PREDICTING URBAN HEAT ISLAND INTENSITY

JING JI

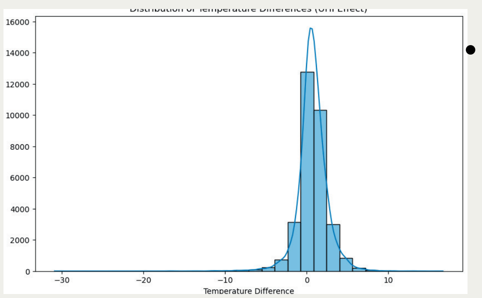
Introduction

- **Urban Heat Islands (UHI) significantly affect climate control, energy consumption, and health in urban areas. Predicting their intensity helps in planning more sustainable cities.**

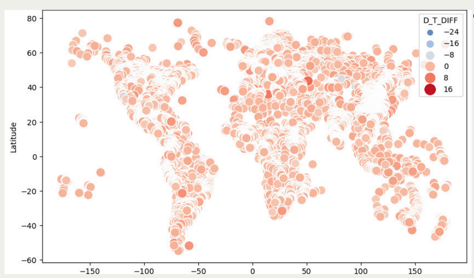
Data Overview

- The Sources: Utilized multiple datasets including urban heat island measurements, global carbon budget data, and land temperature records.
- Preprocessing: Data cleaning involved removing duplicates, handling missing values, and standardizing features for analysis.

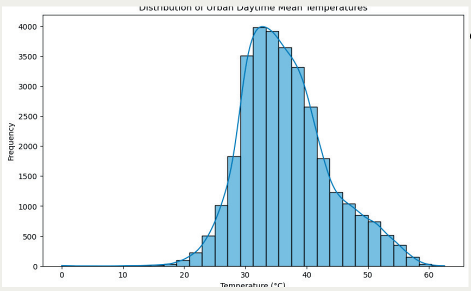
Data Visualizations



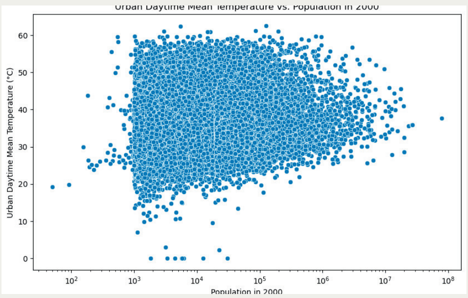
- Histogram showing frequency of temperature differences due to UHI. Centered around zero with notable peaks, indicating significant UHI effects in some urban areas.



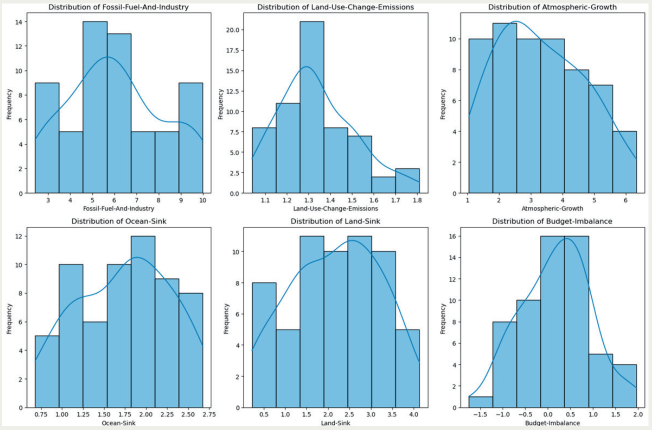
- Scatter map displaying UHI intensity worldwide, with color-coded circles representing the magnitude of temperature differences. Highlights densely populated areas with higher UHI effects.



- Histogram of urban daytime temperatures, mostly ranging from 30 to 40 degrees Celsius. Indicates typical urban warmth, especially in tropical urban settings.

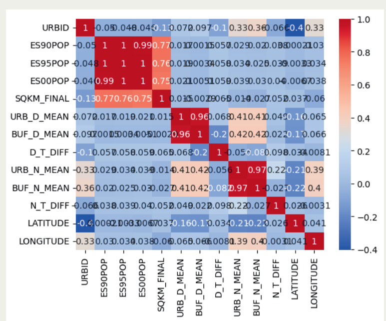


- linking urban temperatures with population sizes in 2000. Larger populations correlate with higher temperatures, showcasing the impact of population density on UHI.



- Fossil Fuels and Industry: Emissions mostly between 5-10 gigatons per year.
- Land-Use Changes: Peak emissions around 1.4 gigatons from deforestation and agriculture.
- Atmospheric Growth: Carbon accumulation in the atmosphere concentrated between 2-4 gigatons annually.
- Ocean Sink: Carbon absorption by oceans, peaking around 2 gigatons.
- Land Sink: Terrestrial ecosystems absorb between 1-3 gigatons of carbon.
- Budget Imbalance: Illustrates years with more emissions than absorption, highlighting carbon budget discrepancies.

Feature Selection



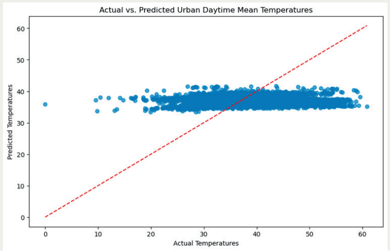
- This matrix shows relationships among variables relevant to Urban Heat Island (UHI) intensity:

- Population Variables: Strong correlations indicate consistent growth over decades.
- Temperature Metrics: High correlation with UHI indicators (D_T_DIFF, N_T_DIFF) emphasizes their role in UHI intensity.
- Geographical Factors: Longitude and Latitude correlate distinctly with urban temperatures, highlighting geographical influences on UHI.
- Carbon Cycle Components: Inverse relationships between emissions and carbon sinks illustrate natural carbon balance dynamics.

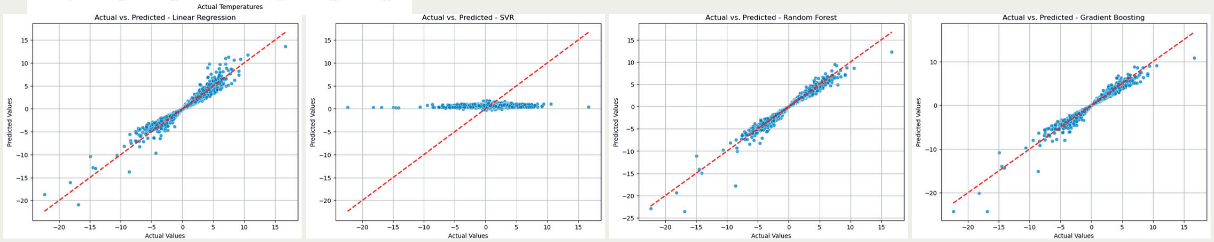
Method

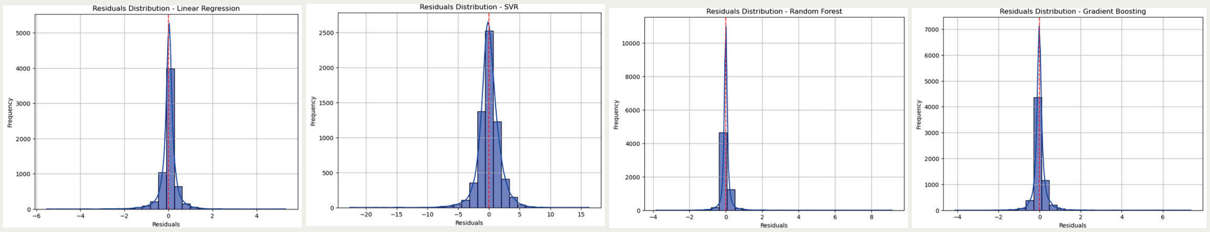
- Data Preparation: Used features like population, square kilometers, latitude, and longitude to predict urban daytime mean temperatures.
- Model Training: Employed Linear Regression, Support Vector Regression (SVR), Random Forest, and Gradient Boosting models.
- Model Evaluation: Assessed models based on Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score, to evaluate accuracy and fit.

Results



- Axes: Actual temperatures on the horizontal and predicted temperatures on the vertical.
- Red Dashed Line: Represents perfect prediction accuracy.
- Observations: The model generally predicts accurately for temperatures between 10°C and 50°C. Deviations at lower temperatures suggest potential areas of underprediction or overprediction.





	Linear Regression	SVR	Random Forest	Gradient Boosting
MAE	0.2279	1.2295	0.1255	0.1744
RMSE	0.4235	1.8101	0.2968	0.3424
R^2 Score	0.9468	0.0282	0.9739	0.9652

- Linear Regression shows good performance with an R^2 Score of 0.9468, indicating that it can explain about 94.68% of the variance in the data. However, it has moderate errors as shown by MAE and RMSE values.
- SVR (Support Vector Regression) performs poorly on this dataset with the lowest R^2 Score of 0.0282, suggesting it only explains about 2.82% of the variance. Its error metrics (MAE and RMSE) are the highest among the models, indicating less accurate predictions.
- Random Forest excels with the highest R^2 Score of 0.9739, demonstrating it can explain about 97.39% of the variance, and it has the lowest MAE and RMSE values, showing high accuracy and reliability.
- Gradient Boosting also shows strong performance with an R^2 Score of 0.9652, and relatively low error metrics, making it a robust choice for modeling this data.

Conclusion and Future Work

- The study successfully demonstrated that ensemble models like Random Forest and Gradient Boosting provide high accuracy in predicting urban daytime mean temperatures. These models significantly outperformed simpler models such as Linear Regression and SVR, showcasing their capability to handle complex datasets and capture the nuanced dynamics of urban heat islands.
- Limitations: 1) SVR's limitations in large datasets and potential overfitting in ensemble models. 2) Limited predictive features may restrict model accuracy.
- Future Work: 1) Explore complex models and additional environmental factors to improve predictions. 2) Apply these models for real-time urban planning.

- Carbon Cost Analysis:
  RAM Consumption: 0.000009 kWh, with a power usage of 3.0 watts.;CPU Consumption: 0.000016 kWh,
  with a total power usage of 5.0 watts. carbon Emissions: 0.0000 kg of CO_2.

- Reference: Openclimatedata. (n.d.). Global-carbon-budget/data at main · openclimatedata/global-carbon-budget. GitHub. https://github.com/openclimatedata/global-carbon-budget/tree/main/data
- Center for International Earth Science Information Network (CIESIN), Columbia University. 2016. Global Summer Land Surface Temperature (LST) Grids, 2013. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H408638T.
- NASA Land Processes Distributed Active Archive Center (LP DAAC). 2002a. MODIS MYD11A1 Version 5. USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (https://lpdaac.usgs.gov), https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/myd11a1