

Impact of Economic Growth on climate injustice and human health

By

Shamia Ferdousy (171014038)

Table of Content

Approval Page	Page 3
Declaration	Page 4
Acknowledgement	Page 5
Abstract	Page 6

List of Tables

1 Coefficient data of linear regression	Page 15
2 Evaluation metrics of regression models	Page 16

List of Figures

1 Environmental Kuznets Curve	Page 8
2 Conceptual Framework	Page 9
3 Rising temperature trend with frequent fluctuations	Page 14
4 Flow diagram	Page 15
5 Per capita carbon emission and income with average temperature	Page 17
6 Temperature in different latitude	Page 17
7 Comparison of performance metrics of different regression models	Page 18

Chapter 1: Introduction

1.1 Introduction	Page 6
1.2 Motivation	Page 8
1.3 Rationale of the Study	Page 8
1.4 Research Questions	Page 10
1.5 Expected Output	Page 10

Chapter 2: Background

2.1 Preliminaries/Terminologies	Page 10
2.2 Related Works	Page 11
2.3 Comparative Analysis and Summary	Page 12
2.4 Scope of the Problem	Page 12
2.5 Challenges	Page 12

Chapter 3: Research Methodology

3.1 Dataset Utilized	Page 13
3.3 Statistical Analysis	Page 13
3.4 Proposed Methodology	Page 14
3.5 Implementation Requirements	Page 16

Chapter 4: Experimental Results and Discussion

4.1 Experimental Results & Analysis	Page 17
4.2 Discussion	Page 18

Chapter 5: Impact on Society, Environment and Sustainability

5.1 Impact on Society	Page 18
5.2 Impact on Environment	Page 19
5.3 Ethical Aspects	Page 19
5.4 Sustainability Plan	Page 19

Chapter 6: Summary, Conclusion, Recommendation and Implication for Future Research

6.1 Summary of the Study	Page 20
6.2 Conclusions	Page 20
6.3 Implication for Further Study	Page 20
References	Page 21
Appendix A	Page 24
Appendix B	Page 26

Approved by -**Bijan Paul**

Senior Lecturer

Department of Computer Science and Engineering

University of Liberal Arts Bangladesh (ULAB)

Syed Akhter Hossain, Ph.D.

Professor and Head

Department of Computer Science and Engineering

University of Liberal Arts Bangladesh (ULAB)

Supervisor

Head of the Department

Declaration

Work in this research titled “Impact of Economic Growth on climate injustice and human health” has been carried out by me in the Department of Computer Science and Engineering, ULAB. The information obtained from literature has been duly acknowledged in the report and in the Reference section. No part of this report has been previously presented for another degree or diploma at this or any other institution.

Acknowledgement

This study is supported by the Department of Computer Science and Engineering, ULAB. The completion of this study could not have been possible without the guidance and help of my advisor, Bijan Paul.

Also, I would like to thank the panel of faculties, Khan Raqib Mahmud and Golam Kibria. Their questions, advice, and guidance helped me to notice the blind spots and improve my work during the whole project.

Abstract

Though on a social and political level, global warming is still a debatable issue, its impact on climate change is undoubtedly intense. With rapid urbanization and industrialization, we can actually feel the changes in temperature trends of both atmosphere and waterbody. Higher temperatures are bringing climate disasters like heat waves, droughts, changing natural meteorological patterns and breaking the ecological balance. Everyday we release thousands of anthropogenic pollutants in the environment through economic activity. The impact of these pollutants on living beings is adverse. We are suffering from different respiratory and cardiac diseases more than ever. Our biodiversity is being extinct. In spite of our economic development, poverty is becoming undefeatable because we are affecting vulnerable areas through development. Every year, more people become climate refugees. Rapid urbanization is also influenced by climate disasters. Hopefully, these situations are avoidable with efficient policy, liability and mass awareness. This study works with historical observational data cornering the global economy, pollution and weather. By analyzing these data, it explores the relationships of different factors with economic growth. It also explains climate injustice, and provides an air quality prediction model.

Chapter 1: Introduction

1.1 Introduction

Over the past three decades, the deaths from chronic respiratory disease increased by 18% [5]. According to the research last year, led by Dr. Min Xie, worldwide, 3.9 million people die from lung diseases [5]. The lead cause behind this is air pollution. It is also the reason for seven million premature deaths. While the advancement of science is prolonging life, mortality is increasing from air-borne diseases such as respiratory infections, lung cancer, obstructive pulmonary disease, stroke, and cardiac diseases.

Increased heat has adverse effects on our environment. Increasing global temperatures is causing the ice sheets to melt. Melting of permafrost due to high temperature is releasing more greenhouse gases into the atmosphere. This is not only affecting the biomes but all of us. Sea level is rising due to the melting of ice caps. This rise will drown half of earth's land area. Every year Bangladesh experiences extreme climate disasters, such as thunderstorms, cyclones, drought, excessive rain, and cold waves. Rising sea level is increasing the salinity in our soil. This salinity intrusion is destroying our forest, causing habitat loss, moreover destroying our agricultural land.

Not only Bangladesh, all the countries close to the ocean are already facing climate disasters. Madagascar, Canada, Philippines, Japan, Fiji are the hotspots most vulnerable to

rising global temperature [6]. Interestingly, their emission rate is much lower than most of the countries. Pollutants from one point become NPS pollution and contaminate the entire atmosphere.

We know the more we use natural resources, the more damage we cause, and we use the most of our resources for manufacturing products, which is the scale to measure our economic growth, our development. Borders may divide our land but we share the same atmosphere. To deal with these environmental, political and ethical issues of concern, climate justice policies are formed. But it is difficult to make informed efficient policies when we do not have enough information about the situation.

This study explores the historical meteorological data in order to understand the factors behind a country's temperature trend and its relationship with climate injustice, biodiversity and human health.

1.2 Motivation

While doing my minor study in Sustainable Development, I noticed that we all talk about how developing countries are contributing to the climate problem. Policymakers and developed countries blame developing countries for environmental pollution. We know that economic development plays a great role in environmental quality.

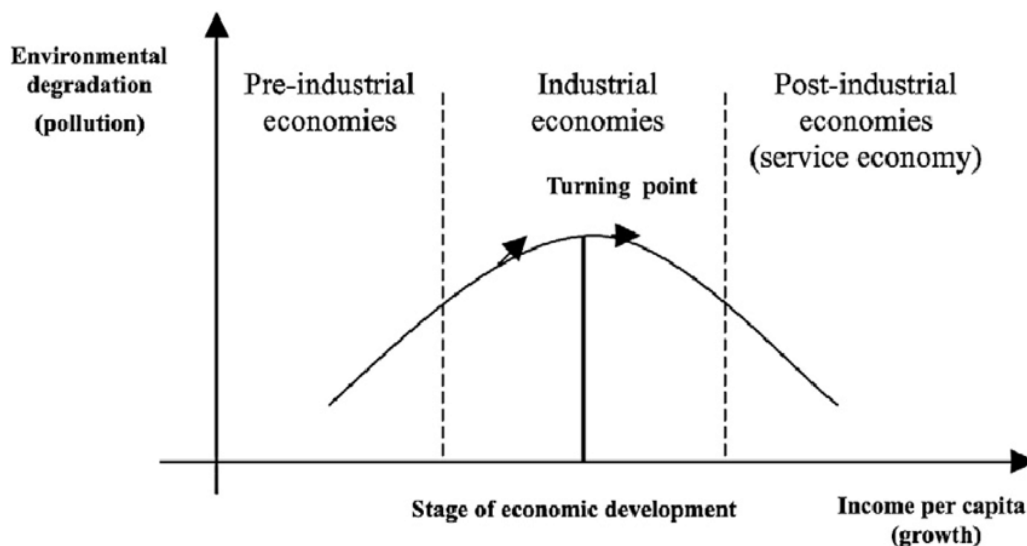


Figure 1. Environmental Kuznets Curve

In 1950, Simon Kuznets showed the hypothesized relationship between environmental quality and economic development [7]. His hypothesis suggests that the degradation of the environment will increase with industrialization and economic development. Environmental degradation will continue up to some level of economic growth, and tend to

decrease after a certain income level is reached [8]. Autoregressive distributed lag testing with data of Bangladesh from 1971 to 2010 confirms a long-run relationship between our environmental quality and economic growth [9]. Considering the global situation, Bangladesh is a nano emitter, but our process of development is leaving carbon footprints significant enough to alter our environment at local level.

Not only Bangladesh, every country going through industrialization and close to the sea is experiencing extreme consequences of contaminated atmosphere. While the development process is creating pollutants, rapid urbanization all over the world is speeding up pollution. Air pollutants in urban areas are causing respiratory and cardiac diseases. An empirical study regarding air pollution found that the level of suspended particulate matter in the atmosphere decreases with the increment of per capita income [10]. Until our income emission takes the u-turn, we are breathing contaminated air. As a result, people living in larger, more polluted cities are 15% to 17% more likely to die a premature death than people living in areas with cleaner air [11].

Industrialization in some parts of the world happened later than others. So economic development will also take time to turn the situation around. Meanwhile, we can see developed countries emit more greenhouse gases than developing countries. These pollutants are largely responsible for global warming, which is causing climate change and contributing to the poverty problem by increasing climate refugees. With these situations in mind, I planned to focus on these issues throughout this study.

Among all the meteorological data, global average temperatures is commonly used as an indicator of climate change. So I explored the relationship between global average temperatures and economic growth, and tried to show its impact on the environment and everything in it.

1.3 Rationale of the Study

The geographical location of Bangladesh creates some disadvantages and threats. For example, we have a 710 km long coast. Out of the 36 most deadly tropical cyclones in recorded history, 26 have been over the Bay of Bengal [12]. Every year we count economic loss worth more than 15% of our total income [13]. On top of that every year thousands of people become climate refugees which increases our poverty rate. By 2050, one in every seven people in Bangladesh will be displaced by climate change [14]. With information about global temperature pattern, pollution, we can make informed decisions to sustain our development. That's why we explored the average temperature of each country.

Pollutants are mostly from agriculture and industries. Along with the pollutants from industrial production (NO_2 , Pb, SO_2), dust, dirt, fog, powdery substances of amorphous carbon are released into the air by natural, industrial, or daily household activities. Oxygen,

nitrogen absorb energy with wavelengths of less than 200 nanometers. Earth reflects some of the energy from sunlight as infrared waves which travel at a wavelength of 700 to 1,000,000 nanometers. The atomic form of carbon allows it to interact with a wide range of wavelengths, between 2,000 and 15,000 nanometers. This overlaps with the wavelength of the infrared waves. The carbons we release in our atmosphere act like a blanket by soaking the energy radiating from the earth's surface as infrared waves. The extra energy causes vibration in molecules. After some time, the molecule releases the extra energy by emitting infrared photons or heat. This increases the temperature of the earth's surface. The more our income grows, the more we consume. This pattern requires more production. And the whole cycle of consumption creates more carbon footprint. So I added per capita income and emission data to our study.

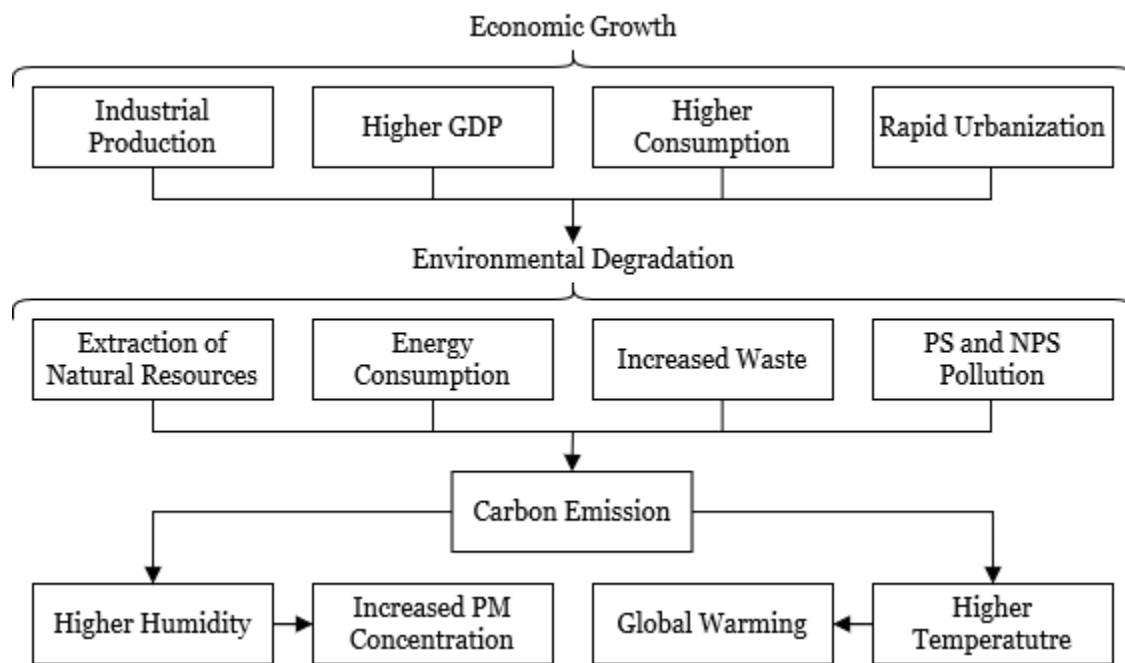


Figure 2. Conceptual Framework

Increased heat evaporates more water from the earth's water body and increases both temperature and humidity. Higher temperature increases health risk for patients with chronic conditions like diabetes, cardiovascular, respiratory, and cerebrovascular disease. Also, increased humidity means more moisture in the air. When humidity is high, our sweat cannot evaporate. Our body needs to work harder to cool. So it increases heart rate, respiration, and blood circulation depth. As a result, we experience excessive sweat and often heat stroke.

Increased humidity also creates more particulate matter. Particulate matters are basically small particles mixed with liquid droplets in air. Moisture in air allows more particles to mix with fog or liquid droplets, and increases the concentration of SPM and PM in air. To

consider the impact of this on our health, we studied meteorological data and air quality in our study.

Increased carbon dioxide, temperature, and humidity increase bio productivity. Increased bioproductivity disrupts the natural food chain, and also creates more natural pollutants. Probably countries with lower emissions are suffering higher heat because of these pollutants. This study focuses on that matter too. By analyzing these mentioned observational data, we will find a relation among the factors, identify the vulnerable hotspots and take action to improve our air quality.

1.4 Research Questions

Throughout this research, we will explore the indicators of a country's temperature trend. We will also see how vulnerable countries are facing climate disasters. What factors play roles in the temperature trend of a country? Which areas are more vulnerable to rising temperature trends? How rising temperature and pollution affect human health?

1.5 Expected Output

From this study I expect to

- explain climate injustice,
- determine the factors behind a country's temperature trend, and
- design a prediction model for AQI prediction to measure impact on our health.

Chapter 2: Background

2.1 Preliminaries/Terminologies

Biome: Biome means a large group of living organisms adapted to a specific pattern of climate.

Equator: Equator is the imaginary line around the middle of the earth. It divides the globe at 0 degree latitude into a Northern Hemisphere and a Southern Hemisphere.

Ensemble: Ensemble is a combination of individual contributing models that form a powerful model together. Better predictions and performance can be achieved with ensembles.

Particulate Matter: Particulate matter is solid particles mixed with liquid droplets in the air. The size of particles are so small that they are not visible, but while we breathe it reaches our respiratory and cardiac system.

Prema frost: Ground that continuously stays under the 0 °C or 32 ° freezes the materials in soil and organic bodies, it is called prema frost.

Tropic of Cancer: The Tropic of Cancer is the northern circle of latitude. Sun is directly overhead on the Tropic of Cancer, this prolongs the day time in June.

Tundra biome: Biome, where bioproductivity is boosted by low temperatures and short growing seasons is called tundra biome.

2.2 Related Works

Jo-Hui Chen and Yu-Fang Huang explored the relationship between carbon emissions and economic growth. Their study showed if economic activity does not involve using and processing natural resources [22], it still increases carbon emission. Any kind of energy consumption leaves carbon footprints that are almost impossible to erase with development.

Gillian L Hughes and his team showed that over the 53 years (1951–2003), monthly minimum temperatures have increased approximately 6.7°C [16]. This increment has an impact on the global average temperature. Climate injustice is understood by asymmetry in cost-benefit and by weakening capabilities of individuals [20].

Numerical models, like WRF-Chem, are widely used for air quality forecasting. Weather Research and Forecasting model coupled with Chemistry(WRF-Chem) investigates air quality by simulating the emission, traffic, chemical reaction and transformation of different gases of the atmosphere.

AQI prediction model prepared by Ma et al. predicted AQI from particulate matter. Though their study was limited to Shanghai, China, it was reliable and useful. Aditya et al. in their work predicted the level of particulate matter from atmospheric components. For data classification, they used logistic regression. This process determined the samples as polluted or clean. To make future predictions based on previous records, they used autoregression.

For AQI prediction, nearest neighbor classification was introduced in Cover and Hart's work [18]. This approach was later advanced by E. G. Dragomir [18]. Castelli et al. and E. G. Dragomir worked with toxic pollutants like CO₂, CO, SO₂, NO_x, PM_{2.5} in order to measure air quality in their separate work [19]. Machine learning approach proposed by Castelli et al. to predict air quality in California introduced Support Vector Machines (SVMs) in their

classification method to identify the optimal separating hyperplane between classes [19]. Hyperlane is the space in between the support vectors.

2.3 Comparative Analysis and Summary

There are several studies related to climate change, its effects, and indicators, but only 11% of that used spatial data, 5% considered multilevel analysis [15]. Most of the air quality prediction models are using traditional statistical prediction. We think the performance of machine learning-based statistical prediction models will be better than the performance of statistical methods. Though G. L. Hughes found a rising trend in minimum temperature, maximum monthly temperature is unchanged [16]. Warmer environments can exceed the tolerance of species. The difference in maximum and minimum temperature can be measured to see the impact on biodiversity.

2.4 Scope of the Problem

Works discussed above have surely created significant impact in solving the problems. But there is always more to contribute. This study will cover historical analysis from 1951, right after the beginning of the third industrial revolution. By analyzing the global temperature data, we hope to find a trend, and explain that trend with economic development. We will also focus on the impact of environmental degradation caused by economic development. To measure the adverse impact on human health this study proposes a machine learning model that predicts air quality index.

2.5 Challenges

Handling missing values was difficult. Though the traditional approach is using predicted values, this could temper the outcome. So we ignored the tuples with missing value. After cleaning the incomplete dataset, the training dataset became smaller. Most cited data sources, like NOAA, were not accessible. A request for a recent dataset was sent to NASA's Earth Observing System Data and Information System(EOSDIS), but the acceptance procedure exceeded our time constraint.

Getting real time meteorological data for prediction was difficult too. Web scraping took a long time to collect data. That's why the prediction system here uses real time data from OpenWeatherMap API calls. This call uses a subscription key registered to specifically my user account. It was not possible to make it available for all to use.

Chapter 3: Research Methodology

3.1 Dataset Utilized

In this work, 5 datasets are used from 5 different sources. Annual per capita income-emission data from 1750 to 2019 is obtained from ourworldindata.org. Dataset was published by “Our World in Data based on the Global Carbon Project” in 2020 [2].

Global average earth surface temperature data is taken from Berkeley Earth. This dataset, affiliated with Lawrence Berkeley National Laboratory is from The Berkeley Earth Surface Temperature Study. It combines 1.6 billion temperature reports starting in 1750 from 16 pre-existing archives [3]. Dataset is publicly available to download on kaggle.com. Geographic location of 195 countries and 50 independent territories are collected from Google Public data. Meteorological data was collected from TuTiempo.com. HTML pages of each month of the year were downloaded, and from those data table datasets of each year were imported to csv files.

Collected dataset had multiple missing and incomplete information. In order to obtain appropriate datasets for this study, raw data was cleaned, constructed and integrated by an iterative process into an acceptable dataset.

First step of the process was data selection. From the gapminder_data dataframe, unnecessary columns such as population were dropped. Uncertainty columns were deleted from the datasets obtained from Berkeley Earth. During the cleaning process, multiple rows with missing carbon emission data were deleted. Preprocessing also ignored the years with less than 100 observations. Unnecessary columns like atmospheric pressure at sea level, fog, rain, storm and snow indicators were deleted. During integration, all the dataset were merged together into one dataframe, except the weather data. Weather data were stored in a different dataframe named aqi_data.

Variable description of each data frame is shown in Appendix A.

3.2 Statistical Analysis

Global average temperature record shows a rising trend with frequent fluctuations. These fluctuations can be the reason for extreme weather events like El Niño–Southern Oscillation. Reason behind this fluctuation can be the difference of carbon emission each year. A sudden rise can be noticed from 1980. To explore the factors behind the temperature trend, a scatter plot of all variables with histogram is created (see Appendix A).

Average temperature of different countries shows negative skewness in the scatter plot. Also, we can see, most of the countries are nano emitters. There are some outliers with high CO2 emission. Though common practise is removing the outliers, we are not removing them. We assume the outliers play an important role in global temperature trends. Per capita income and emission reflects the same pattern. This similarity supports the hypothesis in the conceptual framework presented above.

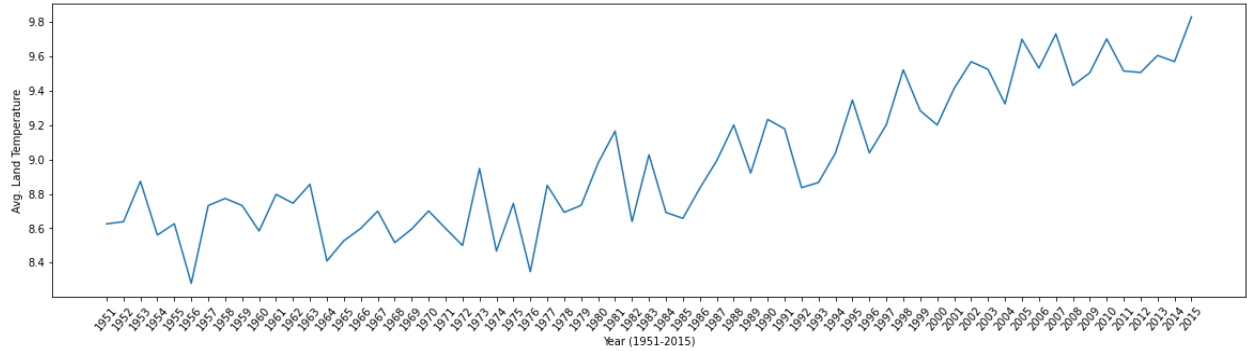


Figure 3. Rising temperature trend with frequent fluctuations

The random distributions in latitude and longitude suggest the data set we chose was large and covered maximum countries. Average temperature sees higher in lower latitude, that explains the higher carbon emission. Carbons here are mostly from bioproductivity. To interpret the relationships of the variables a OLS regression result is created.

Assuming $\hat{\beta}$ = ordinary least squares estimator, x = matrix regressor variable, T = matrix transpose, y = response value variable vector.

$$\hat{\beta} = (xTx)^{-1}xTy$$

The R-squared value in the OLS result is 0.399. A country's per capita carbon emission, income, and geolocation can explain almost 40% of the variance of average temperature.

3.3 Proposed Methodology

The process followed to carry out the research, is shown in the flow diagram. To estimate relationships between independent and dependent variables, Ordinary Least Squares (OLS) is a helpful method. This statistical analysis method is used to estimate the relationship between independent and dependent variables. The relationship is estimated by minimizing the sum of the squares in the difference between the observed and predicted values of the dependent variable configured as a straight line.

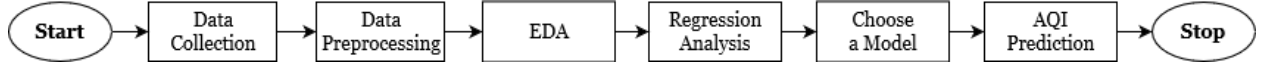


Figure 4. Flow diagram

For the air quality prediction model, we started with linear regression. Coefficient data here suggests that holding all other features fixed, a 1 unit increase in average temperature (T) is associated with decrease of -13.537122 in particulate matter. A 1 unit increase in maximum temperature (TM) is associated with an increase of 3.938286 in particulate matter. A 1 unit increase in minimum temperature (Tm) is associated with a decrease of -3.338939 in particulate matter.

Table 1. Coefficient data of linear regression

T	TM	Tm	H	PP	VV	V	VM
-13.537122	3.938286	-3.338939	-0.267911	-1.110479	3.793134	3.763294	-1.021746

Kth nearest neighbour was used in multiple previous studies mentioned above. We tried k-th nearest neighbour. But the performance of the linear regression tree was better than the KNN model.

Decision tree regression analysis was performed to see if the performance is better than KNN. Decision tree approach builds a model in the form of a tree by breaking the dataset into smaller nodes. Before splitting into nodes it compares the MSE values as performance metrics. This process increases the homogeneity of the result. Performance of the decision tree regression model was better than KNN, but not satisfactory.

$$MSE = \frac{1}{n} \sum_{i=1}^n (observed_i - predicted_i)$$

We know, the accuracy of a predictive model can be boosted in two ways, by embracing feature engineering or by applying boosting algorithms straight away. Gradient boosting approaches consist of three elements- (a) a differentiable loss function to be optimized, (b) weak prediction models, (c) an additive model. Compared to other implementations of gradient boosting, extreme gradient boosting works faster. It uses decision trees as base estimators. To avoid overfitting it allows us to control tree depth.

Extreme gradient boosting algorithm is commonly used for both classification and regression. The concept of boosting came from the idea of modifying relatively weak hypotheses into a better one. Gradient boosting algorithm basically relies on the concept that the best possible next model will minimize the error in prediction while it is combined with previous models. It calculates performance metrics like mean squared error for each possible split and chooses the lowest one to use in the tree.

To see if other supervised models perform better, random forest regression was applied. Accuracy of this model is much higher, but not higher than XGB regression. So XGB regression models were saved and later used for prediction.

Evaluation metrics are stored and compared later to find the perfect model. Evaluation metrics of five regression models are presented in table.

Table 2. Evaluation metrics regression models

Evaluation Metrics	Decision Tree Regression	KNN Regression	Linear Regression	Random Forest Regression	XG Boosting Regression
MAE	45.950376	46.516101	46.737799	34.047186	8.204667
MSE	4693.646048	4693.646048	4263.299194	2268.45924	275.520024
RMSE	67.024351	68.510189	65.293945	47.628345	16.598796

Three locations of Bangladesh were selected to perform the prediction based on the weather pattern. The Tropic of Cancer goes through the centre of Bangladesh, over Dhaka, Khulna and Chittagong. Dhaka is the capital and centre of urbanization, so pollutants here affect the weather. Same temperature pattern is seen in Khulna and Chittagong as an effect on the sun's position. If we divide the map vertically, eastern part has Rajshahi with significant weather attribute, temperature. Western part of Bangladesh has Sylhet with the highest precipitation rate, another significant attribute.

3.4 Implementation Requirements

The implementation of the proposed methodology requires a python environment with multiple modules to run different analyses. A virtual environment is created with the Anaconda environment and package management system.

Numpy 1.19.2 is used to process arrays and objects. While NumPy is memory efficient, Pandas series provides better performance when the dataset is larger. Pandas 1.3.2 is used as a data structure and analysis tool. Pandas DataFrame is used to access the datasets as two dimensional data structures. Data sets are in csv file, to easily handle the columns, Pandas Series is used. To carry out the statistical models 0.5.3 version of patsy. Seaborn is a popular library that works fine with dataframes and the Pandas library. That's why Seaborn 0.11.2 is used for statistical data visualization. To visualize the decision trees in DOT language, PyDotPlus 2.0.2 is used. To handle mathematical problems, scipy 1.6.2 is used. Pickle is used to save the model as a file. Later that file is used to deserialize the model for new predictions. To measure the performance of different models, metrics modules from scikit-learn 0.24.2 are used. To run the web application, the Flask web framework needs to be installed.

Chapter 4: Experimental Results and Discussion

4.1 Experimental Results & Analysis

We see GDP and carbon emission reflects the same pattern in their distribution. We take industrial carbon emission as an indicator of a country's temperature trend.

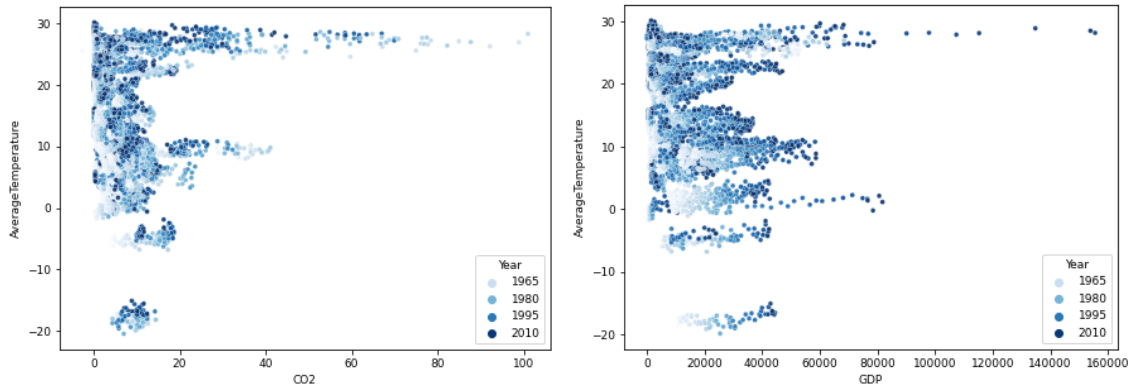


Figure 5. Per capita carbon emission and income with average temperature

While going through factors, we see there is a relationship between temperature of a location and its distance from the equator. The closer a country is to the equator line, the more heat it experiences in its atmosphere. This excessive heat increases the respiration rate, extends the growing season, and increases the overall natural carbon emission throughout the process. Lengthy growing seasons are adverse for biodiversity too. The change in respiration and bioproductivity changes the ecological balance. Sometimes it destroys the food chain, and habitat. That's why species from these locations are more likely to be extinct. This explains why countries with lower emissions still face climate disasters.

From the OLS regression (see Appendix B), we can get this linear model. Here, Avg T = Average temperature. If the carbon dioxide in the atmosphere increases one unit, it will affect the temperature and increase it by 0.0674 units.

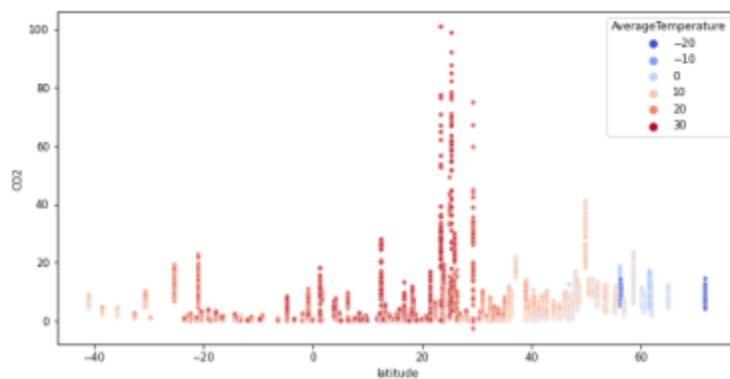


Figure 6. Temperature in different latitude

$$Avg\ T = 24.2364 + (-0.0002 \times GDP) + (0.0674 \times CO_2) + (-0.2007 \times Latitude) + (-0.0066 \times Longitude)$$

Among the five regression models, extreme gradient boosting regression provided the best result. And based on this result, the app was created.

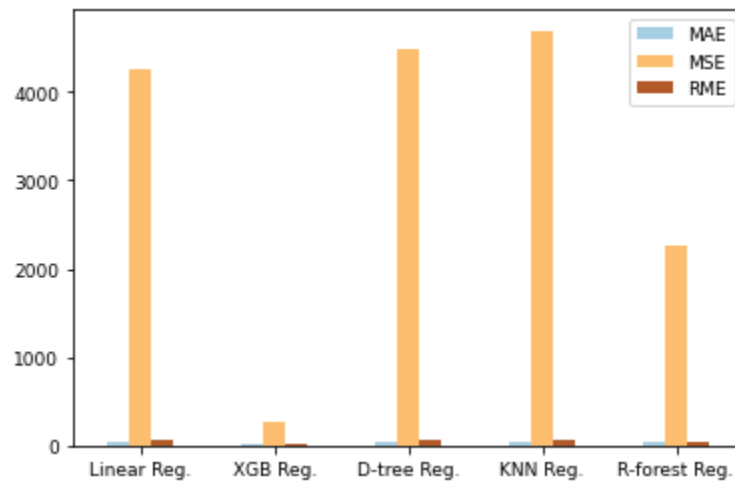


Figure. Comparison of performance metrics of different regression models

4.2 Discussion

Emission data is based on territorial production-based emissions, which do not account for emissions embedded in traded goods. Temperature data in both `global_temperatures` and `country_temperatures` dataframe was collected using mercury thermometers. It is obvious that variation in time impacted the measurements. Weather stations moved due to construction and other causes, mainly for the construction of airports in 1940. Change of location tempered the data collection process. Also use of electronic thermometers instead of mercury thermometers can have cooling bias. Confidence interval of these data is 95%.

Chapter 5: Impact on Society, Environment and Sustainability

5.1 Impact on Society

Stroke, respiratory and cardiac diseases are the most common causes of death in our society. All of them originate from air pollution from industries and household activities. Our waste management system destroys more than it manages. We pollute our environment with our habit more than we imagine. This study contributes to SDG 3, Good Health and Wellbeing, by focusing on the health impact of economic carbon emission. By analyzing the industrial pollution, we show how our industries and infrastructures need innovative modification. This contributes to SDG 9, Industry and Innovation and Infrastructure. This study discusses how our consumption patterns create a chain of

pollution. This discussion pays attention to the SDG 12, Responsible Consumption and Production.

5.2 Impact on Environment

The process of this study is totally electronics device based. The least adverse environmental impact was on the environment while carrying out this project. This study deals with environmental issues and provides ways to seek solutions. With the outcome of this study, we can take necessary actions and minimize environmental impacts of our actions. This study promotes sustainability. Growth only is not sustainable development, not even development, if it is causing environmental degradation in the process. This paper shows and suggests minimizing the pollution from our actions and saving our environment, our climate, and contributes to the SDG 13, climate action.

5.3 Ethical Aspects

All the datasets used in this project are from outside sources. Data of Gapminder Tools downloaded from ourworldindata.org has the creative commons that offers us to use the dataset. Kaggle is open source and public, but datasets downloaded from Kaggle are uploaded by Berkeley Earth, collected in collaboration with Lawrence Berkeley National Laboratory. Though data shared in TuTiempo is not for public use, which may raise an issue. But collected data from HTML pages of TuTiempo site are not used for any commercial use, and also shared data displayed by the website is mostly for public consumption. Real time weather data was collected from registered API calls and have no ethical issue.

5.4 Sustainability Plan

Models used in this work are quite accurate and efficient. This work can be sustained by collecting observational data regularly. With recent data the same model can be run, and new knowledge can be created. This will help us to change or modify our policies, plans and lifestyle.

Chapter 6: Summary, Conclusion, Recommendation and Implication for Future Research

6.1 Summary of the Study

This study explored historical economic, meteorological data and found carbon emission from economic activities is the factor behind a country's temperature trend. With this,

latitude should also be considered as an indicator. Countries with lower latitude suffer from higher temperature and are more vulnerable to climate change. With these insights, this study provides an AQI prediction model based on weather.

6.2 Conclusions

Temperature due to carbon emission is the main attribute focused in this study. Carbon, the most valuable and important element, is becoming deadly, just because we are not watching our actions. Authorities should take Environmental Impact Assessments reports from manufacturing industries more frequently. Policies should make us more liable and responsible for our actions. Without awareness and actions with positive impact from everyone, we will push the earth's limits and destroy the environment. To maintain healthy air quality, NPS pollution should be monitored while controlling PS pollutants. When AQI is high, we can wear masks and avoid outdoor activity.

6.3 Implication for Further Study

The coexistence of higher income and emission puts policymakers in a difficult position. To overcome this challenge, further studies can be carried out. This study only involves carbon emission due to economic activities. Studies that take carbon emission due to indoor or household activities in account should be done. This way, we can actually hold individuals accountable for environmental degradation and change consumers' behaviour. The cycle of production and pollution can be broken. Air quality index prediction model can be used for creating an interactive mobile application that can suggest ways of avoiding pollutants.

References

- [1] G. L. Hughes, S. S. Rao and T. S. Rao. (2006, September 6). "Statistical analysis and time-series models for minimum/maximum temperatures in the Antarctic Peninsula." Royal Society [online]. vol. 463, issue 2077. Available: <https://royalsocietypublishing.org/doi/full/10.1098/rspa.2006.1766>

- [2] P. Friedlingstein, M. O'Sullivan, M. W. Jones et al. (2020). "Global Carbon Budget 2020." *Earth System Science Data* [Online], vol. 12, issue 4. Available: <https://essd.copernicus.org/articles/12/3269/2020/>

- [3] Berkeley Earth. "Climate Change: Earth Surface Temperature Data Exploring global temperatures since 1750," May 2017. Available: <http://berkeleyearth.org/data/>

- [4] W. Adger, S. Hanson, M. Rahman, R. Nicholls. C. Hutton, M. Salehin. (2018). *Ecosystem Services for Well-Being in Deltas*. Available: https://doi.org/10.1007/978-3-319-71093-8_18

- [5] M. Xie, X. Liu, X. Cao et al. (2020, February 11). "Trends in prevalence and incidence of chronic respiratory diseases from 1990 to 2017." *Respiratory Research* [Online], vol. 21. Available: <https://doi.org/10.1186/s12931-020-1291-8>

- [6] IBERDROLA. (2019). COUNTRIES MOST AFFECTED BY CLIMATE CHANGE : Which countries are most threatened by and vulnerable to climate change?. Available: <https://www.iberdrola.com/environment/top-countries-most-affected-by-climate-change>

- [7] A. Aarthi, P. Gayathri, N. R. Gomathi , S. Kalaiselvi , Dr. V. Gomathi. (2020, Mar). "Air Quality Prediction Through Regression Model." *International Journal of Scientific & Technology Research* [Online]. Vol. 9, issue 3, pp 923-928. Available: <http://www.ijstr.org/final-print/mar2020/Air-Quality-Prediction-Through-Regression-Model.pdf>

- [8] E. Nuroglu, R. M. Kunst. (2018, Jan 7). *Industrial Policy and Sustainable Growth* [Online]. Available: https://doi.org/10.1007/978-981-10-5741-0_12

- [9] F. Islam, M. Shahbaz, M. S. Butt. (2013, December). "Is There an Environmental Kuznets Curve for Bangladesh? Evidence from ARDL Bounds Testing Approach". *Bangladesh Development Studies*. Vol. 36, issue.4. Available: <https://ideas.repec.org/a/ris/badest/0563.html>

- [10] S. Dinda, D. Coondoo, M. Pal. (2000). "Air quality and economic growth: an empirical study". *Ecological Economics*. Vol. 34, issue 3. (Pages 409-423). Available: [https://doi.org/10.1016/S0921-8009\(00\)00179-8](https://doi.org/10.1016/S0921-8009(00)00179-8)
- [11] K. Ito, R. Mathes, Z. Ross, A. Nadas, G. Thurston, T. Matte. (2011). "Fine particulate matter constituents associated with cardiovascular hospitalizations and mortality in New York City." *Environ Health Perspect*. Vol. 119. Pages 467-473. Available: <https://www.ncbi.nlm.nih.gov/pubmed/21463978>
- [12] D. N. Pinto. (2021, May 26). *The Weather Channel* [Online]. Available: <https://weather.com/en-IN/india/news/news/2021-05-26-cyclone-yaas-history-bay-of-bengal-storms-80-cyclone-related-deaths>
- [13] Asian Development Bank. (2014) "Bangladesh Could See Climate Change Losses Reach Over 9% Of GDP". Available: <https://www.adb.org/news/bangladesh-could-see-climate-change-losses-reach-over-9-gdp-report>
- [14] Data Commons - Place Explorer. Available: https://datacommons.org/place/country/BGD?utm_medium=explore&mprop=amount&popt=Emissions&cpv=emittedThing%2CCarbonDioxide&hl=en
- [15] Migration Data Portal. (2018). Percentage of methods used for case studies [Online]. Available: https://www.migrationdataportal.org/themes/environmental_migration_and_statistics
- [16] G. L. Hughes, S. S. Rao and T. S. Rao. (2007, January 8). "Statistical analysis and time-series models for minimum/maximum temperatures in the Antarctic Peninsula." *Royal Society* [Online], vol. 463, issue 2077. Available: <https://doi.org/10.1098/rspa.2006.1766>
- [17] E. G. Dragomir. (2010). "Air Quality Index Prediction using K-Nearest Neighbor Technique". *BULETINUL Universității Petrol – Gaze din Ploiești*. Vol. 62, issue 1/2010, pp 103-108. Available: http://www.unde.ro/bmif/docs/20101/pdf_final_12%20EDragomir.pdf
- [18] Rubal, D. Kumar. (2018). "Evolving Differential evolution method with random forest for prediction of Air Pollution". *Procedia Computer Science*[Online]. Volume 132, pp 824-833. Available: <https://doi.org/10.1016/j.procs.2018.05.094>.

- [19] M. Castelli, F. M. Clemente, A. Popovič, S. Silva and L. Vanneschi. (2020, August 4). "A Machine Learning Approach to Predict Air Quality in California." *Complexity* [Online], vol. 2020. Available: <https://doi.org/10.1155/2020/8049504>
- [20] F. Comim. "Climate Injustice and Development: A capability perspective." (2008, August 18). *Development* [Online], vol. 51, pp 344–349. Available: <https://doi.org/10.1057/dev.2008.36>
- [21] C. R. Aditya, C. R. Deshmukh, D. K. Nayana, P. G. Vidyavastu. (2018, May) "Detection and Prediction of Air Pollution using Machine Learning Models." *IJETT* [Online], volume 59, issue 4. Available: <http://ijettjournal.org/archive/ijett-v59p238>
- [22] Jo-Hui Chen and Yu-Fang Huang. (2013, December). "The Study of the Relationship between Carbon Dioxide (CO₂) Emission and Economic Growth." *Journal of International and Global Economic Studies*, vol. 6, issue 2, pp. 45-61.
- [23] J. Ma, Z. Yu, Y. Qu, J. Xu, Y. Cao. (2020). "Application of the XGBoost Machine Learning Method in PM_{2.5} Prediction: A Case Study of Shanghai." *Aerosol and Air Quality Research*, volume 20, pp 128-138.

Appendix A | Data

- Variable descriptions of utilized datasets

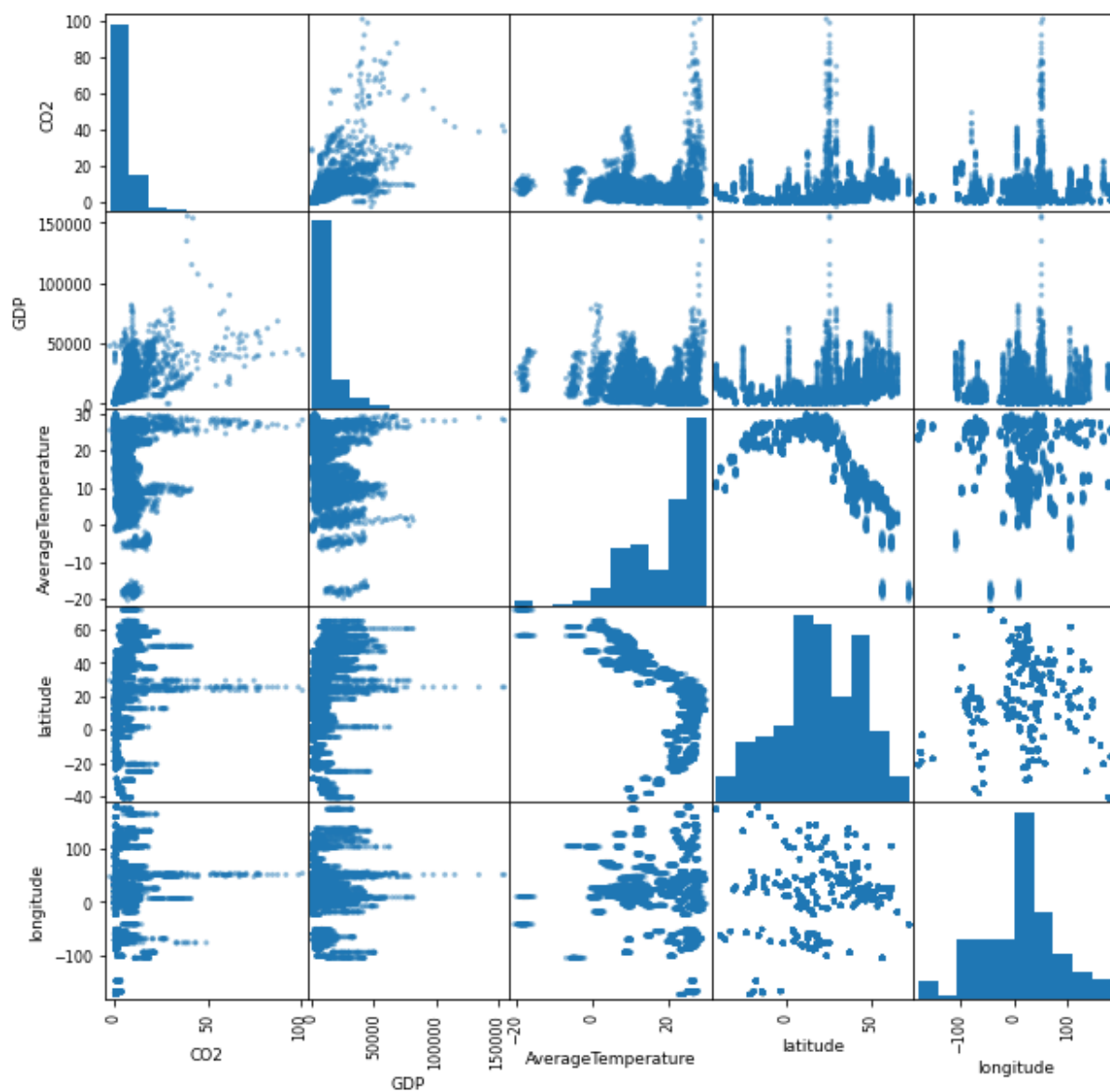
DataFrame		Variables	Unit
1	gapminder_data	Annual CO2 emissions	tonnes per person
		GDP per capita	2011 International-\$
		Total Population	-
2	global_temperatures	Date	dd/mm/yyyy
		Time	hh:mm
		Land Average Temperature	°C
		Land Average Temperature Uncertainty	°C
		Land Max Temperature	°C
		Land Max Temperature Uncertainty	°C
		Land Min Temperature	°C
		Land Min Temperature Uncertainty	°C
		Land and Ocean Average Temperature	°C
		Land and Ocean Average Temperature Uncertainty	°C
3	country_location	Country	-
		Latitude	-
		Longitude	-
		Country Code	-
4	country_temperature	Country	-
		Date	dd/mm/yyyy
		Average Temperature	°C
		Average Temperature Uncertainty	°C
5	aqi_data	Avg. temp	°C
		Max Temp	°C
		Min Temp	°C
		Humidity	
		Precipitation	mm
		Avg Visibility	km
		Wind speed	km/h
		Max sustained wind speed	km/h
		PM2.5 concentration	pm

- **Dataframe sources**

Dataframe	Data	Variables	Source	Link
country_location	Geographical location of countries	Country Code, Latitude, Longitude, Country name	Google Public Data	https://developers.google.com/public-data/docs/canonical/countries_csv
gapminder_data	Per Capita income and carbon emission with population	Entity , Code, Year, CO2, GDP, 145446-annotations, "Total population (Gapminder, HYDE & UN)", Continent	Gapminder Tools	https://ourworldindata.org/grapher/co-emissions-per-capita-vs-gdp-per-capita-internationally-
temp_country	Average Land Temperature in Different Countries	Date, Average Temperature, Average Temperature Uncertainty, Country	Berkeley Earth	https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data?select=GlobalLandTemperaturesByCountry.csv
global_temperatures	Global Average Temperature	Date, Land Avg Temperature, Land Average Temperature Uncertainty, Land Max Temperature, Land Max Temperature Uncertainty, Land Min Temperature, Land Min Temperature Uncertainty, Land And Ocean Average Temperature, Land And Ocean Average Temperature Uncertainty	Berkeley Earth	https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data?select=GlobalTemperatureS.csv
aqi_data	Air pollution	Avg. temp, Max Temp, Min Temp, Humidity, Precipitation, Avg Visibility, Wind speed, Max sustained wind speed, PM2.5 concentration	Tutiempo.net	https://en.tutiempo.net/

Appendix B | Results

- Scatter plot of all variables with histogram



- Ordinary Least Squares Summary

```

=====
                        OLS Regression Results
=====
Dep. Variable:      AverageTemperature      R-squared:                0.399
Model:              OLS                    Adj. R-squared:           0.399
Method:             Least Squares          F-statistic:             1457.
Date:               Wed, 06 Oct 2021        Prob (F-statistic):       0.00
Time:               18:01:47                Log-Likelihood:          -29181.
No. Observations:   8784                   AIC:                     5.837e+04
Df Residuals:       8779                   BIC:                     5.841e+04
Df Model:           4
Covariance Type:    nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
Intercept	24.2364	0.104	232.839	0.000	24.032	24.440
GDP	-0.0002	8.55e-06	-17.650	0.000	-0.000	-0.000
CO2	0.0674	0.013	5.053	0.000	0.041	0.094
latitude	-0.2007	0.003	-62.959	0.000	-0.207	-0.194
longitude	-0.0066	0.001	-5.278	0.000	-0.009	-0.004

```

=====
Omnibus:              1353.073      Durbin-Watson:           0.049
Prob(Omnibus):        0.000        Jarque-Bera (JB):        2628.323
Skew:                 -0.955        Prob(JB):                0.00
Kurtosis:              4.879        Cond. No.                 2.18e+04
=====

```