ScienceDaily. Retrieved July 21, 2021 from [www.sciencedaily.com/releases/2015/05/150528083818.htm](http://www.sciencedaily.com/releases/2015/05/150528083818.htm).
- [2] Arega Bazezew Berlie, Global Warming: A Review of the Debates on the Causes, Consequences and Politics of Global Response. Ghana Journal of Geography Vol. 10 (1), 2018 pages 144–164.
- [3] Silberg, B. (2016). Why a half-degree temperature rise is a big deal – Climate Change: Vital Signs of the Planet. Retrieved 21 July 2021, from <https://climate.nasa.gov/news/2458/why-a-half-degree-temperature-rise-is-a-big-deal/>.
- [4] Kaplan, S. (2021, July 6). *Climate change has gotten deadly. It will get worse.* The Washington Post. <https://www.washingtonpost.com/climate-environment/2021/07/03/climate-change-heat-dome-death/>.
- [5] European Commission. (2017, June 28). *Causes of climate change.* Climate Action - European Commission. [https://ec.europa.eu/clima/change/causes\\_en](https://ec.europa.eu/clima/change/causes_en).
- [6] MacMillan, A., & Turrentine, J. (2021, April 7). *Global Warming 101.* NRDC. <https://www.nrdc.org/stories/global-warming-101#causes>.



- [7] United States Environmental Protection Agency. (2021, April 14). *Overview-of-Greenhouse-Gases*. EPA. <https://www.epa.gov/ghgemissions/overview-greenhouse-gases>.
- [8] Panwar, N. L., Kaushik, S. C., & Kothari, S. (2011). Role of renewable energy sources in environmental protection: A review. *Renewable and Sustainable Energy Reviews*, 15 (3), 1513–1524. <https://doi.org/10.1016/j.rser.2010.11.037>.
- [9] Lund, H. (2007). Renewable energy strategies for sustainable development. *Energy*, 32 (6), 912–919. <https://doi.org/10.1016/j.energy.2006.10.017>.
- [10] Vezmar, S., Spajić, A., Topić, D., Šljivac, D., & Jozsa, L. (2014). Positive and negative impacts of renewable energy sources. *International journal of electrical and computer engineering systems*, 5 (2), 47-55.
- [11] Vicedo-Cabrera, A. M., Scovronick, N., Sera, F., Royé, D., Schneider, R., Tobias, A., Astrom, C., Guo, Y., Honda, Y., Hondula, D. M., Abrutzky, R., Tong, S., Coelho, M. de, Saldiva, P. H., Lavigne, E., Correa, P. M., Ortega, N. V., Kan, H., Osorio, S., ... Gasparrini, A. (2021). The burden of heat-related mortality attributable to recent human-induced climate change. *Nature Climate Change*, 11 (6), 492–500. <https://doi.org/10.1038/s41558-021-01058-x>.
- [12] TWI. (n.d.). *What is Renewable Energy? - Definition, Types, Benefits and Challenges*. Retrieved August 2, 2021, from <https://www.twi-global.com/technical-knowledge/faqs/renewable-energy>.
- [13] Green, E. (2021, June 2). *Wind vs. Solar — Which Power Source Is Better?* Elemental Green. <https://elemental.green/wind-vs-solar-which-power-source-is-better/>.
- [14] Teixeira-Pinto, A. (2021, July 26). *2 K-nearest Neighbours Regression | Machine Learning for Biostatistics*. Biostatistics Statistics Collaboration of Australia. [https://bookdown.org/tpinto\\_home/Regression-and-Classification/k-nearest-neighbours-regression.html](https://bookdown.org/tpinto_home/Regression-and-Classification/k-nearest-neighbours-regression.html).
- [15] scikit-learn developers. (n.d.). *sklearn. neighbors. KNeighborsRegressor*. scikit learn. <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsRegressor.html>.
- [16] Brownlee, J. (2021, February 16). *A Gentle Introduction to XGBoost for Applied Machine Learning*. Machine Learning Mastery. <https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/>.
- [17] Grogan, M. (2020, November 7). *Predicting Weekly Hotel Cancellations with XGBRegressor*. Medium. <https://towardsdatascience.com/predicting-weekly-hotel-cancellations-with-xgbregressor-d73eb74a8624>.
- [18] Brownlee, J. (2020, August 26). *Train-Test split for Evaluating machine learning algorithms*. Machine Learning Mastery. <https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/>.