Project presentation

CO2 Emissions; agri-food sector

Team Members

Justin Tshifaro
Veronicah Sihlangu
Mmapaseka Moswazi
Phumzile Sibiya

Zwiithwaho Mugodo
Nkhubalale Nkadimeng
Akhona Nzama

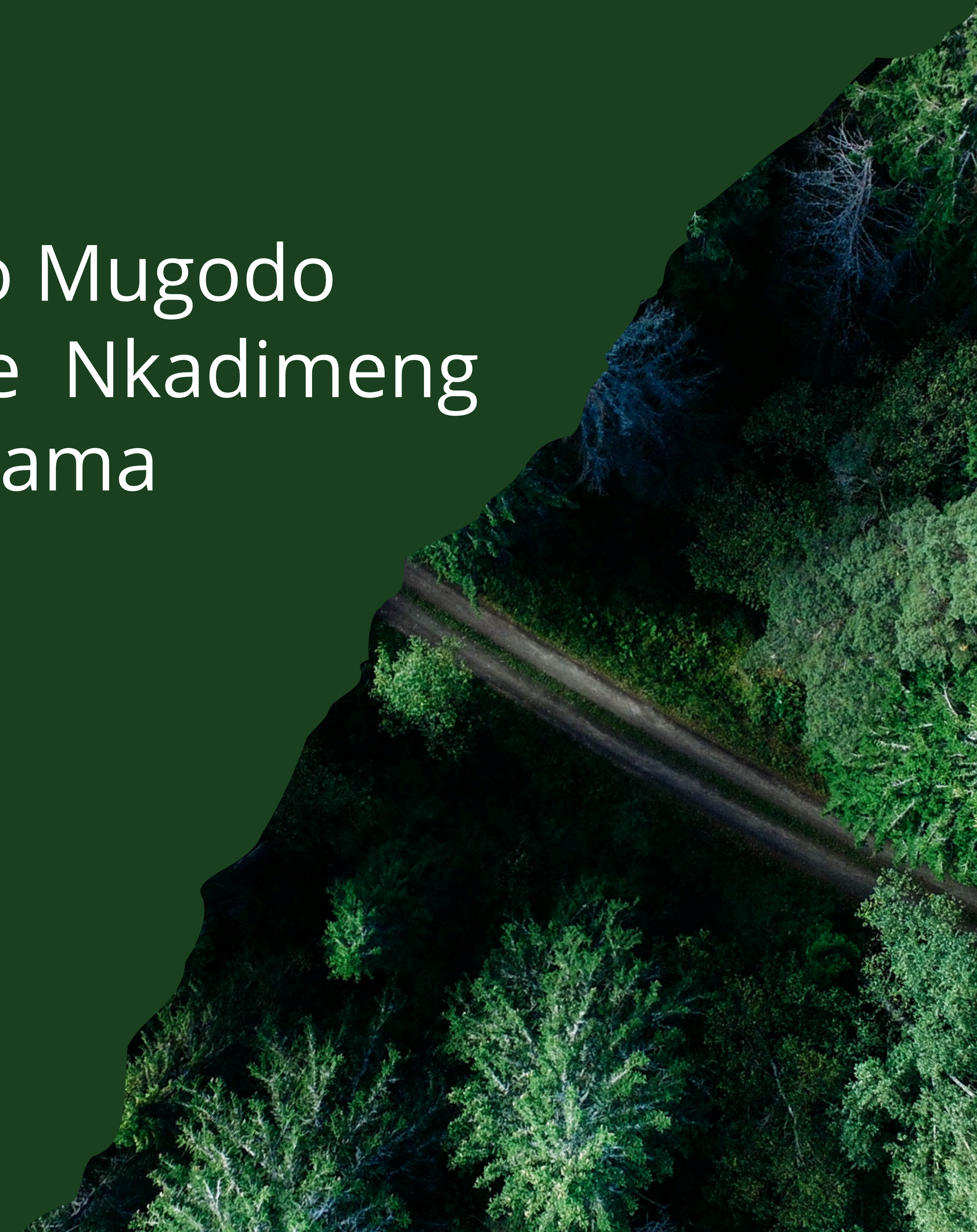


Table of content

Introduction

- We aim to analyze and predict CO_2's impact on agri-food sector climate..
- Stakeholders include policymakers, agricultural businesses, and environmental organizations.
- Project Aim:
  - Analyze CO_2 impact on climate change in agri-food sector, develop sustainability strategies.
- Data Sources:
  - Comprehensive dataset from the Food and Agriculture Organization (FAO) and the Intergovernmental Panel on Climate Change (IPCC).

Problem Statement

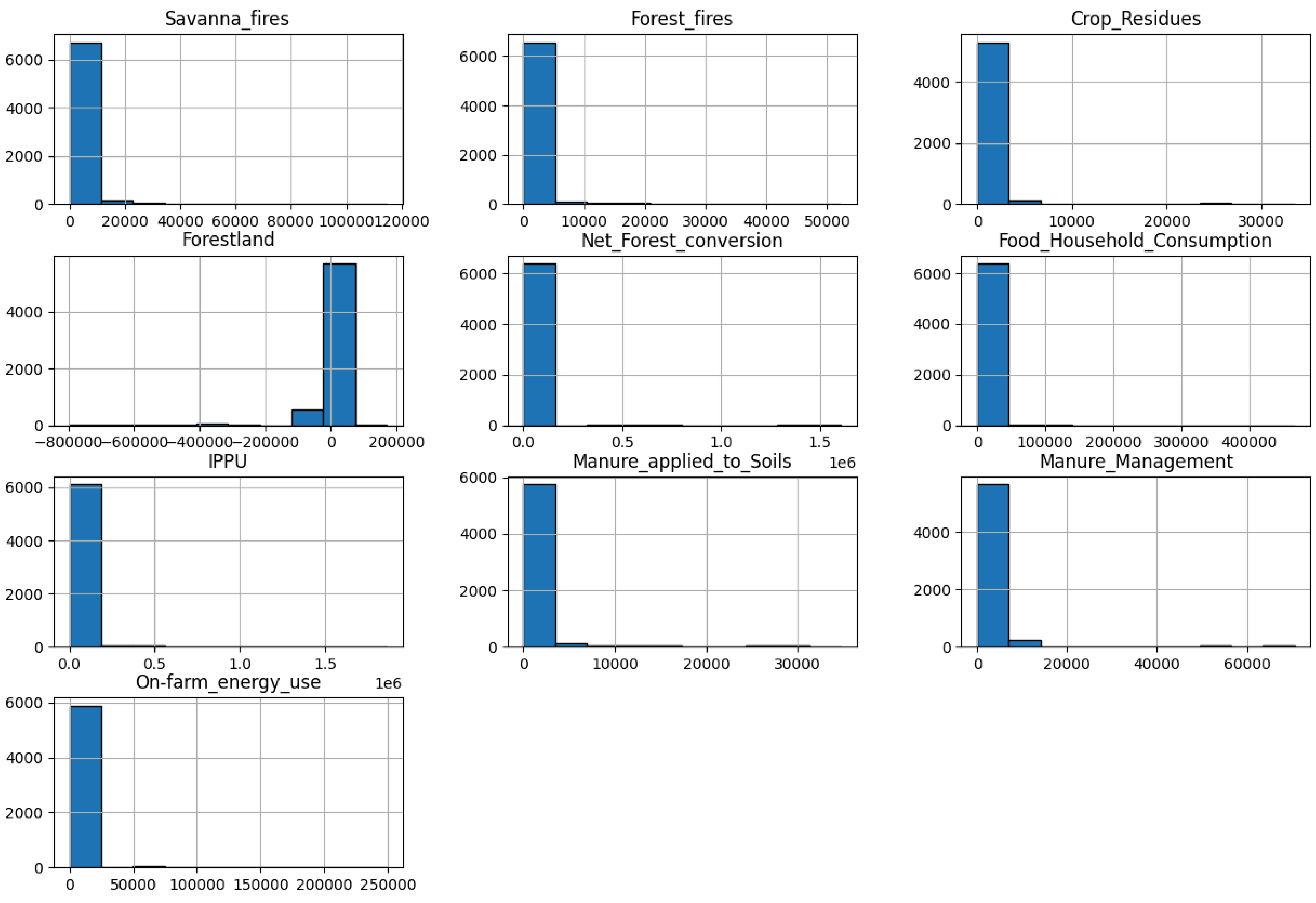
- The agri-food sector significantly contributes to global CO_2 emissions, worsening climate change.
- Key sources, e.g: crop residues, rice cultivation, food transport, and manure management.
- Despite its importance for food production and economic stability, the sector's environmental impact is a major concern.
- The potential of forestland as a carbon sink is underutilized, with sustainable forest management practices not widely adopted.

Objectives

- Identify trends in emission patterns within the agri-food sector.
- Understanding the sources and impacts of carbon emissions related to agriculture and food production.
- Predict Future Average Temperature changes based on CO2 emissions from different sources in the agri-food sector.
- Provide actionable recommendations for implementing sustainable practices within the agri-food sector.

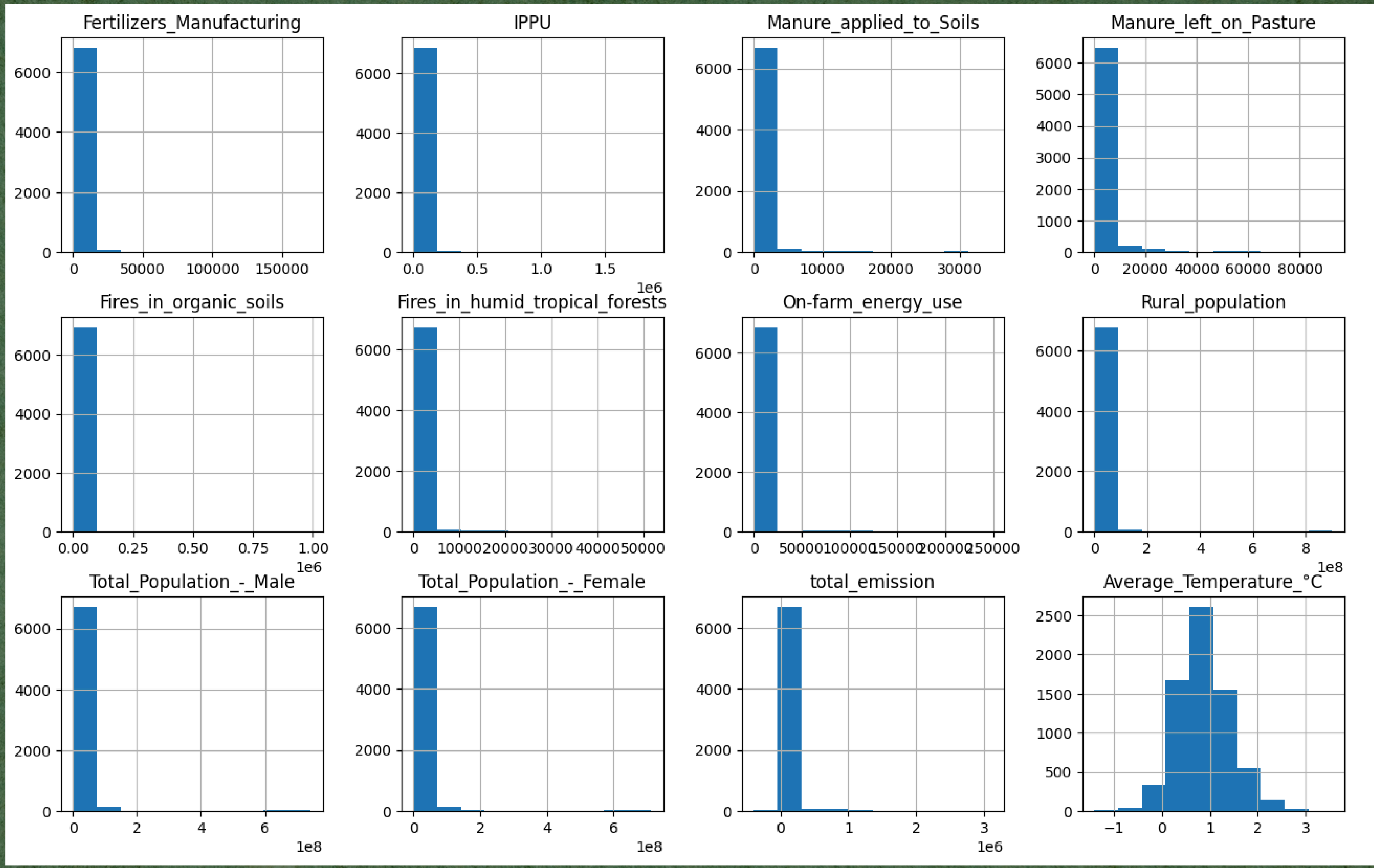
Data cleaning

- The Dataset contained 6965 rows and 31 column
- Removing Space in Columns names
- Some Columns Had missing values, e.g:
  - Savanna_fires: 6934 non-null
  - Forest_fires: 6872 non-null
  - Crop_Residues: 5576 non-null
- The median was used to fill the missing values because the data was numeric and skewed



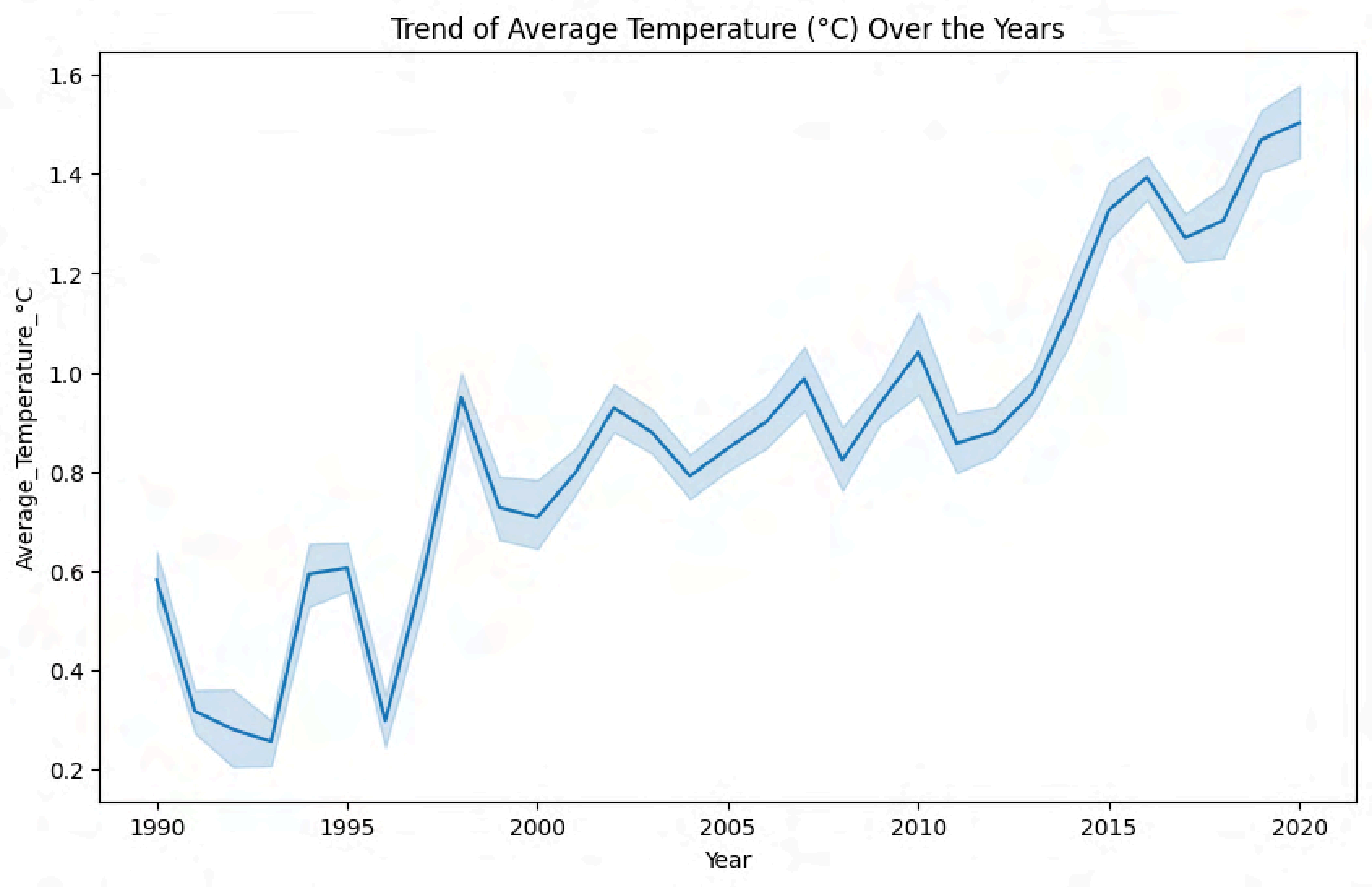
Exploratory Data Analysis (EDA)

- understanding the spread and distribution of data
- provide insights
- Temperature shows a bit of normal distribution



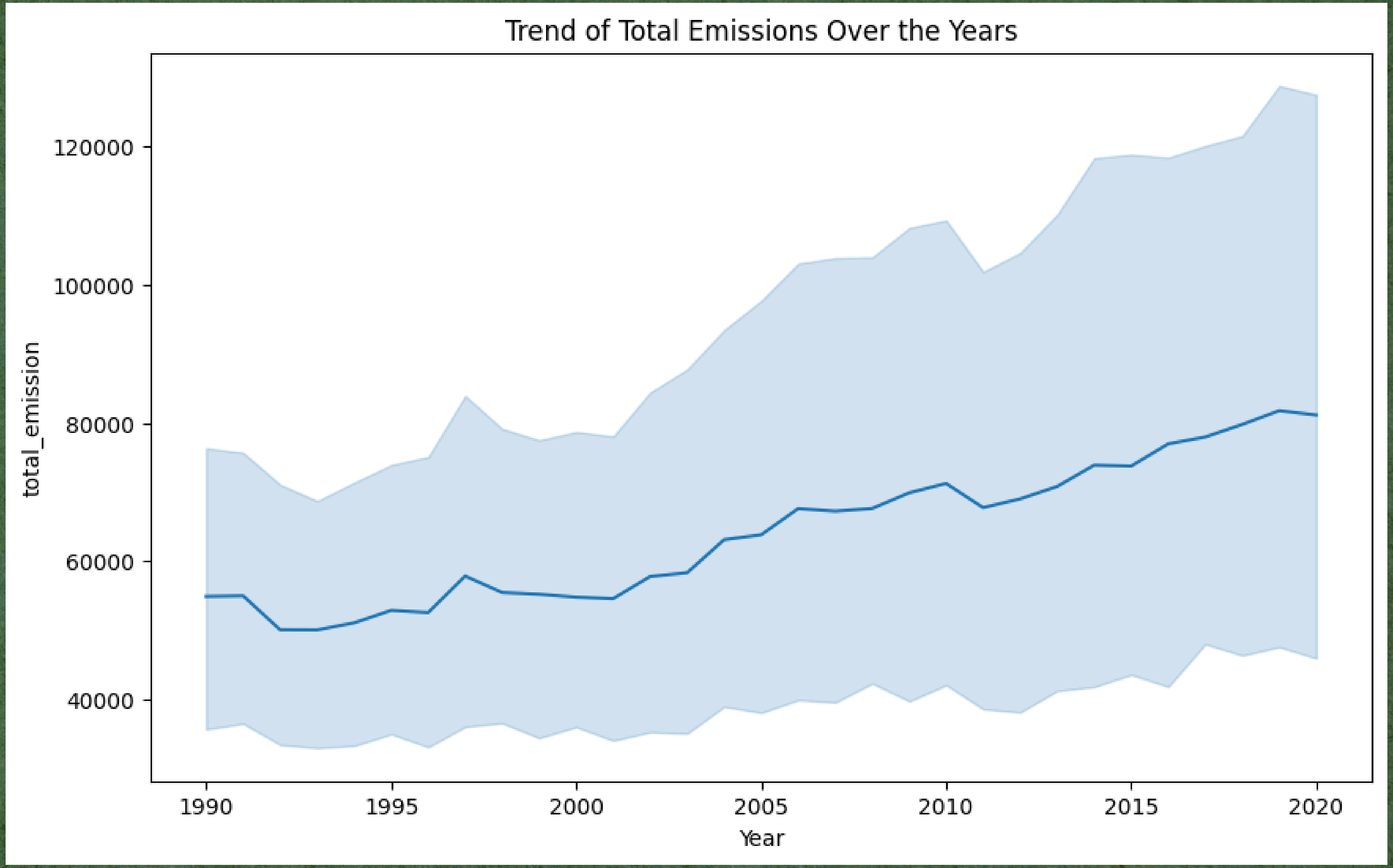
Exploratory Data Analysis (EDA)

- **Average Temperature (1990-2020):**
- **1990: Average temperature around 0.6°C.**
- **1990-1995: Temperature remained relatively stable.**
- **1995-2000: Noticeable decline followed by a rapid increase.**
- **2000-2010: Continued increase in average temperature.**
- **Long-term upward trend starting from 1995 and Substantial rise over the last decade (2010-2020).**



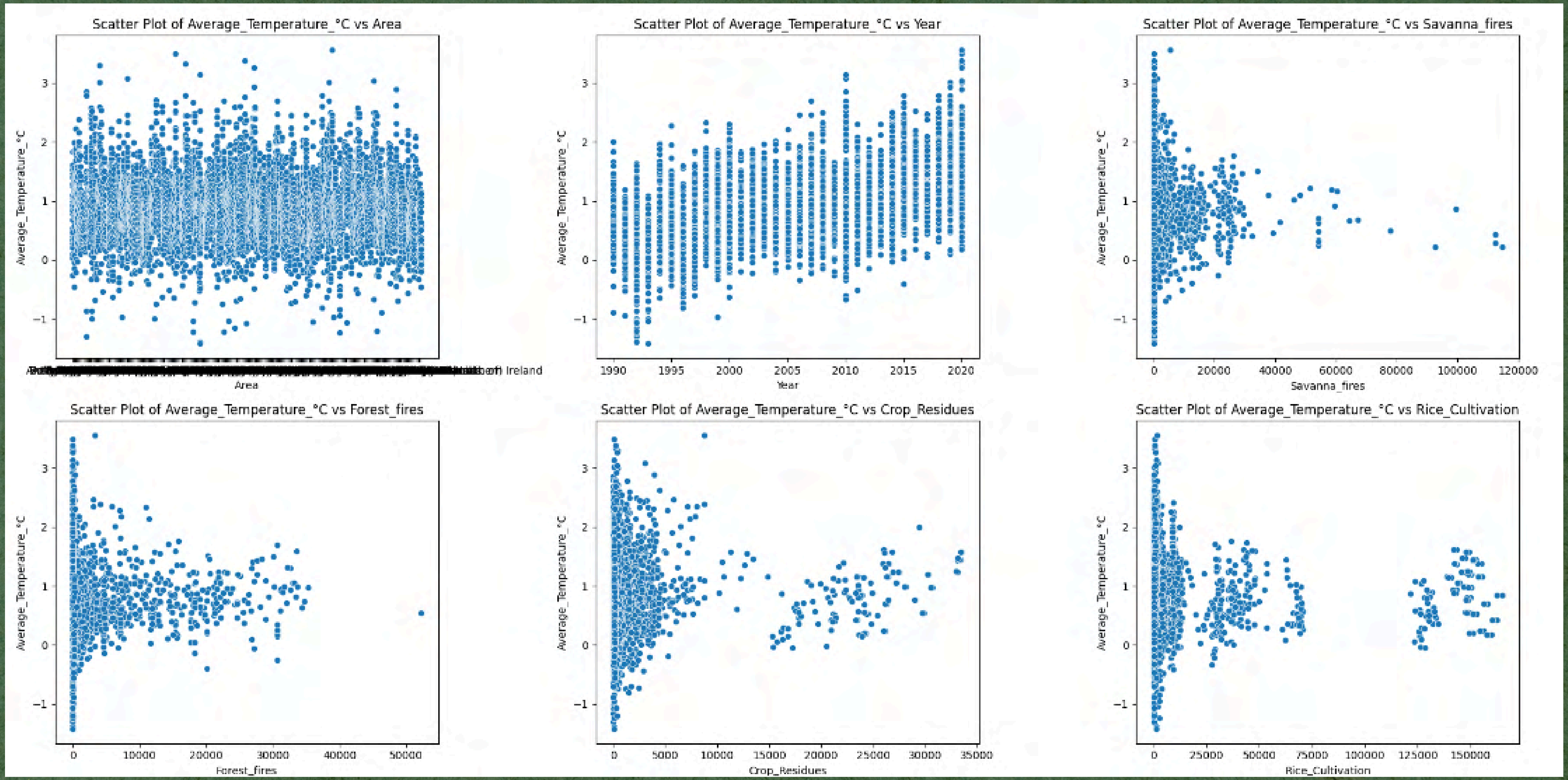
Exploratory Data Analysis (EDA)

- **Total Emissions (1990-2020) Mirrors the trend observed in average temperature.**
- **2000-2010: Continued increase in total emissions.**
- **Reason for Similar Trends:**
  - **Higher CO_2 levels trap more heat in the atmosphere, leading to global warming.**



Exploratory Data Analysis (EDA)

- **Larger agricultural areas are linked to higher temperatures.**
- **Rice farming shows complex temperature impacts, not directly proportional.**
- **Understanding these relationships helps in targeting emission reduction efforts effectively.**

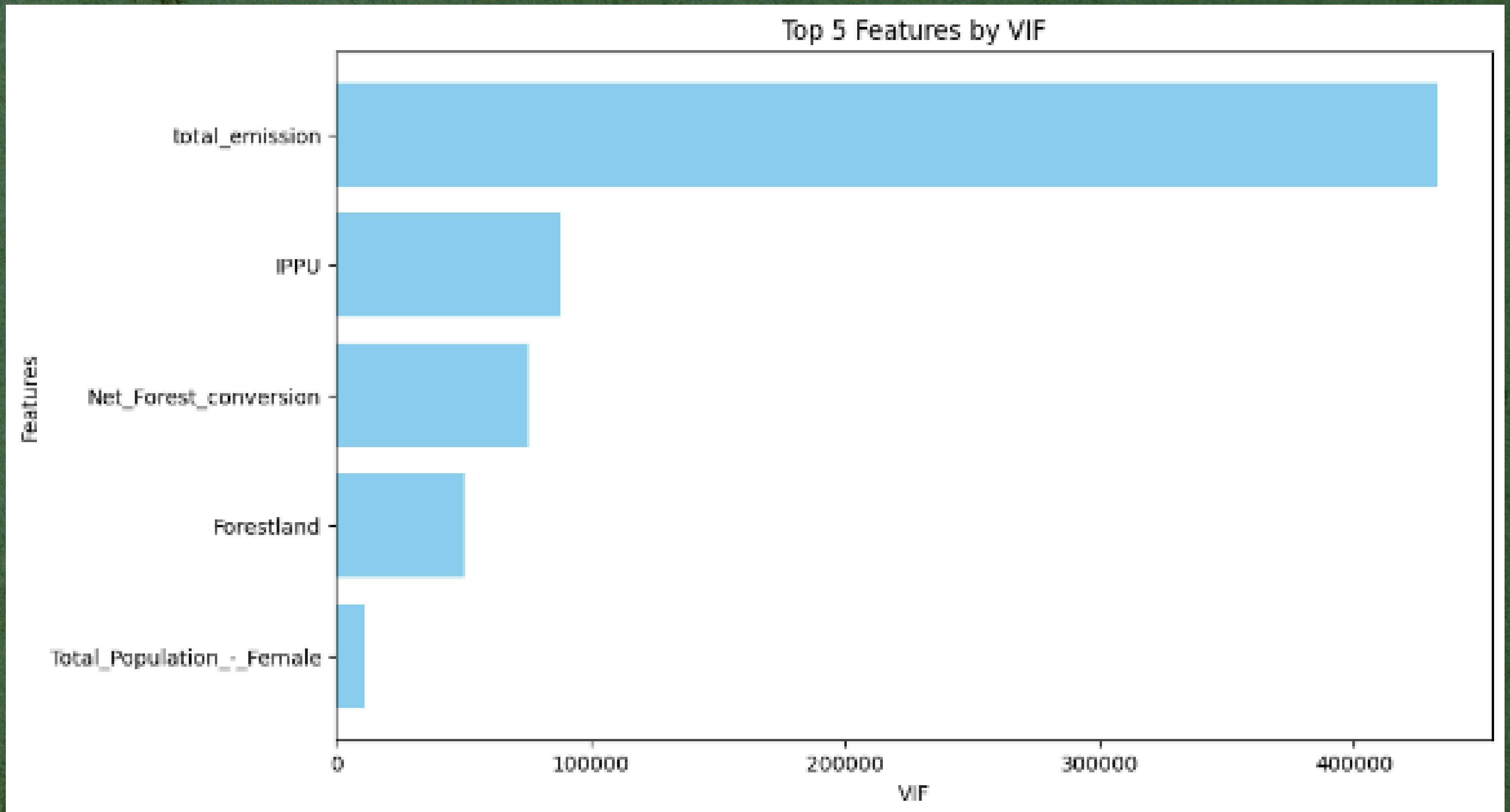


Feature Selection

- Why VIF?
  - Identifies Multicollinearity:
    - inflates the variance of coefficient estimates

  - Improves Model Interpretability:
    - Removing redundant features
    - Overfitting reduced

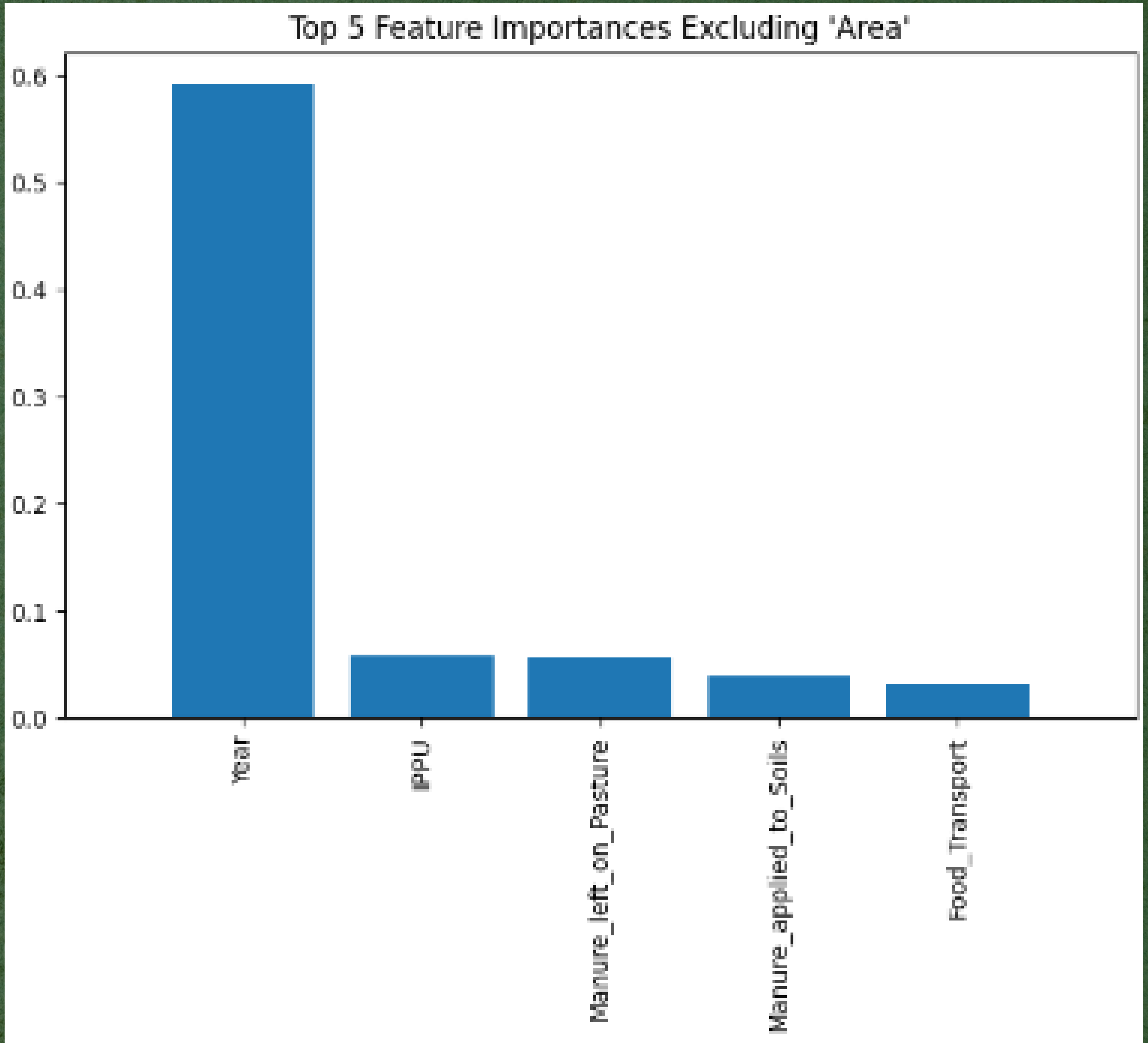
- How to Interpret a VIF Value:
  - Threshold Values:
    - VIF = 1
    - 1 < VIF < 5
    - VIF > 5
    - VIF > 10



- Observed VIF Values and their implications:
  - Extremely high VIF.
  - Risk of removing important features.
  - Conventional VIF thresholds do not apply.

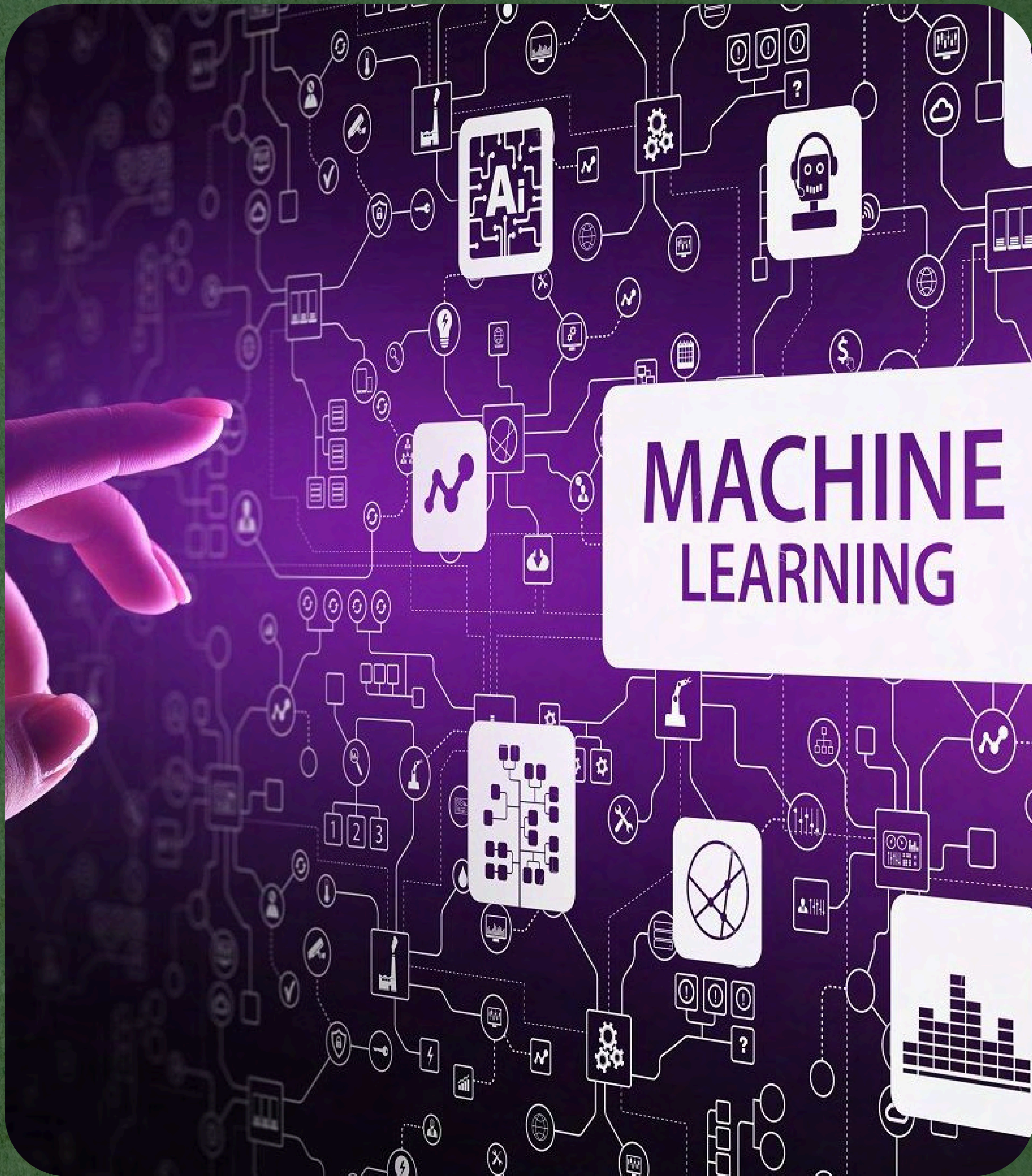
Feature Selection

- **How to proceed with high VIF values?:**
    - **The built-in “Feature Importance” method was Used**



Models

- Extreme Gradient Boosting(XGBoost)
- Optimized with RandomSearchCV
    - Performance:
        - Highly Efficient
        - Superior Predictive Accuracy
        - Kaggle Competitions Winner
    - Built-in Regularization
        - L1 (Lasso)
        - L2 (Ridge)
        - Prevents Overfitting and Improves Generalization
- Random forest -- Not tuned
    - Performance:
        - Robust and Accurate Ensemble Learning Technique
        - Feature Importance Ranking



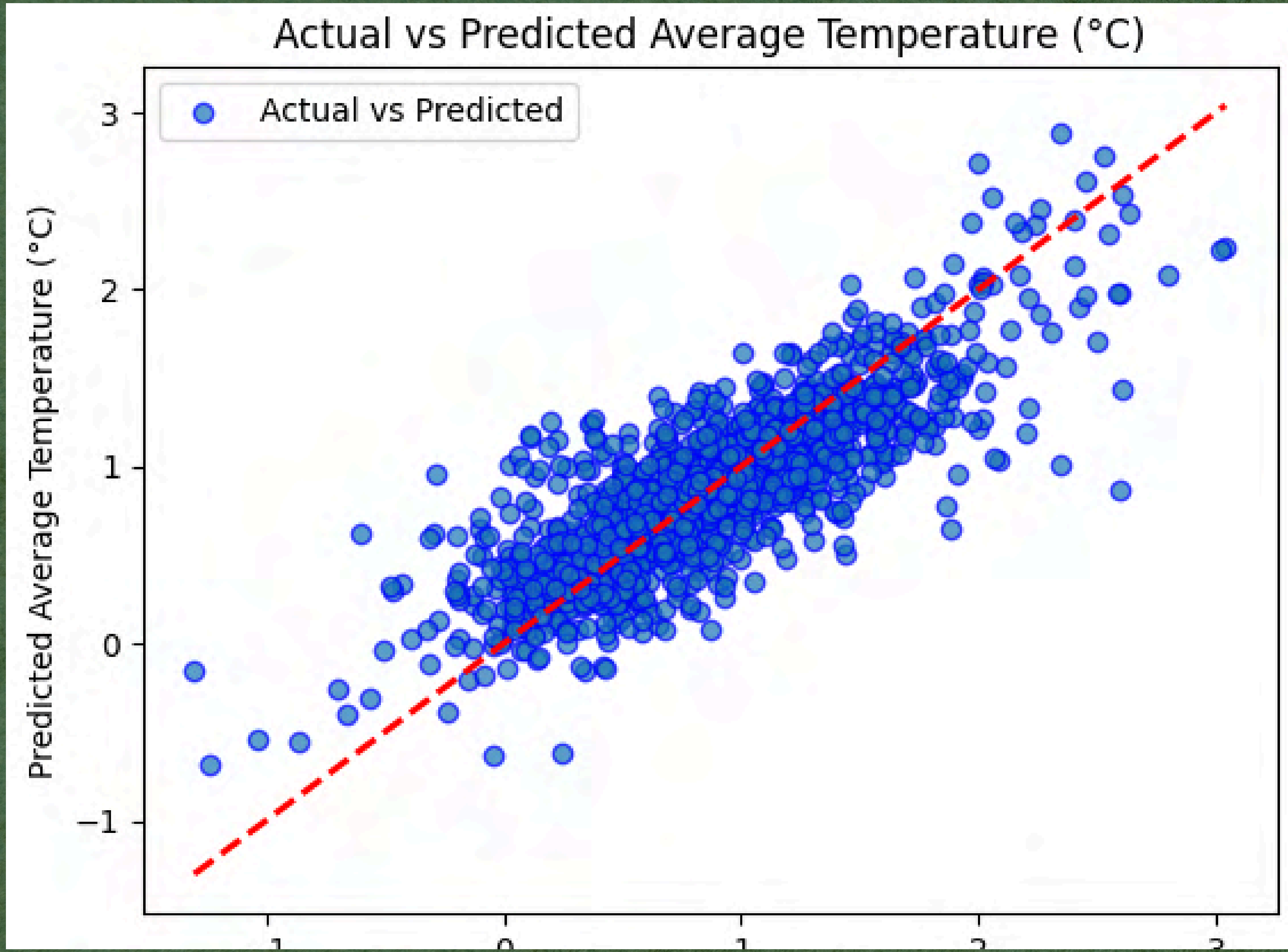
Models

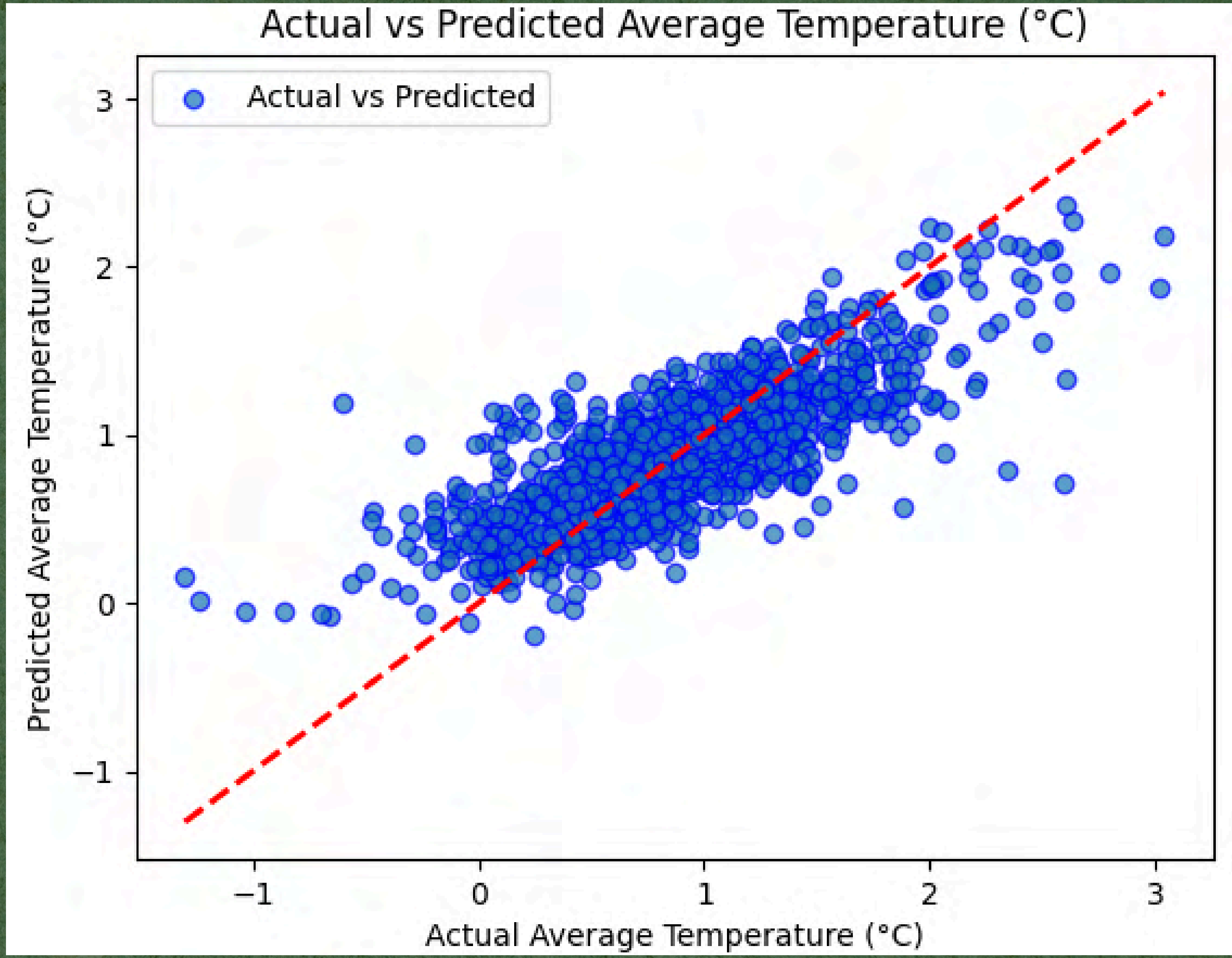
- BaggingRegressor
  - Reduction of Variance
    - Averaging Predictions

  - Robustness to Outliers and Noise
  - This Model was tuned using RandomSearchCV

- Stacking
  - Linear Regression
  - Decision Tree Regressor
  - Ridge Regression
  - Optimized with RandomSearchCV

Models

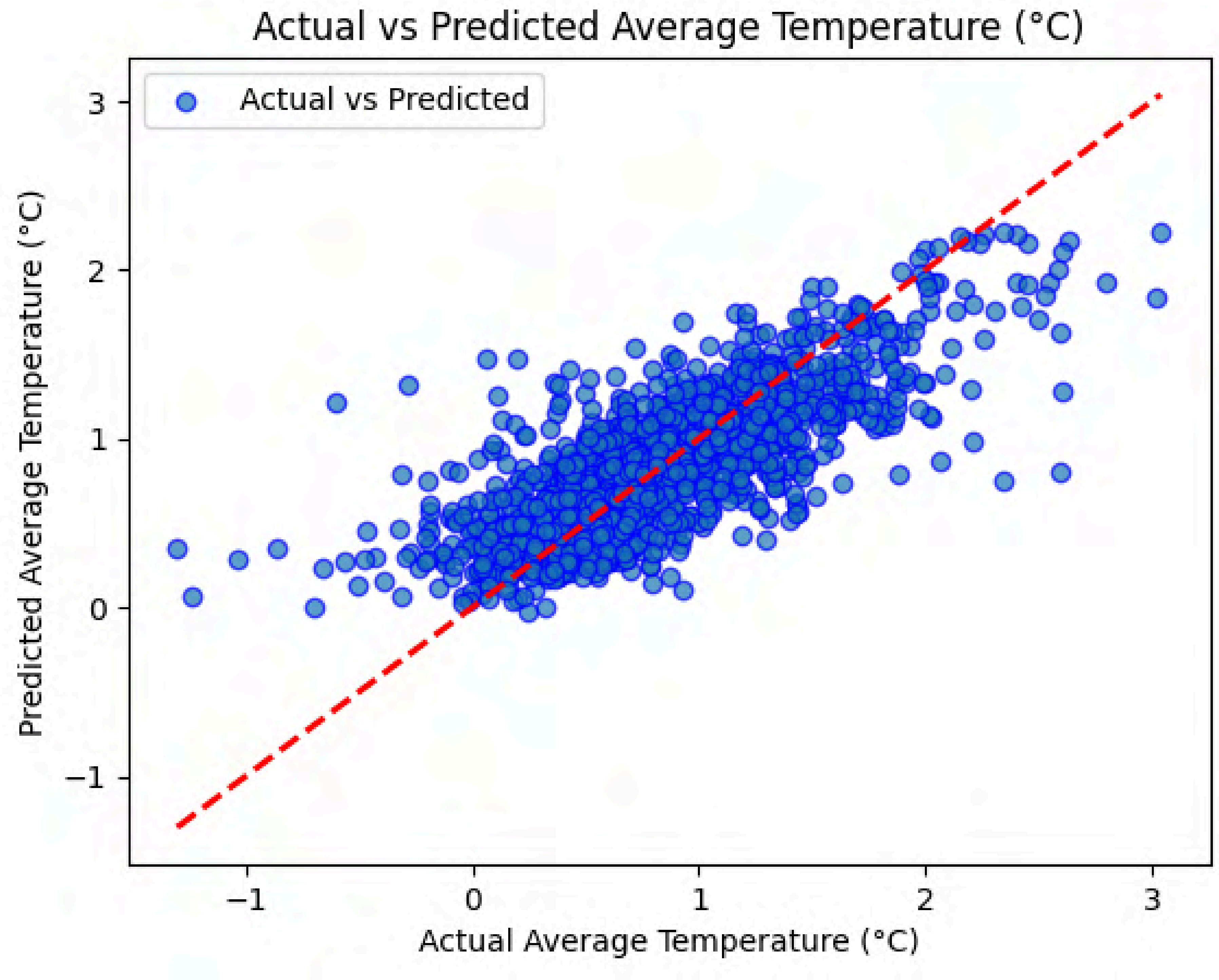


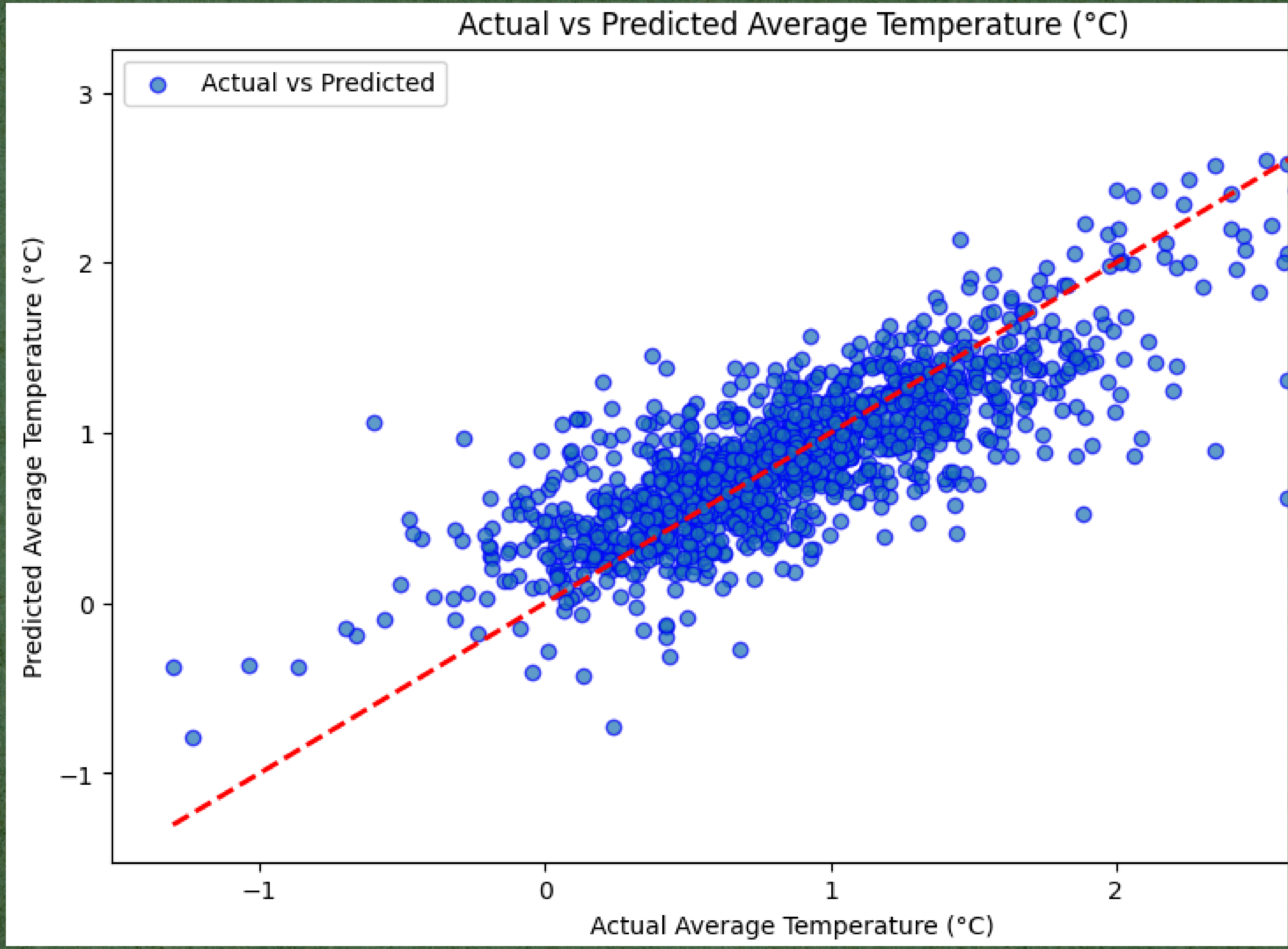


XGBoost

Bagging

Models





Stacked

Rndom-F

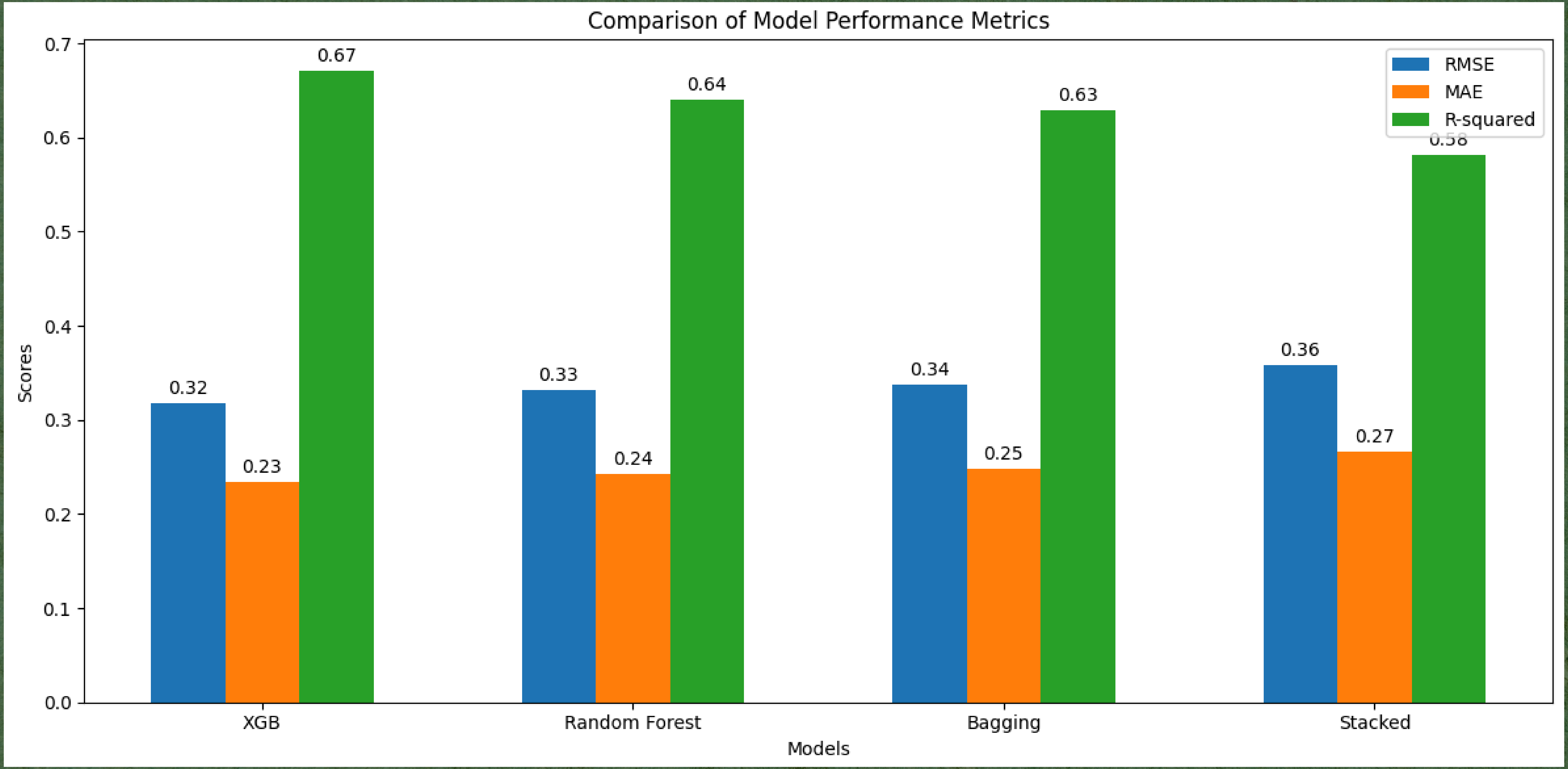
Models Performance

- Evaluation Metrics
  - RMSE (Root Mean Square Error)
    - Penalizes Larger Errors
    - Indicates Model Accuracy
    - Same Units as Input
    - Easier to Interpret

  - MAE (Mean Absolute Error):
    - Easy to Interpret
    - Less Sensitive to Outliers
    - Same Units as Predicted Values

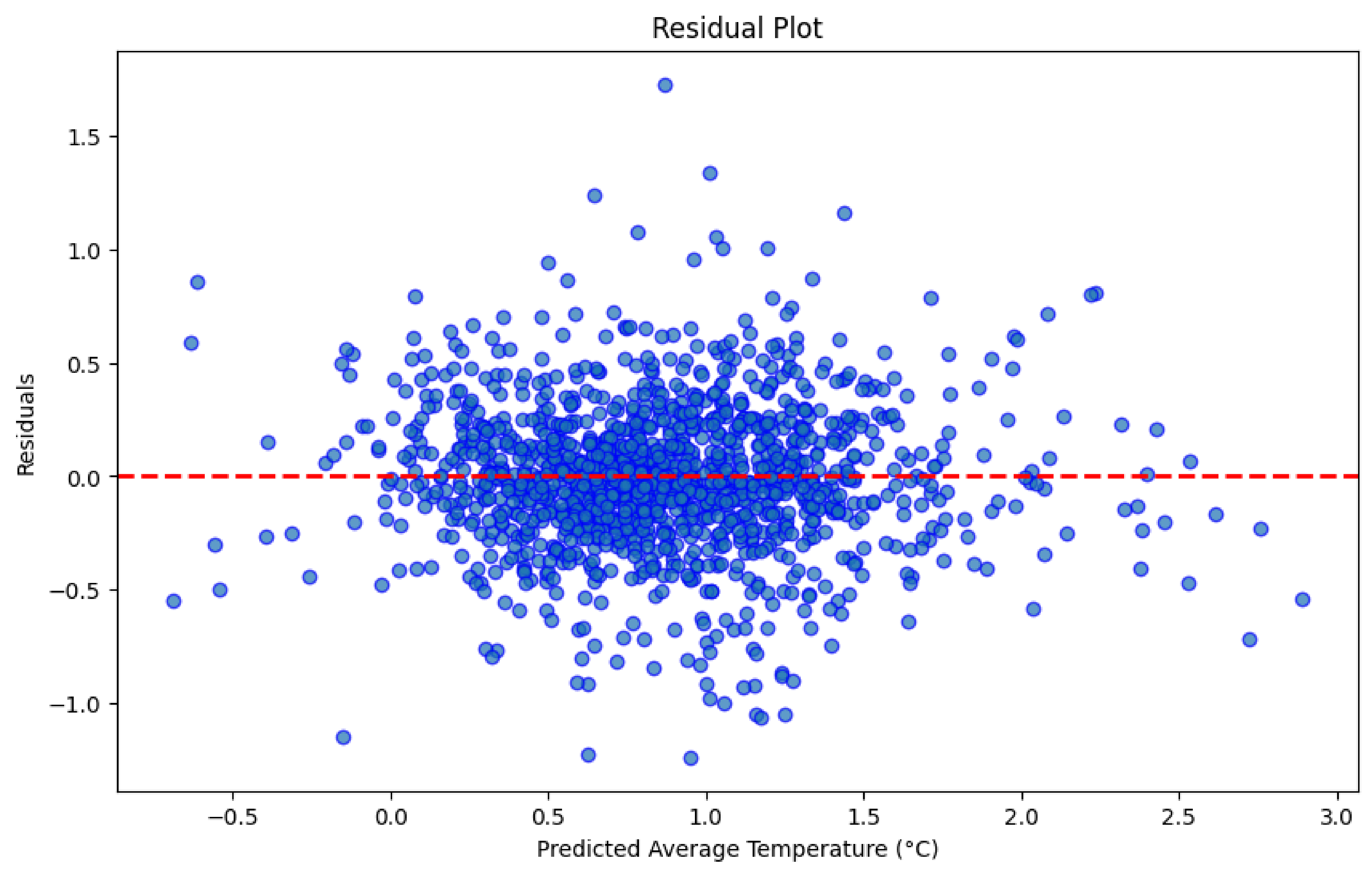
- R-squared
    - Goodness of Fit
    - Indicates Data Fit to Model

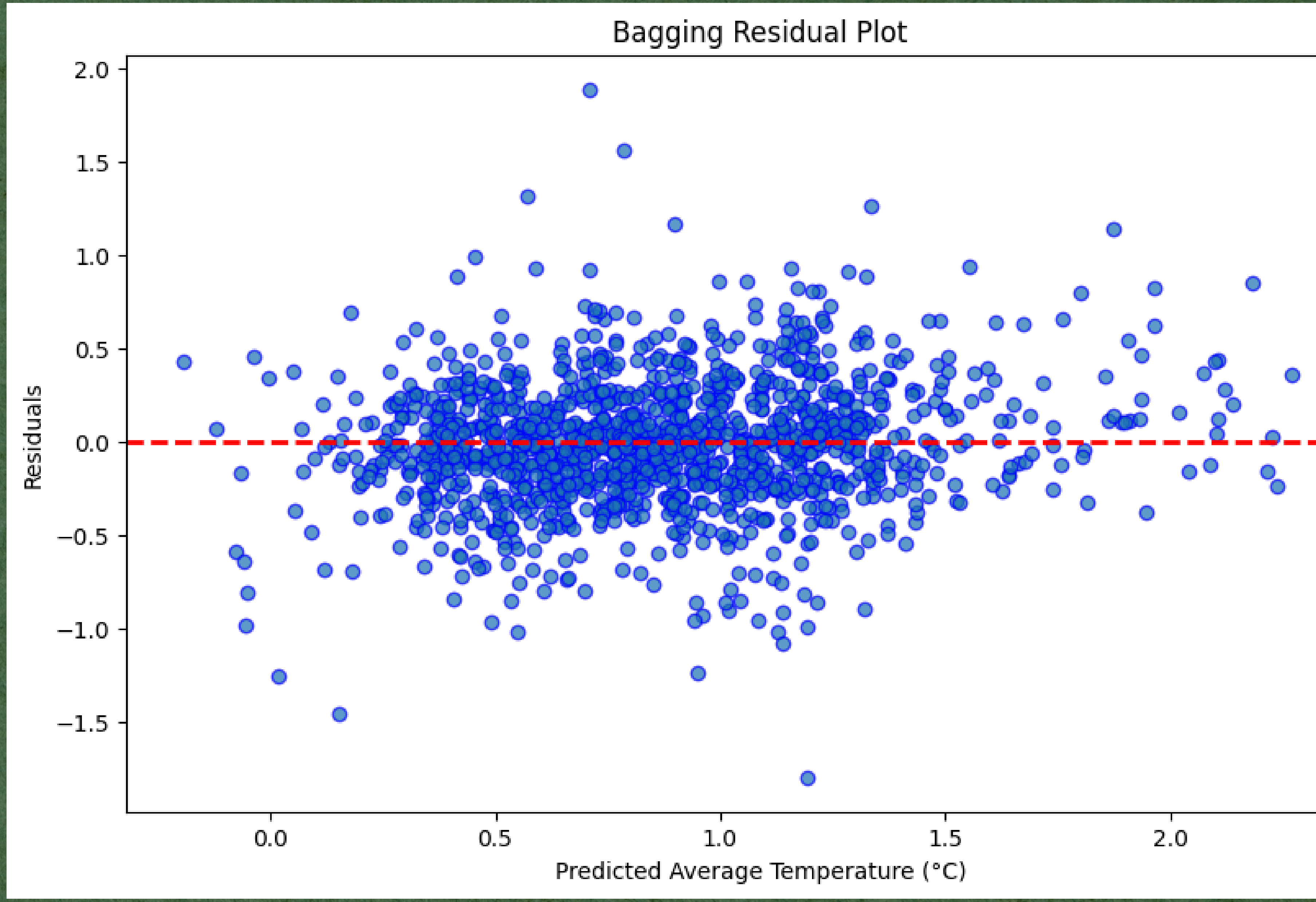
Models Performance



Based on these metrics, the XGBoost model is the most suitable for predicting temperatures in this scenario, outperforming other models in terms of accuracy and reliability.

Models Residuals Plot

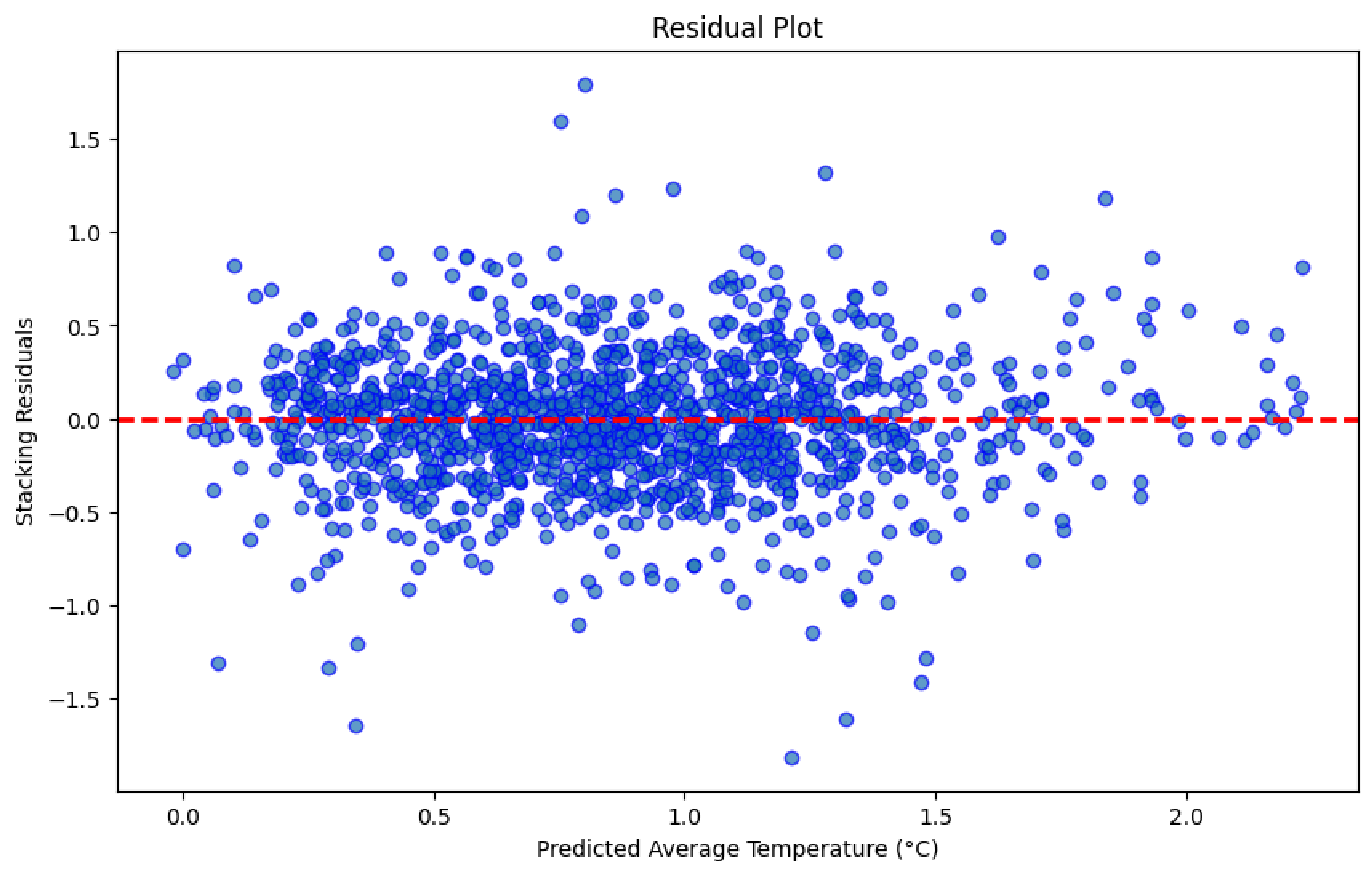




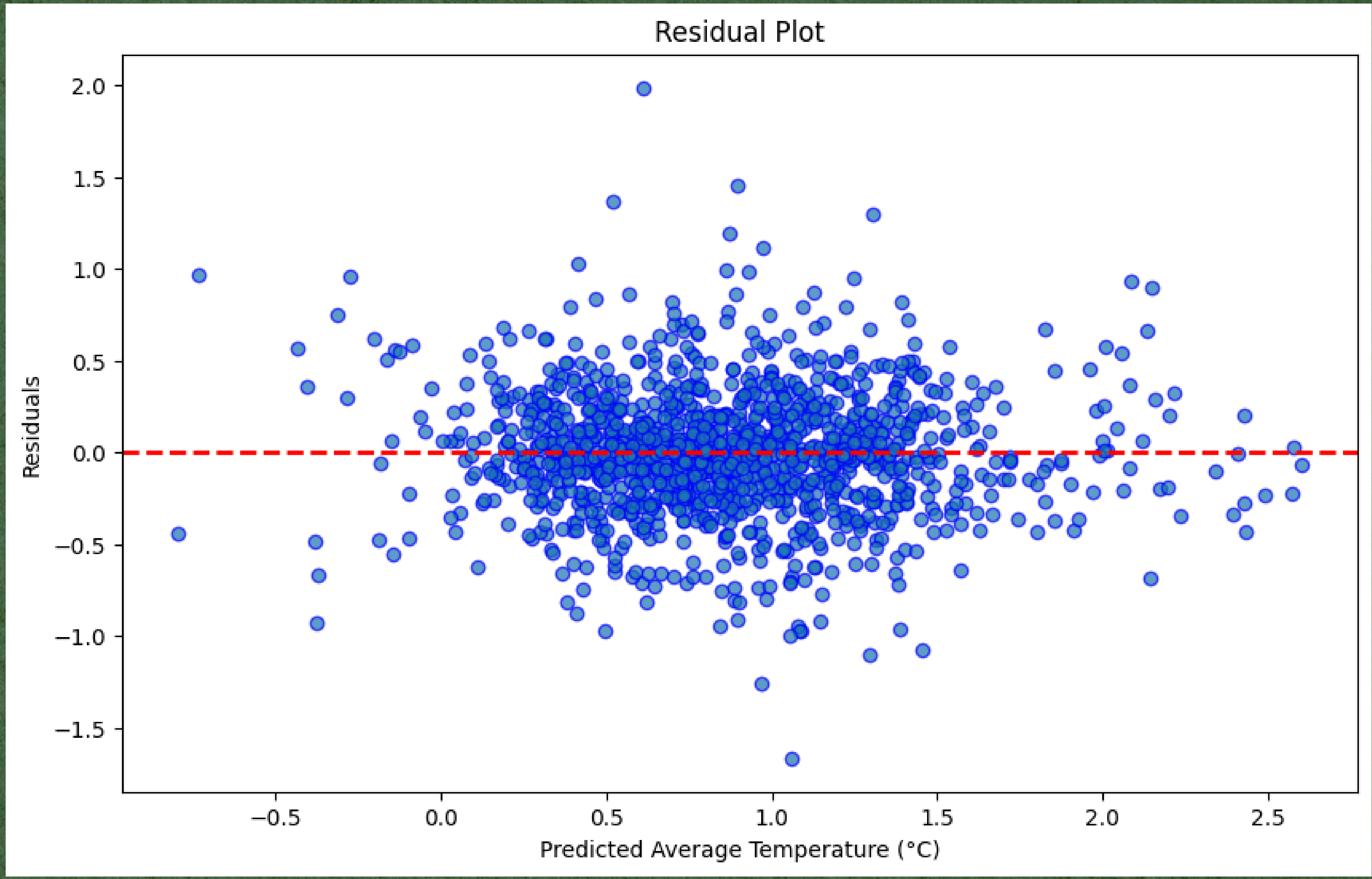
XGBoost

Bagging

Models Residuals Plot



Stacked



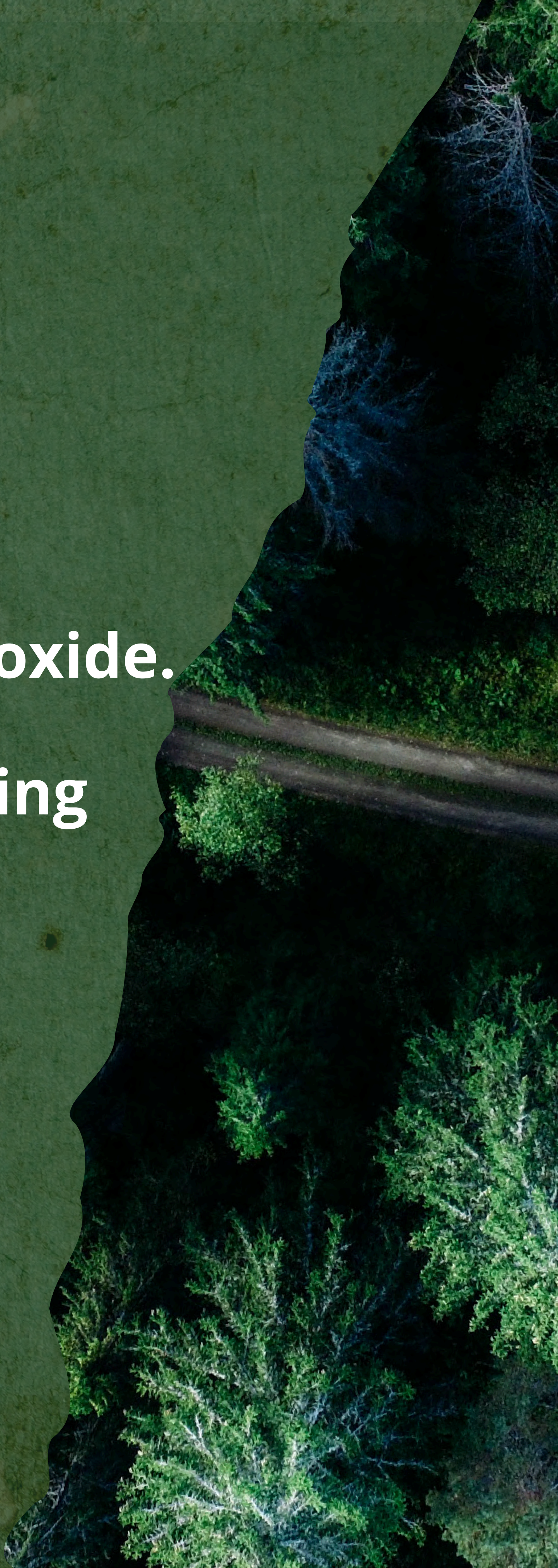
Rndom-F

Conclusion

- Examined CO_2 emissions from the agri-food sector.
- Using the XGBoost model.
- Enhanced predictive capability.

Recommendations

- Use water and fertilizer efficiently to reduce emissions from agricultural activities.
- Plant trees and practice crop rotation to help capture carbon dioxide.
- Educate farmers and consumers about the importance of reducing carbon emissions in agriculture.
- Switch to renewable energy sources for farming operations.



Thank you For Your Attention!

Any Questions, Comments or Concerns?