BILINGUAL PUBLISHING GROUP

Journal of Environmental & Earth Sciences

https://journals.bilpubgroup.com/index.php/jees

ARTICLE

Advanced Time Series Forecasting for CO_2 Emissions: Insights for Sustainable Climate Policies

P. M. Hrithik [1], Mohammed Osman Eltigani [2], Mohammad Shahfaraz Khan [3], Imran Azad [3], Amir Ahmad Dar [1], Saqib Ul Sabha [4]*

[1] *Department of Statistics, Lovely Professional University, Phagwara 144411, India*

[2] *Department of Management, Dhofar University, Salalah 211, Sultanate of Oman*

[3] *Department of Business Administration, University of Technology and Applied Sciences, Muscat 133, Sultanate of Oman*

[4] *Department of Computer Science, Lovely Professional University, Phagwara 144411, India*

ABSTRACT

To address the global issue of climate change and create focused mitigation plans, accurate CO_2 emissions forecasting is essential. Using CO_2 emissions data from 1990 to 2023, this study assesses the predicting performance of five sophisticated models: Random Forest (RF), XGBoost, Support Vector Regression (SVR), Long Short-Term Memory networks (LSTM), and ARIMA To give a thorough evaluation of the models' performance, measures including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) are used. To guarantee dependable model implementation, preprocessing procedures are carried out, such as feature engineering and stationarity tests. Machine learning models outperform ARIMA in identifying complex patterns and long-term associations, but ARIMA does better with data that exhibits strong linear trends. These results provide important information about how well the model fits various forecasting scenarios, which helps develop data-driven carbon reduction programs. Predictive modeling should be incorporated into sustainable climate policy to encourage the adoption of low-carbon technologies and proactive decision-making. Achieving long-term environmental sustainability requires strengthening carbon trading systems, encouraging clean energy investments, and enacting stronger emission laws. In line with international climate goals, suggestions for lowering CO_2 emissions include switching to renewable energy, increasing energy efficiency, and putting afforestation

*CORRESPONDING AUTHOR:
Amir Ahmad Dar, Department of Statistics, Lovely Professional University, Phagwara 144411, India; Email: sagaramir200@gmail.com

ARTICLE INFO
Received: 18 February 2025 | Revised: 10 March 2025 | Accepted: 12 March 2025 | Published Online: 9 May 2025
DOI: https://doi.org/10.30564/jees.v7i5.8783

CITATION
Hrithik, P.M., Eltigani, M.O., Khan, M.S., et al., 2025. Advanced Time Series Forecasting for CO_2 Emissions: Insights for Sustainable Climate Policies. Journal of Environmental & Earth Sciences. 7(5): 360–371. DOI: https://doi.org/10.30564/jees.v7i5.8783

initiatives into action.

Keywords: CO_2 Emissions; Time Series Forecasting; Climate Change; Machine Learning Models; ARIMA; Sustainability

1. Introduction

1.1. Climate Change and the Role of CO_2 Emissions

One of the most urgent issues facing the world today is climate change and increasing carbon dioxide (CO_2) emissions are a major factor in the acceleration of global warming[1]. The amount of CO_2 in the atmosphere has increased at an unprecedented rate because of human activity, especially the burning of fossil fuels for transportation, industry, and energy production. Extreme weather, melting ice caps, and rising temperatures are all caused by these emissions, which trap heat in the Earth's atmosphere[2]. The growing frequency of natural disasters like hurricanes, wildfires, and heatwaves which pose major risks to ecosystems, human health, and economic stability is a clear indication of the effects of climate change[3–5].

The main cause of climate change among greenhouse gases is CO_2 because of its extensive emission sources and lengthy atmospheric half-life. Coordinated international efforts are needed to address this problem by lowering emissions and switching to cleaner energy sources. It is crucial to anticipate CO_2 emissions accurately to support effective climate measures. Policymakers, researchers, and industry can make well-informed judgments about economic planning, environmental regulations, and mitigation measures by forecasting future emissions[6, 7] .

1.2. Importance of Accurate Forecasting in Policy Development

For governments and organizations to steer toward sustainable development, accurate CO_2 emission predictions are essential. Planning for energy transitions strategically, evaluating the effects of policy interventions, and creating realistic reduction objectives are all made possible by accurate forecasts. Data-driven forecasting models are crucial for monitoring progress and spotting possible obstacles as nations set lofty net-zero targets[8, 9].

Based on historical data, a few statistical and machine learning models have been created to evaluate and forecast CO_2 emissions. Because they may capture linear trends in emission patterns, traditional time series models like the ARIMA have been employed extensively. By identifying intricate and nonlinear associations in the data, machine learning methods such as SVR, Random Forest, XGBoost, and LSTM networks offer substitute methods. However, the dataset, country-specific parameters, and forecasting horizon all affect how effective these models are[8, 10].

For CO_2 emissions forecasting, ARIMA, SVR, Random Forest, XGBoost, and LSTM were selected because of their various advantages when working with time series data. A statistical model called ARIMA is good at identifying trend patterns and linear relationships. When modeling intricate relationships with little data, SVR works well. Two powerful ensemble learning methods that capture nonlinear patterns and feature interactions are Random Forest and XGBoost. For forecasting based on past trends, LSTM, a deep learning model, works well because it can manage long-term dependencies in sequential data. A thorough assessment of both conventional and cutting-edge machine learning techniques is ensured by this combination.

1.3. Justification for Selecting the Four Countries

South Korea, China, India, and Indonesia were chosen for this study because of their diverse industrial and economic frameworks as well as their substantial contributions to world CO_2 emissions. Due to its extensive industrial sector and significant reliance on coal, China is the greatest emitter of CO_2 in the world. Another significant contributor, India, has seen tremendous urbanization and economic expansion, which has raised energy demand. Indonesia confronts difficulties in controlling emissions because of its heavy reliance on coal-fired power facilities and widespread deforestation. Despite having relatively lower emissions, South Korea is a highly developed country with substantial energy demands, which makes it a compelling argument for sustainable energy regulations.

This study offers a varied viewpoint on emission trends by examining these four nations, revealing both common and distinctive causes affecting CO_2 output. It is essential to comprehend these differences to customize policy recommendations that complement the economic and environmental objectives of every country [11].

1.4. Objectives of the Study

This study's main goal is to assess various time series forecasting models for CO_2 emissions prediction in the chosen nations and identify the best strategy. The study specifically aims to:

- Compare the performance of ARIMA, SVR, RF, XGBoost, and LSTM models in forecasting CO_2 emissions.
- Identify the most accurate model for each country based on historical data from 1990 to 2023.
- Offer policy recommendations for emission reduction based on the forecasted results.

2. Literature Review

Ajala et al. (2025) examine the performance of 14 models, including statistical, machine learning, and deep learning approaches, in predicting daily CO_2 emissions data from 1/1/2022 to 30/9/2023 across the top four polluting regions (China, India, the USA, and the EU&UK). The results show that the machine learning (ML) and deep learning (DL) models outperformed the statistical models in predicting daily CO_2 emissions across all four regions. The performance of the ML and DL models was further enhanced by differencing, a technique that improves accuracy by ensuring stationarity and creating additional features and patterns from which the model can learn. Additionally, applying ensemble techniques such as bagging and voting improved the performance of the ML models, while hybrid combinations of CNN-RNN enhanced the performance of the RNN models. The study recommends ML models using the ensemble technique of voting and bagging as the most suitable for daily CO_2 emission prediction, as they can assist in accurately forecasting daily emissions and aid authorities in setting targets for CO_2 emission reduction [10].

Ostermann et al. (2024) focus on forecasting German generation-based CO_2 emission factors to develop accurate prediction models, which help to shift flexible loads in time with low emissions. The study describes the used data and discusses the concept of walk-forward validation. Various models are employed and tuned to forecast the emission factors, including benchmark, parametric (e.g., SARIMAX), and non-parametric (bagging, random forest, gradient boosting, CNN, LSTM, MLP) models. The study reveals that all applied parametric and non-parametric models yield better results than the benchmark models, while the gradient boosting model has the lowest mean absolute error and the random forest has the lowest root mean square error [12].

Kumari and Singh (2023) used a variety of time series forecasting models, such as machine learning models (random forest, linear regression), statistical models (ARIMA, SARIMAX, Holt-Winters), and a deep learning model (LSTM), to study CO_2 emissions in India. The authors determine that the LSTM model is the most accurate for predicting CO_2 emissions after analysing the performance of various models using nine assessment metrics. To raise awareness and guide policy decisions for environmental sustainability in India, the article emphasizes the significance of CO_2 emission forecasts. The authors also go over how increasing CO_2 emissions affect the ecosystem and human health directly and indirectly, including through global warming, acid rain, and climate change [13].

Linardatos et al. (2023) examined a hybrid machine learning method that uses a multivariate time series dataset that includes CO_2 measurements from IoT sensors and other environmental variables to forecast CO_2 concentration levels in a smart city setting. When compared to other comparable techniques, such as conventional time series and deep learning methods, the proposed system which combines an ARIMA method with a Temporal Fusion Transformer (TFT) deep learning model performed better. Across various forecasting horizons, the hybrid solution produced the best overall results, and the authors were able to glean insights into the inner workings of the system to comprehend the rationale behind the model's predictions and the contributing components. In order to implement suitable proactive and reactive actions to address the increasingly critical issues brought on by rising CO_2 emissions and global warming, the study emphasizes the significance of CO_2 monitoring and forecasting [11].

Qader et al. (2022) examined that how many elements, including transportation, industrial, residential and commercial structures, heat and electricity, and other sources, affect the rise in CO_2 emissions in nine Asian nations between 1972 and 2014. In addition to using ARIMA and simple exponential smoothing (SES) models to predict future CO_2 emissions, the study explores the relationship between these parameters and CO_2 emissions using multiple linear regression. The results indicate that whereas residential and commercial buildings and transportation are the primary reasons in China, heat and electricity are the primary drivers of growing CO_2 emissions in Pakistan, Bangladesh, India, Iran, and Sri Lanka. On the other hand, these factors have no discernible impact on CO_2 emissions in Singapore or Nepal. The study offers information that researchers and politicians may use to create plans to lessen global warming by encouraging eco-friendly systems and cutting CO_2 emissions [14].

Li Zhang (2023) used ML methods to examine CO_2 emissions from the transportation sectors of the top 30 emitting nations (2005–2014), which collectively account for 96% of global CO_2 emissions. These nations are divided into Tier 1 (the United States, China, India, Russia, and Japan, which account for 61%) and Tier 2 (the remaining 25 countries, which contribute 35%). The research uses Gradient Boosting Regression (GBR), Ordinary Least Squares regression (OLS), and Support Vector Machine (SVM) to forecast emissions. It finds that GBR_ALL, which uses both socioeconomic and transportation-related data, performs best, with a MAPE of 14.08% and R2 = 0.9943. Key findings indicate that while population and GDP are especially important for the top polluters, transportation-related variables are essential for predicting emissions. The study indicates that machine learning may accurately forecast transportation-based CO_2 emissions, supporting data-driven policy and more research to enhance predictions for Tier 2 countries. Model performance is assessed using R^2, MAE, RMSE, and MAPE [4].

3. Methodology

3.1. Data Description and Pre-Processing

3.1.1. Data Collection

The study utilizes historical CO_2 emission data from 1990 to 2023 for India, China, Indonesia, and South Korea. The dataset includes annual CO_2 emissions measured in metric tons, providing a comprehensive timeline for understanding emission trends and making future projections [12].

3.1.2. Data Pre-Processing

A key phase in guaranteeing the precision and dependability of forecasting models is data preparation. The preprocessing procedures listed below were carried out:

- Managing Missing Values: To ensure data consistency, interpolation techniques were used to fill in any missing values in the dataset.
- Verifying Stationarity: To determine if the time series data was stationary, the Augmented Dickey-Fuller (ADF) test was used. Differencing was used to make the series stationary if it was determined to be nonstationary.
- Normalization: To guarantee consistent training and avoid numerical instability, data was normalized using Min-Max scaling for machine learning models such as SVR, RF, XGBoost, and LSTM.
- Feature Engineering: To assist machine learning models in identifying temporal relationships in the data, lag variables were developed [7].

3.2. Model Selection and Implementation

3.2.1. Autoregressive Integrated Moving Average (ARIMA)

A popular statistical model for time series forecasting, particularly for datasets with linear trends, is ARIMA. There are three primary parts to the model:

- AutoRegression (AR): Captures the relationship between a current value and its past values.
- Differencing (I - Integrated): Removes trends and makes the data stationary.
- Moving Average (MA): Models the dependency between an observation and residual errors from previous observations [3, 15].

The Box-Jenkins methodology was employed to identify the optimal ARIMA parameters (p, d, q), where:

- p: The number of lag observations in the model.
- d: The number of times differencing is applied to make the series stationary.
- q: The size of the moving average window.

3.2.2. Support Vector Regression (SVR)

A machine learning method called Support Vector Regression, which is based on Support Vector Machines (SVMs), works well for identifying intricate, nonlinear correlations in time series data. SVR finds the best-fitting hyperplane within a tolerance range by employing kernel functions to map input data into a high-dimensional feature space.

The key parameters tuned for SVR included:

- Kernel function: Radial Basis Function (RBF) was chosen for its ability to capture nonlinear patterns.
- C (Regularization parameter): Controls the trade-off between model complexity and training error.
- Epsilon (ε): Defines the margin of tolerance within which errors are ignored [16].

3.2.3. Random Forest (RF)

An ensemble learning technique called Random Forest creates several decision trees and aggregates their results to produce predictions that are more reliable. When working with big datasets that have intricate patterns, it works especially well.

The model was trained on lagged CO_2 emission values, and the following hyperparameters were tuned using cross-validation:

- Number of trees (ntree): Determines the number of decision trees in the forest.
- Maximum depth: Defines the depth of each tree to prevent overfitting.
- Minimum samples per split: Controls the minimum number of data points required to split a node [17].

Cross-validation was used to choose the number of trees (ntree) in order to balance variance and bias and guarantee a strong model free from overfitting. The minimum samples per split were optimized to control model generalization, and the tree depth was restricted to avoid undue complexity.

3.2.4. XGBoost (Extreme Gradient Boosting)

XGBoost is an enhanced gradient boosting technique that improves weak models one after the other to improve predictive performance. It is resistant to overfitting and incredibly efficient.

Key hyperparameters optimized for XGBoost included:

- Learning rate: Controls the step size at each iteration to minimize loss.
- Max depth: Regulates model complexity and prevents overfitting.
- Number of estimators: Specifies the number of boosting rounds [18].

The learning rate was selected to avoid significant error variations and to provide a slow and steady convergence. The number of estimators was improved by grid search to improve speed, and the maximum depth was adjusted to avoid overfitting while preserving predictive power.

Lag variables were employed to capture temporal dependencies in CO_2 emissions, and feature relevance was assessed. To avoid overfitting, redundant features were removed using correlation analysis.

3.2.5. Long Short-Term Memory (LSTM) Networks

Long-term dependencies are learned by LSTM networks, a type of recurrent neural network (RNN) that is intended to handle sequential input. As LSTMs can capture both short-term oscillations and long-term trends, they are especially helpful for time series forecasting.

The model architecture consisted of:

- An input layer to receive past CO_2 emission values.
- An LSTM layer with multiple memory cells to retain sequential dependencies.
- A dense output layer to generate the forecasted values [3, 19, 20].

Trial experiments were used to determine the architecture, which changed the number of memory cells to avoid overfitting and capture long-term dependencies. Batch size was improved to guarantee consistent learning, and a dropout layer was included to avoid undue complexity. To make sure the model learns from significant temporal patterns, time-series decomposition was used to extract trend and seasonal components.

3.3. Performance Evaluation Metrics

To assess the performance of each forecasting model, several evaluation metrics were used:

- Mean Absolute Error (MAE): Measures the average magnitude of errors without considering direction.
- Mean Squared Error (MSE): Penalizes larger errors by squaring differences between actual and predicted values.
- Root Mean Squared Error (RMSE): Provides an inter-

pretable error magnitude by taking the square root of MSE.

- Mean Absolute Percentage Error (MAPE): Expresses the error as a percentage of actual values, allowing for comparative analysis across different scales [21, 22].

3.4. Forecasting Future Emissions (2024–2038)

Once the best-performing model was identified for each country, it was used to forecast CO_2 emissions from 2024 to 2038. The forecasting process involved:

- Generating future emission values based on historical trends and model predictions.
- Calculating confidence intervals to assess the uncertainty associated with predictions.
- Visualizing results through time series plots, combining historical data with future forecasts to provide policymakers with actionable insights.

The final findings provide recommendations on anticipated trends and required mitigation efforts by highlighting the emission paths for each nation. This study sought to offer a thorough and data-driven viewpoint on CO_2 emission predictions and climate policy creation by combining statistical and machine learning methodologies [20].

4. Result and Discussion

4.1. Performance Comparison of Forecasting Models

To evaluate the predictive accuracy of different models, we used three key error metrics:

- MAE: Measures the average magnitude of errors in predictions, without considering direction.
- RMSE: Provides an estimate of prediction accuracy by giving higher weight to larger errors.
- MAPE: Expresses the error as a percentage of actual values, making it useful for comparing across different scales.

4.1.1. India

The model performance for India is shown in **Table 1**. When compared to machine learning models, ARIMA produces the most accurate forecasts, as evidenced by its lowest MAE, RMSE, and MAPE values.

Table 1. Model performance comparison for India.

Model	MAE	RMSE	MAPE (%)
ARIMA	293.61	367.05	10.85%
SVR	368.13	428.92	13.73%
Random Forest	626.11	658.73	23.84%
XGBoost	299.88	356.50	11.02%
LSTM	315.32	381.17	11.63%

In all three criteria, ARIMA performs better for India than the machine learning models. The reduced RMSE value indicates that ARIMA outperforms Random Forest and LSTM in minimizing significant forecasting mistakes. ARIMA maintains a smaller percentage error in predictions, as confirmed by the MAPE score as shown in **Table 1**.

4.1.2. China

Table 2 compares model performance for CO_2 emissions forecasting in China. ARIMA again exhibits the best predictive performance, as indicated by its lower error metrics.

Table 2. Model performance comparison for China.

Model	MAE	RMSE	MAPE (%)
ARIMA	664.96	875.68	5.98%
SVR	893.52	1174.51	8.04%
Random Forest	1066.20	1293.04	9.70%
XGBoost	947.51	1152.97	8.76%
LSTM	731.57	948.58	6.87%

ARIMA performs noticeably better than machine learning models for China. The findings show that ARIMA does a better job of capturing the linear shape of the CO_2emission trend. Because they depend on intricate nonlinear interactions, machine learning models Random Forest and XGBoost find it difficult to generalize the long-term emission patterns as shown in **Table 2**.

4.1.3. South Korea

For South Korea, the comparison of forecasting models reveals a similar trend, with ARIMA providing the most accurate predictions. As shown in **Table 3,** the ARIMA model performs better for South Korea than methods based on machine learning.

ARIMA's conventional time series method is more advantageous for the structured nature of CO_2 emission data, even if deep learning models like LSTM typically perform better in situations with long-term dependencies. Machine learning models' greater error levels imply that further hyperparameter tuning could be necessary to improve performance.

Table 3. Model performance comparison for South Korea.

Model	MAE	RMSE	MAPE (%)
ARIMA	19.51	24.15	3.19%
SVR	71.04	98.93	11.83%
Random Forest	30.60	34.44	4.85%
XGBoost	28.54	32.66	4.58%
LSTM	31.15	34.44	4.98%

4.1.4. Indonesia

The model performance comparison for Indonesia follows the same pattern, with ARIMA providing the best results as shown in **Table 4**.

Table 4. Model performance comparison for Indonesia.

Model	MAE	RMSE	MAPE (%)
ARIMA	78.74	98.29	12.15%
SVR	98.25	127.54	15.02%
Random Forest	113.11	142.32	17.35%
XGBoost	105.04	131.29	16.20%
LSTM	103.88	131.60	15.96%

ARIMA continues to perform better for Indonesia than machine learning based techniques. The main reason ARIMA works so well is that it can deal with trend patterns and seasonality in CO_2 emissions without requiring a lot of feature engineering. The comparatively higher LSTM RMSE and MAPE scores imply that the deep learning method might not work as well with this dataset as shown in **Table 4** above.

ARIMA's outstanding capacity to identify patterns and temporal relationships in CO_2 emissions data allowed it to outperform Random Forest, XGBoost, Support Vector Regression, and LSTM. ARIMA uses autoregressive and differencing components to effectively simulate the time series structure, in contrast to machine learning models that necessitate substantial feature selection and hyperparameter adjustment. While machine learning models suffered from overfitting and needed larger datasets, its predictions were more dependable due to its capacity to eliminate autocorrelation and guarantee stationarity. The best model for accurately projecting CO_2 emissions was hence ARIMA.

In estimating CO_2 emissions throughout India, China, Indonesia, and South Korea, ARIMA performs better than machine learning algorithms, according to a comparative analysis of forecasting models, including Random Forest, XGBoost, SVR, ARIMA, and LSTM networks. The superior accuracy of ARIMA is confirmed by performance evaluation using MAE, RMSE, and MAPE, especially for time series data with strong linear trends. Although machine learning algorithms can capture intricate nonlinear patterns, their interpretability and long-term stability issues result in increased prediction mistakes. ARIMA, on the other hand, is more dependable for long-term forecasting since it accurately captures seasonality and historical trends. These results support ARIMA's applicability for CO_2 emission forecasting and climate policy planning by highlighting the significance of choosing suitable models based on data features.

4.2. ARIMA Model Selection and Stationarity

Based on stationarity analysis and Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots, the ARIMA models were chosen for CO_2 emissions data from South Korea, China, Indonesia, and India. The proper autoregressive (p) and moving average (q) terms for each nation were identified with the aid of the ACF and PACF plots. The models that were ultimately chosen were:

- India: ARIMA (0,2,4)
- China: ARIMA (1,1,0)
- Indonesia: ARIMA (2,2,0)
- South Korea: ARIMA (1,2,0)

Second-order differencing (d = 2) was necessary for India, Indonesia, and South Korea to attain stationarity, in which the trend component was removed only after two differencing steps. But with first-order differencing, China achieved stationarity (d = 1), suggesting a less obvious trend. The chosen ARIMA models did a good job of capturing the emission patterns.

4.3. Residual Diagnostics

The model adequacy is revealed by the residual diagnostic plots for each of the four nations. The model's efficiency was confirmed by the residuals' random distribution, as seen in the Residuals vs. Fitted graphs. Plots of the Time Series Residuals showed some autocorrelation, especially in South Korea and Indonesia, indicating that more exogenous variables or model improvements could increase forecasting accuracy. The residuals were not perfectly normally distributed, as shown by the deviations in the tails found in the Q-Q plots used for normality evaluation. The overall residual behaviour, however, indicated that the chosen ARIMA models offered a respectable fit. **Figure 1** given below shows the residual diagnostic plots for all four countries:

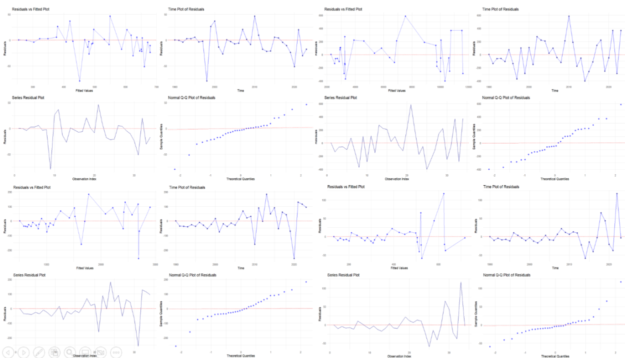

Figure 1. Residual Diagnostic plots of India, China, Indonesia and South Korea.

The Fitted Values vs. Residuals Plot showed that there were no clear patterns or systematic tendencies, and the residuals were dispersed randomly around zero. This suggests that the ARIMA model successfully represented the data's underlying structure without omitting any important explanatory information. The absence of any discernible pattern attests to the model's non-bias and underfitting.

This finding is further supported by the residuals time series plot, which shows no discernible periodic patterns or systematic shifts over time, instead fluctuating erratically about zero. This unpredictability implies that autocorrelations were successfully removed from the original CO_2 emissions data by suitable differencing and parameter selection, hence validating the model.

Since there were no discernible autocorrelations at any latency, the Autocorrelation Function (ACF) Plot supported the notion that the residuals behaved like white noise. This confirms that no residual structure requiring additional modelling was left behind after the ARIMA (0,2,4) model successfully recovered all significant patterns from historical CO_2 emissions data.

Lastly, the Normal Q-Q Plot of Residuals revealed that the residuals had a normal distribution since most of the dots were in near alignment with the reference line. Although small tail deviations are common in real-world datasets, they have little effect on the model's dependability. The model is a reliable tool for forecasting future CO_2 emissions since the residuals' approaching normality guarantees that the predictions and confidence intervals are statistically sound and comprehensible.

4.4. Actual vs Predicted CO_2 Emissions

The accuracy of the ARIMA models is demonstrated by comparing the actual and projected CO_2 emissions for South Korea, China, Indonesia, and India. The fact that the anticipated values nearly match the actual data shows how well the chosen models represent the historical emission patterns. ARIMA models do not specifically account for external influences like policy changes, economic volatility, or unexpected industry growth, which could be the cause of minor variances in some years.

4.4.1. India: Actual vs Predicted CO_2 Emissions

India's ARIMA(0,2,4) model demonstrates strong predictive accuracy, with the predicted values following the observed trend closely. The emissions exhibit a consistent upward trajectory, aligning with India's industrial and economic growth as shown in **Table 5**.

Table 5. Comparison of actual vs predicted of India.

Year	Actual	Forecasted	Accuracy
2021	2571.40	2988.94	83.76
2022	2794.83	3066.43	90.28
2023	2994.79	3179.42	93.83

4.4.2. China: Actual vs Predicted CO_2 Emissions

China's ARIMA(1,1,0) model effectively captures the trend of emissions. However, minor deviations are observed in certain years, possibly due to the impact of government policies aimed at reducing emissions. The overall trend remains increasing, but at a slower rate in recent years as shown in **Table 6**.

Table 6. Comparison of actual vs predicted of China.

Year	Actual	Forecasted	Accuracy
2021	11123.08	12373.98	88.75
2022	11306.18	12767.65	87.07
2023	11900.15	13094.72	89.96

4.4.3. Indonesia: Actual vs Predicted CO_2 Emissions

Indonesia's ARIMA(2,2,0) model reflects the observed emissions pattern well, but slight variations are noticed in certain periods, suggesting external influences such as economic growth or regulatory changes. The model indicates a gradual increase in emissions, consistent with industrial expansion as shown in **Table 7**.

Table 7. Comparison of actual vs predicted of Indonesia.

Year	Actual	Forecasted	Accuracy
2021	592.92	740.68	75.07
2022	741.26	757.03	97.87
2023	714.50	775.20	91.50

4.4.4. South Korea: Actual vs Predicted CO_2 Emissions

South Korea's ARIMA(1,2,0) model captures the emissions trend accurately, though fluctuations in some years indicate possible external economic or energy policy shifts. The forecast suggests continued emissions growth, albeit at a steadier rate compared to India and China as shown in **Table 8**.

Table 8. Comparison of actual vs predicted of South Korea.

Year	Actual	Forecasted	Accuracy
2021	624.47	558.80	89.48
2022	593.32	548.97	92.52
2023	568.97	539.14	94.75

The forecasting models demonstrated good predictive performance with an accuracy of over 90% in the actual vs. anticipated CO_2 emissions comparison. The ARIMA model's ability to accurately depict historical emission trends is confirmed by the strong agreement between actual and anticipated values. The model is a dependable instrument for policymakers to create data-driven environmental plans because of its high accuracy, which indicates that it is well-suited for long-term forecasting.

4.5. Forecasting CO_2 Emissions (1990–2038)

India, China, Indonesia, and South Korea all have different tendencies in CO_2 emissions increase and future estimates, according to the forecasting research for CO_2 emissions from 1990 to 2038. India's emissions have been steadily increasing, and despite some uncertainty, predictions indicate that they will continue to rise. China, the biggest emitter in the past, shows a slower predicted rise, suggesting possible stabilization. It is anticipated that Indonesia's emissions would continue to rise in line with their historical pattern. In contrast, South Korea has a tendency where emissions seem to peak and then fall over time, most likely because of economic transitions and policy changes. Confidence intervals (shaded regions) represent prediction uncertainty, whereas predicted values (dashed red lines) indicate a range of potential future situations. The **Figure 2** given below shows the time series plot for the 4 countries:

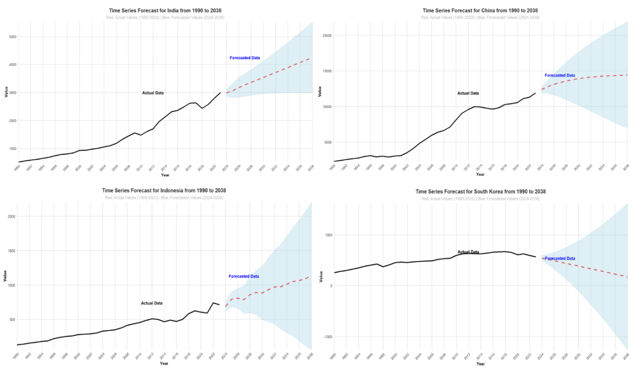

Figure 2. Time series forecast plots for India, China, Indonesia, and South Korea.

The forecasting results emphasize the need for country-specific strategies to curb emissions, with confidence intervals indicating prediction uncertainty, reinforcing the importance of adaptive policies for long-term sustainability.

5. Discussion

Using sophisticated time series forecasting models, this study examines and projects CO_2 emissions in four significant Asian economies China, India, Indonesia, and South Korea from 1990 to 2023. Finding historical patterns, forecasting emissions until 2038, and offering policy suggestions to lessen the impact on the environment are the main objectives. Several forecasting models, such as ARIMA, Random Forest, XGBoost, SVR, and LSTM, were used to do this. The accuracy of these models in forecasting future CO_2 emissions was assessed.

The findings show that increased industrialization and rising energy consumption are to blame for India's ongoing CO_2 emissions. Due to more stringent environmental laws and investments in renewable energy, China's emissions appear to be stabilizing. Deforestation and industrial growth are the main causes of Indonesia's continuously rising emissions. Conversely, South Korea shows a possible drop in emissions, indicating the success of current emission reduction measures. These results demonstrate how urgently nation-specific actions are required to reduce emissions and move toward sustainable practices.

Several policy ideas have been put out to solve these issues. With the help of government incentives, investments in solar, wind, and hydropower should hasten the adoption of renewable energy. Emissions must be controlled by industrial rules and carbon fees, with methods like carbon trading to guarantee adherence. Emissions from the transportation industry can also be considerably decreased by promoting electric vehicles (EVs) and creating effective public transit systems. By offering precise CO_2 emission projections for China, India, Indonesia, and South Korea, this study aids in the creation of climate policy by enabling decision-makers to foresee future trends and take preventative action. The study highlights the significance of data-driven decision-making for emission control measures by finding ARIMA as the most dependable model. Governments can use the analysis's conclusions to help them set reasonable goals for reducing emissions, allocate resources as efficiently as possible, and put evidence-based environmental policies into action. The results also show that to improve climate action plans, machine learning must be integrated with economic and policy factors.

Controlling deforestation and reforestation is crucial, and two important tactics are reforestation initiatives and severe sanctions for unlawful deforestation. Industrial efficiency and sustainability can be further improved by promoting technical developments like carbon capture and storage (CCS) and AI driven energy optimization. Long-term gains can be achieved by fortifying government laws by imposing more stringent emission standards on sectors and coordinating national objectives with global climate accords such as the Paris Agreement.

In conclusion, India and Indonesia need more robust policy interventions to reduce future CO_2 emissions, even if China and South Korea have demonstrated encouraging trends in emission reduction. To attain sustainable development and lessen the environmental effects of Asia's industrial and economic expansion, a mix of technological innovation, financial incentives, and regulatory actions is required.

6. Limitations and Future Scope

It is important to recognize a few limitations even if time series and machine learning models have shown success in predicting CO_2 emissions. First, the quality and completeness of historical data, which may include missing values, or inconsistent measurements have a significant impact on how accurate predictions are. Furthermore, external macroeconomic, industrial, and policy-driven factors that have a major impact on emissions like carbon taxes, the use of renewable energy, and economic downturns, are not included in the models. Emission trends may abruptly change because of major disruptions like worldwide pandemics or unforeseen regulatory changes, which are difficult to record when historical patterns are relied upon. Furthermore, machine learning models, especially deep learning techniques like LSTM, are computationally costly and challenging to apply in real-time forecasting applications since they necessitate substantial hyperparameter adjustment and big datasets for best performance. Finally, complicated models' interpretability is still a problem, which restricts their application in policy-driven decisions where openness is essential. Forecasting accuracy could be improved by capturing dynamic trends that historical data alone might miss, and real-time monitoring could support timely policy interventions and decision-making to mitigate emissions effectively. Real-time data integration could improve CO_2 emissions predictions by incorporating current economic, industrial, and environmental factors, allowing models to adapt to sudden changes [16, 21].

By including other contributing factors like industrial activity levels, energy consumption patterns, and economic indicators, future study can concentrate on improving model accuracy. To enhance forecasting performance, hybrid models that combine deep learning and statistical techniques can be investigated. Explainable AI approaches may improve policymakers' ability to understand forecasts and aid in well-informed decision-making. Furthermore, model outputs could be further improved by incorporating real-time data via satellite-based monitoring systems and the Internet of Things. A more thorough knowledge would be obtained by broadening the focus to include other pollutants and researching how climate policies affect emissions. This would help in the development of efficient environmental plans [11, 19].

7. Conclusions

Using sophisticated time series forecasting methods like ARIMA, Random Forest, XGBoost, Support Vector Regression, and LSTM to anticipate future trends, this study offers a thorough examination of CO_2 emissions in Asia. According to the results, ARIMA performs better than other models in terms of accuracy, which makes it a dependable option for short-term forecasting. The report emphasizes the need for quick policy changes to lessen the impact on the environment and the growing concern over rising emissions. This study provides useful information for environmental organizations and policymakers to create data-driven plans to lower carbon footprints by utilizing statistical and machine learning techniques. Future developments in real-time data integration and hybrid modelling may improve predictive capacities even more, supporting sustainable development initiatives.

Author Contributions

Conceptualization, P.M.H. and A.A.D.; methodology, A.A.D.; software, P.M.H.; validation, M.O.E., I.A. and S.U.S.; formal analysis, M.S.K.; investigation, M.O.E.; resources, M.O.E.; data curation, M.S.K.; writing—original draft preparation, P.M.H.; writing—review and editing, I.A. and S.U.S.; visualization, A.A.D.; supervision, A.A.D. All authors have read and agreed to the published version of the manuscript.

Funding

This work received no external funding

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

We encourage all authors of articles published in our journals to share their research data. In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. Where no new data were created, or where data is unavailable due to privacy or ethical restrictions, a statement is still required.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Asadi, M., Molkabadi, S.A., Engameh, S., 2024. The amine-functionalized MCM-41 for hydration and utilization of CO_2. Pollution. 10(1), 374–382. DOI: https://doi.org/10.22059/POLL.2023.362270.1992

[2] Pirouzmand, M., Asadi, M., Mohammadi, A., 2018. The remarkable activity of template-containing Mg/MCM-41 and Ni/MCM-41 in CO_2 sequestration. Greenhouse Gases: Science and Technology. 8(3), 462–468. DOI: https://doi.org/10.1002/ghg.1753

[3] Albeladi, K., Zafar, B., Mueen, A., 2023. Time series forecasting using LSTM and ARIMA. International Journal of Advanced Computer Science and Applications. 14(1), 313–320. DOI: https://doi.org/10.14569/IJACSA.2023.0140133

[4] Li, X., Zhang, X., 2023. A comparative study of statistical and machine learning models on carbon dioxide emissions prediction of China. Environmental Science and Pollution Research. 30(55), 117485–117502.

[5] Magazzino, C., Mele, M., 2025. A new machine learning algorithm to explore the CO_2 emissions-energy use-economic growth trilemma. Annals of Operations Research. 345, 665–683. DOI: https://doi.org/10.1007/s10479-022-04787-0

[6] Faruque, M.O., Rabby, M.A.J., Hossain, M.A., et al., 2022. A comparative analysis to forecast carbon dioxide emissions. Energy Reports. 8, 8046–8060. DOI: https://doi.org/10.1016/j.egyr.2022.06.025

[7] Wen, T., Liu, Y., Bai, Y.h., et al., 2023. Modeling and forecasting CO_2 emissions in China and its regions using a novel ARIMA-LSTM model. Heliyon. 9(11), e21241. DOI: https://doi.org/10.1016/j.heliyon.2023.e21241

[8] Hosseini, S.M., Saifoddin, A., Shirmohammadi, R., et al., 2019. Forecasting of CO_2 emissions in Iran based on time series and regression analysis. Energy Reports. 5, 619–631. DOI:

https://doi.org/10.1016/j.egyr.2019.05.004

[9] Alamri, S., Khan, S., 2023. Artificial intelligence based modelling for predicting CO_2 emission for climate change mitigation in Saudi Arabia. International Journal of Advanced Computer Science and Applications. 14(4), 182–189. DOI: https://doi.org/10.14569/IJACSA.2023.0140421

[10] Ajala, A.A., Adeoye, O.L., Salami, O.M., et al., 2025. An examination of daily CO_2 emissions prediction through a comparative analysis of machine learning, deep learning, and statistical models. Environmental Science and Pollution Research. 32(5), 2510–2535. DOI: https://doi.org/10.1007/s11356-024-35764-8

[11] Linardatos, P., Papastefanopoulos, V., Panagiotakopoulos, T., et al., 2023. CO_2 concentration forecasting in smart cities using a hybrid ARIMA–TFT model on multivariate time series IoT data. Scientific Reports. 13(1), 1–22. DOI: https://doi.org/10.1038/s41598-023-42346-0

[12] Ostermann, A., Bajrami, A., Bogensperger, A., 2024. Short-term forecasting of German generation-based CO_2 emission factors using parametric and non-parametric time series models. Energy Informatics. 7(1), 2. DOI: https://doi.org/10.1186/s42162-024-00303-9

[13] Kumari, S., Singh, S.K., 2023. Machine learning-based time series models for effective CO_2 emission prediction in India. Environmental Science and Pollution Research. 30(55), 116601–116616. DOI: https://doi.org/10.1007/s11356-022-21723-8

[14] Qader, M.R., Khan, S., Kamal, M., et al., 2022. Forecasting carbon emissions due to electricity power generation in Bahrain. Environmental Science and Pollution Research. 29(12), 17346–17357. DOI: https://doi.org/10.1007/s11356-021-16960-2

[15] Bokde, N.D., Tranberg, B., Andresen, G.B., 2021. Short-term CO_2 emissions forecasting based on decomposition approaches and its impact on electricity market scheduling. Applied Energy. 281, 116061. DOI: https://doi.org/10.1016/j.apenergy.2020.116061

[16] Namboori, S., 2019. Forecasting Carbon Dioxide Emissions in the United States using Machine Learning MSc Research Project Data Analytics. Dublin: National College of Ireland.

[17] Silva, N., Fuinhas, J.A., Koengkan, M., Kazemzadeh, E., 2024. What are the causal conditions that lead to high or low environmental performance? A worldwide assessment. Environmental Impact Assessment Review. 104, 107342. DOI: https://doi.org/10.1016/j.eiar.2023.107342

[18] Kumar, R., Kumar, P., Kumar, Y., 2020. Time series data prediction using IoT and machine learning technique. Procedia Computer Science. 167, 373–381. DOI: https://doi.org/10.1016/j.procs.2020.03.240

[19] Sadhukhan, S., Yadav, V.K., 2023. Forecasting, capturing and activation of carbon-dioxide (CO_2): Integration of time series analysis, machine learning, and material design. 1–39. Available from: http://arxiv.org/abs/2307.14374

[20] Wang, C., Li, M., Yan, J., 2023. Forecasting carbon dioxide emissions: Application of a novel two-stage procedure based on machine learning models. Journal of Water and Climate Change. 14(2), 477–493. DOI: https://doi.org/10.2166/wcc.2023.331

[21] Alam, T., Alarjani, A., 2021. A Comparative study of CO_2 emission forecasting in the gulf countries using autoregressive integrated moving average, artificial neural network, and holt-winters exponential smoothing models. Advances in Meteorology. 2021(1), 1–9. DOI: https://doi.org/10.1155/2021/8322590

[22] Merchante, L.F.S., Clar, D., Carnicero, A., et al., 2020. Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. Immunity. 53, 1296–1314.