Proceedings of Book Series PBS -Vol 42, pp.29-36 **Advances in Energy Transition and Environmental Impacts**

Machine Learning Approaches for Predicting Transport-Related CO_2Emissions in North Africa

Najah Kechiche [#1], Jawher Bouaziz [#2], Mustapha Lamine [3], Walid Hassen[#4]

[#] *Laboratory of Metrology and Energy Systems, National Engineering School of Monastir, Monastir University, 5019 Monastir, Tunisia*

[1]*Higher Institute of Transport and Logistics of Sousse, Sousse University, 4023 Sousse, Tunisia*
kechiche2000@gmail.com

[2]*The Higher National Engineering School of Tunis (ENSIT), Tunis University, 1008 Tunis, Tunisia.*
bouazizjawher1997@gmail.com

[3]*Laboratory of Mechanics and High Energy Physics, Faculty of Sciences Aïn-Chock, Hassan II University of Casablanca, P.O. Box 5366 Maarif Casablanca 20100, Morocco*
mustapha.lamine@gmail.com

[4] *National Engineering School of Monastir, Monastir University, 5019 Monastir, Tunisia*
hassen.walid@gmail.com

***Abstract*— In this study, we address the gap in TCO_2emissions prediction in the North African region by developing a machine learning framework for Algeria, Morocco, and Tunisia. The correlation analysis revealed that population and urbanization rate are the main positive drivers of transport-related CO_2emissions, whereas renewable energy contributes to emission reduction. Four machine learning algorithms (RF, SVR, XGBoost, and MLP) were developed and evaluated for TCO_2 prediction. The developed models achieved strong predictive performance during training, with R^2 values exceeding 0.994 and very low prediction errors. On test dataset, SVR provided the best performance for Algeria (R^2 = 0.98), RF showed the highest accuracy for Morocco (R^2 = 0.96), and MLP outperformed the other models for Tunisia (R^2 = 0.99). These results demonstrate the capability of machine learning techniques to predict TCO_2emissions and emphasize the need for selecting the most appropriate model to each country.**

***Keywords*— Transport-related CO_2 emissions; Machine learning; Carbon emission prediction; North Africa; Sustainable transportation**

I. INTRODUCTION

Climate change, primarily caused by carbon emissions, poses a significant global environmental challenge [1]. In this regard, the transport sector accounts for about a quarter of the total greenhouse gas emissions [2], indicating the significance of developing accurate prediction models for the implementation of the mitigation strategies and Paris Agreement objectives [3, 4]. Different methodologies, such as gray prediction and statistical techniques and various artificial intelligence methods, including Support Vector Regression (SVR), Random Forest (RF), the Extreme Gradient Boosting (XGBoost) method, Multiple-Layer Perceptron (MLP), and many other models, have been used to predict CO_2emissions from the transport sector (TCO_2) [5-10]. The effectiveness of these models for predicting TCO_2emissions and their contribution to the development of sustainable transport policies has been shown in several studies [5, 6]. On the other hand, Africa contributes only a small share of overall greenhouse gas emissions. Rapid urbanization, industrialization, and economic expansion are accelerating TCO_2emissions. South Africa, Nigeria, Egypt, and Algeria are the largest carbon emitters on the continent [11]. Despite this growing challenge, studies on TCO_2prediction in Africa remain scarce. To our knowledge, only one recent study has applied machine learning to predict transportation-related GHG emissions in Africa's major emitters, identifying CatBoost as the best-performing model (R^2 = 0.960, MAPE = 4.9%, rRMSE = 6.7%) and highlighting passenger cars in use and GDP as the main driving factors [12]. In North Africa, a study conducted in Morocco used ANN and SVR models to forecast transport energy consumption (TEC) and CO_2emissions under six economic and electric-vehicle adoption scenarios. The ANN model predicted a decline in TEC and CO_2emissions by 2030, whereas SVR projected values close to those of 2020, highlighting the strong influence of economic growth and the important role of electric vehicles in reducing future transport emissions [13]. To address this gap, the present study investigates transport-related CO_2 (TCO_2) emissions in Tunisia, Algeria, and Morocco using machine learning techniques. The study first examines the relationships between TCO_2emissions and a broad set of demographic, economic, energy,

Copyright © 2026
ISSN: 2961-6611

transport, and environmental variables through correlation analysis. It then develops and compares the predictive performance of four machine learning algorithms to identify the most accurate model for each country. By providing the first comparative machine learning framework for TCO_2prediction across multiple North African countries, this study contributes to the existing literature and provides valuable insights for sustainable transport planning and climate change mitigation.

II. METHODOLOGY

In the present study, we develop a machine learning model to predict total TCO_2in Algeria, Morocco, and Tunisia. Our methodology consists, first, of a correlation analysis that was conducted to identify the relationships between TCO_2emissions and key features, including demographic, energy, and climatic variables. Second, after the preprocessing process, we developed four machine learning models: Random Forest (RF), Support Vector Regression (SVR), Extreme Gradient Boosting (XGBoost), and Multi-Layer Perceptron (MLP). All approaches were optimized using grid search with 5-fold cross-validation. We divided it into 80% for training and 20% for testing, and then we evaluated model performance using classic metrics such as coefficient of determination (R^2), mean absolute error (MAE), and relative root mean square error (rRMSE). Finally, we selected and compared the best-performing model for each country based on regression and its predictive accuracy.

A. Data Collection

The dataset used in this study was collected from Our World in Data (OWID) [14], and covers the period between 1990 and 2021 for the three North African countries. All variables are collected by the site Our World in Data. The selected predictors include population (POP), renewable energy consumption (REN), fossil fuel combustion energy consumption, average surface temperature (TEMP), and urbanization rate (URB). These variables were selected due to their influence on TCO_2emissions. Table 1 presents the target variable (TCO_2) and the independent variables used in our study.

TABLE I
DESCRIPTION OF THE TARGET AND INDEPENDENT VARIABLES USED FOR TCO_2 PREDICTION.

Feature	Abbreviation	Domain	Unit / Description
Total transport-related CO_2 emissions	TCO_2	Environmental	$MtCO_2$
Population	POP	Socioeconomic	Number of inhabitants
Renewable Energy Consumption	REN	Energy	% of total energy consumption
Energy Combustion	EC	Energy	TJ (consumed energy from combustion)
Surface Temperature	TEMP	Meteorological	°C
Urbanization Rate	URB	Socioeconomic	% of urban population

B. Performance evaluation metrics

To evaluate the performance of the prediction models, this study employed three of the most used metrics, including: the coefficient of determination (R^2) mean absolute error (MAE), and the relative Root Mean Square Error (rRMSE) as show in Table 2.

TABLE III
DEFINITIONS AND FORMULAS OF THE PERFORMANCE EVALUATION METRICS.

Metric	Definition	Formula
R^2	Measures the proportion of variance in the observed TCO_2 values explained by the model. Higher values (closer to 1) indicate better predictive performance [15].	$R^2 = \left(1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2}\right)$
MAE	Measures the average absolute difference between observed and predicted TCO_2 values. Lower values indicate better accuracy [16].	$MAE = \frac{1}{N}\sum_{i=1}^{N}\lvert y_i - \hat{y}_i\rvert$

Copyright © 2026
ISSN: 2961-6611

Proceedings of Book Series PBS -Vol 42, pp.29-36 **Advances in Energy Transition and Environmental Impacts**

rRMSE	The relative Root Mean Square Error (rRMSE) ranges between 0 and 100%. It is obtained by dividing the RMSE by the mean of the actual (observed) values and multiplying by 100 to express it as a percentage. The closer of the rRMSE value is to 0, the better the model's predictive performance [17].	$rRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N} y_i - \hat{y}_i}}{\bar{y}_i} \times 100$

III. RESULTS AND DISCUSSION

A. Correlation Analysis

This section presents the correlation analysis between TCO_2emissions and the selected explanatory variables for each country (Algeria, Morocco, and Tunisia). In Algeria, the population has exhibited the highest positive correlation with TCO_2emissions (r = 0.95), indicating a strong influence of population growth on the rise in transport emissions. Urbanization has also demonstrated a very strong positive correlation (r = 0.94), suggesting that it is closely related to a rise in transportation demand and transportation-related energy requirements. Surface temperature displays a moderate positive correlation with transport emissions (r = 0.56), indicating that climate variation may affect transport emissions indirectly by changing energy requirements and travel behaviour. Energy combustion has shown a weak positive correlation (r = 0.35), indicating that fossil fuel consumption remains a significant factor for emissions, but change in time is not as closely related to TCO_2as in the case of population and urbanization. On the other hand, renewable energy shows a relatively strong negative correlation with TCO_2(r = -0.82) as it highlights the significance of renewable energy development in reducing transport-related CO_2emissions, as shown in Fig.1.a.

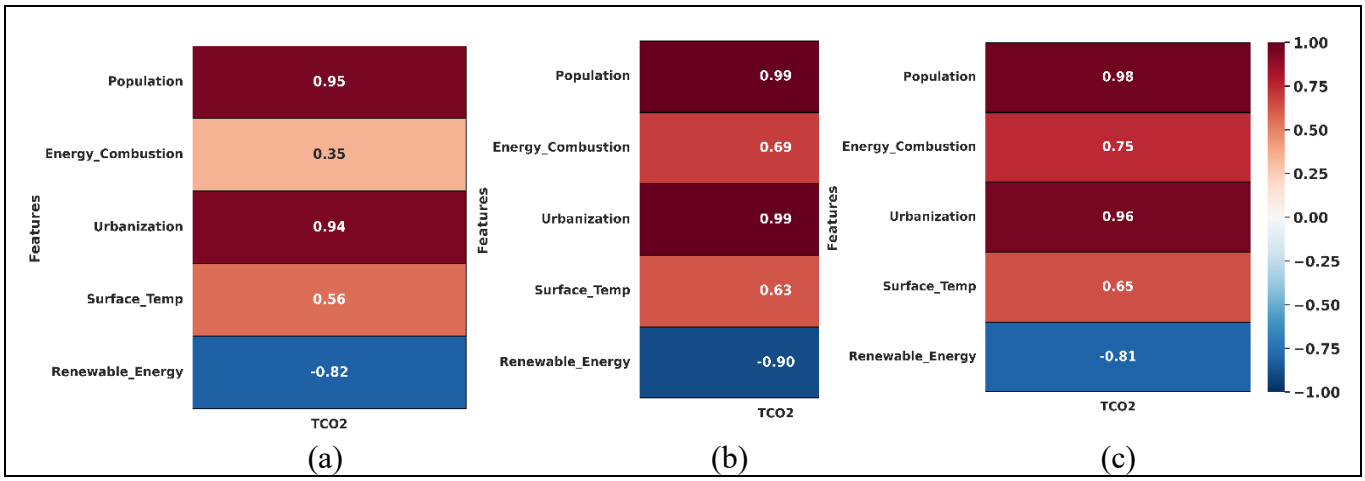


Fig. 1 Correlation analysis between TCO_2emissions and selected features for (a) Algeria, (b) Morocco, and (c) Tunisia.

When we compared it to the other two countries of this study, Morocco presents extremely strong correlation patterns. The results affirm that both population and urbanization have very high correlation values with TCO_2(r = 0.99). This suggests that the growth in population and urbanization is closely correlated with the increase in transport emissions. Energy combustion also shows a positive correlation but is not equal to that of the population and urbanization (r = 0.69). This implies that fossil fuel resources continue impacting levels of emissions. Surface temperature has shown a moderate positive association: r = 0.63, which points out that there might be some sort of relationship between transport energy demand and climatic conditions. On the contrary, renewable energy has shown the strongest negative correlation in the analysis (r = -0.90). This suggests that the development of renewable energy leads to a decrease in transport CO_2 emissions (Fig.1.b).

For Tunisia, the population variable has a very strong correlation coefficient with TCO_2 emissions (r = 0.98), which shows the importance of demographic trends in the formation of transportation emissions. Urbanization also has a very high coefficient of correlation (r = 0.96), which means that urban growth is related to the increase in transport movement. In addition, energy combustion shows a higher strength of relationship as compared to Algeria and Morocco (r = 0.75), indicating that fossil fuel consumption is more significantly influencing transport emissions in Tunisia. Surface temperature shows a moderate positive

Copyright © 2026
ISSN: 2961-6611

correlation (r = 0.65), while renewable energy shows a significant negative correlation (r = -0.81), demonstrating the benefits of adopting renewable energy sources in reducing CO_2 emissions in transportation, as shown in Fig.1.c. Across the three North African countries, Population and Urbanization consistently show the strongest positive relationships with TCO_2emissions, demonstrating the dominant influence of demographic pressure and urban expansion on transport carbon intensity. Renewable Energy is negatively correlated with emissions in all cases, with the strongest mitigation association observed in Morocco (r = -0.90), followed by Algeria (r = -0.82) and Tunisia (r = -0.81). The stronger correlation between Energy Combustion and TCO_2in Tunisia (r = 0.75) compared with Algeria (r = 0.35) and Morocco (r = 0.69) suggests a higher sensitivity of Tunisian transport emissions to fossil fuel consumption patterns. Overall, the correlation analysis confirms that demographic evolution, urbanization, energy structure, and renewable energy penetration are key factors associated with transport-related CO_2emissions variations in North African countries.

B. Prediction results

Table 3 presents the training results of the four machine learning approaches (RF, SVR, XGBoost, and MLP) for Algeria, Morocco, and Tunisia. All the models produced an excellent predictive fit with R^2 values between 0.994 and 1.000, indicating that all approaches used in the training phase capture the relationship between independent variables and TCO_2. Among all models, XGBoost was always the best model, reaching perfect training accuracy (R^2 = 1.000) with almost no errors (MAE = 0.000-0.001; rRMSE = 0.000) for all study countries. Overall, all models performed successfully with R^2 values greater than 0.994 along with low MAE (0.092-0.427) and rRMSE (0.0%-2.7%) on the training dataset.

TABLE IIIII
MODEL PERFORMANCE ON THE TRAINING DATASET.

Country	Model	R^2	MAE	rRMSE (%)
Algeria	RF	0.998	0.427	2.1
	SVR	0.997	0.254	2.1
	XGBoost	1.000	0.001	0.0
	MLP	0.999	0.188	1.0
Morocco	RF	0.997	0.187	2.3
	SVR	1.000	0.092	0.9
	XGBoost	1.000	0.000	0.0
	MLP	0.997	0.192	2.2
Tunisia	RF	0.994	0.097	2.7
	SVR	0.997	0.093	1.9
	XGBoost	1.000	0.000	0.0
	MLP	0.995	0.098	2.6

For Algeria, the Support Vector Regression (SVR) model provided the best performance, achieving the highest R^2 value (0.985), the lowest MAE (1.179), and the lowest relative RMSE (4.7%). Random Forest (RF) also showed strong predictive capability (R^2 = 0.976, MAE = 1.505, rRMSE = 6.0%), while XGBoost achieved slightly lower accuracy (R^2 = 0.961, rRMSE = 7.6%). The Multilayer Perceptron (MLP) model obtained the lowest performance among the tested algorithms (R^2 = 0.920, MAE = 2.433, rRMSE = 10.9%), indicating a weaker generalization ability for Algeria's testing data.

For Morocco, the four models exhibited relatively similar performances, with R^2 values ranging from 0.961 to 0.967. The RF model achieved the highest determination coefficient (R^2 = 0.967) and a low rRMSE of 6.1%, demonstrating good predictive stability. However, SVR provided the lowest MAE value (0.560), followed closely by MLP (0.590), indicating accurate prediction errors despite slightly lower variance explanation. XGBoost showed comparable results (R^2 = 0.963, MAE = 0.706, rRMSE = 6.5%), confirming the robustness of ensemble-based approaches for Morocco as shown in Table 4.

For Tunisia, the best prediction accuracy was obtained using the MLP model, which achieved the highest R^2 value (0.991), the lowest MAE (0.111), and the lowest rRMSE (2.5%) among all evaluated models and countries. RF also demonstrated excellent performance (R^2 = 0.988, MAE = 0.140, rRMSE = 2.9%), while SVR and XGBoost showed slightly lower but still satisfactory results ($R^2 \approx 0.95$ and rRMSE < 6%). These

Copyright © 2026
ISSN: 2961-6611

Proceedings of Book Series PBS -Vol 42, pp.29-36 **Advances in Energy Transition and Environmental Impacts**

results suggest that nonlinear learning approaches, particularly neural networks, are highly effective in capturing the specific patterns governing TCO_2emissions in Tunisia.

In general, the testing results indicate that model performance is country-dependent, reflecting differences in emission drivers, energy structures, transportation systems, and socioeconomic conditions. SVR was the most accurate model for Algeria, RF provided the most consistent performance for Morocco, and MLP achieved the highest predictive capability for Tunisia. The low rRMSE values (<11% for all cases and <7% for most models) confirm the reliability of the developed machine learning framework for TCO_2emission prediction across North African countries. Table 4 presents the predictive performance of the four machine learning models on the independent testing dataset for Algeria, Morocco, and Tunisia.

TABLE IVV
MODEL PERFORMANCE ON THE TESTING DATASET.

Country	Model	R^2	MAE	rRMSE (%)
Algeria	RF	0.976	1.505	6.0
	SVR	0.985	1.179	4.7
	XGBoost	0.961	1.922	7.6
	MLP	0.920	2.433	10.9
Morocco	RF	0.967	0.660	6.1
	SVR	0.961	0.560	6.6
	XGBoost	0.963	0.706	6.5
	MLP	0.961	0.590	6.6
Tunisia	RF	0.988	0.140	2.9
	SVR	0.950	0.236	5.9
	XGBoost	0.951	0.286	5.8
	MLP	0.991	0.111	2.5

Fig.2 presents the comparison plots between observed and predicted TCO_2emissions for Algeria, Morocco and Tunisia which confirms the results reported in Tables 3 and 4.

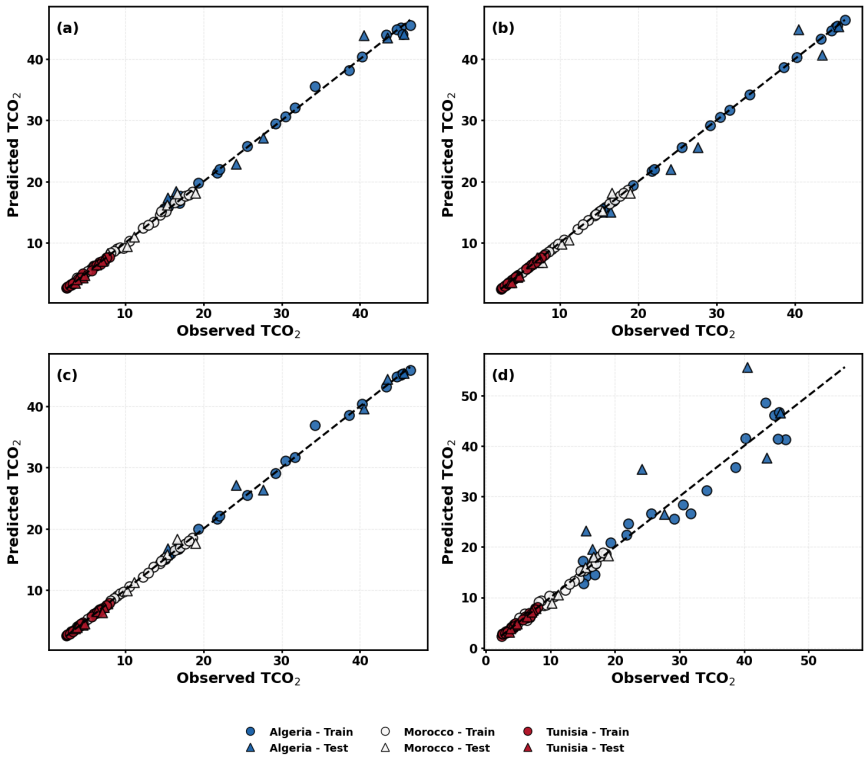


Fig. 2 Regression plots of the machine learning models: (a) Random Forest (RF), (b) Support Vector Regression (SVR), (c) XGBoost, and (d) Multilayer Perceptron (MLP).

Copyright © 2026
ISSN: 2961-6611

Proceedings of Book Series PBS -Vol 42, pp.29-36 **Advances in Energy Transition and Environmental Impacts**

For Algeria, both training and testing phases show good agreement, with slight underprediction and overprediction mainly observed in the middle and higher emission ranges. Particularly for the MLP model, an underprediction tendency is observed during the training phase for emission levels above 30 $MtCO_2$, while some overprediction occurs at lower ranges above 20 $MtCO_2$. In the testing phase, the model shows a tendency to overestimate emissions above 50 $MtCO_2$.

For Morocco and Tunisia, the predicted values closely match the observed emissions, confirming the reliability of the models and the statistical results.

Despite these deviations, the low errors and high accuracy metrics indicate strong model performance.

IV. Conclusions

This study developed and assessed machine learning models for predicting transportation-related CO_2 emissions (TCO_2) in Tunisia, Algeria, and Morocco. The correlation analysis indicated that urbanization and population increase were the primary factors affecting emission levels, and that renewable energy had a negative association with TCO_2, indicating that it may help reduce emissions. Additionally, there was a positive association between energy combustion and transportation emissions due to the use of fossil fuels.

The comparison of RF, SVR, XGBoost, and MLP models revealed that country-specific features, such as energy systems, transportation patterns, and socioeconomic situations, affect prediction performance. For Algeria, SVR performed the best, RF produced the most reliable results for Morocco, and MLP demonstrated the best predictive power for Tunisia. The suggested machine learning framework's robustness was validated by the low rRMSE values and the strong agreement between observed and anticipated emissions. These results show how machine learning approaches can be useful tools for TCO_2 predicting and for promoting climate mitigation and sustainable transportation policies in North African countries.

Acknowledgment

The authors would like to sincerely thank the Laboratory of Metrology and Energy Systems, National Engineering School of Monastir, University of Monastir, 5019 Monastir, Tunisia, for providing the research environment, and financial support.

References

[1] Fan, J., Meng, X., Tian, J., Xing, C., Wang, C., & Wood, J. (2023). A review of transportation carbon emissions research using bibliometric analyses. *Journal of Traffic and Transportation Engineering (English Edition)*, *10*(5), 878-899.

[2] International Energy Agency. CO2 Emissions in 2022; International Energy Agency: Paris, France, 2023.

[3] Agreement, P. (2017). United nations framework convention on climate change (unfccc).(2015). Climate Change Secretariat: Bonn, Germany.

[4] Marrakech Partnership. Climate Action Pathway: Transport, Vision and Summary; United Nations Climate Change: Bonn, Germany, 2020.

[5] Kazancoglu, Y., Ozbiltekin-Pala, M., & Ozkan-Ozen, Y. D. (2021). Prediction and evaluation of greenhouse gas emissions for sustainable road transport within Europe. Sustainable Cities and Society, 70, 102924.

[6] Huang, S., Xiao, X., & Guo, H. (2022). A novel method for carbon emission forecasting based on EKC hypothesis and nonlinear multivariate grey model: Evidence from transportation sector. Environmental Science and Pollution Research, 29(40), 60687-60711

[7] Guo, X., Kou, R., & He, X. (2024). Towards Carbon Neutrality: Machine Learning Analysis of Vehicle Emissions in Canada. Sustainability, 16(23), 10526.

[8] Wang, H., Wang, X., Yin, Y., Deng, X., & Umair, M. (2024). Evaluation of urban transportation carbon footprint− Artificial intelligence based solution. Transportation Research Part D: Transport and Environment, 136, 104406.

[9] Li, X., Ren, A., & Li, Q. (2022). Exploring patterns of transportation-related CO2 emissions using machine learning methods. Sustainability, 14(8), 4588. »

[10] Çınarer, G., Yeşilyurt, M. K., Ağbulut, Ü., Yılbaşı, Z., & Kılıç, K. (2024). Application of various machine learning algorithms in view of predicting the CO2 emissions in the transportation sector. *Science and Technology for Energy Transition*, *79*, 15.

[11] World Bank (2024) World development indicators. https://databank.worldbank.org/. Accessed 05 Dec 2024.

[12] Adam, A. M., Yusuf, A. M., Amir, I. Y., Arab, O. A., & Ali, O. A. (2025). Predicting transportation-related greenhouse gas emissions in Africa's top emitters: a machine learning approach. Environmental Sciences Europe, 37(1), 166.

Copyright © 2026
ISSN: 2961-6611

Proceedings of Book Series PBS -Vol 42, pp.29-36 **Advances in Energy Transition and Environmental Impacts**

[13] Oubnaki, H., Haouraji, C., Mounir, B., Mounir, I., & Farchi, A. (2024). Energy use and CO2 emissions of the Moroccan transport sector. International Journal of Electrical and Computer Engineering (IJECE), 14(1), 86-98.
[14] https://ourworldindata.org/
[15] Kayakuş, M., Terzioğlu, M., Erdoğan, D., Zetter, S. A., Kabas, O., & Moiceanu, G. (2023). European Union 2030 carbon emission target: The case of Turkey. Sustainability, 15(17), 13025.
[16] Pramanik, P., Jana, R. K., & Ghosh, I. (2024). AI readiness enablers in developed and developing economies: Findings from the XGBoost regression and explainable AI framework. Technological Forecasting and Social Change, 205, 123482.
[17] Ağbulut, Ü. (2022). Forecasting of transportation-related energy demand and CO2 emissions in Turkey with different machine learning algorithms. Sustain Prod Consum 29: 141-157.

Copyright © 2026
ISSN: 2961-6611