Journal of Sustainability Perspectives

journal homepage: https://ejournal2.undip.ac.id/index.php/jsp/

# AI-Driven Data Analysis for Sustainable Development

*R.E. Azizov, N.T. Ismayilova**

*Azerbaijan State Oil and Industry University, Azerbaijan*

*Corresponding author: nigar.ismailova@asoiu.edu.az

## Article Info

**Received:**
14 April 2025
**Accepted:**
15 September 2025
**Published:**
30 December 2025

**DOI:**
10.4170/jsp.2025.29834

**Abstract.** Sustainable development is a global challenge which requires an innovative approach merging environmental science, economics, policy-making, and artificial intelligence. The data-driven approach using intelligent methodologies is valuable for evaluating and mitigating environmental impacts. This study exploits data from different sources and machine learning methods to analyze key sustainability indicators, focusing on $CO_2$ emissions, ecological footprint, and load capacity factor. The analysis emphasizes advanced feature selection techniques and predictive modelling to identify the most significant economic, industrial, agricultural, and environmental factors that affect sustainability. Comparative analysis shows differences between the importance of indicators established through expert-driven decisions across various scientific fields and AI-driven assessments. The research attempts to solve the problem following a multi-step process: (1) clustering of countries based on environmental indicators to identify patterns and classify according to similar performance; (2) evaluation of the socio-economic and environmental factors' impact on $CO_2$ emissions using machine learning; (3) predicting future trends in emissions and sustainability metrics through high-level artificial intelligence techniques such as Hidden Markov models. This study will potentially serve policymakers, enabling data-driven decision-making to promote sustainable development efforts. The results demonstrate the value of interdisciplinary approaches to deal with sustainability challenges and to stimulate a balanced path toward economic growth and environmental protection.

## Keywords:

Clustering analysis, data-driven decision making, ecological footprint, environmental indicators, feature selection, predictive modelling, sustainable development

## 1. Introduction

Artificial Intelligence (AI) and the possibility of processing data from various sources and nature using Machine Learning (ML) techniques has opened new perspectives in the

Sustainable Development (SD). In recent years, environmental degradation, climate change, biodiversity loss, socio-economic disparities, and unbalanced economic growth has become an issue of great importance. Previous research based on traditional economic models typically focused on growth metrics like GDP, demonstrated that there is the need for comprehensive strategies and innovative technologies for understanding the intricate web of environmental and social factors that underpin true sustainability [1,2].

The complexities involved from the interconnectedness of natural ecosystems, economic systems, and human societies are known challenges in evaluation and implementation of effective strategies for SD. On the other hand, availability of extensive and heterogeneous data sources expounds additional problems for policymakers and researchers seeking to understand sustainability dynamics. From this point of view, implementation of data-driven approaches powered by advancements of AI and ML allows us to analyse the complex relationship between environmental and social factors for understanding sustainable dynamics [3]. The integration of satellite data processing plays a crucial role in sustainable development by providing real-time, accurate and large-scale environmental monitoring, which supports informed decision making and policy implementation for reducing carbon emissions and preserving natural resources [4, 5].

Uncovering hidden patterns, disclosure of key drivers affecting environmental and socio-economic change, as well as intelligent support for decision making processes can be achieved primarily through integration and analysis of large datasets using AI and ML methodologies [6]. The main objective of this work is to provide both descriptive and predictive insights into sustainable dynamics via several ML techniques like clustering, ensemble learning, time series analysis, etc.

The experimental work demonstrated here is divided into three parts: clustering of countries based on a set of environmental and socio-economic indicators using unsupervised ML techniques, evaluating the impact of various indicators on $CO_2$ emissions through application of supervised machine learning models and predictive modelling to forecast future trends in emissions and sustainability metrics via time series analysis.

Evaluation of sustainability performance of countries requires a multidimensional analysis using a wide range of indicators covering climate change, health, poverty, economic resilience, agriculture, industry, technology, transport, social equity and resource management. Previous research in this field typically only investigated the fixed indicators for countries' evaluation or analysis of variables related to special regions and countries [7-16]. How nations are balancing environmental management with economic and social situations is the main question for understanding the current global sustainability development circumstances and for shaping future strategies. In order to address mentioned question grouping of countries using k-means clustering and birch clustering based on indicators representing climate change such as $CO_2$ emissions, arable land, cereal yield, energy use, forest area, economics - GDP, area, population, fossil fuel energy resources, renewable energy resources, sociology - birth rate, death rate, urban population, rural population etc. was carried out. This approach will serve as a more efficient alternative to traditional classification of countries based on isolated economic indicators, sector specific attributes or regional approaches [17-19].

The second phase of the work involved the impact evaluation of the socio-economic and environmental indicators for understanding the drivers of climate change and designing effective mitigation strategies. Feature importance analysis using Lasso regression, random forest and Extreme Gradient Boosting algorithm (XGBoost) allows to uncover hidden patterns,

detect non-obvious correlations in contrast with traditional statistical methods assuming linear relationships by passing transmission of the complex, non-linear and interactive effects that define emission patterns across countries [20].

Monitoring and forecasting of $CO_2$ emissions represents an important topic to study for supporting international commitments such as Paris agreement, the United Nations Sustainable Development Goals and the recent COP29 declarations, which emphasized the urgent need for science-driven strategies to limit global temperature rise and accelerate the transition to low-carbon economies [21, 22]. A potential solution on capturing the complex, non-linear and multivariate nature of $CO_2$ emissions dynamics involved time series prediction using deep learning models such as LSTM (long short-term memory), GRU (gated recurrent units) or transformers which outperforms traditional statistical methods [23]. In contrast, HMM provides a probabilistic framework well suited to model dynamic, state-switching behaviour observed in emissions as result of policy changes, economic rises or technological interventions. The third part of the work focused on application of the mentioned methodologies for analysing $CO_2$ change dynamics.

This study performs a comprehensive analysis of several indicators using global data from the World Bank, encompassing a wide range of countries. The objective is to uncover patterns and relationships relevant to environmental performance and $CO_2$ emissions through the application of machine learning techniques. Although the analysis maintains a global perspective, Azerbaijan, where authors of the paper are based, is used as a case study to contextualize and exemplify the findings. This country-specific focus does not imply a regional limitation of the study but rather serves to provide a more detailed interpretation of results in a familiar national context.

## 2. Methodology

This section outlines data collection, data pre-processing, application and evaluation of the supervised and unsupervised machine learning algorithms for decision making used within this research. For realization of research World Bank Data are chosen according to the high reliability and accuracy provided by national statistical offices [24]. An overview of the main methods used for achieving the goals of the current work described here with their advantages and disadvantages according to the results of the experiments. This part of the paper comprises three sections. Firstly, an overview of unsupervised machine learning methods used for grouping of the countries based on several numerical parameters about them are given. The second part of the section describes methods used for extraction of the key features affecting the target parameter, followed by details of the time series analysis for prediction of the $CO_2$ emissions according to the historical data.

### 2.1. Data Collection and Pre-processing

Data used in this study have been obtained from the World Bank by downloading and combining data sheets representing values of indicators during years 1960-2022. After excluding the parameters and countries with many missing values (when more than 30 % of values are missing) dataset consisted of 343 parameters about 266 countries, some of parameters are listed below: total $CO_2$ emissions, $CO_2$ emissions from transport, GDP, area, population, birth rate, death rate, forest area, fossil fuel energy consumption, rural population, urban population, total annual freshwater withdrawals, agriculture, forestry and fishing, Gini index, access to electricity, etc. Dataset was scaled using the standard scaling

algorithm to ensure that all features in a dataset will contribute to the supervised and unsupervised models. The importance of dimensionality reduction is easily seen from the shape of the dataset; existence of the parameters with the high positive and negative correlation shows the necessity to reduce dimension by selecting the most informative features. For this reason, PCA (principal component analysis) was applied to the dataset for reducing overfitting and sensitivity of irrelevant features, at the same time to increase the speed of the data processing. After transformation of the original dataset by using PCA, 50 uncorrelated parameters are found for use in the next step - clustering of the countries.

### 2.2. Clustering Algorithms

Clustering methods attempt to partition a dataset into clusters, where the objects in the same cluster are more like each other than data points in the other clusters. These algorithms represent typical solutions for data analysis problems such as customer and image segmentation, anomaly detection, market basket analysis, pollution monitoring, etc. Several popular clustering methods are available for addressing the mentioned problems: partition-based, hierarchical, density-based, model-based, grid-based, constraint-based, graph-based and fuzzy clustering methods. Grouping of the countries' dataset consisting of values for climate change, social, economic and other development indicators has been implemented using k-means and BIRCH clustering methods correspondingly classified as partition-based and hierarchical.

K-means clustering algorithm aims to partition dataset $X = \{x_1, x_2, \ldots, x_N\}, x_i \in R^d$, where $i = 1, 2, \ldots, N, N -$ number of objects in dataset and $d -$ is dimension i.e., number of features in the dataset, into k disjoint subsets (clusters) by minimizing the variance of objects in each cluster [25]. Algorithm starts with the choosing the number of clusters k which is the main hyperparameter affecting the performance of the method. The elbow method for choosing the optimal number of clusters was used in this work. Performance evaluation of the k-means clustering algorithm is performed using silhouette score [26]. Silhouette score measures how accurate each data point was assigned to its cluster compared with other clusters. After determination of the number of clusters k algorithm initializes the centroids of clusters randomly and assigns each data point to the nearest cluster. Recalculation of the centroids as the mean of the assigned data points and assignment of the points to new clusters is repeating until convergence.

The next clustering algorithm applied in this work for clustering of the countries is BIRCH (balanced iterative reducing and clustering using hierarchies) - hierarchical clustering algorithm which builds the compact cumulative structure called clustering features tree. The main advantage of the BIRCH method compared to the k-means is that BIRCH is more robust to outliers due to summarization in clustering features tree. In addition, Mean Shift clustering algorithm, which is a centroid-based and non-parametric clustering algorithm was used for the grouping of the countries according to the collected dataset.

### 2.3. Methods Used for Features Selection

The ability to automatically detect features that strongly influence predictions, especially in the high-dimensional datasets where relationships are not obvious, is one of the advantages of ML methods. Linear models and tree-based models are currently the most used approaches for evaluation of the feature importance. The Lasso (least absolute shrinkage and selection operator) and random forest methods are chosen to extract the important

indicators affecting the environmental indicator, such as total $CO_2$ emissions. The Lasso method assigns zero to unimportant features in the dataset and keeps only the most informative features. This method evaluates the features by adding the L1 regularization term to the cost function of the model, which is usually linear.

Another model used in this work for extraction of most important features affecting climate change parameters is the random forest method. It is an ensemble learning algorithm, which builds multiple models for classification and regression tasks, where outputs are combined using majority voting or mean value for improvement of accuracy, overfitting elimination and effectively handling high-dimensional data. The random forest method evaluates the features based on their contribution to the improvement of the model's accuracy. This method calculates feature importance scores based on criteria like impurity reduction or permutation impact. Consequently, parameters strongly affecting target variables (in the current dataset total $CO_2$ emissions or $CO_2$ emissions from transport) based on ML algorithms are determined. XGBoost algorithm is another ensemble learning algorithm, which is preferred for its speed and performance. The main advantage of this method is that it builds trees sequentially by correction of errors in each iteration.

### 2.4. Time Series Analysis for Prediction Climate Change Indicators

However classical statistical methods have become a standard benchmark for analysing historical data and predicting future patterns, deep networks such as LSTM are more useful for modelling highly non-linear relationships [27]. The ability to maintain memory of the input of LSTM which is a powerful type of neural network makes it suitable for solving problems involving sequential data like a time series. LSTM architecture consists of three gates: the input gate, the forget gate and the output gate, which control the memory cell. Determination of data which must be added, removed or sent as output are also realized by these gates. In this work, LSTM is used for learning patterns in collected data to predict target indicators.

HMM is one of the basic examples of variable-based machine learning models which is very suitable for processing discrete time series using transition and emission probabilities [28, 29]. The implementation of HMM for time series analysis to predict $CO_2$ emissions is sufficiently valuable in terms of discovering underlying emission states, prediction of state shifts and forecasting of the target assuming current policy and situation. Comparison of performance metrics' values as result of mentioned models for the testing step of the collected data demonstrates the potential superiority of HMM model with the ability of identification of hidden patterns and detection of unusual situations.

## 3. Results and Discussions

Figures and Tables are placed in groups of text and annotated. The figure is followed by This section summarizes and discusses the main findings of data processing for climate change. A total of 344 features were used for clustering of 266 countries and regions (table 1). After pre-processing (filling the missing values and PCA) 266 objects with 50 uncorrelated features were clustered using k-means clustering method, BIRCH method and MeanShift clustering algorithms. For evaluation of the applied clustering methods and optimization of the parameter in the k-means algorithms silhouette score was used. After implementation of the k–means clustering algorithms with the k values in range from 3 to 20 and calculation corresponding values for the silhouette score k = 13 was defined as the optimal number of the countries' clusters with the highest silhouette score. From the received 13 clusters 3

clusters are out of interest, because they contain less than 3 countries. One of the interesting clusters is cluster 8 containing Azerbaijan, these are other countries from the same cluster: Algeria, Egypt, Iran, Iraq, Jordan, Kyrgyzstan, Libya, Burma (Myanmar), Mongolia, Pakistan, Philippines, Syria, Tajikistan, Turkmenistan, Tunisia, Uzbekistan, Yemen. Analysis of the ranking list of the countries based on their carbon footprint, i.e., $CO_2$ emissions for the 2022-year (this is the newest announced data) countries from this cluster with the lowest carbon footprint are Kyrgyzstan, Tajikistan and Yemen [30].

Table 1. Description of Selected Indicators

| No. | Class of indicators | Name of indicators |
|---|---|---|
| 1 | Agriculture & Rural Development | Surface area (sq. km) |
| 2 | | Forest area (% of land area) |
| 3 | | Arable land (% of land area) |
| 4 | | Rural population |
| 5 | | Agriculture, forestry, and fishing, value added (current US$) |
| 6 | | Annual freshwater withdrawals, agriculture (% of total freshwater withdrawal) |
| 7 | | Cereal production (metric tons) |
| 8 | Climate Change | Urban population (% of total population) |
| 9 | | Annual freshwater withdrawals, total (billion cubic meters) |
| 10 | | Access to electricity (% of population) |
| 11 | | Energy use (kg of oil equivalent per capita) |
| 12 | | Mortality rate, under-5 (per 1,000 live births) |
| 13 | | Population in urban agglomerations of more than 1 million (% of total population) |
| 14 | | Renewable electricity output (% of total electricity output) |
| 15 | Economy & Growth | GDP per capita (current US$) |
| 16 | Education | Labor force, total |
| 17 | | Unemployment, total (% of total labor force) (modelled ILO estimate) |
| 18 | | Unemployment, female (% of female labor force) (modelled ILO estimate) |
| 19 | Energy & Mining | Fossil fuel energy consumption (% of total) |
| 20 | | Renewable energy consumption (% of total final energy consumption) |

| No. | Class of indicators | Name of indicators |
|---|---|---|
| 21 | Environment | Carbon dioxide ($CO_2$) emissions (total) excluding LULUCF (Mt $CO_2e$) |
| 22 | | Carbon dioxide (CO2) emissions from Transport (Energy) (Mt $CO_2e$) |
| 23 | | Plant species (higher), threatened |
| 24 | Health | Birth rate, crude (per 1,000 people) |
| 25 | | Death rate, crude (per 1,000 people) |
| 26 | Poverty | Gini index |
| 27 | Infrastructure | Rail lines (total route-km) |

As result of application BIRCH clustering algorithm cluster containing Azerbaijan is more extended than cluster received after application of the k-means clustering algorithm and contains countries as Albania, Latvia which are in the bottom of the $CO_2$ emissions by country list. Given that countries grouped within the same cluster have similar environmental, economic, and social parameters information about the countries with the lowest carbon footprint indicator offers a significant potential for policy transfer and strategy adaptation. Countries with big carbon footprint which can be measured by the carbon emissions can study and emulate the policy frameworks, technological implementations, social management, and sustainability solutions of their counterparts with the low carbon emissions. Such intra-cluster knowledge transfer is particularly valuable and guarantees the successful implementation. Accuracy of the clustering was measured using the silhouette score, which is equal to for k-means clustering 0.1926 and 0.2342 for the BIRCH clustering. The comparative analysis of methods demonstrates that BIRCH may be more suitable for uncovering subtle patterns in the dataset.

Additionally, experimental results indicate that Mean Shift clustering is not effective for this dataset, as is assigned most countries to a single dominant cluster while producing some additional clusters containing only one, two or three countries each. This imbalance can be explained by the algorithm's sensitivity to bandwidth selection and the presence of dominant high-density regions in the feature space.

The extraction of the indicators strongly affecting the changing of the $CO_2$ emissions was performed using Lasso regression and the XGBoost algorithm. Lasso regression uses input parameters to predict the target variable by applying penalty to avoid overfitting, it is the form of regularization for linear regression models. In this approach feature importance is defined by assigning coefficients of the less relevant features to exactly zero. Existence of many parameters representing environmental, climate change, economic, and other characteristics leads to the application of not informative or not relevant features for the prediction. These features do not make significant contribution to the predictive performance, which makes necessary to reduce dimensionality in the dataset. For this reason, parameters representing indicators for the years in the dataset used for the clustering were deleted and only columns defining indicator values for the 2022 and 2023 years were used for the prediction $CO_2$ emissions. $R^2$ score – statistical measure of how well the regression predictions approximate the real data, was used for the evaluation of the predictive performance. R2 score was calculated for the comparison of predicted data for the test dataset (20% of the used dataset) and real values, and it was 0.9886 for the Lasso regression.

Although the only small percent of collected parameter was used Lasso model achieved a high R2 score on the test dataset, demonstrating excellent predictive performance. Consequently, according to the results of Lasso model, the most significant 10 features affecting the prediction performance are 1. urban population (% of total population) (2022), 2. rural population (2022), 3. birth rate, crude (per 1,000 people) (2023), 4. energy use (kg of oil equivalent per capita) (2022), 5. population in urban agglomerations of more than 1 million (% of total population) (2022), 6. renewable energy consumption (% of total final energy consumption) (2022), 7. mortality rate, under-5 (per 1,000 live births) (2023), 8. cereal production (metric tons) (2022), 9. GDP per capita (current US$), 10. surface area (sq. km) (2022).

For the further assessment of the features' significance XGBoost ensemble learning model was applied due to its high predictive performance and essential ability to assess feature importance. XGBoost evaluates the feature importance using several metrics as information gain, cover and frequency. Application of this model helps to determine the subset of input variables which have high importance scores as well as features contributing minimally to the predictive process. Since the underlying data structure is predominantly linear, simple linear model as Lasso is more flexible than the tree-based approach. This fact can be explained by the comparatively low R2 score which was equal to 0.8816 for the testing of the XGBoost model. Despite that the results demonstrate the potential superiority of the linear Lasso model, the most important features list determined as the result of the XGBoost model has not a big difference from the list received after the application of the Lasso model: 1. surface area (sq. km) (2022), 2. birth rate, crude (per 1,000 people) (2023), 3. urban population (% of total population) (2022), 4. rural population (2022), 5. population in urban agglomerations of more than 1 million (% of total population) (2023), 6. energy use (kg of oil equivalent per capita) (2022), 7. cereal production (metric tons) (2022), 8. forest area (% of land area) (2022), 9. agriculture, forestry, and fishing, value added (current US$) (2022), 10. mortality rate, under-5 (per 1,000 live births) (2023).

While some indicators, such as urban population, may intuitively seem to have a strong impact on $CO_2$ emissions, data analysis using applied ML methods provides a systematic and objective way to validate such assumptions and uncover less obvious patterns. Application of the ML algorithms for evaluation of the feature importance not only rank the importance of variables based on their actual influence, but also uncover cases where significant indicators, in fact, have minimal predictive power. This highlights the critical role of data-driven approaches in avoiding biases and ensuring that policy and decision making is based on evidence instead of assumption.

Here we evaluate experimental evaluation of LSTM model and HMM for analysing $CO_2$ emissions and their prediction based on the values for 251 countries and regions for 65 years. The LSTM model is trained with sequences of shape [samples, timesteps = 3, features = 1], where each sequence contains three years of emission data, is optimized by the Adam optimizer with mean squared error. Figure 1 demonstrates dependency between real and predicted $CO_2$ emissions for Azerbaijan.

Application of HMM for prediction of the $CO_2$ emissions allows to identify latent regimes characterizing different statistical patterns in the time series (figure 2). By segmenting $CO_2$ emissions data into distinct states, where each of these states represents a specific pattern such as stable, increasing, peak, or reducing emissions, this approach allows to encode effectively structural changes and transitions [21, 30]. Plotted graph of the states received as result of extraction of the initial and transitional probabilities in case of Azerbaijan

corresponds to different time periods and explains $CO_2$ dynamics in these periods: relatively stable emissions during the early period (1960 – early 1980s), peak in emissions around mid-1980s, declining the emissions phase in the late 1980s and early 1990s reflecting significant changes in environmental conditions, which can be related to socio-economic transformations occurring during that period following by industrial decline and political instability.

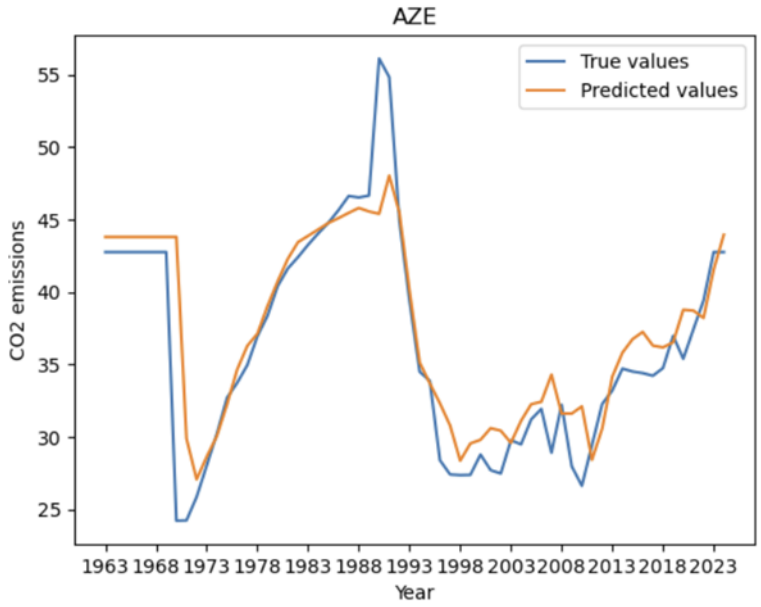


Figure 1. Dependency between real and predicted $CO_2$ emissions for Azerbaijan

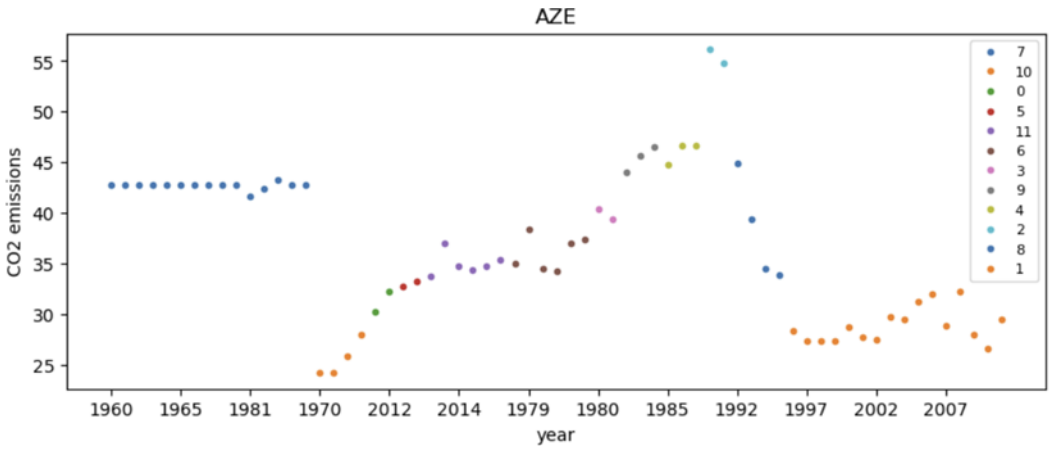


Figure 2. Distribution of hidden states of $CO_2$ emission change received as result of HMM model in case of Azerbaijan

## 4. Conclusions and Future Perspectives

Examination a wide range of environmental, economic, energy and social indicators from 266 countries and regions can be a significant step forward comprehensive exploration of a climate change and $CO_2$ emissions using data science and machine learning approaches. The results demonstrate how data-driven techniques can uncover patterns and relationships that support sustainable development policy and planning.

The clustering analysis using supervised learning algorithms such as k-means, BIRCH and MeanShift algorithms allowed to group countries with similar characteristics. These clusters provided insights into shared environmental profiles and development footprints prompting foundations for comparative analysis. For instance, countries like Azerbaijan were clustered with different countries from Asia, Europe and Africa, suggesting common structural, social and policy factors influencing their emission trajectories. Eventually these clusters can inform policy borrowing, where countries with higher emissions learn from their lower-emitting associates. Specified cross-national learning supports regional cooperation and increases the effectiveness of climate mitigation strategies.

Identification of the most significant features influencing $CO_2$ emissions using machine learning models highlighted the importance of factors such as urbanization, energy use, land use, population dynamics and economic development. These findings provide a deeper understanding of driving force of $CO_2$ emissions and emphasize most impactful areas of interventions. Results of this part supports more targeted and resource-efficient policy-design shrinking the list of influential indicators.

By incorporating temporal modelling using LSTM and HMM for exploration of the dynamic nature of emissions over time, this work provides more accurate analysis compared with traditional statistical analysis. The LSTM model enabled forecasting based on historical trends, offering ability to foreseen future emission levels. The HMM, on the other hand, helped identify distinct regimes called states in the emissions trajectory – such as growth, stability or decline, which offer a structured understanding of historical transitions. For example, periods of declining emissions were linked to major socio-economic transformations, including political shifts and industrial decline, providing more saturated richer context for interpreting the data.

Collaborative application of intelligent methods demonstrate possibility to support evidence-based decision making using environmental data. The potential application of analysing past trends, identifying influential indicators and predicting of future emissions is providing valuable tools for policymakers and researchers for evaluating the progress toward sustainability goals. Several interesting aspects may be extracted by integration of satellite-based remote sensing data. This opportunity allows to improve spatial resolution and relevance of environmental monitoring. Additionally, usage of real-time processing through indicators such as energy consumption, deforestation rates and urban expansion can improve efficiency and timeliness of predictive models. The full potential of application AI and ML for addressing global sustainability challenges can be proved by growing of data availability and computational capacity. Finally, providing effective tools for guiding evidence-based and adaptive policies that aligned with long-term climate goals of United Nations and climate agreements such as COP29.

## Acknowledgment

The authors gratefully acknowledge the support of ASOIU HPC Center for providing the computational resources used in the study's data analytics.

## Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Authors Contribution

Both authors made substantial contributions to the conception, analysis, and preparation of this work. **R.A.** conceived the research idea and designed the methodology. **N.I.** carried out the experiments and data collection. **N.I.** conducted the data analysis and interpretation. **R.A.** contributed to the manuscript writing and critical revision. All authors reviewed and approved the final version of the manuscript.

## References


1. Arora NK, Mishra I. United Nations Sustainable Development Goals 2030 and environmental sustainability: race against time. Environmental Sustainability. 2019;2(4):339–42.
2. United Nations Sustainable Development Goals [Internet]. Available from: https://sdgs.un.org/
3. de Souza JT, de Francisco AC, Piekarski CM, Prado GF do, de Oliveira LG. Data mining and machine learning in the context of sustainable evaluation: a literature review. IEEE Latin America Transactions. 2019;17(03):372–82.
4. Andries A, Morse S, Murphy R, Lynch J, Woolliams E, Fonweban J. Translation of Earth observation data into sustainable development indicators: An analytical framework. Sustainable Development. 2019;27(3):366–76.
5. Pande CB, Egbueri JC, Costache R, Sidek LM, Wang Q, Alshehri F, et al. Predictive modeling of land surface temperature (LST) based on Landsat-8 satellite data and machine learning models for sustainable development. Journal of Cleaner Production. 2024;444:141035.
6. Yao Z, Lum Y, Johnston A, Meija-Mendoza LM, Zhou X, Wen Y, et al. Machine learning for a sustainable energy future. Nature Reviews Materials. 2023;8(3):202–15.
7. Poli G, et. al. A Data-Driven Approach to Monitor Sustainable Development Transition in Italian Regions Through SDG 11 Indicators. In: Gervasi Osvaldo and Murgante B and GC and TD and CRAMA and FLMN, editor. Computational Science and Its Applications – ICCSA 2024 Workshops. Cham: Springer Nature Switzerland; 2024. p. 337–55.
8. Jabbari M, Shafiepour MM, Ashrafi K, Abdoli G. Differentiating countries based on the sustainable development proximities using the SDG indicators. Environment, Development and Sustainability. 2020;22(7):6405–23.
9. da Silva J, Fernandes V, Limont M, Rauen WB. Sustainable development assessment from a capital's perspective: Analytical structure and indicator selection criteria. Journal of Environmental Management. 2020;260:110147.
10. Chen IC. Predicting regional sustainable development to enhance decision-making in brownfield redevelopment using machine learning algorithms. Ecological Indicators. 2024;163:112117.
11. Molina-Gómez NI, Rodríguez-Rojas K, Calderón-Rivera D, Díaz-Arévalo J,L López-Jiménez PA. Using machine learning tools to classify sustainability levels in the development of urban ecosystems. Sustainability (Switzerland). 2020 Apr 1;12(8).
12. El-Aal MFA. The relationship between $CO_2$ emissions and macroeconomics indicators in low and high-income countries: using artificial intelligence. Environment, Development and Sustainability. 2024.
13. Sinaga KP, Hussain I, Yang MS. Entropy K-Means Clustering with Feature Reduction under Unknown Number of Clusters. IEEE Access. 2021;9:67736–51.

14. Elghazel H, Aussem A. Unsupervised feature selection with ensemble learning. Machine Learning. 2015;98(1–2):157–80.
15. Li S, Siu YW, Zhao G. Driving Factors of $CO_2$ Emissions: Further Study Based on Machine Learning. Frontiers in Environmental Science. 2021;23:9.
16. Li X, Ren A, Li Q. Exploring Patterns of Transportation-Related $CO_2$ Emissions Using Machine Learning Methods. Sustainability (Switzerland). 2022;14(8:4588
17. Castelli T, Mocenni C, Dimitri GM. A machine learning approach to assess Sustainable Development Goals food performances: The Italian case. Plos one. 2024;19(1):e0296465.
18. García-Rodríguez A, Nuunez M, Robles M, Govezensky T, Barrio R, Gershenson C, et al. Sustainable visions: unsupervised machine learning insights on global development goals. PloS one. 2025;20(3):e0317412.
19. Mathrani A, Wang J, Li D, Zhang X. Clustering analysis on sustainable development goal indicators for forty-five asian countries. Sci. 2023;5(2):14.
20. Li J, Irfan M, Samd S, Ali B, Zhang Y, Badulescu D, et al. The Relationship between Energy Consumption, $CO_2$ Emissions, Economic Growth, and Health Indicators. International Journal of Environmental Research and Public Health. 2023;20(3):2325
21. United Nations Climate Change. The Paris agreement [Internet]. Available from: https://unfccc.int/process-and-meetings/the-paris-agreement
22. COP29 Baku Azerbaijan. COP29 Declarations on Green Digital Action [Internet]. Available from: https://cop29.az/en/pages/cop29-declaration-on-green-digital-action
23. Kumari S, Singh SK. Machine learning-based time series models for effective $CO_2$ emission prediction in India. Environmental Science and Pollution Research. 2023;30(55):116601-116616.
24. World Bank Group. World Bank Data [Internet]. Available from: https://data.worldbank.org/
25. Likas A, Vlassis N, Verbeek JJ. The global k-means clustering algorithm. Pattern recognition. 2003;36(2):451-461.
26. Januzaj Y, Beqiri E, Luma A. Determining the Optimal Number of Clusters using Silhouette Score as a Data Mining Technique. International Journal of Online & Biomedical Engineering (iJOE). 2023;19(4):174-182
27. Yadav A, Jha CK, Sharan A. Optimizing LSTM for time series prediction in Indian stock market. Procedia Computer Science. 2020;167:2091-2100.
28. Hanif M, Sami F, Hyder M, Ch MI. Hidden Markov model for time series prediction. Journal of Asian Scientific Research. 2017;7(5):196-205.
29. Cao W, Zhu W, Demazeau Y. Multi-Layer Coupled Hidden Markov Model for Cross-Market Behavior Analysis and Trend Forecasting. IEEE Access. 2019;7:158563–74.
30. Worldometer. $CO_2$ emissions by country [Internet]. Available from: https://www.worldometers.info/co2-emissions/co2-emissions-by-country/