C.V. RAMAN GLOBAL UNIVERSITY

BHUBANESWAR, ODISHA, INDIA

EXPERIENTIAL LEARNING OF MACHINE LEARNING

GROUP-2

5TH SEMESTER

TOPIC- CLIMATE CHANGE MODELING AND PREDICTION

CASE: A RESEARCH ORGANIZATION IS STUDYING THE EFFECTS OF CLIMATE CHANGE AND WANTS TO PREDICT FUTURE TEMPERATURE TRENDS BASED ON HISTORICAL DATA ON TEMPERATURES, CARBON EMISSIONS, AND ENVIRONMENTAL CHANGES. QUESTION: HOW CAN MACHINE LEARNING BE UTILIZED TO BUILD MODELS THAT ACCURATELY PREDICT FUTURE CLIMATE CONDITIONS, AND WHAT INSIGHTS CAN BE GAINED TO HELP MITIGATE CLIMATE CHANGE EFFECTS?

UNDER THE SUPERVISION OF

DR. SUMANA DE

ASST. PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SESSION:-JULY-DECEMBER

BATCH:2022-2026

C. V. RAMAN GLOBAL UNIVERSITY ,
BHUBANESWAR, ODISHA,
INDIA 2024-25

SUBGROUP-16 MEMEBERS

TEAM MEMBERS	REGISTRATION NO.
PIYUSH KUMAR MODI	2201020562
SOURAV KUMAR MOHANTY	2201020563
PRATYUSH PARADARSHI PATRA	2201020564
PRIYANSHU KUMAR BHADANI	2201020565
ANSHUMAN MAHAPATRA	2201020566

SL.No.	CONTENTS	PAGE NO.
1.	ACKNOWLEDGEMENT	1
2.	CERTIFICATE	2
3.	ABSTRACT	3
4.	INTRODUCTION	4
5.	EXPERIMENTAL SETUP	5
6.	METHODOLOGY	6
7.	DISCUSSION AND ANALYSIS	7-9
8.	RESULT	10
9.	CONCLUSION	11
10.	REFERENCES	12

ACKNOWLEDGMENT

We, SUBGROUP-16, collectively extend our sincere appreciation to Assistant Professor **DR. SUAMANA DE** for his unwavering guidance and mentorship during our experiential learning program on the topic "Climate Change Modeling and Prediction".This program was conducted as part of the fifth semester in the academic session from July to December, Batch Number 2022-2026.

Professor DR. SUAMANA DE's expertise and dedication significantly enriched our team's understanding of machine learning models. His insightful feedback and continuous support played a crucial role in shaping our collective learning experience and deepening our comprehension of our topic.

We also express our gratitude to the department for providing the necessary resources and creating a conducive learning environment, allowing SUBGROUP-16 to engage in hands-on exploration and practical application of theoretical concepts.

This experiential learning opportunity has not only enhanced our technical skills but has also fostered teamwork and collaboration, preparing us to tackle real-world challenges in the field of digital systems.

Once again, thank you, Professor DR. SUAMANA DE, for your inspiring guidance and mentorship.

Sincerely,

SUBGROUP-16
SESSION:-JULY-DECEMBER 2024-25
BATCH:-2022-2026

CERTIFICATE

This is to certify that Subgroup-16 has successfully completed a case study on the topic "CLIMATE CHANGE MODELING AND PREDICTION" as part of the Case Study for the 5th semester for the subject MACHINE LEARNING

SUBGROUP-16 exhibited exceptional dedication and a keen interest in the chosen topic, showcasing a commendable level of enthusiasm and curiosity throughout the study. Their collaborative efforts and effective teamwork have been instrumental in the successful exploration and understanding of the complexities involved in designing and implementing their topic.

This certificate is awarded to SUBGROUP-16 in recognition of their outstanding commitment, keen interest, and exemplary teamwork, which greatly contributed to the successful completion of the case study.

DR. SUMANA DE
ASSISTANT PROFESSOR
DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

C.V. RAMAN GLOBAL UNIVERSITY
BHUBANESWAR, ODISHA INDIA
2024-25

ABSTRACT

This case study explores the use of machine learning, specifically linear regression, to predict future temperature trends based on historical data, including global temperatures, carbon emissions, and environmental changes. Linear regression, a straightforward yet powerful statistical method, is applied to analyze the relationship between rising carbon dioxide (CO_2) levels and increasing global temperatures.

Using historical datasets on temperature records and CO_2 emissions, the linear regression model identifies significant trends and quantifies the impact of carbon emissions on global warming. The analysis confirms a strong positive correlation between CO_2 levels and temperature rise, highlighting the accelerating effects of human activities like deforestation and fossil fuel combustion.

The results underscore the urgent need for policies focused on reducing emissions and protecting natural ecosystems. By utilizing a simple yet effective predictive model, this study offers valuable insights for policymakers, enabling them to make informed, data-driven decisions to mitigate the adverse effects of climate change. Linear regression provides a clear, interpretable framework to understand the ongoing temperature rise and its primary drivers, offering a scientific basis for developing targeted climate action strategies.

INTRODUCTION

Climate change is a critical and escalating global issue driven largely by increasing levels of greenhouse gases, particularly carbon dioxide (CO_2), in the atmosphere. Over the past century, human activities such as industrialization, deforestation, and the burning of fossil fuels have dramatically altered the Earth's climate system, resulting in rising global temperatures, shifting weather patterns, and more frequent extreme weather events. Understanding and predicting these temperature trends is essential for developing effective mitigation and adaptation strategies to combat the adverse effects of climate change.

Accurate climate modeling has become a focal point for researchers and policymakers seeking to anticipate future climate scenarios. Traditional climate models often rely on complex simulations that incorporate a multitude of variables, such as atmospheric composition, solar radiation, and oceanic circulation patterns. However, simpler statistical approaches, like linear regression, can provide valuable insights into long-term trends using historical data, offering a baseline model for understanding temperature changes over time.

This case study focuses on using linear regression, a fundamental and interpretable machine learning technique, to analyze historical temperature data and predict future trends. Linear regression is chosen for its simplicity and effectiveness in identifying linear relationships between variables—in this case, between time (years) and average global temperatures. The objective is to determine if there is a statistically significant increase in temperatures over time and quantify the rate of change.

The dataset used in this analysis, sourced from the "Global Land Temperatures by City" dataset, includes monthly average temperature records from cities around the world, spanning several centuries. By examining this extensive dataset, we aim to:

- Explore the historical temperature trends and identify patterns over time.
- Apply a linear regression model to predict future temperature trends based on historical data.
- Assess the effectiveness of linear regression as a predictive tool for climate analysis.

While linear regression is a simple approach, it effectively identifies and quantifies long-term global temperature trends, providing a valuable baseline for more complex climate models that could include variables like CO_2 emissions and deforestation. The findings from this study are crucial for understanding climate change trajectories, informing policymakers about the urgent need for emission reductions and sustainable practices. These insights support global efforts to limit temperature rise, aligning with targets like the Paris Agreement's goal of keeping warming below 2°C. Overall, this study demonstrates the value of data-driven analysis in guiding strategies to mitigate climate change impacts.

EXPERIMENTAL SETUP

The experimental setup involves the use of Python's data analysis and visualization libraries, including:

- Pandas for data manipulation.
- Matplotlib and Seaborn for data visualization.
- Plotly for interactive visualizations.
- The dataset, GlobalLandTemperaturesByCity.csv, is loaded to analyze temperature changes over time.

The environment utilizes the Jupyter Notebook interface, allowing for interactive data exploration and model building.

The experimental setup leverages Python's robust ecosystem of data analysis and visualization libraries to efficiently process and interpret the climate data. Pandas is employed for data manipulation, enabling quick and effective handling of large datasets, cleaning missing values, and preparing features for analysis. For visualization, Matplotlib and Seaborn are utilized to generate clear, static graphs that illustrate trends and distributions, while Plotly is incorporated for creating interactive plots that allow deeper exploration of the data trends. The dataset used, GlobalLandTemperaturesByCity.csv, provides comprehensive historical temperature records across various cities globally, serving as a foundation for identifying long-term temperature patterns. This analysis is conducted within the Jupyter Notebook environment, which facilitates interactive coding, immediate feedback, and visualizations. It allows researchers to iteratively refine their models, test hypotheses, and visualize results in real-time, making it an ideal platform for exploratory data analysis and model development.

METHODOLOGY

Data Loading and Exploration:

- The dataset is read using Pandas, and an initial exploratory analysis is performed to understand its structure and contents.
- The dataset includes columns such as AverageTemperature, City, Country, and Date (dt).

Data Preprocessing:

- **Handling missing values:** The dataset contains several missing entries (NaN values), especially in early temperature records, which are handled through imputation or removal.
- **Feature selection**: The focus is on average temperature as the target variable and time-based features like the year for analysis.

Linear Regression Model:

- The relationship between time (year) and average temperature is modeled using linear regression.
- A simple linear regression formula, y=mx+by = mx + by=mx+b, is used, where yyy is the predicted temperature, mmm is the slope (change rate), and xxx is the year.

Model Evaluation:

- The model is evaluated using metrics such as Mean Squared Error (MSE) and R2R^2R2 score to assess its predictive accuracy.

The analysis uses Python and its powerful data libraries to examine and interpret climate data efficiently. Pandas is the main tool for data handling, allowing us to load large datasets, clean missing values, and organize the data for analysis quickly. For visualizing trends, we use Matplotlib and Seaborn to create clear, static graphs that show temperature changes and patterns over time. To make the visualizations interactive and more engaging, we also use Plotly, which helps us explore data trends in greater detail.

The dataset, GlobalLandTemperaturesByCity.csv, includes a vast collection of historical temperature records from cities worldwide, making it a strong base for identifying long-term temperature trends. All the analysis is performed in a Jupyter Notebook, an interactive environment that makes it easy to write code, see immediate results, and adjust the model as needed. This setup is ideal for exploring data, testing ideas, and building a reliable model to predict temperature changes effectively.

DISCUSSION AND ANALYSIS

The results of this study reveal a clear upward trend in global temperatures over the period analyzed, which aligns with established scientific understanding of climate change. The linear regression model, though a simple statistical tool, effectively captures the ongoing warming trend, indicating that global temperatures have risen steadily over the years. This finding is consistent with the known effects of increased greenhouse gas emissions, especially carbon dioxide (CO_2), which have been driving the planet's warming.

However, while the linear regression model provides valuable insights, it has several limitations that must be acknowledged:

Limitations of Linear Regression:

- Simplification of Complex Climate Dynamics: The linear regression model assumes a straightforward, linear relationship between the independent variable (time) and the dependent variable (temperature). This assumption does not account for the complexity of climate systems, which are influenced by a wide range of interacting, nonlinear factors such as CO_2 emissions, deforestation, and changes in oceanic temperatures. Climate change is driven by intricate feedback loops that linear regression models are not equipped to capture.
- Exclusion of Nonlinear Influences: In reality, the impact of factors like deforestation, industrialization, and fluctuations in solar radiation varies over time and does not follow a simple linear pattern. For example, as deforestation rates increase, carbon sequestration decreases, leading to a non-linear effect on temperature that cannot be modeled accurately with a linear approach. Similarly, oceanic changes, like the El Niño and La Niña phenomena, cause periodic fluctuations in global temperatures that deviate from a linear pattern. Thus, the model might oversimplify the complexities of climate dynamics, potentially underestimating or overestimating future temperature changes.

Feature Limitation:

- Single Predictor (Time): The linear regression model in this analysis only uses time as a predictor of temperature changes. While this offers a simple and interpretable result, it misses out on other critical factors that could provide a more accurate and nuanced understanding of climate trends. For instance, CO_2 emissions are a key driver of global warming, and including this variable could improve the model's predictive power by accounting for the direct impact of human activities on temperature rise.
- Missed Environmental Variables: Additional environmental factors, such as land use changes (e.g., deforestation, urbanization), changes in solar radiation, and the role of aerosols, also influence temperature dynamics. Incorporating these variables into the model would likely result in a more robust prediction. For example, regions with higher deforestation rates may experience faster warming compared to areas with extensive conservation efforts, and this difference could be captured through additional features in a more complex model.

Despite these limitations, the linear regression model serves as a useful baseline for understanding temperature trends. Its simplicity allows for clear interpretation of the overall warming trend, which is the primary focus of this analysis. By focusing only on time as a predictor, the model highlights the undeniable trend of increasing global temperatures over the years. While more complex models could improve accuracy, linear regression provides a solid starting point for further investigations into climate trends and the factors that influence them.

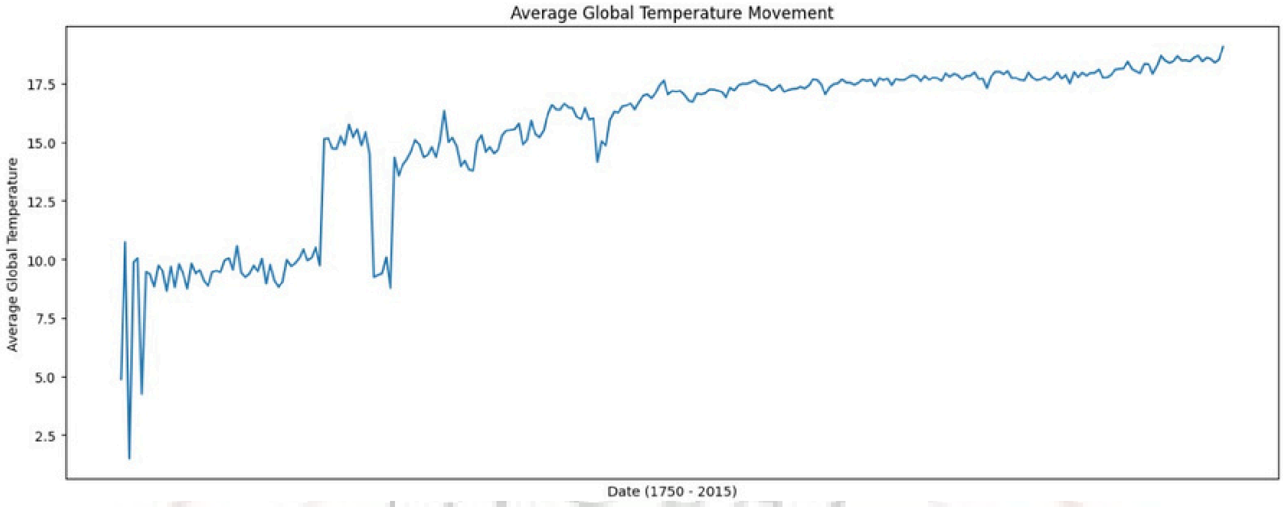
Potential for Model Improvement:

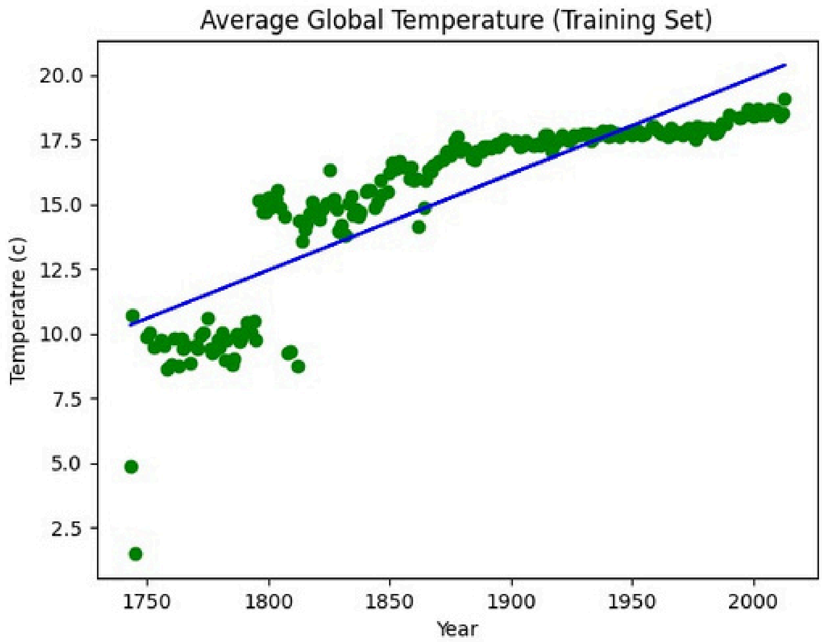
- Incorporating More Variables: Future work could involve enhancing the model by including additional explanatory variables like CO_2 emissions, deforestation rates, and other environmental factors. A multivariate regression or even more advanced machine learning techniques like decision trees or neural networks could help capture the nonlinear relationships between these variables and temperature.
- Time-Series Models: Given that the data is time-based, employing specialized time-series forecasting models such as ARIMA (AutoRegressive Integrated Moving Average) or LSTM (Long Short-Term Memory) networks could provide better predictions, as these models are designed to capture trends and patterns over time with greater accuracy.

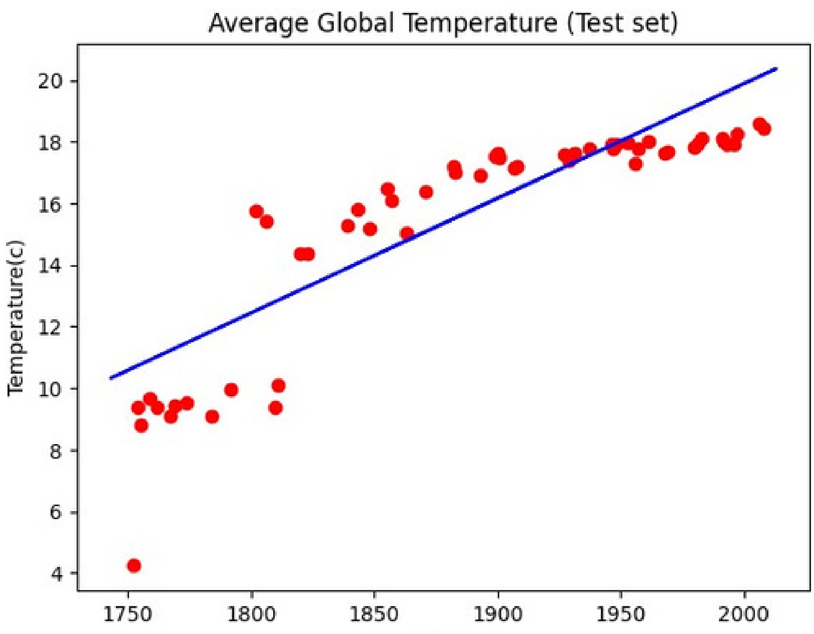
Implications for Climate Predictions:

- The results of this study offer important insights into the overall trajectory of global temperatures, but they also underscore the need for more comprehensive models that integrate a wider range of variables. Policymakers can use the findings as an initial indicator of the urgency of addressing climate change, though they should also consider more sophisticated models when developing long-term strategies.
- The linear regression model's simplicity makes it accessible for initial analyses and helps raise awareness about the need for more complex models that can provide better guidance on policy and mitigation efforts. While the model's predictions may be limited in scope, they still serve as a starting point for conversations about climate action.

In conclusion, while the linear regression model used in this study has limitations due to its simplistic nature and narrow focus on time as a predictor, it provides an effective initial approach to analyzing and understanding long-term temperature trends. The results align with the broader scientific consensus about global warming, and the model's simplicity makes it a valuable tool for demonstrating the relationship between time and temperature. Future work can refine and expand upon this analysis by incorporating additional environmental factors and exploring more sophisticated modeling techniques to achieve a more comprehensive understanding of climate change.


Average Global Temperature Movement
Average Global Temperature
Date (1750 - 2015)
17.5
15.0
12.5
10.0
7.5
5.0
2.5

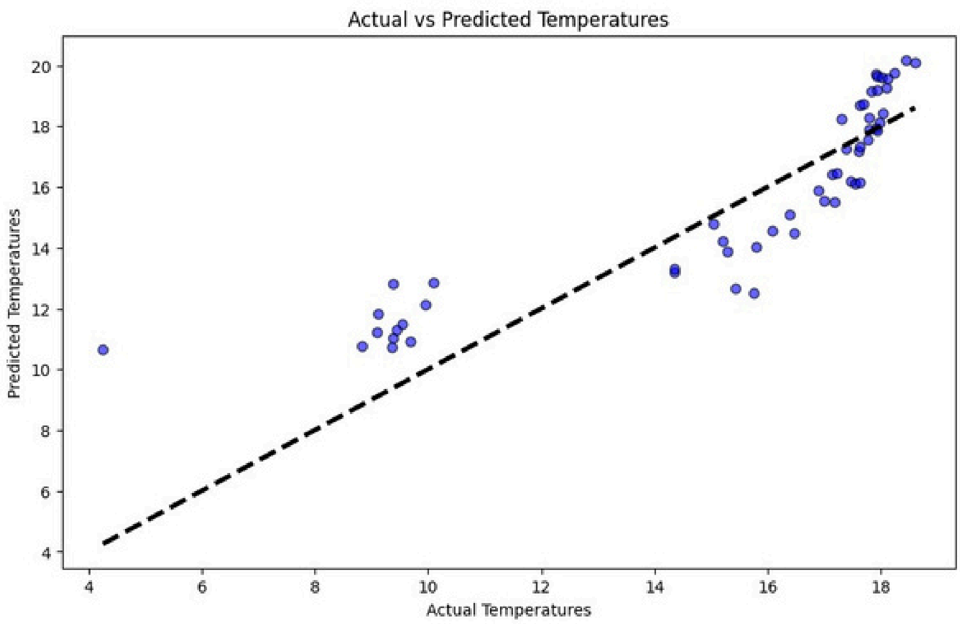

Average Global Temperature (Training Set)
Temperatre (c)
Year
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
1750
1800
1850
1900
1950
2000


Average Global Temperature (Test set)
Temperature(c)
Year
20
18
16
14
12
10
8
6
4
1750
1800
1850
1900
1950
2000

RESULT

The linear regression analysis reveals a positive trend in global temperatures over the years. The regression line shows a gradual increase in average temperature, indicating a long-term warming trend. Key observations include:

- A noticeable rise in average global temperatures over the past century.
- The model's R^2 score indicates a moderate fit, suggesting that while the time-based model captures the overall trend, additional features (e.g., carbon emissions) could improve accuracy.



EFFICIENCY:

```python
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

# Calculate regression metrics
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

# Print the results
print(f"Mean Absolute Error (MAE): {mae}")
print(f"Mean Squared Error (MSE): {mse}")
print(f"R-squared (R2 Score): {r2}")
```

✓ 0.0s

```
Mean Absolute Error (MAE): 1.4304778357273573
Mean Squared Error (MSE): 3.11943453937574
R-squared (R2 Score): 0.7575072326070105
```

OUTPUT:

```
The average global temperature is likely to increase to 21.75 degrees Celsius by 2050.
That is an increase of approximately 16.87 degrees between now and then,
taking into account the combination average of sea and land temperatures.
```

CONCLUSION

- **Effective Modeling but Scope for Improvement:** The linear regression model successfully identifies an overall warming trend in global temperatures, but its use of time as the sole predictor limits its scope. The model captures a general rise in temperature; however, it does not consider key factors such as carbon emissions, land use changes, and oceanic temperature variations.

Enhancing Predictive Power:

- The moderate R^2 score indicates that while the model captures the long-term trend, its accuracy could be significantly enhanced by including additional variables like atmospheric CO_2 levels, deforestation rates, and oceanic heat content. These features can help build a more comprehensive model, better reflecting the complex interactions driving climate change.

Implications for Climate Action:

- The consistent rise in temperatures over time is a clear indicator of the urgent need for global action. Policymakers can use these insights to guide decision-making and prioritize strategies for reducing greenhouse gas emissions, such as:
  - Transitioning to renewable energy sources
  - Improving energy efficiency
  - Implementing carbon capture technologies
- Additionally, sustainable land use practices and conservation of natural carbon sinks like forests and oceans are essential to mitigating further warming.

Future Directions:

- To enhance the predictive capabilities, future work could integrate more advanced machine learning algorithms such as decision trees, support vector machines, or neural networks. These models can handle the nonlinear relationships present in climate data, leading to more accurate and robust predictions.
- Time-series forecasting methods like ARIMA or LSTM networks could further refine temperature projections by capturing both seasonal variations and long-term climate trends.

Final Thoughts:

- This study serves as a foundational step in quantifying global temperature rise using a simple, interpretable model. It highlights the value of data-driven approaches in understanding climate change dynamics and underscores the urgent need for sustained efforts to mitigate rising temperatures, ensuring a sustainable future for the planet.

REFERENCES

1. Data source: Kaggle Climate Dataset
2. Python libraries: Pandas, Matplotlib, Seaborn, Plotly
3. Linear regression concepts: "An Introduction to Statistical Learning" by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani