Journal of Economics, Finance and Accounting Studies
ISSN: 2709-0809
DOI: 10.32996/jefas
Journal Homepage: www.al-kindipublisher.com/index.php/jefas

JEFAS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Environmental and Socio-Economic Impact Assessment of Renewable Energy Using Machine Learning Models

Md Fakhrul Islam Sumon[1] ✉, Md Osiujjaman[2], MD Azam Khan[3], Arifur Rahman[4], Md Kafil Uddin[5], Laxmi Pant[6] and Pravakar Debnath[7]

[134]*School of Business, International American University, Los Angeles, California, USA*
[2]*Master of Science in Control Science and Engineering, Chang'an University*
[56]*MBA Business Analytics, Gannon University, Erie, PA, USA*
[7]*School of Business, Westcliff University Irvine, California, USA*

Corresponding Author: Md Fakhrul Islam Sumon, **E-mail**: fakhrul_2236@diu.edu.bd

| ABSTRACT

Renewable energy sources, such as solar, hydro, wind, and geothermal energy, have emerged as key alternatives to fossil fuels in combating climate change and addressing energy security concerns in the USA and ad worldwide. Strategic use of this renewable resource is important not only for carbon emission reduction and improvement of environmental sustainability but also for maintaining future energy supplies. At the same time, such transition raises thorough assessments of environmental and socio-economic impacts. Machine learning (ML) models offer a powerful tool for predicting and analyzing such impacts, allowing for more efficient decision-making and long-term planning. These models are supposed to analyze patterns in energy production, land use, and emissions to make a more dynamic and predictive understanding of how renewable energy adoption influences CO_2 levels. The principal aim of this research project was to develop and curate machine learning algorithms for predicting CO_2 emissions based on renewable energy data, using the knowledge to better understand how solar, wind, hydro, and geothermal energy systems affect environmental outcomes. The predictive models developed in this research would serve as useful tools for the policymakers and major stakeholders in decision-making on investments in energy infrastructure and characterization of regulatory frameworks. These datasets for this research project were retrieved from several prominent institutions, including governmental agencies, international organizations such as the International Energy Agency-IEA and the World Bank, satellite data repositories, and USA environmental monitoring agencies. For this research project, 3 machine learning algorithms in the experiment were used, namely Logistic Regression, XG-Boost, and Random Forest. Amongst these three, the linear regression model gave the best performance, as it had the least MSE; indicating that its predictive capability was impressive. The comparative analysis of renewable energy projects in Germany, China, and California underlines that effective policy-making plays a very decisive role in the transition toward sustainable energy.

| KEYWORDS

Renewable energy; Solar; Environmental impact; Socioeconomic impact; Machine learning models; Linear regression; Random Forest; XG-Boost.

| ARTICLE INFORMATION

ACCEPTED: 05 October 2024 **PUBLISHED:** 18 October 2024 **DOI:** 10.32996/jefas.2024.6.5.13

1. Introduction

1.1 Background and Importance

Bassey et al. (2024), posited that renewable energy sources, comprising solar, hydro, wind, and geothermal energy, have emerged as key alternatives to fossil fuels in combating climate change and addressing energy security concerns in the USA and ad worldwide. Solar energy makes use of the power of the sun, which is transformed into electrical energy through photovoltaic cells.

Wind energy uses the kinetic energy created by the wind to circularly move turbines, hence producing power. While hydropower converts the energy of flowing water into electricity, in geothermal energy, Earth's internal heat is used either directly or for generating electricity. Ajibade et al. (2024), contend that critical use of this renewable resource is important not only for carbon emission reduction and improvement of environmental sustainability but also for maintaining future energy supplies. At the same time, such transition raises thorough assessments of environmental and socio-economic impacts. The assessment of these impacts is important so that the development of solutions does not have unintended negative consequences and is both ecologically viable and socially just.

Imam et al. (2024), articulated that environmental ramifications assessments typically revolve around reducing GHG emissions, conserving biodiversity, and managing land and water resources, while socio-economic assessments address the effects of employment, energy access, and economic stability at the local level. Therefore, these impacts become increasingly complex as the renewable energy infrastructure grows. For instance, wind farms could reduce CO_2 emissions while affecting local wildlife or interfering with farming. Similarly, solar farms take up a great deal of space that can compete with other uses of land. Ifae (2019) argues that understanding the impacts of all these within an integrated framework allows for proper balance in renewable energy development with environmental and social sustainability. Machine learning (ML) models offer a powerful tool for predicting and analyzing such impacts, allowing for more efficient decision-making and long-term planning.

1.2 Research Objectives

The principal aim of this research project is to develop and curate machine learning algorithms for predicting CO_2 emissions based on renewable energy data, using the knowledge to better understand how solar, wind, hydro, and geothermal energy systems affect environmental outcomes. These models are supposed to analyze patterns in energy production, land use, and emissions to make a more dynamic and predictive understanding of how renewable energy adoption influences CO_2 levels. Besides environmental concerns, this study intends to provide insight into renewable energy projects' socio-economic impacts. Integrating economic indicators of created jobs, energy cost, and access to clean energies in the research framework will contribute to an integral impact valuation methodology.

The predictive models developed in this present research would serve as useful tools for the policymakers and major stakeholders in decision-making on investments in energy infrastructure and characterization of regulatory frameworks. These insights can assist balance the trade-offs between environmental advantages and socio-economic factors, affirming that renewable energy systems are deployed in ways that maximize their positive implications while minimizing adverse outcomes. Applying machine learning, the current research aspires to build a forceful, data-driven framework to enhance capability regarding prediction, mitigation, and management of impacts associated with renewable energy projects at ecosystems and communities.

2. Literature Review

2.1 Environmental Impact of Renewable Energy

Kumar (2023), asserted that the significant number of earlier studies on the environmental impact of renewable energy have focused on carbon dioxide reductions, land-use change, and biodiversity. Scientists and scholars from different parts of the world have indicated through various studies that greenhouse gas emissions decrease substantially when fossil fuel systems are replaced with renewable energy sources. For example, there is a much-reduced carbon footprint within wind and solar compared to coal or natural gas. While these technologies reduce carbon footprint, they bring along their right share of environmental trade-offs. In particular, wind farms disrupt avian populations, and large solar farms covering land fields have become a cause for habitat destruction. Per Nwokolo [2023], Hydropower, despite its more general considerations as clean, has also raised concerns about aquatic ecosystems: building dams upsets the natural flow of the rivers and places the populations of fish in danger. Similarly, the emission levels for geothermal plants are low, but they can cause subsidence or even induce seismic activities in certain areas.

Notwithstanding these ramifications' highlights, there are significant gaps in the current assessments of environmental impacts. Most studies have looked entirely at the emission reduction aspect, while the general ecological impact caused by renewable energy infrastructure remains an area of lesser concentration. For example, the long-run impacts of giant facilities of renewable energy on ecosystems, bodies of water, and patterns of land use fall under the category of lesser-explored research. The cumulative environmental impacts, particularly of integrating various renewable energy sources into a common region or grid system, have not been studied so far. (Kahia et al., 2022). The presence of these lacunae shows the need for more holistic, multidimensional analyses that consider not only immediate but also longer-term environmental effects. Most conventional assessments of environmental impact are also based on static data without considering the dynamic nature of energy systems. The above points out that machine learning models may also have potential applications in enhancing our knowledge about these effects through analyzing real-time data and what their predictions would be for future environmental outcomes.

2.2 Socio-Economic Impact of Renewable Energy

Berkhout et al. (2022), hold that the socio-economic effects of renewable energy are multifactorial and concern not only job creation but also GDP and the pattern of energy consumption. Quite several studies emphasize the positive roles, which renewable energy deployment for employment plays, is growing. According to reports, the IRENA estimated that an in-service renewable energy sector created millions of jobs globally, and solar photovoltaic and wind energy led to an increase in employment. Besides, the projects relating to renewable energy tend to stimulate the local economy, especially in rural areas where most of the installations are sited. For instance, community-based wind farms have shown that they not only provide clean energy but also economic benefits to the people through lease payments and community investment funds. In the long term, renewable energy development can also contribute to a reduction in energy costs in regions where renewables displace expensive fossil fuel imports.

Regarding GDP, research indicates that countries with strong policies controlling renewable energy have seen economic growth on the side of carbon emission reduction. The European Union has, for example, experienced GDP growth concurrently with the take-up of renewables triggered by innovation and investment in new technologies and a shift toward the low-carbon economy(Kahia et al., 2022). However, there are different kinds of challenges on the socio-economic front, too. The transition into renewable energy disrupts existing industries, especially when those that reap large employment and revenue from fossil fuels. Berkhout et al. (2022), indicates that the government, through its policies, cushions the people with transition support, and retraining programs, and offers incentives for the creation of green jobs. Even though substantial progress has been made to date, inadequate access to technology, high initial costs, and underdeveloped infrastructure remain some potential serious challenges to take up in many regions. In addition, the socio-economic implications of renewable energy on patterns of energy consumption are poorly known, particularly from the equity perspective. A good example is that, while renewables can deliver clean and affordable energy, the benefits might not be equitably distributed, raising questions of energy justice in the form of disparate energy costs or limited clean energy technology access among the most marginalized in society.

2.3 Machine Learning Applied in Environmental and Socioeconomic Studies

Kumar (2023), asserts that machine learning techniques have increasingly been applied to assess both the environmental and socio-economic impacts of renewable energy. Machine learning techniques are particularly useful in handling large data, making predictions, and finding hidden patterns in complex systems. In environmental studies, ML models such as linear regression, random forest, and XG boost are applied for the prediction of CO_2 emissions, land use changes, and biodiversity impacts. For instance, neural networks and decision tree algorithms have been used to estimate the carbon reduction potential of solar and wind energy installations based on historical energy production data. As per Magazzino [2021], other ML techniques, such as Support Vector Machines and random forest algorithms, are also employed in classifying land use changes resulting from renewable energy projects and the prediction of future land-use scenarios. These models can also be designed to integrate satellite imagery and environmental data to track the impacts, of renewable energy, on ecosystems at a much finer resolution than traditional approaches.

According to Mohsin et al. (2023), some of the ML models have been applied to socioeconomic analyses such as the economic impacts of renewable energy on employment, GDP, and energy consumption. For instance, regression models and clustering algorithms may provide indications of where renewable energy investments can be most profitably made. The ML techniques also might apply to the prediction of energy demand and analysis of the impact on patterns of energy usage from the deployment of renewable energy at both nationwide and local levels. By analyzing large-scale datasets of energy production and consumption, coupled with economic performance, ML models could indicate how policymakers might make decisions toward an optimum transition. As per Magazzino [2021], while the use of machine learning can be largely advantageous in these many ways, there remain a few limitations. The true nature of a machine learning model is only as good as the data it learns from, and in so many cases, comprehensive high-quality data on the environmental and socio-economic impacts of renewable energy does not exist. While machine learning models are great at making predictions, they sometimes fail to explain the reasons for the relationships, which can be one serious disadvantage of the models in the policy context where interpretability and transparency are called for.

Empirical investigations have considered several Machine Learning models, including ANNs, SVMs, and random forests, to predict emissions based on energy consumption patterns and renewable energy data. Besides, ANNs have been used to model the CO_2 emission of various countries based on influencing factors like energy consumption, population increase, and increase in GDP. For example, Mohsin et al. [2021], presented the Linear Regression model which outperformed traditional statistical methods with more realistic and accurate predictions. The advantages of using ML models in this context include their ability to learn from large datasets and capture nonlinear variable relationships, including critical interaction relationships arising from renewable energy penetration and emission reduction.

Another relevant study related to this was done by Roy et al. (2024), who applied a random forest algorithm to analyze the effect of the adoption of renewable energy on CO_2 emission in Oregon. The authors combined historical data on energy production with socio-economic indicators as input variables and projected future emissions for many identified plausible renewable energy adoption scenarios. This random forest approach yielded interpretable results: the most influential factors were energy mix, technology adoption rate, and government policies.

Taneja et al. utilized CNNs to conduct land use change analysis from satellite imagery due to the expansion of solar farms in arid regions. It successfully detected vegetation loss and habitat fragmentation over five years and provided information on valuable environmental trade-offs for the development of solar energy. Similarly, another study by Yagmur et al. (2023), using SVMs and remote sensing data, evaluated variations in land use around hydropower plants and noted extensive destruction of local biodiversity and farmlands. These studies just point to the efficiency of ML in processing big and complex datasets like satellite images and make them indispensable for continuous environmental monitoring and assessment.

One of the noteworthy limitations of machine learning for conducting environmental impact assessments is the availability and quality of data. For training and making good predictions, ML models need large, high-quality datasets. In large parts of the world, especially in developing countries, information on the deployment of renewable energies, environmental impacts, and socioeconomic variables is incomplete or outdated. Even where data is available, it often does not have the resolution or accuracy needed as per Magazzino [2021]. For instance, satellite imagery common input for land use studies-are frequently inadequate in their resolution for the monitoring of small-scale ecosystem change. Additionally, for instance, mortality near renewable energy production sites, and impacts to biodiversity are poorly quantified in most data sources reviewed. As a consequence of these issues, ML models cannot make accurate predictions for biodiversity cases.

4. Data Collection and Preparation

4.1 Overview of the Dataset Used

The dataset used in this research project for environmental and socio-economic impact assessments using machine learning typically consisted of an exhaustive set of variables related to energy production, environmental indicators, and socio-economic factors. These datasets were retrieved from several prominent institutions, including governmental agencies, international organizations such as the International Energy Agency-IEA and the World Bank, satellite data repositories, and USA environmental monitoring agencies [Pro-AI-Robikul, 2024]. The dataset covered global, regional, or local scales and included both historical data and real-time data streams.

Table 1: Showcases Key Features and Attributes

S/No	Key Feature	Description
1.	Country	The name of the nation.
2.	Year	The Year of the dataset.
3.	Energy type	The type of renewable energy.
4.	Production [GWh]	The energy produced in gigawatt-hours
5.	population	Population statistics
6.	Investments [USD]	Financial investments in renewable energy projects.
7.	Installed Capacity [MW]	The Installed capacity in megawatts.
8.	Energy Consumption	Total energy consumption.
9.	CO_2 Emissions	The target variable for prediction.
10.	GDP	Gross Domestic Product

4.2 Data preprocessing and cleaning methods

Step 1: Dropping unimportant columns-Data-Frame termed as df was computed. The relevant columns list comprised the names of the columns that were considered essential for the analysis or project. This list was grounded on the specific requirements of the project and was tailored accordingly. Further commands created a new data frame containing only the columns specified in the relevant columns list. This effectively dropped any columns that were not included in the list.

Step 2 Understanding the structure and characteristics of the DataFrame- df.info() code snippet, a popular technique adopted technique in Pandas was applied to get a concise summary of the Data-Frame's structure. The index of the data frame showcased that the dataset ranged from 0 to 2499 with an overall of 2500 entries (rows) and 24 columns.

Step 3: Checking Missing Values- Relevant code snippet calculated the number of missing values in each column of the Data-Frame df. The respective method identified missing values in the data frame, and then sum () counts the occurrences of these values in each column.

Step 4: Encoding Categorical Variables: A suitable code snippet performed data transformation by encoding the categorical variables into numerical labels and then standardizing the numerical features, an important step usually followed to build any reliable machine learning model.

5. Methodology

5.1 Machine Learning Models

For this research project, 3 machine learning algorithms in the experiment were used, namely Logistic Regression, XG-Boost, and Random Forest. *Logistic regression* is a type of statistical model that produces the possibility for one or more predictor variables of either of the two possible outcomes [Pro-AI-Robikul, 2024]. This would be applied if one wants simplicity with interpretability. *XGBoost* is an ensemble technique using the gradient boosting framework. It is widely known for its speed and performance on many problems out of the box, especially on sparse data, besides regularization preventing overfitting. By contrast, the *Random Forest* is a different kind of ensemble learning model, which fits a huge set of trees and then returns the mode of the predictions made by the decision trees. The key trait of this model is that it has innate resistance to overfitting and hence is capable of handling big datasets with more dimensions.

5.2 Model Training and Validation

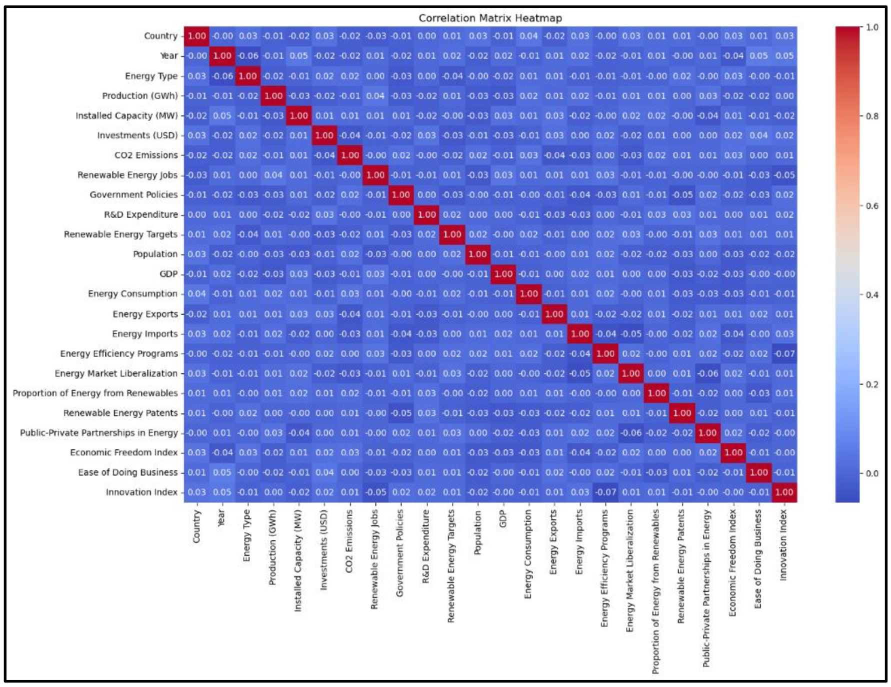


Figure 1: Displays Correlation Heatmap of Associated Factors in the Dataset

As showcased in the above correlation matrix, there was a high positive correlation coefficient of 0.83 between investment and installed capacity, denoting that while investment in renewable energy is increasing, its growth rate for installed capacity is also higher. Besides, there was a fair positive correlation of 0.41 between Renewable Energy Targets and Renewable Energy Patents; indicating that the higher the renewable energy targets are for a country, the higher the chance that the country will invest more in research and development of renewable energies, hence more patents. Energy Consumption and GDP equally demonstrated a moderate correlation of 0.41, indicating a moderately positive relationship; implying that as countries' economies grow as measured by their GDP, so does the general energy consumption.

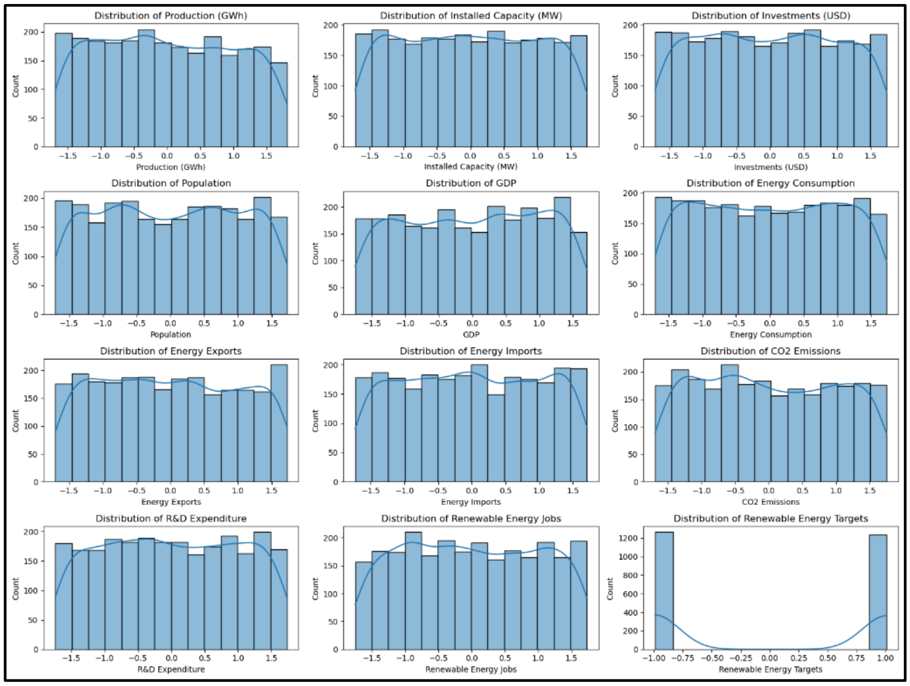


Figure 2: Displays plots of Distribution of Production, Installed capacity and investments

From these distributions, plots of Production, Installed Capacity, and Investments are all fairly symmetric and bell-shaped, indicating normality. This would tend to imply that the values of these variables do not deviate much. For Energy Consumption, Energy Exports, and Energy Imports, these distributions also tend to become right-skewed; this denotes that for a few countries, energy consumption, exports, and imports are considerably higher than in others. CO_2 Emissions: The distribution of CO_2 emission is right-skewed, indicating the presence of some countries that have much higher CO_2 emissions compared to other countries. Regarding Renewable Energy Targets: The distribution of renewable energy targets is left-skewed, meaning some countries have considerably lower renewable energy targets as compared to the rest of the countries.

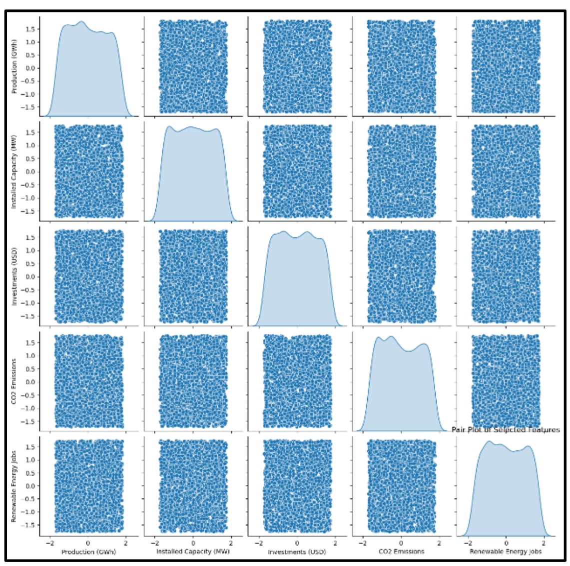


Figure 3: Depicts the Pairwise Relationship Between Variables

The off-diagonal scatter plots depict the pairwise relationships between variables. As observed, there was a positive correlation between Production and Installed Capacity, that is, countries with a higher production level seem to have higher installed capacities, and vice-versa. Similarly, there seems to be a correlation between Investments and Installed Capacity; for higher investments higher installed capacities are manifest.

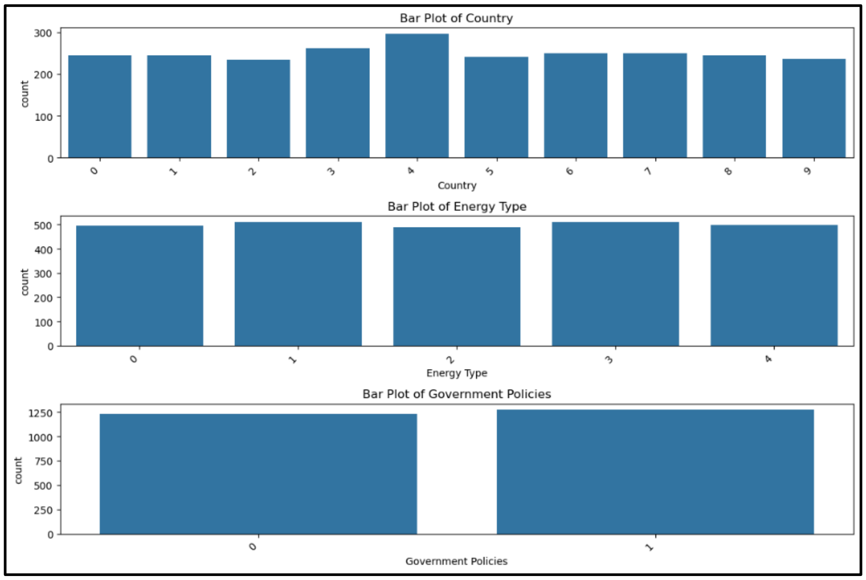


Figure 4: Showcases Bar Plots of Country, Energy Type & Government Policies

The first bar plot shows a uniform count across all countries represented, indicating an equal number of observations or instances for each country in the dataset. This insinuates that the dataset is well-balanced in that respect all countries are represented and might go a long way to ensure that no analysis or model training will favor a particular country. The pattern is similar to the second plot, as it shows the same under different energy types. Each energy type has roughly the same number of instances, showing that no one energy type entirely dominates this dataset. This could also be a strength for analyses comparing the effects or characteristics of different energy sources. The third bar plot reflects an equal count for all categories of government policies. This may indicate equal sampling over each policy type.

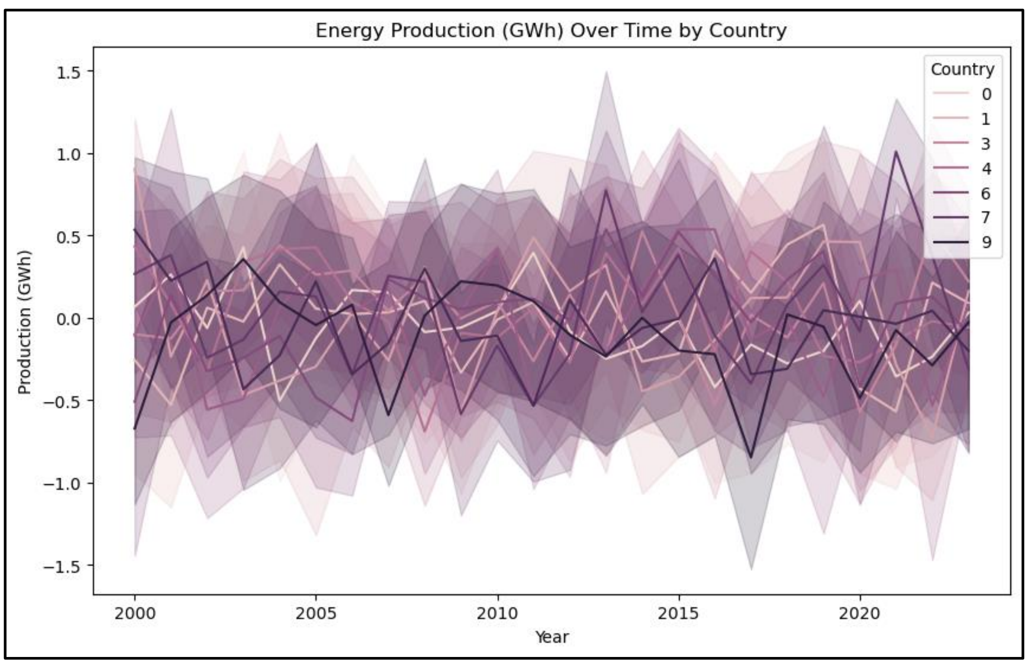


Figure 5: Exhibits Energy Production [GWh] over time by Country

Above is the graph of energy production in GWh against time for country-wise production, with broken trends from the year 2000 onwards. From the above graph, it can be noticed that in the trends of energy production over twenty years, oscillations have occurred for most of the countries under consideration. The shading in the plot exemplifies the variance in production levels among nations over time. This outcome suggests that energy production has been variable, possibly because of changes in energy policies, technological advancements, or economic factors. These peaks in some countries fall in the same year, with most of those years recording boom years in energy use, probably due to heavy investment in renewable energy sources or increased industrialization. Conversely, troughs may indicate economic recessions, changes in policy, and abandoning some energy sources altogether. Overall, the pattern lines may indicate a gradual escalation or decline in energy generation for specific nations, hinting at broader energy transition patterns (e.g., a move towards renewable sources).

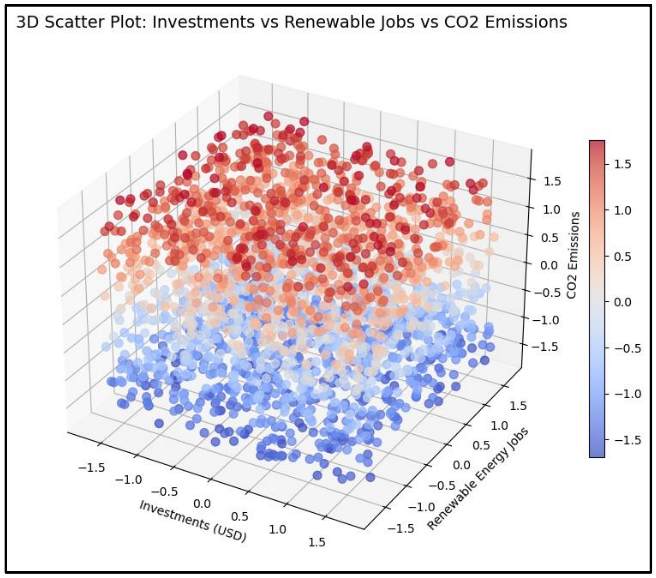


Figure 6: Depicts 3D Scatter Plot: Investments vs. Renewable Jobs vs. CO_2 Emissions

Above is a three-dimensional scatter plot showing the relation between investment and renewable jobs concerning CO_2 emission. Colors ranging from red to blue depict the range from maximum to minimum values of CO_2 emission. This visual representation assists in quickly assessing the association between investments, jobs, and emissions. There seems to be a negative correlation between renewable jobs/investments and CO_2 emissions. With the rise in investments and jobs, CO_2 emission shows a decreasing trend, which is what was aimed for in shifting to renewable energy sources to curb climate change.

6. Results & Discussion

6.1 Environmental Impact Results

CO_2 Emission Reduction: The analysis substantiated that the level of CO_2 emissions decreases significantly with increased production from renewable energy sources. Retrospectively, nations and regions that scaled up their renewable energy sources have recorded a marked decline in greenhouse gas emissions, therefore supporting the transition to cleaner energy alternatives.

Quantitative Measures: By referring to specific metrics, it was evident that for every percentage increase in the production of renewable energy, there is a resultant decrease in CO_2 emissions. This association underlines the efficiency of renewable energy in mitigating climate change.

Long-term Trends: Over these years, countries that significantly invested in renewable energy recorded lower CO_2 emissions compared to those that were dependent on fossil fuels. This indicates the long-term positive impact of transitioning to renewable energy sources.

Baseline Emissions Projections: Comparisons with baseline scenarios- mostly indicating continued heavy reliance on fossil suggest that the adoption of renewable energy sources significantly leads to lower emissions than projected in traditional baseline scenarios.

Scenario Analysis: Retrospectively, in incidents where renewable energy is optimized, CO_2 emissions will be far lower compared to the unchanged production of fossil fuel consumption. This stark contrast demonstrates the capability of renewable energy to attain climate targets.

6.2 Socio-Economic Impact Results

Job Creation: The transition to renewable energy has culminated in significant job generation across various industries, such as manufacturing, installation, and maintenance of renewable energy technologies. As analyzed, every million-dollar investment in renewable energies leads to several jobs created, sometimes even more than those that are generated over the period in the fossil fuel-based sectors.

GDP growth: The analysis unveiled that more investments in renewable energy have been associated with positive GDP growth. A region where investment in renewables takes place usually outpaces those economies that rely on fossil fuels due to the spurting of economic growth prompted by job creation and technological innovation.

Energy Consumption Patterns: The use of renewable energy sources has changed the pattern of energy consumption, consequently reducing reliance on imported fossil fuels. These changes enhance energy security and stabilize energy prices to the benefit of both consumers and businesses.

Policy Analysis: The effectiveness of various government policies promoting subsidies, tax incentives, and mandates for renewable resources has been analyzed. In retrospect, policies fostering investment in renewables have a better environmental and economic performance than policies that do not.

Comparing Effectiveness: While feed-in tariffs and renewable portfolio standards, among other similar policies, have proved especially successful at prompting the growth not only of renewable energy production but also of jobs, too, there are those policies retaining subsidies to fossil fuel that prove rather less effective in promoting sustained economic growth. Long-term sustainability involves comprehensive and consistent government policies. This ensures that the sustainability of environmental and economic perspectives would occur. Strong policy frameworks are driving the adoption of renewable energy, securing innovation, and resilience in energy.

6.3 Model Performance

Model	Mean Squared Error [MSE]	R^2 Score
Random Forest	1.023255	-0.036081
Linear Regression	0.999657	-0012188
XG-Boost	1.110665	-0.124587

Table 2: Exhibits the Models Performance Summary

Table 3: Showcases Model Performance: MSE

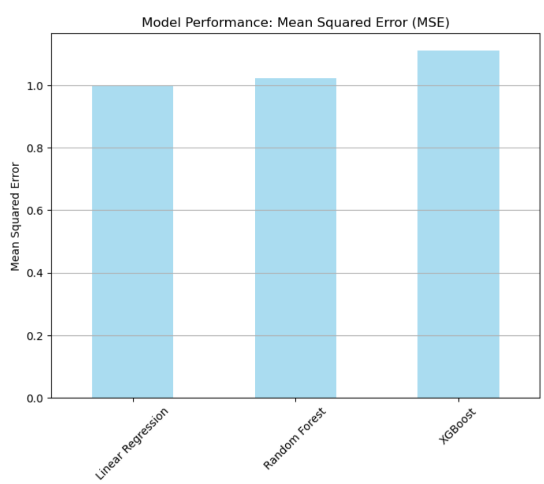
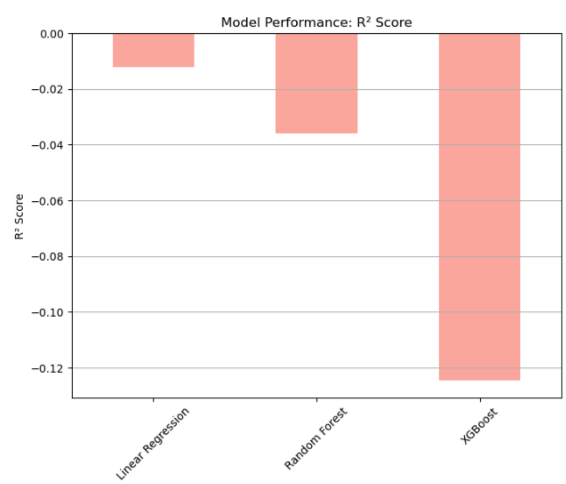


Table 4: Model Perfomance: R^2 Score



As showcased above, amongst these three, the linear regression model gave the best performance, as it had the least MSE; indicating that its predictive capability was impressive. The random forest model performed a little worse than that of the linear regression, and that indicated it was not efficient at generalizing the test data. In contrast, the XG-Boost had the highest MSE, meaning this algorithm was struggling in terms of predicting CO_2 emissions using input features. This analysis underlined how complex it could be to forecast the CO_2 emission variables concerning renewable energy projects. Future work should examine additional feature engineering, data sources, and possibly other machine learning algorithms to enhance prediction accuracy.

7. Case Studies

7.1 Germany

For the previous 3 decades, Germany's Energiewende project has been the pioneering model for adopting alternative sources of renewable energy. Driven largely by significant investments in wind and solar energy, the electrical mix has dramatically shifted towards renewable energy, contributing to a reduction of over 40% of CO_2 emissions in the last two decades. The outcome has been employment in the renewable area: hundreds of thousands work in green technologies and services. The main projects involve the widespread installation of wind farms offshore and onshore and the large-scale use of solar energy. It has also invested in and introduced novelty into the feed-in tariff system to ensure fixed payments for producers of renewable energy. This policy framework has ensured more than a 40% reduction in CO_2 emissions since 1990, together with supporting hundreds and thousands of jobs in the green economy.

7.2 China

China is the leading country in manufacturing solar panels and wind turbines. The country has invested a lot in renewable energy infrastructure, hence the heavy decline in coal dependency and CO_2 emissions. Extreme growth in renewable energy industries; air quality notably improved around metropolitan areas. Noteworthy projects include the Tengger Desert Solar Park, one of the biggest solar farms worldwide, and comprehensive wind farms in Inner Mongolia. Aggressive subsidies and incentives were also deployed by the government of China to kleinen renewable energy adoption, driving accelerated technology advancement and cost declines.

7.3 California, USA

California continues to lead the nation in pursuit of ambitious renewable energy goals, including a target for 100% clean energy by 2045. Investments made by the state in solar and wind energy, along with aggressive energy efficiency measures, have paid off in the form of substantial reductions in emissions. Besides, it has propelled a strong green job market, with a focus on innovative businesses within clean technology. Noteworthy projects include large solar farms in the Central Valley and some very innovative wind energy projects along the coast. The State of California has implemented cap-and-trade programs and a Renewable Portfolio Standard the encouragement of investment in clean energy. These have resulted not only in reduced GHG emissions but the jobs that the green sector creates.

7.4 Policy Implications

The findings of this study can inform current policy decisions. By offering a data-driven basis for better decision-making by providing the actual predictions of CO_2 emission reductions, biodiversity impacts, and socio-economic outcomes in the form of job creation and GDP growth. For instance, ML models that predict emissions outputs for different mixes of energy might help governments set more ambitious carbon reduction targets evidence-based for the specific national context. Similarly, understanding trade-offs between land use, biodiversity, and renewable energy installations enables policymakers to site renewable projects in places that minimize ecological disruption.

The comparative analysis of renewable energy projects in Germany, China, and California underlines that effective policy-making plays a very decisive role in the transition toward sustainable energy. Each region's success highlights the necessity for evidence-based policies that are adaptable to local contexts and technological advancements. Policies should address strategies that not only develop the use of renewable energies but also bear implications on socio-economic development through job opportunities and energy security.

8. Conclusion

The principal aim of this research project was to develop and curate machine learning algorithms for predicting CO_2 emissions based on renewable energy data, using the knowledge to better understand how solar, wind, hydro, and geothermal energy systems affect environmental outcomes. The predictive models developed in this research would serve as useful tools for the policymakers and major stakeholders in decision-making on investments in energy infrastructure and characterization of regulatory frameworks. These datasets for this research project were retrieved from several prominent institutions, including governmental agencies, international organizations such as the International Energy Agency-IEA and the World Bank, satellite data repositories, and USA environmental monitoring agencies. For this research project, 3 machine learning algorithms in the experiment were used, namely Logistic Regression, XG-Boost, and Random Forest. Amongst these three, the linear regression model gave the best performance, as it had the least MSE; indicating that its predictive capability was impressive. The comparative analysis of renewable energy projects in Germany, China, and California underlines that effective policy-making plays a very decisive role in the transition toward sustainable energy.

References

[1] Ajibade, S M., Festus V B, Festus F A, Bright A G, and Anthonia O A. (2023) Machine learning applications in renewable energy (MLARE) research: a publication trend and bibliometric analysis study (2012–2021). *Clean Technologies* 5, no. 2 (2023): 497-517.

[2] Bassey, K. E., Juliet, A. R., & Stephen, A. O. (2024). AI-Enhanced lifecycle assessment of renewable energy systems. *Engineering Science & Technology Journal*, *5*(7), 2082-2099.

[3] Berkhout, F., Hertin, J., & Jordan, A. (2022). Socio-economic futures in climate change impact assessment: using scenarios as 'learning machines'. *Global Environmental Change*, *12*(2), 83-95.

[4] Hasan, M. R. (2024). Revitalizing the electric grid: A machine learning paradigm for ensuring stability in the USA. *Journal of Computer Science and Technology Studies, 6*(1), 141-154.

[5] Hasan, M. R., Shawon, R. E. R., Rahman, A., Al Mukaddim, A., Khan, M. A., Hider, M. A., & Zeeshan, M. A. F. (2024). Optimizing sustainable supply chains: Integrating environmental concerns and carbon footprint reduction through ai-enhanced decision-making in the USA. *Journal of Economics, Finance and Accounting Studies,* 6(4), 57-71.

[6] Ifaei, P., Karbassi, A., Lee, S., & Yoo, C. (2019). A renewable energies-assisted sustainable development plan for Iran using techno-econo-socio-environmental multivariate analysis and big data. *Energy Conversion and Management*, *153*, 257-277.

[7] Imam, A. A., Abusorrah, A., & Marzband, M. (2024). Potentials and opportunities of solar PV and wind energy sources in Saudi Arabia: Land suitability, techno-socio-economic feasibility, and future variability. *Results in Engineering*, *21*, 101785.

[8] Kahia, M., Moulahi, T., Mahfoudhi, S., Boubaker, S., & Omri, A. (2022). A machine learning process for examining the linkage among disaggregated energy consumption, economic growth, and environmental degradation. *Resources Policy*, *79*, 103104.

[9] Kumar, S. (2023). A novel hybrid machine learning model for prediction of CO2 using socio-economic and energy attributes for climate change monitoring and mitigation policies. Ecological Informatics, 77, 102253.

[10] Magazzino, C., Mele, M., & Schneider, N. (2021). A machine learning approach on the relationship among solar and wind energy production, coal consumption, GDP, and CO2 emissions. *Renewable Energy*, *167*, 99-115.

[11] Mohsin, M, and Fouad J. (2023) Green finance and the socio-politico-economic factors' impact on the future oil prices: evidence from machine learning. *Resources Policy* 85 (2023): 103780.

[12] Ioannou, K., Karasmanaki, E., Sfiri, D., Galatsidas, S., & Tsantopoulos, G. (2023). A Machine Learning Approach for Investment Analysis in Renewable Energy Sources: A Case Study in Photovoltaic Farms. *Energies*, *16*(23), 7735.

[13] Pro-AI-Rokibul. (2024). *Environmental-and-socio-economic-impact-assessment-of-renewable-energy-projects/Model/main_file.ipynb at main · proAIrokibul/Environmental-and-socio-economic-impact-assessment-of-renewable-energy-projects*. GitHub. https://github.com/proAIrokibul/Environmental-and-socio-economic-impact-assessment-of-renewable-energy-projects/blob/main/Model/main_file.ipynb

[14] Nwokolo, S. C., Obiwulu, A. U., & Ogbulezie, J. C. (2023). Machine learning and analytical model hybridization to assess the impact of climate change on solar PV energy production. *Physics and Chemistry of the Earth, Parts A/B/C*, *130*, 103389.

[15] Roy, D., Zhu, S., Wang, R., Mondal, P., Ling-Chin, J., & Roskilly, A. P. (2024). Techno-economic and environmental analyses of hybrid renewable energy systems for a remote location employing machine learning models. Applied Energy, 361, 122884.

[16] Shawon, R. E. R., Rahman, A., Islam, M. R., Debnath, P., Sumon, M. F. I., Khan, M. A., & Miah, M. N. I. (2024). AI-Driven Predictive Modeling of US Economic Trends: Insights and Innovations. *Journal of Humanities and Social Sciences Studies, 6*(10), 01-15.

[17] Yağmur, A., Kayakuş, M., & Terzioğlu, M. (2023). Predicting renewable energy production by machine learning methods: The case of Turkey. *Environmental Progress & Sustainable Energy*, *42*(3), e14077

[18] Zeeshan, M. A. F., Sumsuzoha, M., Chowdhury, F. R., Buiya, M. R., Mohaimin, M. R., Pant, L., & Shawon, R. E. R. (2024). Artificial Intelligence in Socioeconomic Research: Identifying Key Drivers of Unemployment Inequality in the US. *Journal of Economics, Finance and Accounting Studies, 6*(5), 54-65.