Predictive Modeling of CO2 Emissions in the Philippines: A Machine Learning Approach

Isabel Joy D. Adriatico
University of San Carlos
Cebu City, Cebu, Philippines
17101842@usc.edu.ph

Rolan Dave I. Pacampara
University of San Carlos
Cebu City, Cebu, Philippines
18101459@usc.edu.ph

Jameson Adriel S. Perez
University of San Carlos
Cebu City, Cebu, Philippines
19103259@usc.edu.ph

Shannen S. Du
University of San Carlos
Cebu City, Cebu, Philippines
18400199@usc.edu.ph

Nikki Jel M. Pan
University of San Carlos
Cebu City, Cebu, Philippines
19103491@usc.edu.ph

Daryll L. Restauro
University of San Carlos
Cebu City, Cebu, Philippines
19103225@usc.edu.ph

Abstract—This paper presents a machine learning approach to predict CO2 emissions in the Philippines. The study utilizes a dataset containing various features such as year, population, GDP, and different sources of CO2 emissions. The predictive model is implemented using linear regression, and its performance is evaluated. The results indicate the effectiveness of the proposed approach in predicting CO2 emissions.

Keywords—CO2 emissions, machine learning, predictive modeling, linear regression, feature scaling

I. INTRODUCTION

The escalation in carbon dioxide (CO2) emissions represents an enduring menace to the environment, leading to irreversible consequences such as global warming, elevated sea levels, and shifts in climate [1]. In addition to the adverse effects on the environment, it has also been associated with detrimental consequences for human health [2].

As a developing nation, the Philippines faced the impacts of climate change, exemplified by events like Super Typhoon Yolanda, contributing to its 5th position on the long-term Climate Risk Index [3]. The Philippines signed the Paris Agreement in late 2016 and committed to a conditional target of 70% CO2 emission reduction below business as usual by 2030 in its Intended Nationally Determined Contribution (INDC); the reduction comes from energy, transport, waste, forestry, and industry. The Paris Agreement strengthened global action against climate change to enhance UNFCCC implementation [4]. Nonetheless, the Philippines uses over 88% of fossil fuels in its overall usage of primary energy [5]. With the expectation of driving economic growth, there is a projected increase in energy consumption. This is compounded by the Department of Energy's plans to amplify coal-based energy generation and augment the capabilities of coal-fired power plants until 2040. Given the Philippines' heavy reliance on fossil fuels as a catalyst for economic development, it remains to be seen how the government will achieve its objectives [6].

This paper focuses on predicting CO2 emissions in the Philippines using a machine learning approach. The study aims to provide insights into the relationships between various features and CO2 emissions.

II. METHODS

A. Data Collection

The dataset used in this study is sourced from the repository of online scientific publication Our World In Data. It includes features such as year, population, GDP, and different types of CO2 emissions (coal, oil, gas). The dataset serves as the foundation for understanding the relationships and patterns that contribute to CO2 emissions in the Philippines.

B. Data Preprocessing

The dataset undergoes preprocessing steps, including handling missing data, feature scaling, extracting relevant features and the target variable, and splitting it into training and test sets.

C. Modeling

A linear regression model is chosen for CO2 emission prediction due to its fundamental assumption of a linear relationship between the independent and dependent variables [7]. Using this model, our dependent variable will only be CO2 emissions while our independent variables will include economic indicators, such as population and Gross Domestic Product (GDP), and CO2 emissions specifically from coal, oil, and gas consumption in the Philippines from 1910 to 2018. CO2 emissions from coal, oil, and gas consumption was used to explore how the changes in each type of energy consumption relate to changes in CO2 emissions. Year will also be included as an independent variable as we are dealing with a time-series data which will allow the model to account for potential trends or seasonality in CO2 emissions over time.

This choice is deemed suitable as all of the said independent variables may exhibit linear relationships with CO2 emissions. These are then utilized to train the model. The model's performance is then rigorously evaluated on both the training and testing sets, allowing for a comprehensive assessment of its ability to capture and predict CO2 emissions patterns. The interpretability of linear regression coefficients further aids in understanding the impact of each selected feature on CO2 emissions, providing valuable insights into the driving factors of environmental impact.

D. Training the Model

Once the model is selected, it needs to be trained on a dataset. Training involves fitting the model to the historical data, allowing it to learn the underlying patterns and relationships.

E. Prediction

Once the model is trained, it is employed to make predictions on the test dataset. This phase assesses the model's ability to generalize to unseen data. The predictions provide insights into the model's performance and its capacity to accurately estimate CO2 emissions based on the selected features.

F. Evaluation

The model's performance is rigorously evaluated using key metrics such as mean squared error (MSE) and R-squared (R2) score. The MSE quantifies the average squared difference between predicted and actual values, while the R2 score indicates the proportion of variance in the target variable that is predictable from the features. These metrics offer a comprehensive assessment of the model's predictive capabilities.

III. RESULTS AND DISCUSSION

A. Mean Squared Error (MSE) & R-squared score (R2)

Figure 1 showcases the results of the MSE and R2 of the linear regression model. MSE reflects the overall accuracy of the model across all data points with lower MSE value (i.e. closer to 0) indicating a better predictive performance while R2 is a measure of how well the independent variables explain the variability of the dependent variable. It ranges from 0 to 1, with 1 indicating a perfect fit.

In this case, having a low MSE and a high R2 value implies that there is minimal error in the model's ability to accurately predict CO2 emissions, and the chosen independent variables are highly effective in explaining the variation in CO2 emissions.

Mean Squared Error: 0.112070784413155553
R-squared (R2) Score: 0.9998762669252846

Fig 1. The Mean Squared Error and R-Squared Score

B. Linear Regression of Actual Versus Predicted CO2 Emissions

Figure 2 shows that almost all points lie on the diagonal line, indicating that the predicted values match the actual values. The figure is consistent with the overall performance of the model as reflected by the Mean Square Error and R-squared score.

Deviations from the line suggest areas where the model can be improved or where additional investigation may be needed.

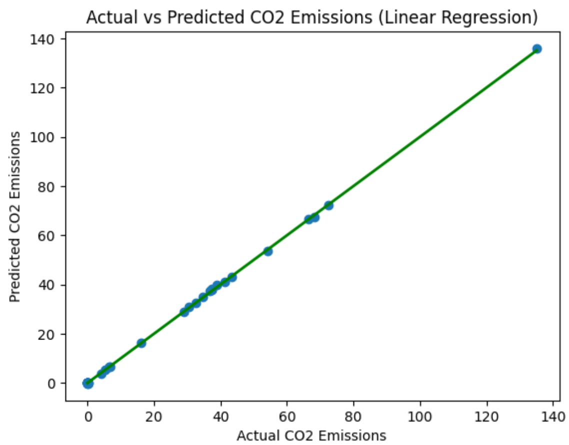

Fig 2. Scatter plot of actual versus predicted CO2 emissions

C. Linear Regression Residual Plot

Figure 3 depicts the residuals (i.e. differences between actual and predicted values) is centered around zero. A horizontal line at y = 0 in the residual plot represents this scenario. A random scatter suggests that the model is capturing the underlying patterns in the data. Residuals are approximately normally distributed. These characteristics suggest that the model is performing well and meeting its assumptions.

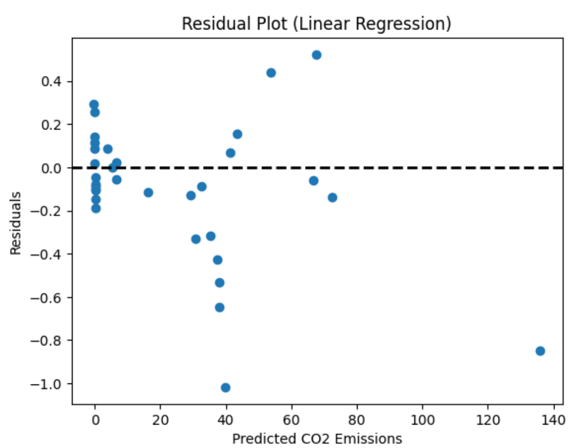

Fig 3. Residual plot of predicted CO2 emissions

IV. CONCLUSION

In conclusion, the analysis of the linear regression model's performance, as illustrated in Figures 1, 2, and 3, provides compelling evidence of its effectiveness in predicting CO2 emissions. The low Mean Squared Error (MSE) indicates that, on average, the model's predictions closely align with the actual values, showcasing a high level of accuracy. The high R-squared value further emphasizes the model's capability to explain a significant proportion of the variability in CO2 emissions based on the selected independent variables.

Figure 2 visually reinforces the model's accuracy, with the majority of points lying along the diagonal line, indicating a close match between predicted and actual values. Deviations from the line, if any, may signal areas for improvement or the need for further investigation.

Figure 3, depicting the residuals centered around zero with a random scatter, suggests that the model captures the underlying patterns in the data. The approximately normal distribution of residuals aligns with the assumptions of linear regression, indicating that the model is performing well.

Overall, the combination of these analyses provides strong support for the reliability and robustness of the linear regression model in predicting CO2 emissions. However, continuous monitoring and validation on new data are recommended to ensure the model's continued performance and generalization to unseen scenarios. The model's success in meeting assumptions and delivering accurate predictions underscores its potential utility in understanding and managing CO2 emissions based on the specified independent variables.

REFERENCES

[1] S. Solomona, G.K. Plattnerb, R. Knuttic and P. Friedlingsteind, "Irreversible climate change due to carbon dioxide emissions," Proceedings of the National Academy of Sciences of the United States of America, vol. 106, no. 6, p. 1704–1709, 2008.

[2] S. Chaabouni and K. Saidi, "The dynamic links between carbon dioxide (CO2) emissions, health spending and GDP growth: A case study for 51 countries," Environmental Research, vol. 158, pp. 137–144, 2017.

[3] S. Kreft, D. Eckstein, and I. Melchior, “Global climate risk index 2017: Who suffers most from Extreme weather events? Weather-related loss events in 2015 and 1996 to 2015,” 2016.

[4] Republic of the Philippines, “Intended Nationally Determined Contributions (INDCs),” 2015.

[5] BP Statistical Review of World Energy, “BP

Statistical Review of World Energy 2018," London, UK, 2018.
[6] Department of Energy, "Coal Roadmap 2017-2040," 2017. [Online]. Available: https://www.doe.gov.ph/pep/coal-roadmap2017-2040. [Accessed: 13-Dec-2023].
[7] JMP, Introduction to Statistics. https://www.jmp.com/en_us/statistics-knowledge-portal/what-is-regression/simple-linear-regression-assumptions.html