Analysis of the Effect of CO2 Emissions on Average Temperature Change

Prepared for

Dr. Muhammad Farrukh Shahid

Associate professor of AI and data science,

FAST-NUCES, Karachi

AND

Ms. Mehak Mazhar

Lecturer of AI and Machine Learning,

FAST-NUCES, Karachi

By

Ayan Bin Rizwan (21K-3058)

Student of Artificial Intelligence, FAST-NUCES, Karachi

AND

Muhammad Ayan Badar (21K-3065)

Student of Artificial Intelligence, FAST-NUCES, Karachi

December 13, 2023

Abstract

This project explores the intricate relationship between CO_2 emissions and average temperature change using the 'Agrofood_co2_emission' dataset. With 31 features and 6965 rows, the dataset encompasses various emission sources, measured in kilotons. Leveraging a comprehensive literature review, meticulous data cleaning, and feature engineering, we employ ensemble techniques in our methodology to uncover non-linear trends. Linear models' limitations are underscored, emphasizing the effectiveness of ensemble methods. Visualizations aid in conveying key patterns, and discussions explore the broader implications within climate science. The study concludes with recommendations for future research and practical interventions. This report contributes valuable insights to the understanding of the complex interplay between CO_2 emissions and temperature change, informing both scientific inquiry and policy decisions in the realm of climate change mitigation.

Index

1. Introduction

Climate change, driven by human-induced factors such as CO2 emissions, poses a critical threat to global ecosystems. This project seeks to unravel the intricate relationship between CO2 emissions from various sources and the consequential changes in average temperature. The dataset 'Agrofood_co2_emission' from Kaggle, boasting 31 features and 6965 rows, presents a unique opportunity to delve into the complexities of these interactions. By deciphering the nuances of this relationship, we aim to provide valuable insights for climate science and inform sustainable development policies.

2. Literature Review

A comprehensive literature review reveals the intricate web of interconnections between CO2 emissions and temperature change. Past research highlights the diverse sources of CO2 and their varying impacts on climate. While some studies focus on specific emission sources like savanna and forest fires, others delve into the consequences of agricultural practices such as rice cultivation and pesticide manufacturing. Understanding the findings of these studies aids in shaping the research questions and refining the methodology for our analysis.

3. Data Collection

The 'Agrofood_co2_emission' dataset acts as the backbone of our investigation. This rich dataset encapsulates emissions data from a multitude of sources, providing a holistic view of the anthropogenic impact on the environment. Each feature, measured in kilotons, represents a unique facet of CO2 emissions, making it a robust foundation for our analysis. Alongside, the inclusion of temperature change records enables us to explore correlations and causations within the data.

4. Data Cleaning and Preprocessing

To ensure the reliability of our analysis, we meticulously address issues within the dataset. Missing values, outliers, and inconsistencies are tackled through rigorous data cleaning. By fostering a clean and well-organized dataset, we lay the groundwork for precise and accurate analysis, minimizing potential biases that could affect the credibility of our results.

5. Feature Engineering

Feature engineering becomes imperative to distill meaningful insights from the dataset. Aggregating similar columns helps streamline the information, reducing redundancy and enhancing the dataset's effectiveness. This step not only optimizes computational efficiency but also lays the groundwork for uncovering hidden patterns and relationships within the data.

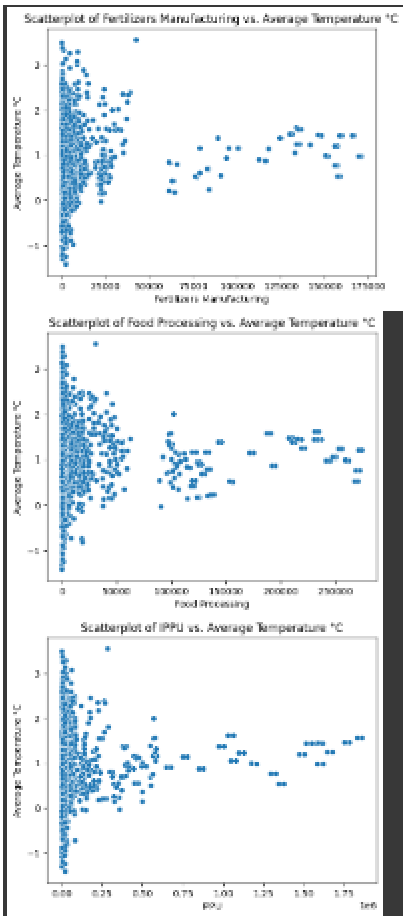
6. Methodology

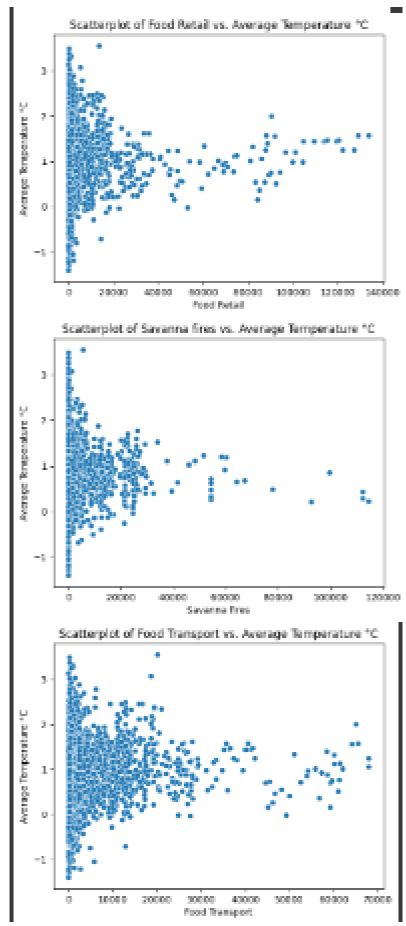
The methodology section meticulously outlines our approach to analyzing the data. Pearson Coefficient Correlation emerges as the method of choice for feature selection, enabling us to identify variables with significant relationships. The categorical feature 'Area' undergoes Label Encoding to facilitate its incorporation into our analysis, and standardization with Standard Scaler ensures a level playing field for all features. Initial attempts with Linear Regression reveal the need for more sophisticated models, prompting the adoption of ensemble techniques to capture non-linear trends.

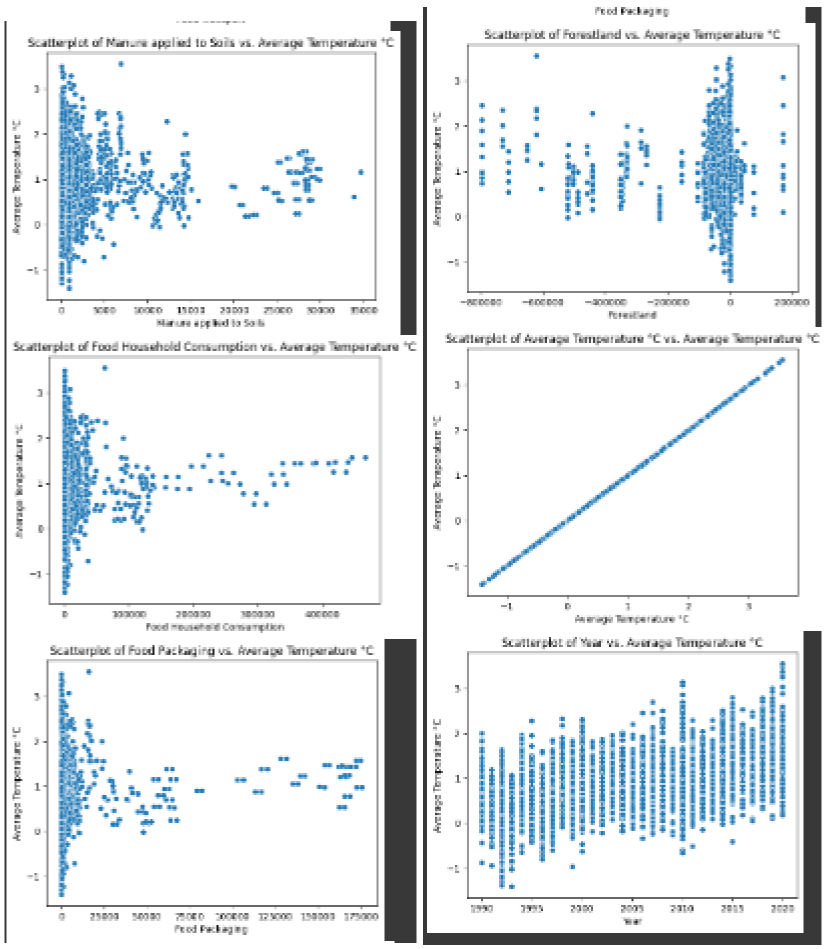
Correlation of input features with target variable

```
Forestland                          -0.048015
Savanna fires                       -0.046508
Forest fires                        -0.037813
Fires in humid tropical forests     -0.034495
Net Forest conversion               -0.030748
Fires in organic soils              -0.023731
Rice Cultivation                    -0.022532
Rural population                    -0.019764
Manure left on Pasture              -0.015928
Area                                -0.012524
Total Population - Male              0.003623
Total Population - Female            0.005456
Agrifood Systems Waste Disposal      0.008995
On-farm Electricity Use              0.009081
Crop Residues                        0.017827
total_emission                       0.019043
Pesticides Manufacturing             0.027960
Drained organic soils (CO2)          0.029030
Manure Management                    0.031886
Urban population                     0.036263
On-farm energy use                   0.038966
Food Packaging                       0.040767
Fertilizers Manufacturing            0.041462
Manure applied to Soils              0.041498
Food Processing                      0.053083
Food Household Consumption           0.053663
IPPU                                 0.060064
Food Retail                          0.073404
Food Transport                       0.075724
Year                                 0.545932
Average Temperature °C               1.000000
Name: Average Temperature °C, dtype: float64
```

Scatter plot of input features with target variable







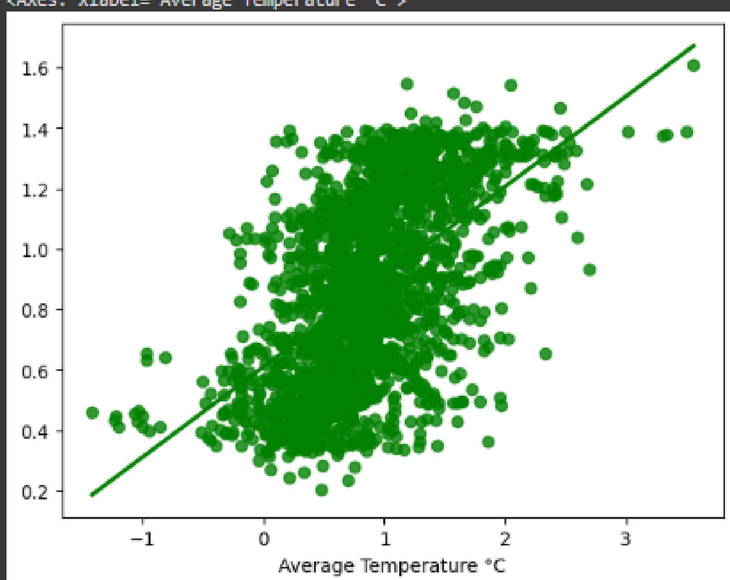
This show that the relationship between the target variable and the input features in non linear.

7. Data Analysis and Results

Linear Regression, our initial attempt, underscores the complexity of the data as the model struggles to establish meaningful correlations. The pivot to ensemble techniques—Random Forest, XGBoost, Gradient Boost, and AdaBoost Regressor—yields promising results. These models prove adept at capturing the intricate non-linear relationships within the dataset, offering more accurate predictions of the impact of CO_2 emissions on average temperature change.

Linear Regression

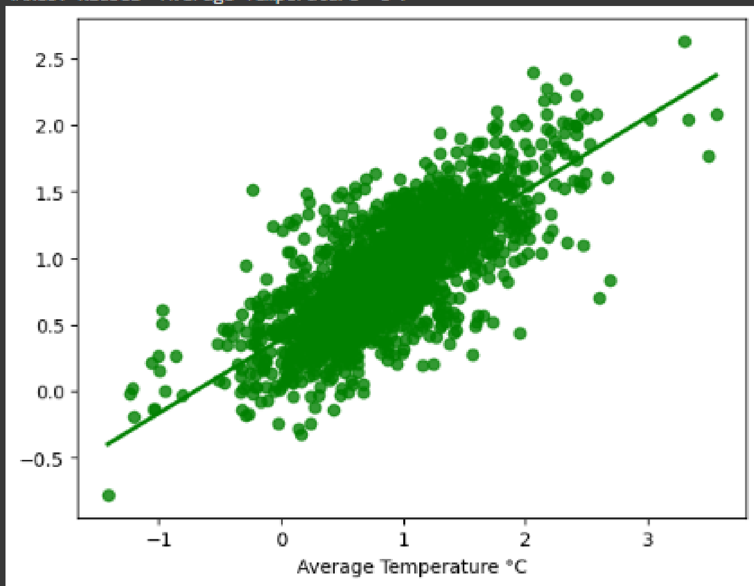
```
<Axes: xlabel='Average Temperature °C'>
```




```
[30] result = calculate_regression_metrics(y_test, pred)
     print(result)

{'MSE': 0.22425357430747966, 'RMSE': 0.4735541936330832, 'MAE': 0.35791519347024253, 'R2': 0.3010297576010328}
```

Random Forest Regression

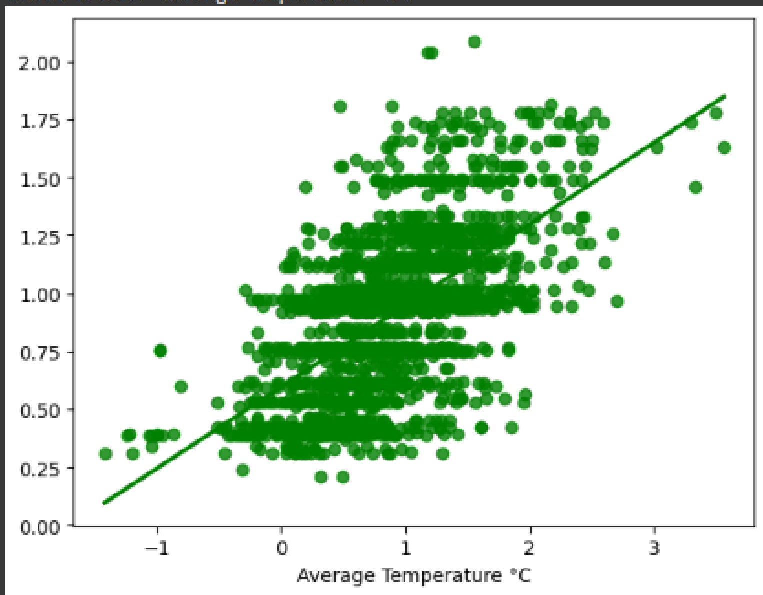
```
<Axes: xlabel='Average Temperature °C'>
```




```
[33] result = calculate_regression_metrics(y_test, pred)
     print(result)

{'MSE': 0.13906894378065324, 'RMSE': 0.3729194869950526, 'MAE': 0.26987587679356906, 'R2': 0.5665395584230409}
```

AdaBoost Regression

```
<Axes: xlabel='Average Temperature °C'>
```

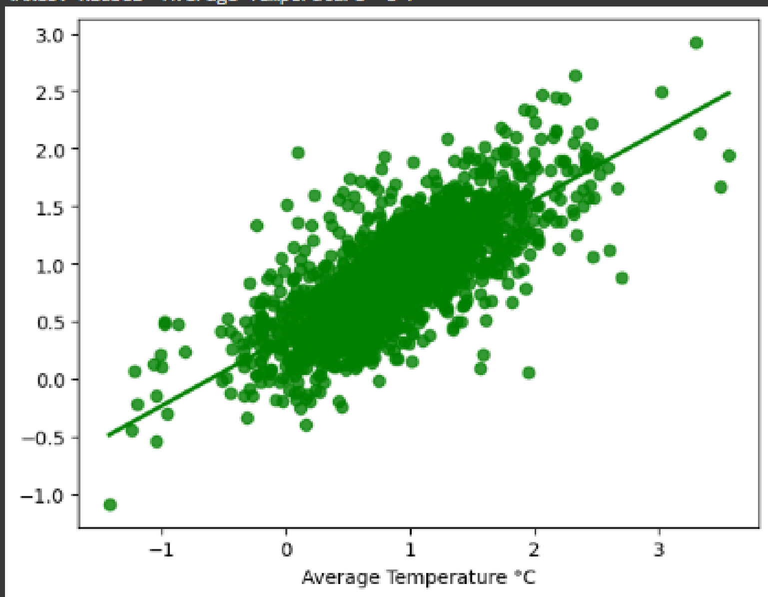



```
[36] result = calculate_regression_metrics(y_test, pred)
     print(result)
```

```
{'MSE': 0.20694842075387096, 'RMSE': 0.4549158391987148, 'MAE': 0.34439072041029944, 'R2': 0.3549677490532108}
```

XGB Regression

```
<Axes: xlabel='Average Temperature °C'>
```

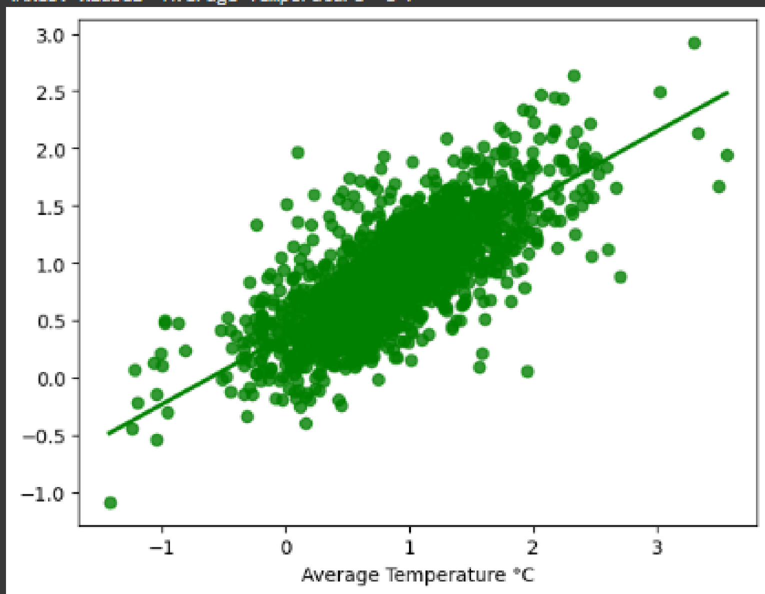



```
[39] result = calculate_regression_metrics(y_test, pred)
     print(result)
```

```
{'MSE': 0.13982992616090753, 'RMSE': 0.3739383988853078, 'MAE': 0.27237887241723324, 'R2': 0.5641676718636839}
```

Gradient Boosting Regression

```
<Axes: xlabel='Average Temperature °C'>
```



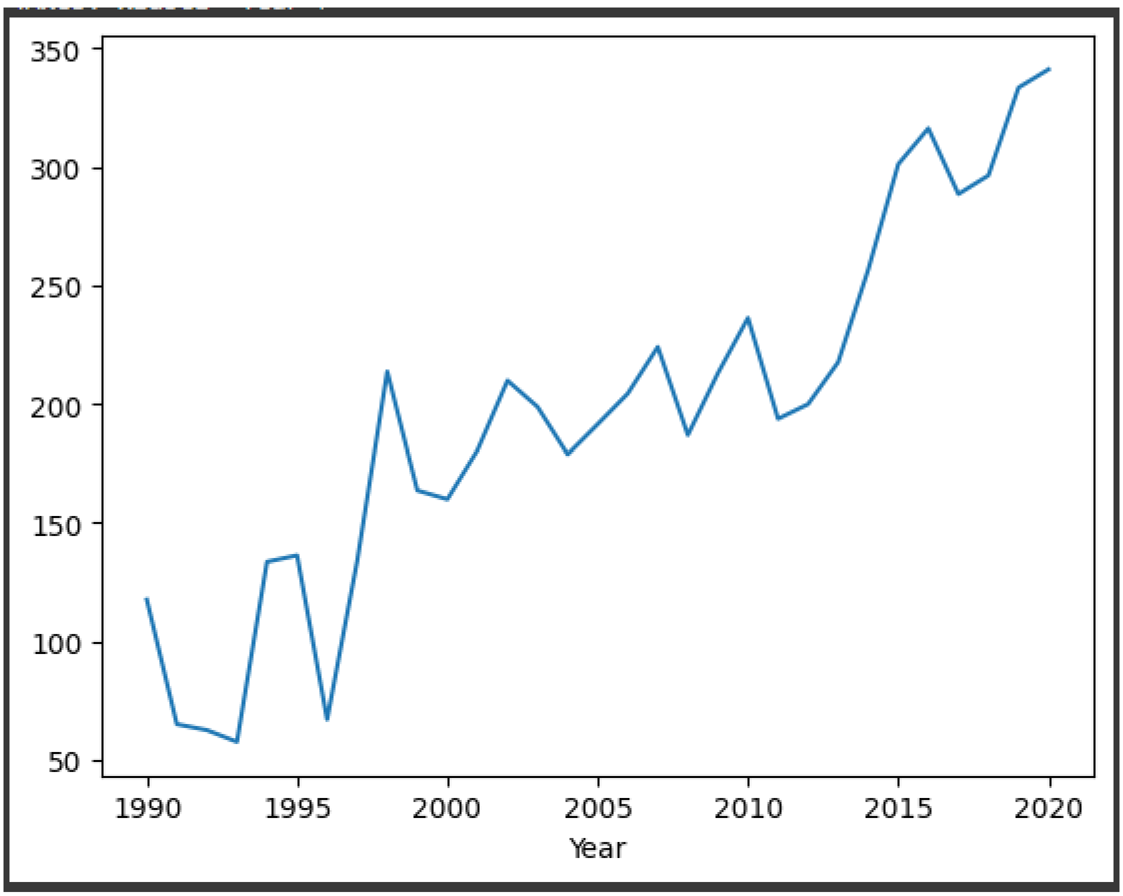

```
[42] result = calculate_regression_metrics(y_test, pred)
     print(result)

{'MSE': 0.13982992616090753, 'RMSE': 0.3739383988853078, 'MAE': 0.27237887241723324, 'R2': 0.5641676718636839}
```

The accuracy of simple Linear Regression is low as compared to that of the advanced ensemble learning regression models. Still the best accuracy is around 56% which is still low but this indicates some interesting patterns about the impact of CO2 on Temperature rise.

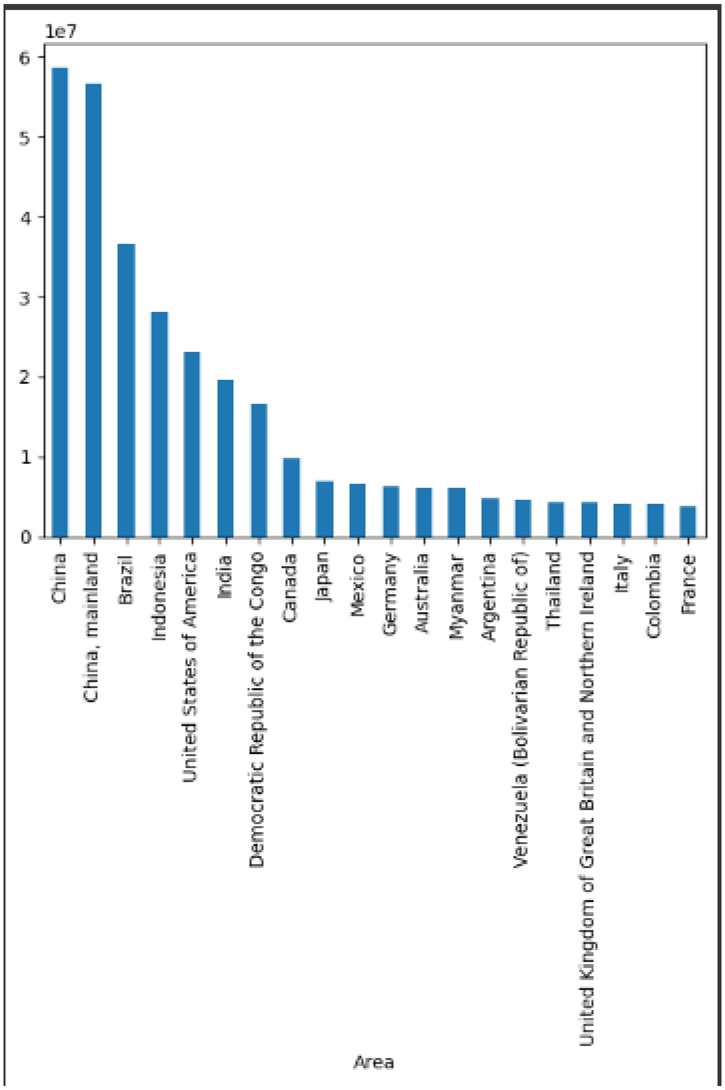
8. Visualization

Visual representations, such as graphs and charts, breathe life into our findings. Graphical aids are employed to elucidate complex patterns, making the data more accessible and digestible. These visualizations serve as crucial tools for communicating our discoveries to a broader audience, enhancing the transparency and interpretability of our results.

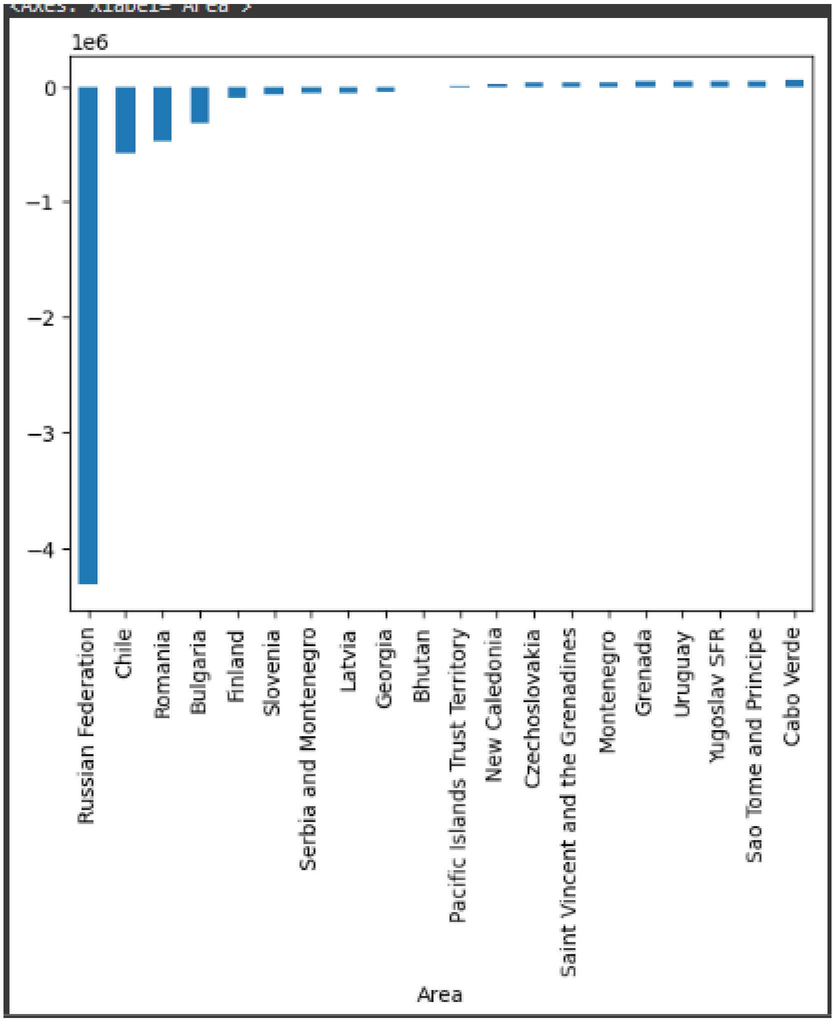
Time Series plot of Average Temperature Change (C)



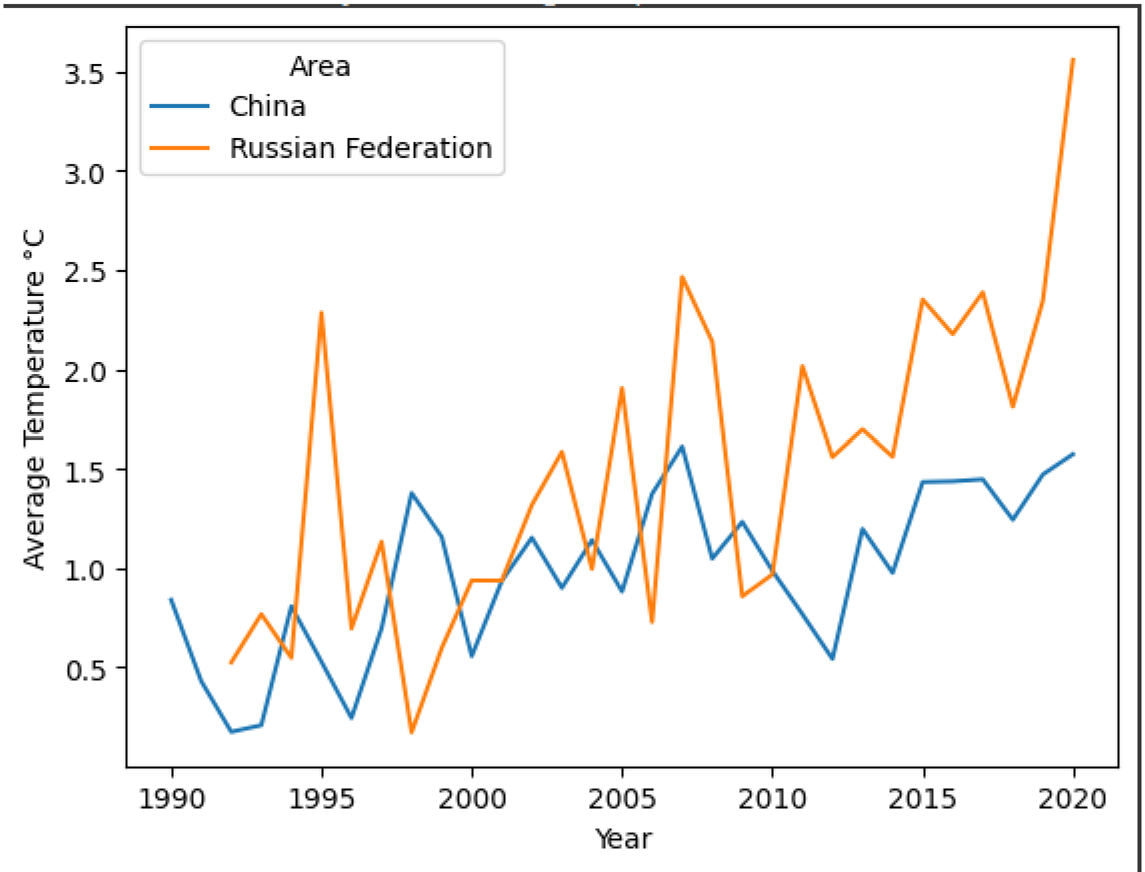
Bar graph of countries with highest contribution in the emission of CO2



Bar graph of countries with Lowest contribution in the emission of CO2



Comparison of China and Russia



9. Discussion

In the discussion section, we delve into the implications of our findings. The complex interplay between different emission sources and their impacts on average temperature change is explored. We contextualize our results within the broader landscape of climate science, discussing potential ramifications and shedding light on the intricate relationships that govern our environment.

10. Conclusion

In conclusion, our project brings to light the multifaceted relationship between CO2 emissions and temperature change. The study not only unveils the limitations of linear models but also showcases the effectiveness of ensemble techniques in capturing the complex dynamics within the dataset. By this we can conclude that the CO2 alone does not have an impact on the average temperature rise and there must be some other factors as well. By contributing to our understanding of these relationships, this project provides a steppingstone for further research and informs policy decisions aimed at mitigating climate change.

11. Recommendations

As we look to the future, recommendations emerge for further research and practical applications. Suggested avenues for exploration include more nuanced analyses of specific emission sources, the integration of additional environmental variables, and the consideration of regional variations. On a practical level, recommendations for policymakers are offered, emphasizing the need for targeted interventions to mitigate the impact of CO2 emissions on average temperature change.