Development of a Regression Model to Predict Global Warming with Machine Learning

GIL-VERA V. D.[1], QUINTERO-LÓPEZ C.[2]
[1]Department of Engineering,
[1]Information Systems and Knowledge Society Research Group,
Luis Amigo Catholic University,
Transversal 51A # 67B 90,
COLOMBIA

[2]Basic and Applied Neurosciences Research Group,
Luis Amigo Catholic University,
Transversal 51A # 67B 90,
COLOMBIA

Abstract: - Global warming is a phenomenon caused by the increase of greenhouse gases, affecting the global climate, ecosystems, and human health. The alteration of climate patterns and the occurrence of extreme phenomena affect the natural habitats of various species, causing forced migrations, population reduction, and extinction of species. This research uses a simple linear regression (SLR) model based on Machine Learning (ML) to predict the global average temperature (°C) in the short, medium, and long term. Based on historical data and temporal forecasting techniques, the model allows for forecasting future scenarios and assessing possible environmental risks. The developed SLR model performed well (R^2=0.7383), the results underline the importance of accurate predictions for creating effective climate change mitigation policies and strategies.

Key-Words: - Climate change, Environmental impact, Global warming, Greenhouse gases, Machine learning, Regression model, Temperature prediction.

Received: June 26, 2024. Revised: December 29, 2024. Accepted: February 22, 2025. Published: March 26, 2025.

1 Introduction

Global warming is when the earth's temperature and the atmosphere's layers near the planet increase artificially due to the increase of gases produced due to various human activities, which are qualified as greenhouse gases in the atmosphere, [1]. It is generally demonstrated that the climate is worsening globally, and the devastating environmental damages increase as the climate changes, [2].

Thanks to ML, it is possible to analyze large amounts of climate data to understand how various factors contribute to global warming and predict future climate scenarios with greater accuracy, [3]. This work aimed to build a SLR to predict global warming. A ML model, specifically an SLR model was trained to predict the average temperature (°C) 5, 10, 50, or 100 years from now, to map these temperatures to the devastating effects of even a small amount of average temperature change.

Data collected by Berkeley Earth, an independent US non-profit organization dedicated to the science and analysis of environmental data, was used, [4]. Global warming is affecting the earth and the main reason is the increase in temperature. In this work, the increase and decrease of global temperature are analyzed. The data used considers the temperatures of developed and developing countries.

ML is useful for predicting global warming because it can analyze large volumes of data and detect complex patterns that traditional methods do not easily identify. It can process massive data from various sources (satellite images, temperatures, gas emissions) and build accurate predictive models optimized over time, [5]. In addition, it allows predictions to be made at different scales and timescales, supporting decision-making on climate change mitigation and adaptation strategies.

In summary, ML improves predictions and facilitates the design of data-driven policies, aiming to raise awareness of the future consequences of global warming, based on temporal forecasting techniques and ML models, specifically RLS.

2 Problem Formulation

Global climate change refers to the increase in the earth's temperature, caused by human factors, [6]. It is caused by the greenhouse effect of certain gases in the atmosphere such as carbon dioxide (CO_2) or methane (CH_4) that block the escaping heat, [7]. The concentration of these gases has increased dramatically due to human impact since the mid-20th century, with the burning of fossil fuels (oil and gas) and deforestation being the main causes of this increase, [8]. Observed and expected effects include increasingly prolonged periods of drought, forest fires, and greater extreme weather events, [9].

Global warming is the long-term warming of the planet's overall temperature, which has increased significantly over the last hundred years due to the burning of fossil fuels, [10]. As the human population has increased, so has the volume of fossil fuels burned (coal, oil, and natural gas), and their burning causes what is known as the "greenhouse effect" in the Earth's atmosphere, [11]. Global warming of land and sea continues to increase, and the levels of warming have steadily increased decade after decade. Each of the last three decades has been warmer than any previous decade since 1850, [12].

Climate change affects temperatures and has a direct impact on the planet's ecosystems and biodiversity, [13]. This loss of biodiversity has serious consequences for the balance of ecosystems, as each species plays a crucial role in its environment. The change in biodiversity affects the availability of natural resources and can impact economic activities that depend on them, such as fishing and agriculture, [14].

In addition, climate change has significant effects on human health. Higher temperatures contribute to the spread of vector-borne diseases, such as dengue, malaria, and Zika, as they create conditions conducive for vectors, such as mosquitoes, to reproduce and survive in areas where they could not before, [15].

Likewise, the increase in extreme weather events, such as heat waves and floods, generates emergencies that affect public health infrastructure and increase the risk of respiratory diseases, dehydration, and other health problems, [16].

The global economy also faces substantial risks from climate change. Sectors such as agriculture, energy, and tourism are vulnerable to changes in climate, directly affecting food production, energy supply, and recreational activities, [17].

Prolonged droughts and changing rainfall patterns can reduce agricultural yields, increase food prices, and lead to food insecurity, [18]. On the other hand, melting glaciers and rising sea levels put at risk coastal areas, where a significant part of the global infrastructure and human population is concentrated, [19].

3 Problem Solution

Predicting global warming is of vital importance to mitigate the devastating effects caused by climate change, which leads to taking measures to reduce greenhouse gas emissions and mitigate the severity of climate change, [20]. Analyzing how global warming will progress allows assessing the risks to various sectors such as agriculture, water resources, coastal infrastructure, and public health, which allows formulating strategies to cope with changing conditions, [21].

Climate change has a significant impact on ecosystems and biodiversity. Predicting global warming helps to take measures to protect vulnerable species and habitats, [22]. Global warming can cause extreme weather events, heat waves, and the spread of diseases. Predicting these changes helps to implement public health measures and safety protocols, [23].

This environmental problem often disproportionately affects marginalized communities; by predicting global warming, governments can better plan for equitable distribution of resources and protections. Predictive modeling applied to global warming incentivizes international cooperation and agreements on preventive climate actions as countries work together to address the global challenge, [24]. In short, predicting global warming helps humanity prepare for the future and take measures to limit its impact on the planet and its inhabitants.

4 Background

In [25], they discussed the importance of ML in addressing climate change, highlighted high-impact problems where ML can make a difference, and called on the data science community to join the global effort against climate change. These authors emphasize the value of the recommendations to diverse audiences, recognize the need for conceptual innovation, and encourage researchers and engineers to apply their expertise to solve pressing societal problems related to climate change.

In [26], they highlighted the value of linear regression in providing information on climate change patterns, while recognizing the need for more sophisticated modeling approaches. This involves the use of linear regression to analyze historical temperature trends, establish correlations with environmental indicators, identify global temperature variations, and predict future temperature trajectories.

In [27], they used climate data to carry out statistical model building for assessing the causes of global warming. It can be noticed that random forest gave us a better result than other ML algorithms. Carbon dioxide (CO_2) is by far the dominant and most impactful influencer of temperature change, with methane and nitrous oxide following well behind. The analysis stressed the need to look at the impacts of all three gases.

In [28], they have assessed the possibilities of ML approaches in climate change risk assessment (CCRA), outlining diverse existing techniques and algorithms, as well as their performance compared to some other real-valued datasets remote sensing data.

They also remarked that climate change projections and composite hazards are poorly addressed in the scientific literature. Likewise, in [29], the authors explained this potential to harness ML against climate change and qualified its possible power for forecasting, optimization, and scientific experimentation in global warming analysis.

5 Methodology

In this work an SLR model was built, the objective of this ML technique is to try to explain the relationship between a dependent variable Y (dependent variable), and a set of independent variables $X1,...,Xn$ (independent variables).

The relationship between the response variable Y (Average Global Temperature °C) and a single explanatory variable X (Combined Global Temperature °C) is explained in an SLR model.

Using the regression techniques of a variable Y on a variable X, a function that is a good approximation of a cloud of points (xi,yi) (Figure 1).

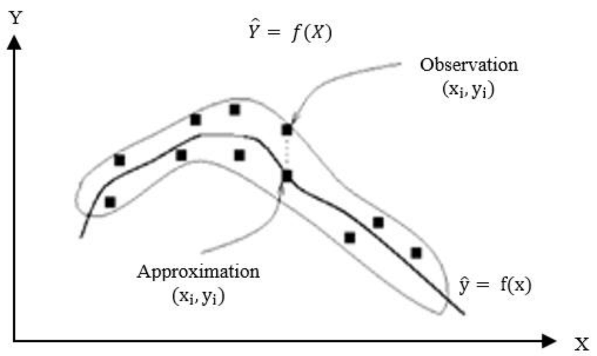

Fig. 1. Distribution diagram

The SLR model has the following expression:

$$Y = \alpha + \beta X + \varepsilon \quad (1)$$

where α is the ordinate at the origin, β is the slope of the line and ε is the error. X and Y are random variables, so an exact linear relationship between them cannot be established.

The data used in the construction of RLS models contain temperature readings for a wide range of dates, from 1750 to 2015 (Figure 2). This dataset specifically covers the combined global temperature, rather than just sea or land temperature. This option was chosen because of the number of comparative analyses available on the combined global temperature rather than the other two options separately. The data used are available in [30].

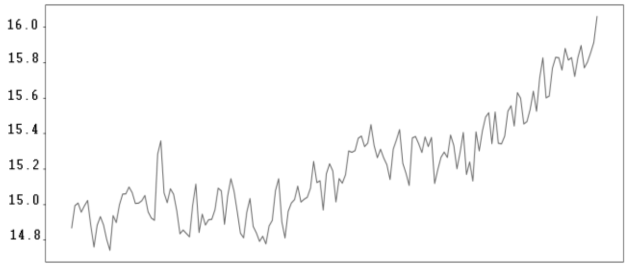

Fig. 2. Average Global Temperature VS Date (1750-2015)

To ensure data quality, preprocessing processes were performed, such as data cleaning (elimination of outliers and missing data by linear interpolation) and normalization (data were standardized to facilitate model comparison and training). The data were divided into a training set (80%) and a test set (20%) by random selection, maintaining the temporal distribution to preserve historical patterns.

To evaluate the effectiveness of the SLR model compared to other Machine Learning techniques, additional models were implemented, including Naïve-Bayes (classification-based probabilistic model), Random Forest (nonlinear ensemble model with multiple decision trees), multilayer perceptron (ANN-MLP) (neural network model capable of capturing nonlinear relationships). Each model was trained and evaluated using the same data and metrics, allowing a fair comparison.

Although the SLR model provides an initial insight into global warming trends, it has inherent limitations. First, the assumption of linearity, climate change effects are often non-linear and influenced by multiple interdependent variables. Second, external factors are not considered, the model does not include variables such as CO_2

emissions, land use, or precipitation patterns, which can significantly influence temperature trends.

Devastating environmental damage will only increase as the climate continues to change. It is proven that the climate is changing for the worse, but that is not the main problem that was addressed in this analysis. In this analysis, I wanted to be able to train a ML model that would tell us what the average temperature will be 10, 50, or 100 years from now so that we can relate these temperatures to the devastating effects of even a small amount of average temperature change.

6 Results

Fig. 3 presents the linear regression line that models the change in average global temperature over time, using a test data set. Years are presented on the X-axis, from about 1750 to 2015. The Y-axis (Temperature in °C), indicates the average global temperature in degrees Celsius.

The line in Figure 3 is a linear regression line indicating an increasing trend in average global temperature over time. The positive slope of the line shows that, on average, the temperature has increased. The dots represent the observed data. Although some points are scattered around the trend line, most of them line up relatively close to it, suggesting a positive correlation between time and temperature.

The dispersion of the points around the line suggests that, although there is a general upward trend, the temperature doesn't rise uniformly each year; in some years there are significant deviations from the trend line.

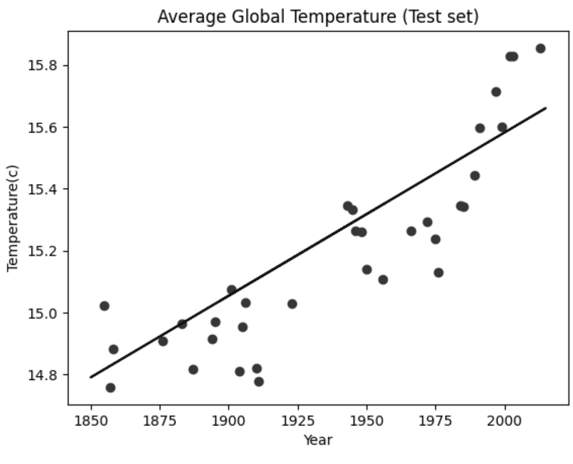

Fig. 3. Year VS Temperature °C

The slope of the regression line is approximately 0.0047. This indicates that, on average, the global temperature has increased by 0.0047 °C per year throughout the record. The intercept is approximately -0.5381, representing the estimated average temperature value in year zero. This value is not interpretatively meaningful in this context, as it does not reflect an actual year.

The Coefficient of Determination (R^2) measures how well a regression model represents the observed data. It is calculated using the following formula:

$$R^2 = 1 - \left(\frac{SSE}{SST}\right) \tag{2}$$

where:
- SSE (Sum of Squared Errors): The sum of the squared differences between the observed values and the values predicted by the model.

$$SSE = \sum(y_i - \hat{y_i})^2 \tag{3}$$

where y_i are the observed values and $\hat{y}_i$ are the values predicted by the model.
- SST (Sum Total of Squares): The sum of the squared differences between the observed values and the mean of the observed values.

$$SST = \sum(y_i - \bar{y})^2 \tag{4}$$

where $\bar{y}$ is the mean of the observed values.

The value of R^2 varies between 0 and 1 (if it is 1 the model explains 100% of the variability in the data, if it is 0 the model explains none of the variability in the data. In simple terms, the R^2 indicates the percentage of the variability in the data that can be explained by the regression model.

The R^2 value was 0.7383. This means that about 73.8% of the variability in the average global temperature can be explained by the linear regression model over time. Although it shows an increasing trend, the relatively low value suggests that factors other than year contribute significantly to temperature variability. In summary, the results indicate an upward trend in average global temperature over the years, although other factors also have a significant influence on year-to-year temperature changes.

Table 1 presents a general comparison of the coefficient of determination (R^2) with other more sophisticated machine learning models.

Table 1. General comparative

Model	**Coefficient of Determination (R^2)**
SLR	0.7383
Naïve-Bayes	0.7925
Random Forest	0.8832
ANN-MLP	0.9015

Although the SLR model obtained the lowest R^2, it was greater than 0.7, which indicates that the model can explain the variability of the average global temperature moderately, indicating its limitations in modeling inherent complexities. The ANN-MLP model obtained the highest performance among the four models that were compared; thus, it can generate more accurate predictions when compared to the Random Forest, Naïve-Bayes, and SLR models.

7 Discussion

A SLR model analysis identified a general trend of increasing average global temperature from 1750 to 2015. Although the value of the coefficient of determination (R^2=0.7383) indicates a limited ability of the model to explain the total variability of the data, this result underscores the multifactorial complexity of climate change and the importance of considering additional variables in future analyses.

At the same time, the positive global panorama of a worldwide temperature alteration considers mankind's booming different action back in October 2023 that has affected worldwide temperature weakening soon after consuming fossil power and clearing woodlands. The results confirm that greenhouse gases including carbon dioxide (CO_2) and methane (CH_4) are the main contributors to warm temperatures over land. Nevertheless, Land use patterns, industrial emissions, and nature variability can also play a role in the model developed during this study however their exclusion made sense given that previous studies describe each parameter of interest as well.

More than anything, the findings highlight the need for practical solutions to develop new policies and technology that radically reduce greenhouse gas emissions.

While very useful in describing the trend over time model provides information on historical trend models, its methodological limitations also present an opportunity to implement more sophisticated ML-based approaches i.e., deep neural networks or random forest models. Such methods may be able to capture nonlinear relationships, as well as correlations among multiple factors that the SLR model does not consider.

Lastly, the results-based assessments also prompt some questions regarding decision-making relevance. Using linear trends to predict future temperature, for example, may greatly underestimate the nonlinear dynamics of climate systems. It is thus important to couple these models with qualitative analyses and scenario studies incorporating climate feedback, and social and economic impacts. This work highlights the possibilities of ML concerning climate change modeling and provides a foundation for more research that can inform actual tools for climate adaptation and mitigation.

8 Conclusion

Due to the modeling of global warming based on ML temperature rise and the influencing variables that will help in predicting the temperature rise effect, preventive measures can be taken to prevent huge environmental and social forces like heat waves extreme weather events, and public health crises from taking place.

Predicting climate change helps to develop impacts and policies that reduce greenhouse gas emissions and improve the resilience of vulnerable communities. Our research shows that ML can be an effective climate analysis tool to provide high-resolution forecasts, essential for the long-term adaptation and mitigation of the impacts of a changing climate.

The SLR model proposed in this paper shows that taken alone, the passage of time is correlated with average global temperature (and insolation), but other SLR models that take additional factors into account may be required to improve predictive power.

This result supports the need to frame climate change in a multidimensional manner, incorporating data-driven analyses with prevention strategies responsive to both social and environmental functions. The model explains little variability so we encourage future work to pursue complex adaptive approaches that can lead to more effective and accurate climate planning.

Future research can expand the dataset with information from satellite sources and climate simulations and perform sensitivity analyses to assess the impact of different climate factors on the predictions.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

The author used ChatGPT 3.5 to improve the manuscript's language and readability. After using this software, the author edited the content as needed and took full responsibility for the publication's content.

References:

[1] K. Akerlof, E. W. Maibach, D. Fitzgerald, A. Y. Cedeno, and A. Neuman, Do people 'personally experience' global warming, and if so how, and does it matter?, *Global Environmental Change-Human and Policy Dimensions*, Vol. 23, No. 1, pp. 81–91, Feb. 2013, doi: 10.1016/j.gloenvcha.2012.07.006.

[2] J. Joireman and R. L. Liu, Future-oriented women will pay to reduce global warming: Mediation via political orientation, environmental values, and belief in global warming, *Journal of Environmental Psychology*, Vol. 40, pp. 391–400, Dec. 2014, doi: 10.1016/j.jenvp.2014.09.005.

[3] R. A. Winter, Innovation and the dynamics of global warming, *Journal of Environmental Economics and Management*, Vol. 68, No. 1, pp. 124–140, Jul. 2014, doi: 10.1016/j.jeem.2014.01.005.

[4] Berkeley Earth, "Data," Berkeley Earth, 2024, [Online]. https://berkeleyearth.org/data/ (Accessed Date: November 14, 2024).

[5] T. Addiscott, Global Warming. Not just the business of climatologists? Facts about global warming. Rational or emotional issue?, *Energy & Environment*, Vol. 22, No. 4, pp. 537–538, Jun. 2011, doi: 10.1260/0958-305X.22.4.537.

[6] W. Shao, B. D. Keim, J. C. Garand, and L. C. Hamilton, Weather, climate, and the economy: Explaining risk perceptions of global warming, 2001-10, *Weather, Climate, and Society*, Vol. 6, No. 1, pp. 119–134, Jan. 2014, doi: 10.1175/WCAS-D-13-00029.1.

[7] T. H. Woo, Climate change analysis in energy-mix with non-carbon emission energy incorporated with pandemic society, Environment, *Development and Sustainability*, Vol. 25, No. 10, pp. 11723–11733, Oct. 2023, doi: 10.1007/s10668-022-02551-9.

[8] D. A. Brown, The importance of expressly examining global warming policy issues through an ethical prism, *Global Environmental Change-Human and Policy Dimensions*, Vol. 13, No. 4, pp. 229–234, Dec. 2003, doi: 10.1016/S0959-3780(03)00053-0.

[9] T. Bodenmann, What does it mean to claim that 'humans cause global warming'? A philosophical analysis of causal claims in the IPCC's arguments for anthropogenic global warming, *GAIA-Ecological Perspectives for Science and Society*, Vol. 17, No. 2, pp. 205–212, 2008, doi: 10.14512/gaia.17.2.8.

[10] W. Shao, Weather, climate, politics, or God? Determinants of American public opinions toward global warming, *Environmental Politics*, Vol. 26, No. 1, pp. 71–96, 2017, doi: 10.1080/09644016.2016.1223190.

[11] X. Zhao, Is global warming mainly due to anthropogenic greenhouse gas emissions?, *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, Vol. 33, No. 21, pp. 1985–1992, 2011, doi: 10.1080/15567030903515013.

[12] J. Bohr, Is it hot in here or is it just me? Temperature anomalies and political polarization over global warming in the American public, *Climatic Change*, Vol. 142, No. 1–2, pp. 271–285, May 2017, doi: 10.1007/s10584-017-1934-z.

[13] L. Tang, M. Higa, N. Tanaka, and N. Itsubo, Assessment of global warming impact on biodiversity using the extinction risk index in LCIA: a case study of Japanese plant species, *International Journal of Life Cycle Assessment*, Vol. 23, pp. 314–323, 2018. doi: 10.1007/s11367-017-1319-6.

[14] F. Busato, B. Chiarini, G. Cisco, and M. Ferrara, Do people really care about global warming? *Economics and Business Letters*, Vol. 11, No. 1, pp. 24–32, Mar. 2022, doi: 10.17811/ebl.11.1.2022.24-32.

[15] H. Olosutean and M. Cerciu, Water sustainability in the context of global warming: A bibliometric analysis, *Sustainability*, Vol. 14, No. 14, pp. 8349-8363, Jul. 2022, doi: 10.3390/su14148349.

[16] A. Gettelman, M. W. Christensen, M. S. Diamond, E. Gryspeerdt, P. Manshausen, P. Stier, D. Watson-Parris, M. Yang, M. Yoshioka, T. Yuan, Has reducing ship emissions brought forward global warming? *Geophysical Research Letters*, Vol. 51, No. 15, pp. 1-8, Aug. 2024, doi: 10.1029/2024GL109077.

[17] P. Panja, T. Kar, and D. K. Jana, Impacts of global warming on phytoplankton-zooplankton dynamics: A modeling study, *Environment, Development and Sustainability*, Vol. 26, No. 5, pp. 13495–13513, May 2024, doi: 10.1007/s10668-023-04430-3.

[18] M. Helbling and D. Meierrieks, Global warming and urbanization, *Journal of Population Economics*, Vol. 36, No. 3, pp.

1187–1223, Jul. 2023, doi: 10.1007/s00148-022-00924-y.

[19] M. J. Kishi, Global warming effects on marine ecosystem, *Research in Marine Sciences*, Vol. 8, No. 3, pp. 233–252, Sep. 2023, [Online]. https://resmarsci.com/wp-content/uploads/2023/10/Research_in_Marine_Sciences_24-83-3.pdf (Accessed Date: November 30, 2024).

[20] P. Donti, How machine learning can help tackle climate change, *XRDS*, Vol. 27, No. 2, pp. 58–61, Dec. 2020, doi: 10.1145/3433142.

[21] Q. Lu and R. B. Lund, Simple linear regression with multiple level shifts, *Canadian Journal of Statistics,* Vol. 35, No. 3, pp. 447–458, Sep. 2007, doi: 10.1002/cjs.5550350308.

[22] G.-V. Anghelache, E. Bugudui, A. Manole F.C.P. Lilea, and C. C. Sava, Using the linear simple regression model in the analysis of trade activity, *Metalurgia International*, Vol. 18, No. 2, pp. 220-222, 2013.

[23] S. Wang, W. Chen, and R. Yang, Fisher information in ranked set sampling from the simple linear regression model, *Communications in Statistics - Simulation and Computation*, Vol. 53, No. 3, pp. 1274–1284, Mar. 2024, doi: 10.1080/03610918.2022.2044053.

[24] K. A. Marill, Advanced statistics: Linear regression, Part I: Simple linear regression, *Academic Emergency Medicine*, Vol. 11, No. 1, pp. 87–93, Jan. 2004, doi: 10.1197/S1069-6563(03)00600-6.

[25] D. Rolnick, P. L. Donti, L. H. Kaack, K. Kochanski, A. Lacoste, K. Sankaran, A. S. Ross, N. Milojevic-Dupont, N. Jaques, A. Waldman-Brown, A. S. Luccioni, T. Maharaj, E. D. Sherwin, S. K. Mukkavilli, K. P. Kording, C. P. Gomes, A. Y. Ng, D. Hassabis, J. C. Platt, F. Creutzig, J. Chayes, and Y. Bengio, Tackling Climate Change with Machine Learning, *ACM Comput. Surv*., Vol. 55, No. 2, pp. 1–96, Feb. 2022, doi: 10.1145/3485128.

[26] A. Vishwakarma, Unsustainable Management of Plastic Wastes in India: A Threat to Global Warming and Climate Change, in Contemporary Environmental Issues and Challenges in Era of Climate Change, P. Singh, R. Singh, and V. Srivastava, Eds., Singapore: Springer Singapore, 2020, pp. 235–244. doi: 10.1007/978-981-32-9595-7_13.

[27] H. Zheng, Analysis of Global Warming Using Machine Learning, *Computational Water, Energy, and Environmental Engineering*, Vol. 07, No. 3, pp. 127–141, 2018, doi: 10.4236/cweee.2018.73009.

[28] F. Zennaro, E. Furlan, C. Simeoni, S. Torresan, S. Aslan, A. Critto, and A. Marcomini, Exploring machine learning potential for climate change risk assessment, *Earth Science Reviews*, Vol. 220, pp. 103752-103771, 2021, doi: 10.1016/j.earscirev.2021.103752.

[29] S. K. Prion and K. A. Haerling, Making sense of methods and measurements: Simple linear regression, *Clinical Simulation in Nursing*, Vol. 48, pp. 94–95, Nov. 2020, doi: 10.1016/j.ecns.2020.07.004.

[30] V. D. Gil-Vera, "Global Temperatures Dataset," GitHub, 2024, [Online]. https://github.com/victorgil777/Global_Warming/blob/main/GlobalTemperatures.csv (Accessed Date: November 30, 2024).

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Gil-Vera, V.D. carried out the simulation and optimization, implemented the algorithm, organized and executed the experiments in Section 4, and was responsible for the statistics. Quintero-López, C. contributed to the proofreading of the final version of the manuscript.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

The Luis Amigo Catholic University financed this research - SISCO Research Group, Cost Center 0502020965.

Conflict of Interest

The author has no conflicts of interest to declare.