IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

ANALYSIS OF SECTORAL OUTPUT AND ENERGY RELATED CO_2 EMISSION TREND BASED ON STATISTICAL ANALYSIS AND MACHINE LEARNING

[1]**Joshua Peter Inyang**
[1]University of Cross River State, Calabar, Nigeria.

Abstract : This study examines the impact of industrialization and energy-associated CO_2 emission drivers in the Nigerian manufacturing sub-sector from 2000 to 2020. The *Logarithmatic Mean Divisia Index (LMDI)* method was employed to decomposed the key driving factors, including labour effect, income effect, energy intensity, and energy structure effect. Additionally, a *stacking regression* was implemented, combining *linear regression* and *Huber regression* with a final *linear regression estimator* to develop a machine learning model for forcasting CO_2 emissions based on historical and real-time data. The results indicate that among the manufacturing subsectors, the *Chemical and Pharmaceutical (CHPM) industry* was the largest contributor to CO_2 emissions, accounting for *13.67% of total energy consumption* during the study period. This trend is further exacerbated by Nigeria's heavy reliance on fossil fuels. While the CHPM subsector exhibited the highest emissions, other subsectors also contributed significantly, with emission-driving factors flunctuating across industries. The machine learning model forecasts a continued *increase in CO_2 emission over the next 30years*, emphasizing the urgent need for sustianable energy policies and industrial strategies to mitigate enviromental impact.

1. INTRODUCTION

Global warming and climate change are driven by rising greenhouse gas (GHG) emissions, industrialization, and reliance on carbon-intensive energy. The industrial sector is a major contributor to CO_2 emissions, necessitating efficient energy demand modeling. Machine learning and artificial intelligence provide data-driven solutions to analyze energy patterns, improve efficiency, and mitigate emissions. This research evaluates statistical and machine learning models to forecast CO_2emissions from Nigeria's manufacturing subsectors, aiming to develop targeted mitigation strategies for sustainable industrial operations. Rising CO_2 emissions, driven by fossil fuel dependence in industrial activities, pose a major global challenge. A key issue is the lack of firm-level emissions data, hindering precise carbon management. While extensive research exists on macro-level energy efficiency and GHG emissions, sector-specific studies in developing countries like Nigeria remain scarce. Existing studies have analyzed CO_2emission drivers but lack advanced statistical and machine learning models for deeper insights and forecasting. This research integrates statistical and machine learning techniques to analyze emission drivers and predict future CO_2 emissions in Nigeria's industrial subsectors, enabling targeted mitigation strategies for sustainable development.The aim of this study involves industrial energy demand modelling of manufacturing output and CO_2 emission change based on statistical analysis and machine learning**.** The specific objectives include analyze firm-specific and sectoral data to determine the primary factors influencing CO2 emissions, generate forecasts of CO2 emissions based on historical data and predictive models, providing insights into future emission trends and potential areas for intervention and provide recommendations for policymakers and industry stakeholders to improve regulatory frameworks and industrial practices, leveraging insights from the research to support sustainable development and effective climate change mitigation.

This research contributes to the existing literature by proposing an innovative approach for modelling industrial energy demand that considers the complex interactions between manufacturing sub-sector output, climate change factors, and energy consumption dynamics [1][2]. By leveraging advanced computational techniques such as machine learning and statistical modelling and utilizing comprehensive datasets, this study aims to enhance the accuracy and reliability of energy demand forecasts in the industrial sector [3][4]. By addressing the limitations of conventional energy demand forecasting methods and incorporating a multidimensional analysis approach, this study aims to bridge the gap between theory and practice in industrial energy management [3][4]. The findings and recommendations from this research can inform evidence-based decision-making and policy formulation, ultimately contributing to more sustainable and resilient industrial energy systems [5][6]. The industrial sector stands as a crucial player in the global energy landscape, contributing significantly to both energy consumption and greenhouse gas emissions [7][8]. The study utilized historical data on industrial activities, climate variables, and energy consumption patterns to develop predictive models that captured the dynamic relationships between these variables. The results demonstrated the efficacy of ML techniques in accurately predicting industrial energy demand under different scenarios, highlighting the importance of considering both manufacturing sub-sector output and climate change impacts in energy demand modeling. Similarly, research by [9] utilized statistical analysis to model industrial energy demand in the context of climate change mitigation strategies . The study integrated comprehensive datasets on industrial activities, energy consumption patterns, and climate change variables to develop ML models that could forecast energy demand dynamics with high accuracy. The findings underscored the significance of incorporating climate change factors into energy demand modeling frameworks to support informed decision-making and policy development aimed at achieving sustainability goals.These data-driven approaches offer valuable insights into the complex interactions between industrial activities, energy consumption patterns, and environmental factors, facilitating the development of effective strategies for energy management and climate change mitigation in the industrial sector. Table 2.1 shows a list of some related work in this field, this tends to pinpoint the contribution and research path that some individuals have ventured into and the method used.Table 2.1, below provides insights into various studies focusing on CO_2 emissions, encompassing different methodologies such as statistical analysis and machine learning. However, it is noteworthy that none of the listed studies specifically address the context of CO_2 emission modelling using both machine learning and statistical analysis within Nigeria. This gap presents an opportunity for us to dive into this area of concern and conduct an in-depth exploration tailored to the Nigerian context. By combining the strengths of machine learning algorithms and statistical methods, we aim to develop a comprehensive understanding of CO_2 emission patterns in Nigeria, contributing valuable insights to the field of environmental sustainability and policy formulation within the country.

Table 2.1: list of previous industrial-related studies

Reference	Method Used	Country/Sector	CO_2 Influence	Remark
[10]	Extended Logarithmic Mean Divisia Index (LMDI) decomposition model	China's manufacturing industry data	- Industrial activity effects - Energy intensity - Investment intensity - R&D efficiency - R&D intensity	- Different driving factors in different periods - Related to economic development and government policies
[11]	LMDI decomposition method	Turkish manufacturing industry data	- Main factors for CO_2 changes	- Changes in total industrial activity and energy intensity
[12]	Statistical Analysis	Manufacturing Output Data	impact of output changes on CO_2	Utilized migration models for Analysis
[13]	Machine Learning	Energy Consumption	predicted CO_2 trends based on energy data	Applied neural network models for predictions
[14]	Statistical analysis	Production Efficiency data	Relationship between efficiency and CO_2	Examined correlation between efficiency and CO_2
[15]	Machine Learning	Sensor Data	Predicted CO2 levels based on sensors	Employed decision tree algorithms for prediction
[16]	Statistical Analysis	Industry Reports	Impact on industrial change on CO2	Analyzed sector-wise CO_2 variations
[17]	Machine Learning	Energy Usage Patterns	predicted CO2 based on energy patterns	Used clustering algorithms for pattern recognition

2. **METHODOLOGY**

2.1 Data Collection Procedure

The study utilize published data of registered manufacturing firm's/enterprise annual financial reports listed in the directory of the Manufacturing Association of Nigeria (MAN) and National Bureau of Statistics (NBS), between 2000-2020. Data obtained in these directories is expected to account for the labour force (man-hours worked), financial factors (sales, assets, equity and cost of goods sold (COGS)), energy consumption, and production output. Sectoral output will be measured in monetary terms as value-added in million Naira. While Firms output will be measured as sales quoted in billion Naira. Other firm's characteristics considered include: Labour Force: The total man-hours worked; Financial Factors: Sales, assets, equity, and cost of goods sold (COGS); Energy Consumption: The amount of energy utilized by the firms; Production Output: Measured in monetary terms as value-added in million Naira, and firm output measured as sales quoted in billion Naira. The total annual energy consumption that denotes the quantity of energy expended or utilized annually in GJ was estimated using Eq. (2.1) at firm level.

2.2 The calculation method for carbon emission

The study adopted three categories of energy consumption in the various manufacturing firms: electricity, fuel-oil, and diesel. $Ei = E_C + F_C + D_C$ (2.1) Where E_C is the cost of electricity consumed by the firm, F_C is the cost of fuel-oil consumed by the firm and D_C is the cost of diesel consumed by the firm. According to the IPCC Guidelines for National Greenhouse Gas Emission Inventory, this study uses energy consumption data and emission coefficients to calculate carbon dioxide emissions of industrial manufacturing firms [18]. As shown in eqn. (2.2): $\mathrm{C} = \sum_{i}^{8}\sum_{j}^{4}\mathrm{C_{ij}} = \sum_{i}^{8}\sum_{j}^{4}\mathrm{K_{ij}} \times \mathrm{f_i}$ (2.2) where C denotes the CO_2 emissions of energy consumption, *i* denotes the eight different sub-industries; *j* denotes the fuel types of energy consumption; $\mathrm{K_{ij}}$ denotes the consumption of energy j by sub-industries i; $\mathrm{EF_j}$ denotes the CO_2 emissions coefficients of j fuel types, and it is estimated according to the carbon content, carbon oxidation rate, and net calorific value provided by IPCC. The reference for the calculation of the CO_2 emission coefficient of electricity is shown in Eq. (2), and EFF describes the electricity emission factor (7.88 tCO_2/104 kWh) [20]. $\mathrm{C_{EL}} = \mathrm{C_{Ec}} \times \mathrm{EFF}$ (2.3). According to the expanded Kaya identity and LMDI model (Ang, 2015), the total industrial firms CO_2 emissions of energy consumption can be expressed based on Eq. (2.4):

$$C = \sum_i\sum_j \mathrm{C_{ij}} = \sum_i\sum_j \frac{\mathrm{C_{ij}}}{\mathrm{E_{ij}}} \times \frac{\mathrm{E_{ij}}}{\mathrm{E_i}} \times \frac{\mathrm{E_i}}{\mathrm{G_i}} \times \frac{\mathrm{G_i}}{\mathrm{LA_i}} \times \mathrm{LA_i} = \sum_i\sum_j \mathrm{EF_{ij}} \times \mathrm{ES_{ij}} \times \mathrm{E_i} \times \mathrm{IVA_i} \times \mathrm{LA_i} \quad (2.4)$$

where $\mathrm{EF_{ij}}$ represent the emission factor, $\mathrm{ES_{ij}}$ denotes energy structure, $\mathrm{E_i}$ is the energy intensity, $\mathrm{IVA_i}$ represent the income effects and $\mathrm{LA_i}$ is the labour effects respectively. To quantitatively analyze the influencing factors in the total CO_2 emissions, the additive LMDI method was used to decompose Eq. (3.4). The changes in total CO_2 emissions from the base year (C^0) to the target year (C^t) can be decomposed as the following five effects: $\Delta\mathrm{C^T} = \Delta\mathrm{C^t} - \Delta\mathrm{C^0} = \Delta\mathrm{C^{EF}} + \Delta\mathrm{C^{ES}} + \Delta\mathrm{C^{EI}} + \Delta\mathrm{C^{IVA}} + \Delta\mathrm{C^{LA}}$(2.5) where $\Delta\mathrm{C^{EF}}, \Delta\mathrm{C^{ES}}, \Delta\mathrm{C^{EI}}, \Delta\mathrm{C^{IVA}}$ and $\Delta\mathrm{C^{LA}}$ denote the effects of emission coefficient, energy structure, energy intensity, per capita income, and labour, respectively.The above-mentioned five effects can be written as follow$\Delta\mathrm{C^{EF}} \sum_i\sum_j \frac{\mathrm{C_{ij}^t - C_{ij}^o}}{\mathrm{InC_{ij}^t - inC_{ij}^o}} \mathrm{In}\frac{\mathrm{EF_{ij}^t}}{\mathrm{EF_{ij}^0}}$ (3.6) $\Delta\mathrm{C^{ES}} \sum_i\sum_j \frac{\mathrm{C_{ij}^t - C_{ij}^o}}{\mathrm{InC_{ij}^t - inC_{ij}^o}} \mathrm{In}\frac{\mathrm{ES_{ij}^t}}{\mathrm{ES_{ij}^0}}$ (2.7) $\Delta\mathrm{C^{EI}} \sum_i\sum_j \frac{\mathrm{C_{ij}^t - C_{ij}^o}}{\mathrm{InC_{ij}^t - inC_{ij}^o}} \times \mathrm{In}\frac{\mathrm{EI_i^t}}{\mathrm{EF_i^0}}$ (2.8) $\Delta\mathrm{C^{IVA}} \sum_i\sum_j \frac{\mathrm{C_{ij}^t - C_{ij}^o}}{\mathrm{InC_{ij}^t - inC_{ij}^o}} \mathrm{In}\frac{\mathrm{IVA_i^t}}{\mathrm{IVA_i^0}}$ (2.9) $\Delta\mathrm{C^{LA}} \sum_i\sum_j \frac{\mathrm{C_{ij}^t - C_{ij}^o}}{\mathrm{InC_{ij}^t - inC_{ij}^o}} \mathrm{In}\frac{\mathrm{LA_i^t}}{\mathrm{LA_i^0}}$ (2.10)

$$\mathrm{L}\,(\mathrm{C_{ij}^t}, \mathrm{C_{ij}^0}) = \begin{cases} \frac{\mathrm{C_{ij}^t - C_{ij}^o}}{\mathrm{InC_{ij}^t - inC_{ij}^o}}, & \mathrm{C_{ij}^t} \neq \mathrm{C_{ij}^0} \\ \mathrm{C_{ij}^t} \text{ or } \mathrm{C_{ij}^0}, & \mathrm{C_{ij}^t} = \mathrm{C_{ij}^0} \end{cases} \qquad (2.11)$$

The CO_2 emissions coefficients of various fuels are all assumed to be fixed when calculating industrial firm's CO_2 emissions, which has trivial or no contributions at all to the changes in CO_2 emissions. That is to say, the $\Delta\mathrm{C^{EF}}$ in Equation (4) should be 0 [19].

The performance of our prediction model is evaluated based on the following models:

$$\text{Root Mean Squared Error (RMSE)} = \sqrt{\tfrac{1}{n}\textstyle\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (2.12)$$

$$\text{Mean Absolute Error (MAE)} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (2.13)$$

$$\mathrm{R-squared}\ (\mathrm{R}^2) = \ 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \qquad (2.14)$$

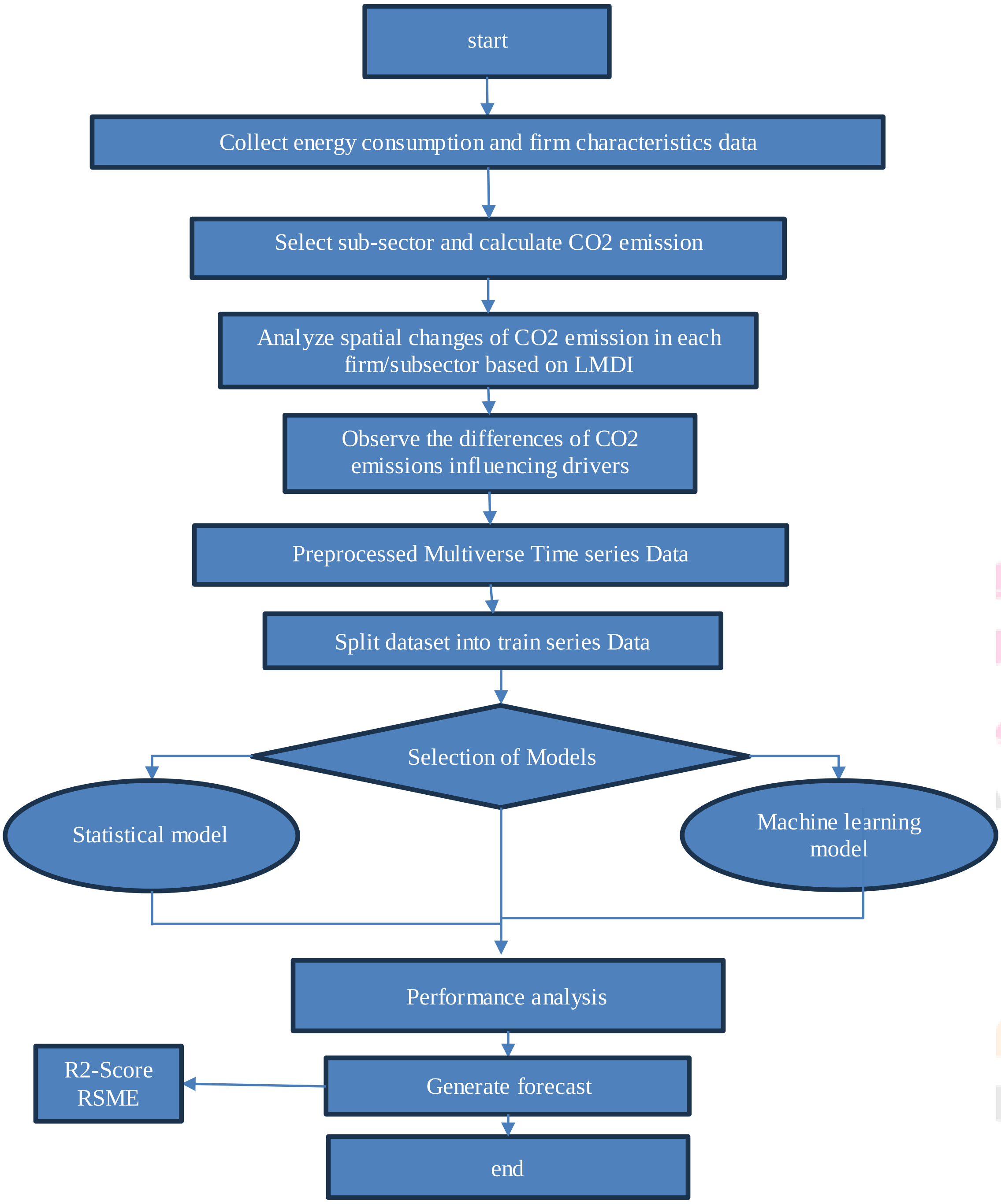


Fig 2.1 Flow Chart

3. **RESULTS AND DISCUSSION**

3.1 Characteristics of CO_2 emissions

3.1.1 Carbon emissions from different sub-industries

From figure 3.1a, the industry that produced the most CO_2 emissions overall was the chemical and pharmaceutical (CHPM) sector. This was followed by the other manufacturing (OTM), non-metallic products (NMP), textile and footwear (TAF), hotels and catering services (HCS), food and beverage (FB), livestock, and agro-allied (LA) sectors. This was due to the fact that CHPM energy consumption made up the greatest portion of the total energy consumption, which was projected to be 13.67% between 2000 and 2020. The CO_2 emissions (share) of OTM increased from 1.81×10^9 $MJCO_2e$ (4.60%) in 2000 to 1.95×10^9 $MJCO_2e$ (4.60%) in 2020.

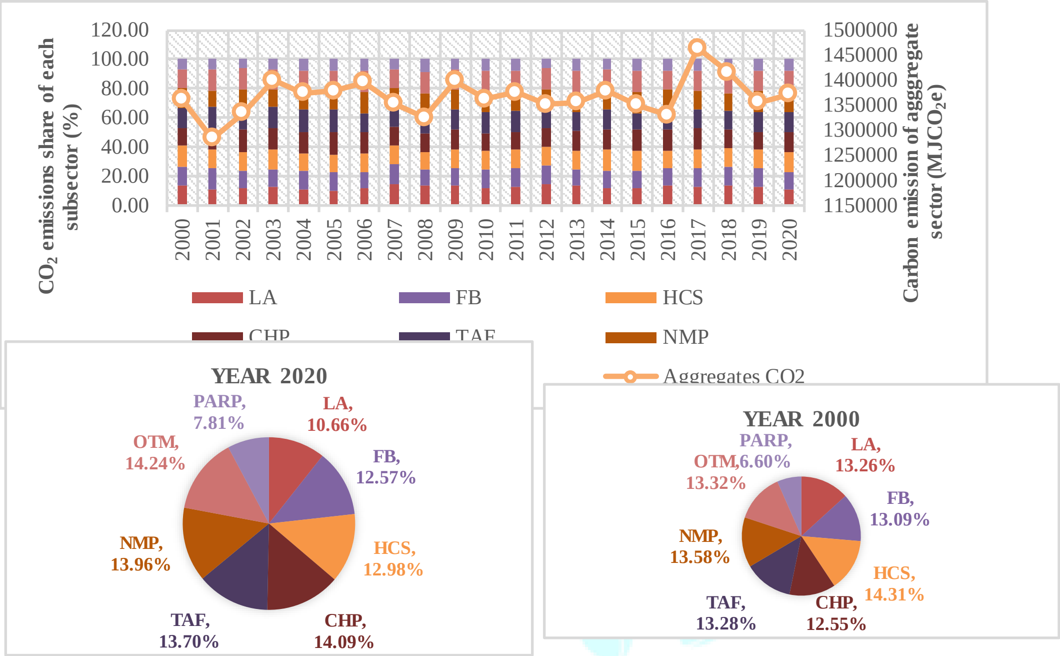
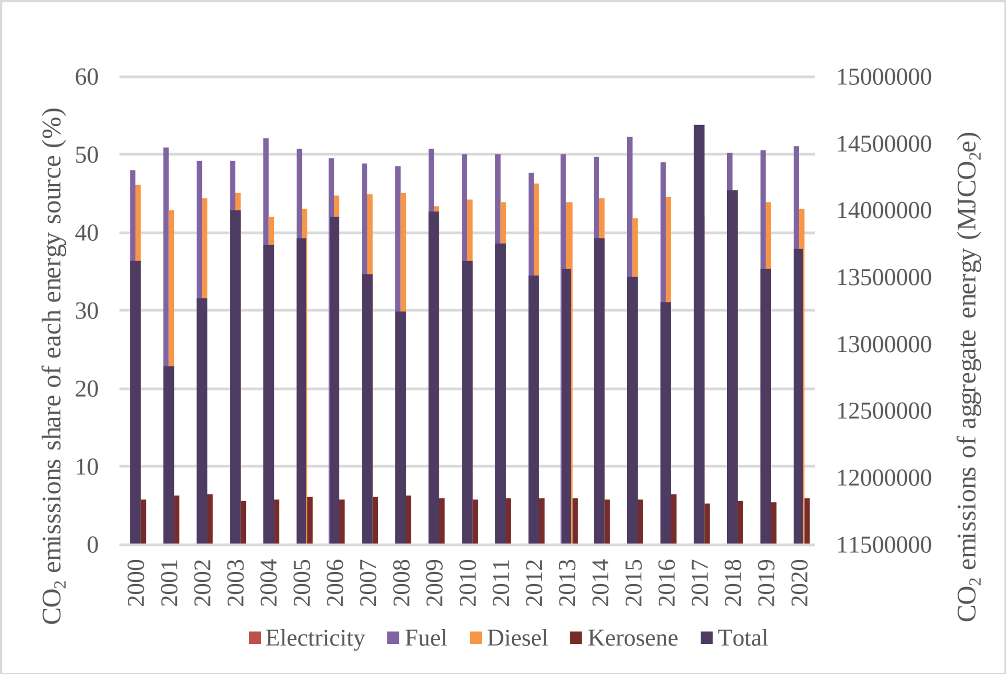
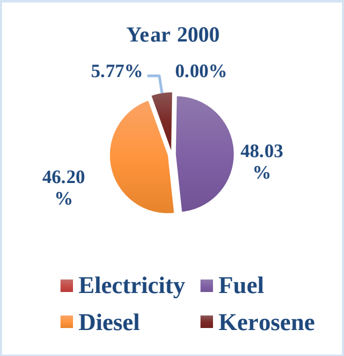


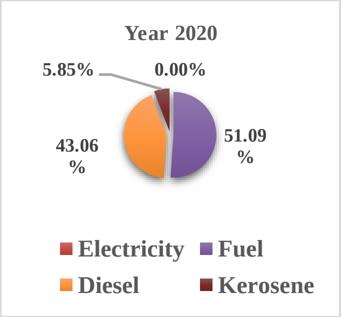
Fig. 3.1a. Change in sub-industrial CO_2 emissions and percentage (piechart) from 2000 to 2020.

PARP's CO_2 emissions increased significantly from 8.99 × 10^8 $MJCO_2e$ (4.20%) in 2000 to 1.07 × 10^9 $MJCO_2e$ (5.01%) in 2020, with an average annual growth rate of 19.12%, the highest among sub-industries.In contrast, ELE's emissions declined from 1.95 × 10^9 $MJCO_2e$ (5.30%) to 1.78 × 10^9 $MJCO_2e$ (4.84%), experiencing the second-largest decrease at -8.66% annually. LA recorded the largest drop, reducing from 1.81 × 10^9 $MJCO_2e$ (5.01%) to 1.46 × 10^9 $MJCO_2e$ (4.06%), with an average annual decline of -19.00%. FB's emissions saw a slight decrease from 1.78 × 10^9 $MJCO_2e$ (5.03%) to 1.72 × 10^9 $MJCO_2e$ over the same period.(4.86%) during 2000-2020 (Fig. 4.1a), which revealed that FB's CO_2 As shown in Fig 3.1b, the fuel oil (PMS) generated the highest amount of CO_2 emissions between 2000 and 2020, with an average of 1.44 x 1011 $MJCO_2e$, as indicated in Fig. 3.1b. Subsequently, 1.26 × 1011 MJCO2e was produced by diesel (AGO) and household kerosene (HHK), respectively. During the same time frame, the CO_2 emissions from the production of power were negligible. Additionally, Fig. 3.1b shows that in 2020, the percentage of diesel (AGO) dropped from 46.20% in 2000 to 43.06%. This data showed that Nigeria's manufacturing sectors' energy consumption structure had improved.According to Figure 3.1b the fuel oil (PMS) portion of CO_2 emissions grew from 6.54 x 109 $MJCO_2e$ (48.03%) in 2000 to 7.01 × 109 $MJCO_2e$ (51.09%) in 2020. It overtook diesel (AGO) as the primary source of CO_2 emissions, HHK produced the least CO_2 emissions and the smallest increment between 2000 and 2020, with its CO_2 emissions (ratio) rising from 7.86 × 108 $MJCO_2e$

(5.77%) to 8.02 × 108 MJCO2e (5.85%) (Fig. 3.1b). Fig.3.1b.During the same assessment period,







electricity's CO_2 emissions were not negligible.

3.2. Drivers' decomposition of CO_2 emissions from a multi-sectoral standpoint

Table 3.1 displays the cumulative contribution rate of each effect across the six sub-industries using the established LMDI. However, to investigate the precise causes of the variations in CO_2 emissions across the various sub-industries, the CO_2 emissions decomposition findings from 2000 to 2020 were computed and are displayed in Fig. 3.3.

Table 3.1

Cumulative contribution rates (%) of CO_2 emissions from each sub-industry during 2000-2020.

Sub-industries	LA effect	IN effect	EI effect	ES effect	CEC effect
Livestock/Agro-allied	-68.76	490.30	-509.40	-12.14	0.00
Food & Beverages	-70.15	62.58	-65.82	-26.61	0.00
Hotels & catering services	56.76	-557.01	422.01	-21.76	0.00
Chemical & Pharmaceutical	22.17	-540.62	612.34	6.11	0.00
Textile & Footwear	79.03	-1108.67	1127.82	1.81	0.00
Non-metallic Products	-11.81	-631.76	729.44	14.13	0.00
Other manufacturing	-18.62	-495.83	621.64	-7.19	0.00
Pulp & Paper Production	-13.98	-136.03	238.15	11.86	0.00
Total	**-141.25**	**-2638.51**	**2942.16**	**-62.39**	**0.00**

3.3. Drivers' decomposition of energy-related CO_2 emissions from a multi-stage standpoint.

Fig. 3.3a presents the year-by-year decomposition results based on the established LMDI model. However, Fig. 3.12 displays the breakdown findings for each level. Table 3.2 displays the cumulative contribution rates of each influence at various phases. It is evident that while energy intensity demonstrated the driving impacts, the effects of energy labor, income, and energy structure reduced CO_2 emissions. Between 2000 and 2020, the carbon emission coefficient had negligible impact on changes in carbon emissions.

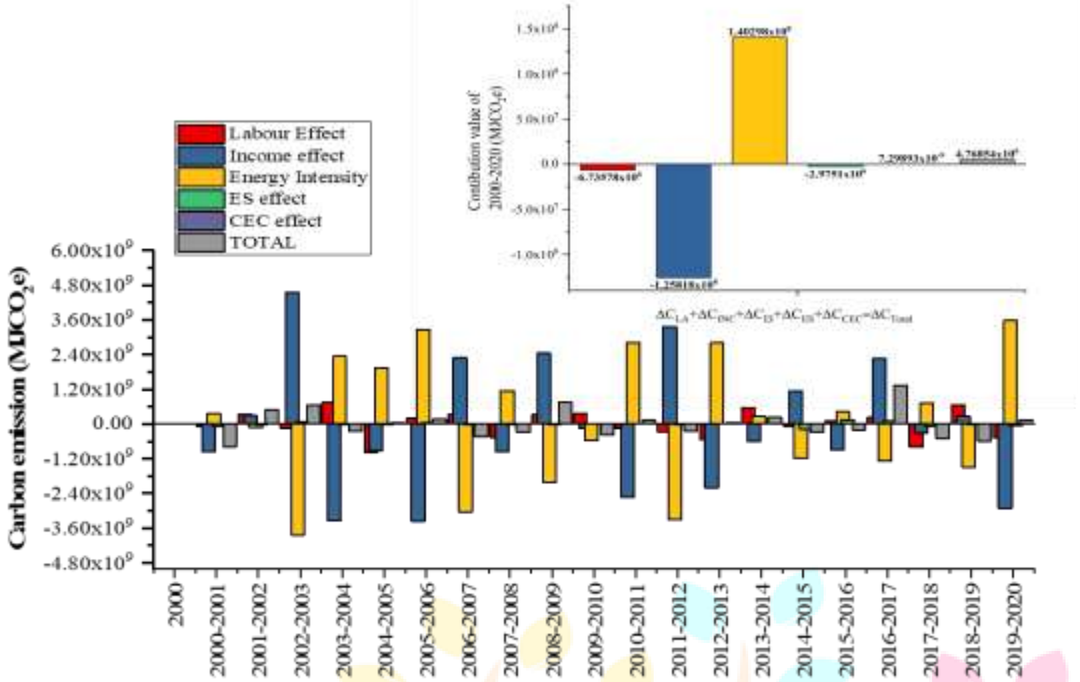


Fig. 3.3a. The Year-by-year decomposition results of CO_2 emissions from the aggregate manufacturing sector during 2000-2020.

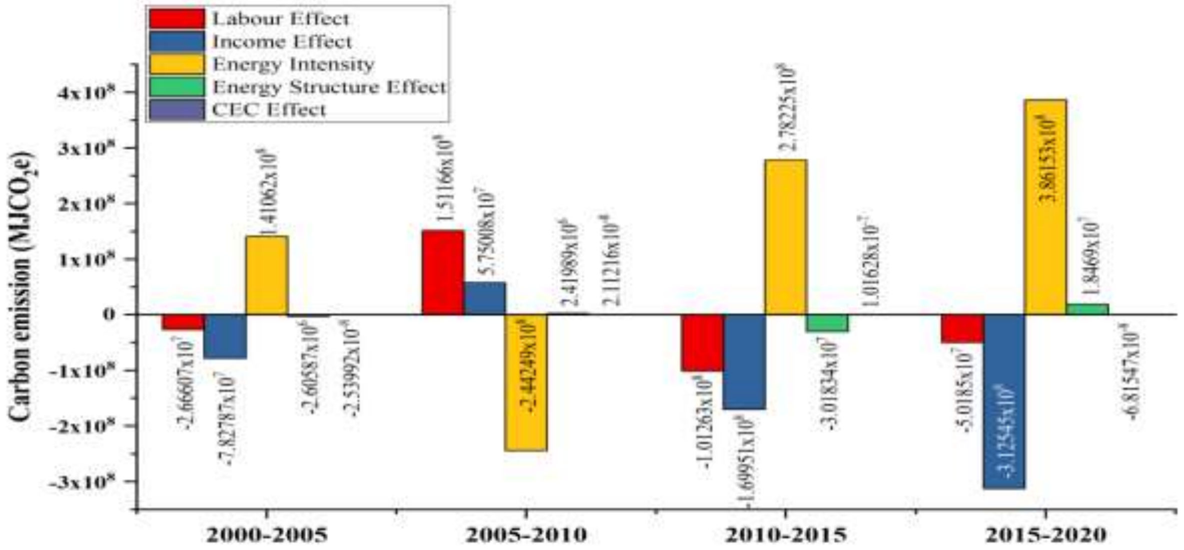


Fig. 3.3b. Changes in total CO_2 emissions caused by the five drivers at different stages

Table 3.2

Cumulative contribution rates (%) of CO_2 emissions change at different stages during 2000-**2020.**

Stages	**Labour Effect**	**Income effect**	**Energy Intensity**	**ES effect**	**CEC effect**
2000-2005	-79.54	-233.55	420.87	-7.77	0.00
2005-2010	455.84	173.39	-736.53	7.30	0.00
2010-2015	-436.99	733.39	1200.63	-130.25	0.00
2015-2020	-119.79	-746.06	921.77	44.09	0.00
2000-2020	-141.25	-2638.51	2942.16	-62.39	0.00

3.3.1. Income effect

From 2000 to 2020, the income effect's total contribution value was -1.25818× 108 MJCO2e (Fig. 3.3c and Table 3.2). The primary cause of the ongoing reduction in carbon emissions is the ongoing loss in industrial value-added. This is the underlying explanation. The key element contributing to the decrease in CO_2 emissions from industrial enterprises in Nigeria between 2000 and 2020 was the cumulative contribution rate, which was -2638.51%. The economic growth of industrial firms decreased at an average rate of -3.35% over the course of the study period (see Fig. 3.3c).

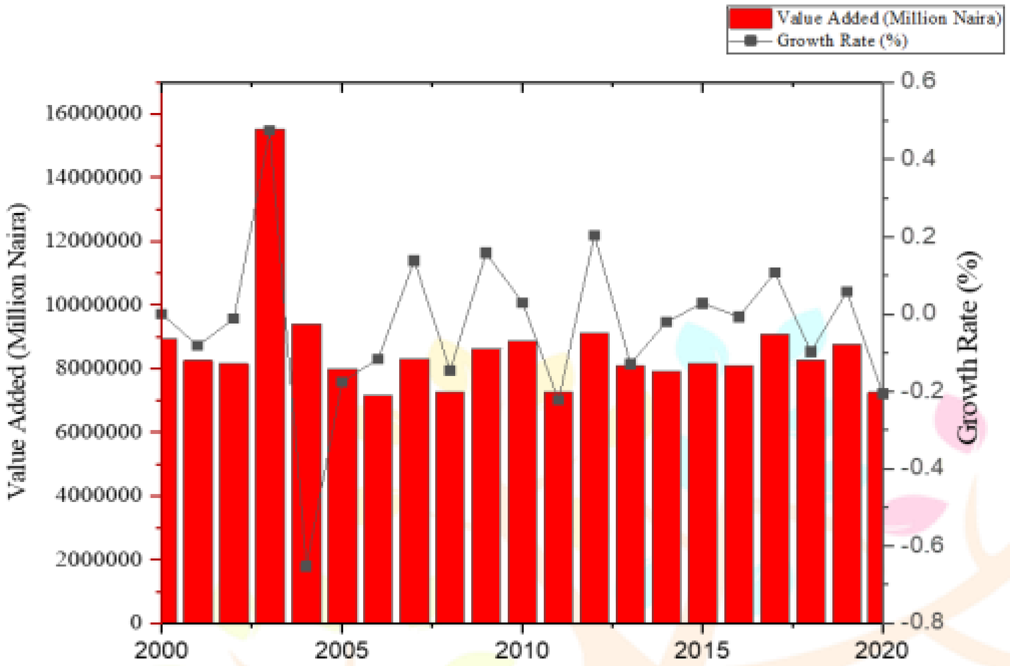


Fig. 4.3c Share of industrial firm's value-added in Nigeria's manufacturing sector.

For each of the four stages, the cumulative contribution values (rates) of the income effect were -7.8278 × 10^7 MJCO$_2$e (-233.55%), 5.75008 × 10^7 MJCO$_2$e (173.39%), -1.69951 × 10^8 MJCO$_2$e (733.39%) and -3.12545 × 10^8 MJCO$_2$e (-746.06%), respectively), which showed that the income effect was the leading factor inhibiting CO_2 emissions.

3.3.2. Labour effect

The second contributing component to the reduction of CO_2 emissions by industrial enterprises over the study period is the man-hours worked, which is expressed as the labor effect. Between 2000 and 2020, the labor effect's total contribution value (rate) was -6.73578 x 108 MJCO2e, or -141.25% (Table 3.2). The labor effect exhibited a fluctuating tendency, with an average decrease of -0.19% from 69,945 hours in 2000 to 68,453 hours in 2020. -2.66607 × 107 MJCO2e (-79.54%), 5.75008 × 107 MJCO2e (455.84%), -1.69951 × 108 MJCO2e (-436.99%), and -3.12545 × 108 MJCO2e (-119.79%) were the cumulative contribution values (rates) from the first to fourth stages, respectively. The first factor that could be causing the fluctuating pattern is the reduction in working hours, especially in the third and fourth stages when there was a mandatory lockdown because of the COVID-19 epidemic. In this period, economic activity was suppressed, which led to capacity underutilization as most businesses reduced staff and began operating remotely.

3.3.3. Energy structure Carbon emission coefficient effect

Between 2000 and 2020, the energy structure effect's total contribution value (rate) was -2.9751 x 108 MJCO2e, or -62.39 percent. The fuel percentage dropped from 46.1% in the first stage to 45.9% in the second stage, and then climbed from 46% in the third stage to 46.9% in the fourth stage, as shown in Fig. 4.14. On the other hand, the shares of kerosene and diesel dropped from 38.2% and 5.3% in the first stage to 37.6% and 5.1% in the fourth stage, respectively, while the part of electricity increased from 10.4% in the first stage to 10.5% in the fourth stage and was maintained by 10.5% in the future stages. In the four stages, the cumulative contribution values (rates) of energy structure were -2.60587 × 10^6 $MJCO_2e$ (-7.77%), 2.41989 × 10^6 $MJCO_2e$ (7.30%), -3.01834 × 10^7 $MJCO_2e$ (-130.25%) and 1.8469 × 10^7 $MJCO_2e$ (44.09%), respectively. Thus, it can be said that although the energy structure has been somewhat optimized, it has typically shown to have a negligible restraining influence on the increase in CO_2 emissions between 2000 and 2020.

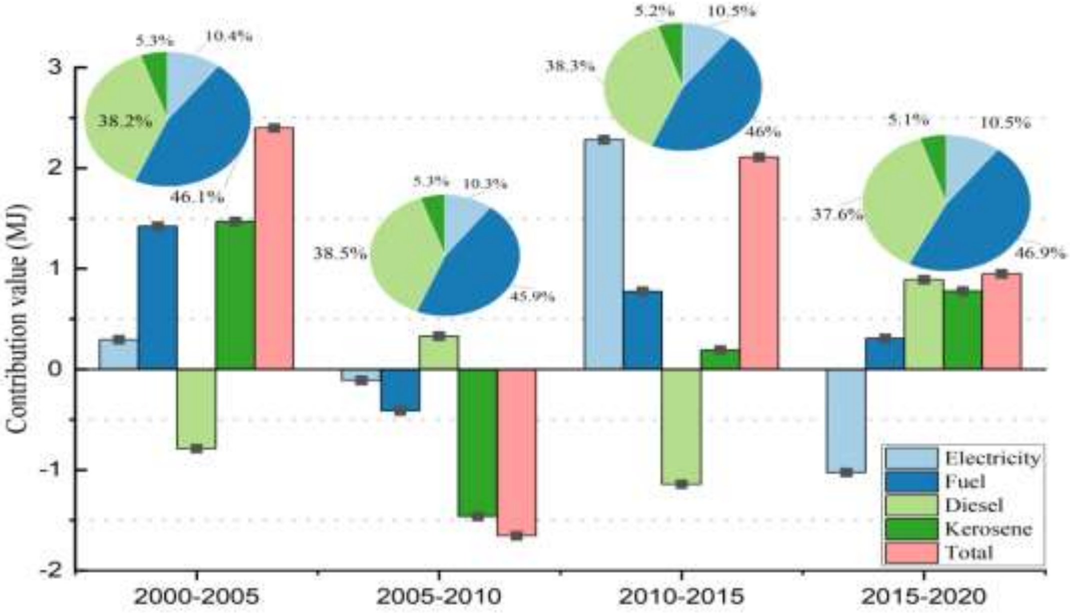


Fig. 3.3d Energy structure at each stage (the pie chart) and contribution values of different energies at different stages.

Considering Energy intensity effect From the perspective of the entire study period, energy intensity (EI) was the most significant driving factor of CO_2 emissions in the industrial firms. The EI effect annually increased by 24.56%, from 0.06 MJ/Naira in 2000 to 0.07 MJ/Naira in 2020 (see Fig. 4.3e). This remarkable change substantially increased CO_2 emissions by 1.40298 × 10^8 $MJCO_2e$ with a cumulative contribution rate of 2942.16% (Fig. 4.3e and Table 3).

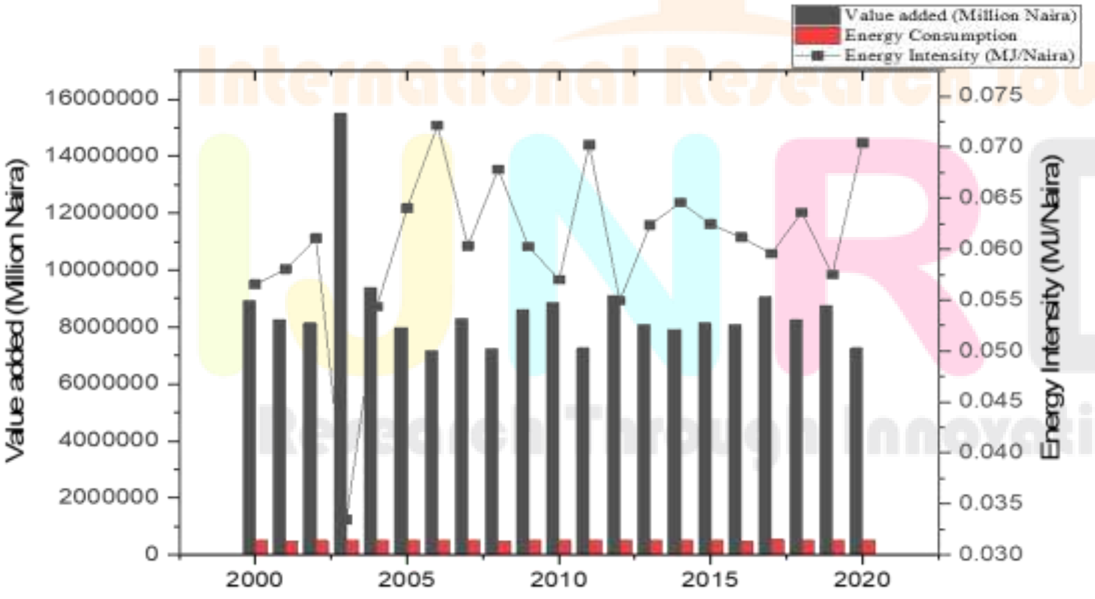


Fig. 3.3e. Energy intensity of the aggregate industrial firms during 2000-2020.

When looking at the sub-periods, energy intensity only had an impact on CO_2 emissions in the second stage, contributing a cumulative value (rate) of -2.44249× 108 MJCO2e (-736.53%). Energy intensity raised CO2 emissions by 1.41062 × 108 MJCO2e, 2.78225 × 108 MJCO2e, and 3.846153 × 108 MJCO2e in the first, third, and fourth stages, respectively, with cumulative rates of 420.87%, 1200.63%, and 921.77%. This exception is strongly linked to the rise in energy intensity that occurred at that time in the textile and footwear sub-industries, as well as in the chemical and pharmaceutical industries. The energy intensity of the chemical and pharmaceutical subindustry increased emission by 6.80319 × 107 $MJCO_2e$ (612.34%) between 2000 and 2020. From the first to the fourth stage, the corresponding values were 9.12729 × 107 $MJCO_2e$, -1.38764 × 108 $MJCO_2e$, 1.54844 × 108 $MJCO_2e$, and 1.64774 × 108 $MJCO_2e$, respectively. The driving effect of energy intensity in the textile and footwear sub-industry increased CO_2 emission in the first stage and peaked in the fourth stage, contributing 2.61048 × 108 $MJCO_2e$ and 7.26309 × 107 $MJCO_2e$, respectively.

3.4 Data Preprocessing and Feature Selection for machine learning

In this study, the dataset comprises historical records of CO_2 emissions, energy consumption, and economic value added across various sub-sectoral industries. To prepare this data for analysis, several preprocessing steps were undertaken:

Conversion to DateTime Format: The 'YEAR' column in the dataset was converted to a datetime format and designated as the index. This transformation allows for more effective time series analysis and ensures that the data is organized chronologically.

- Handling Missing Values: To address any gaps in the dataset, missing values were managed using interpolation methods or by filling them with mean values. This approach helps maintain the integrity of the dataset and ensures that subsequent analyses are based on complete information.
- Feature Standardization: The features within the dataset were standardized to have a mean of 0 and a standard deviation of 1. Standardization is crucial for machine learning models as it ensures that all features contribute equally to the model's performance, preventing any single feature from disproportionately influencing the results due to its scale.

These preprocessing steps are essential for preparing the data for machine learning algorithms, ensuring that the dataset is clean, consistent, and ready for model training and evaluation. To determine the most pertinent features for the prediction model, Recursive Feature Elimination (RFE) was utilized. RFE is an iterative method that removes the least significant features based on their weights from a linear regression estimator, thereby identifying the most influential variables for model performance.

The final selected features for the prediction model were:

- Total Energy (MJ)
- Value Added

The coefficients for these features were as follows:

Total Energy (MJ): This coefficient indicates a positive relationship between total energy consumption and CO_2 emissions. Specifically, for each unit increase in total energy consumption, CO_2 emissions increase by 30,768.63 units, highlighting that higher energy usage is associated with greater CO_2 emissions; Value Added: This coefficient shows a negative relationship between economic value added and CO_2 emissions. For each unit increase in value added, CO_2 emissions decrease by 0.69 units. This suggests that as the economic value added by the sector increases, CO_2 emissions tend to decrease.To understand the percentage impact of these coefficients on CO_2 emissions, consider the following calculation for the impact of a 1% change. A 1% increase in total energy leads to a 30,768.63-unit increase in CO_2 emissions. If we assume an initial total energy consumption of 1,000,000 MJ, a 1% increase would be 10,000 MJ. Therefore, the corresponding increase in CO_2 emissions would be 10,000 × 30,768.63 = 307,686,300 units. As a percentage of the original emissions, this represents a substantial increase, highlighting the significant impact of energy consumption on CO_2 emissions. A 1% increase in value added results in a 0.69 unit decrease in CO_2 emissions. If the initial value added is 100,000 units, a 1% increase would be 1,000 units. Thus, the decrease in CO_2 emissions would be 1,000 × 0.69 = 690 units. This percentage decrease reflects a relatively smaller but still notable reduction in emissions as economic value-added increases. These coefficients provide insight into how changes in energy consumption and economic value added can influence CO_2 emissions, aiding in the development of targeted mitigation strategies.

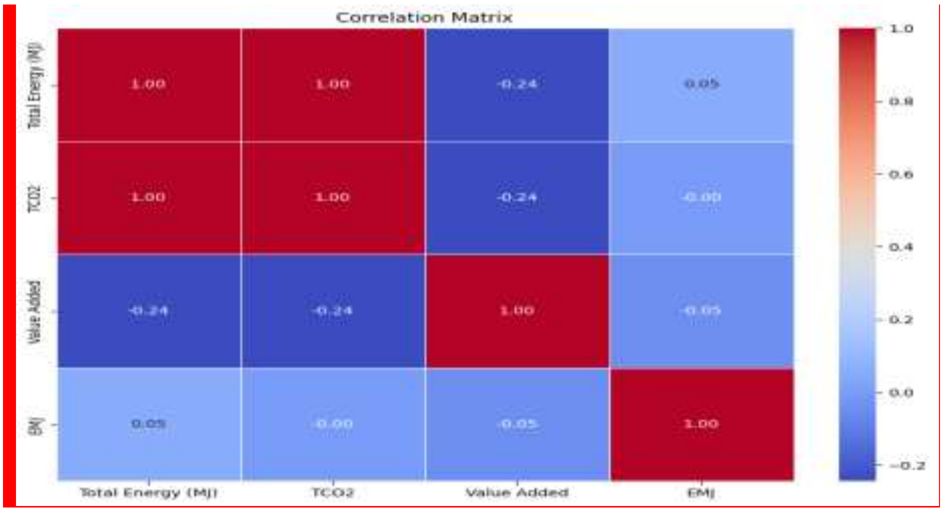
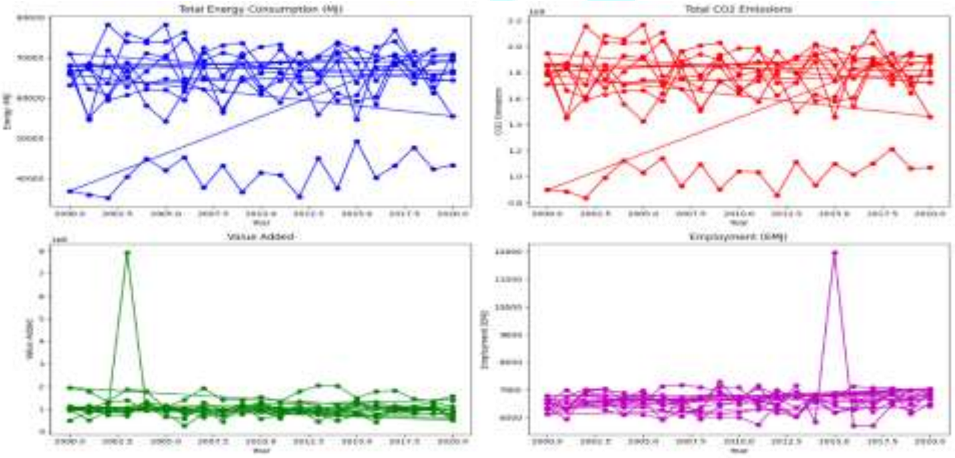


Fig 3.6 Correlation Matrix



In developing the machine learning models for predicting CO_2 emissions, two base models were selected to form the foundation of the stacking regressor:

Linear Regression: This model applies a straightforward linear approach to establish a relationship between the selected features (Total Energy and Value Added) and the target variable (CO_2 emissions). Linear Regression assumes a linear relationship between the dependent and independent variables, which makes it simple yet effective for modelling the data when the relationship is approximately linear.Huber Regressor: This robust regression technique is designed to be less sensitive to outliers compared to traditional linear regression methods. The Huber Regressor combines the properties of both linear regression and robust methods. It uses a linear loss for residuals smaller than a specified threshold and a quadratic loss for larger residuals. This characteristic helps to mitigate the impact of outliers and provides a more reliable estimate of the model parameters. Stacking Regressor: To enhance the predictive performance of the models, a stacking regressor was implemented by combining the two base models—linear regression and Huber regression—with a final linear regression estimator. This ensemble technique leverages the strengths of each base model to improve the overall forecast accuracy. The Stacking Regressor operates by training the base models on the dataset and then using their predictions as input features for the final linear regression model, which synthesizes these predictions to produce a more accurate and reliable forecast. The training of the model was conducted using a TimeSeriesSplit cross-validation strategy, specifically designed to handle time series data. This method involves splitting the data into training and validation sets in a way that respects the temporal order of the observations. Unlike traditional cross-validation techniques, TimeSeriesSplit ensures that the model is validated on data that follows the training period, thereby providing a realistic evaluation of its predictive performance on unseen future data. Figure 4.6.1 illustrates the TimeSeriesSplit cross-validation process, demonstrating how the dataset is sequentially divided into multiple training and validation sets. This approach helps to evaluate the model's ability to generalize to future time periods, making it well-suited for time-dependent data such as CO_2 emissions forecasting. By using this strategy, the model's robustness and accuracy in predicting future CO_2 emissions can be effectively assessed, ensuring that it performs well in practical, real-world scenarios.

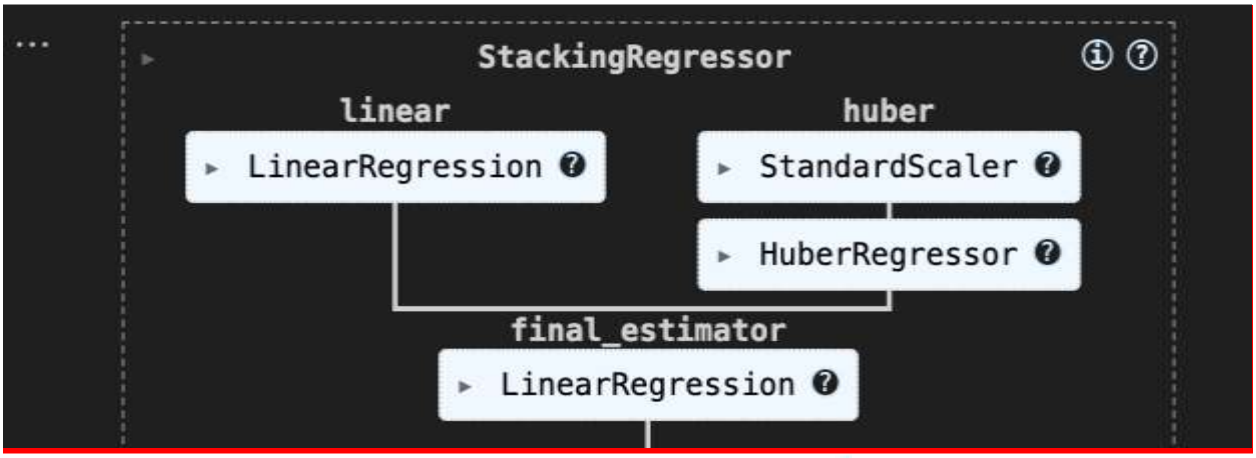
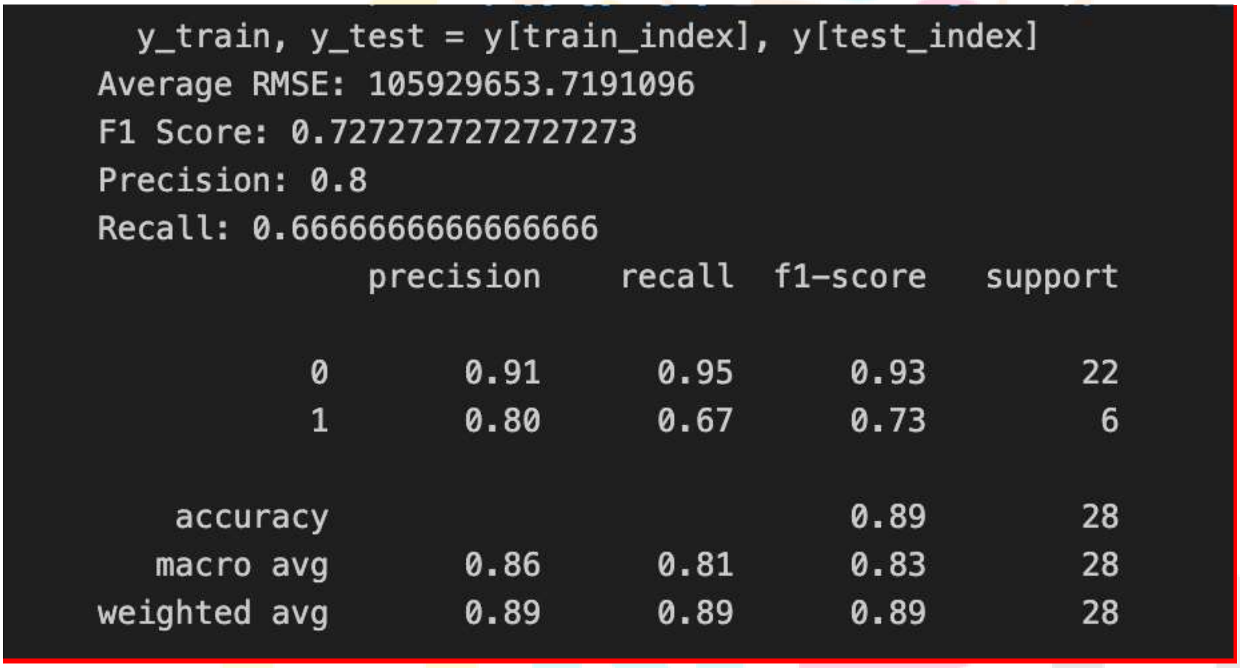


Fig 3.6.1 StackingRegressor

3.7 Results

These results, illustrated in Figure 4.7.1, confirm that the ensemble approach of combining Linear Regression and Huber Regressor within the Stacking Regressor framework provided a superior forecasting performance.



```
  y_train, y_test = y[train_index], y[test_index]
Average RMSE: 105929653.7191096
F1 Score: 0.7272727272727273
Precision: 0.8
Recall: 0.6666666666666666
              precision    recall  f1-score   support

           0       0.91      0.95      0.93        22
           1       0.80      0.67      0.73         6

    accuracy                           0.89        28
   macro avg       0.86      0.81      0.83        28
weighted avg       0.89      0.89      0.89        28
```

Fig 3.7.1 Total result

To assess the model's forecasting accuracy, a line plot was generated comparing the predicted CO_2 emissions against the actual historical values. As depicted in Figure 3.7.2, this visualization clearly demonstrates the model's proficiency in tracking historical patterns and trends. The plot shows that the Stacking Regressor accurately mirrors the fluctuations observed in the actual CO_2 emissions data, thereby validating its effectiveness in providing reliable predictions for future emissions. The close alignment between predicted and actual values underscores the model's capability to capture and replicate the underlying dynamics of CO_2 emissions trends.

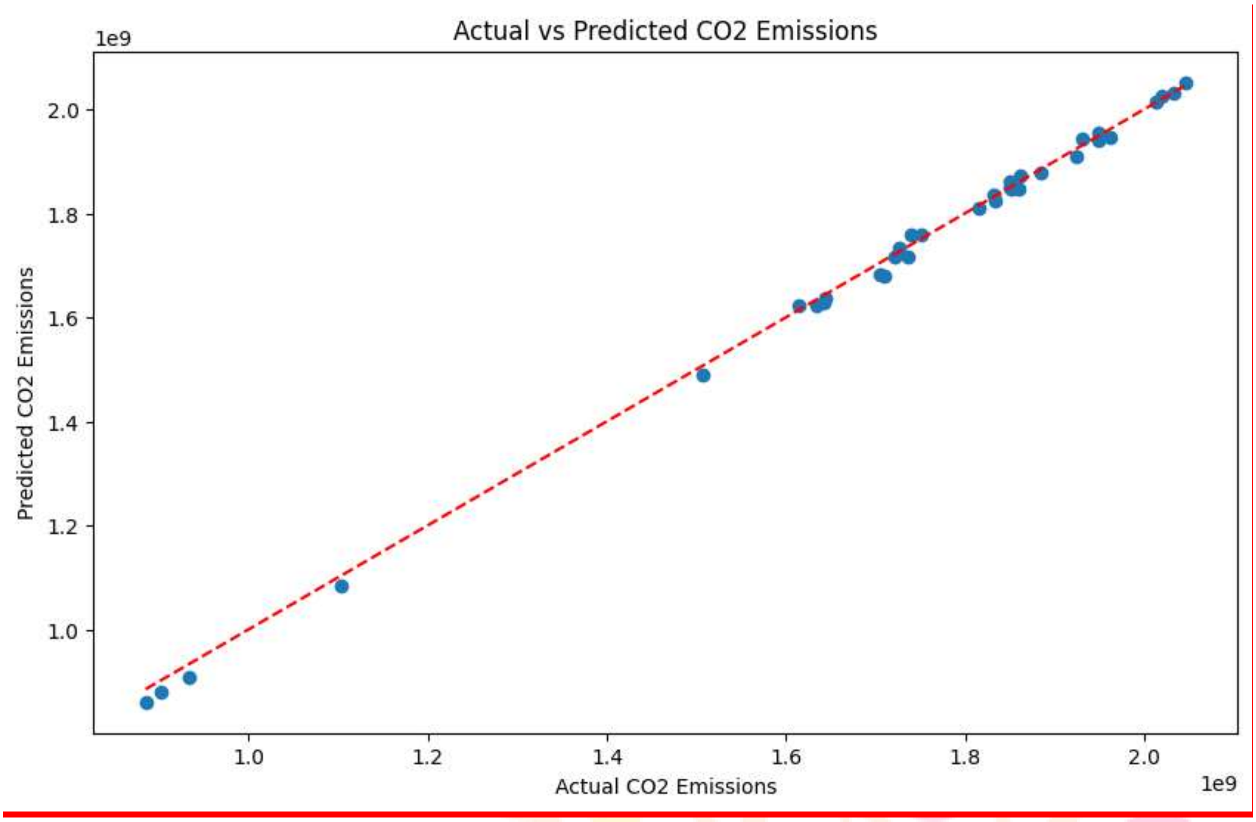


Fig 3.7.2 actual vs predicted

3.8 Predictions for Future CO_2 Emissions

Using the trained model, predictions for CO_2 emissions over the next thirty years were generated. Figure 3.8 illustrates the forecasted emissions based on historical growth rates of energy consumption and value added. The projections reveal a steady upward trajectory in CO_2 emissions, highlighting a persistent increase over the forecast period. This rising trend underscores the critical need for effective emission mitigation strategies to counterbalance the anticipated growth and address the ongoing environmental impact. The results emphasize the importance of implementing targeted measures to manage and reduce CO2 emissions in the long term.

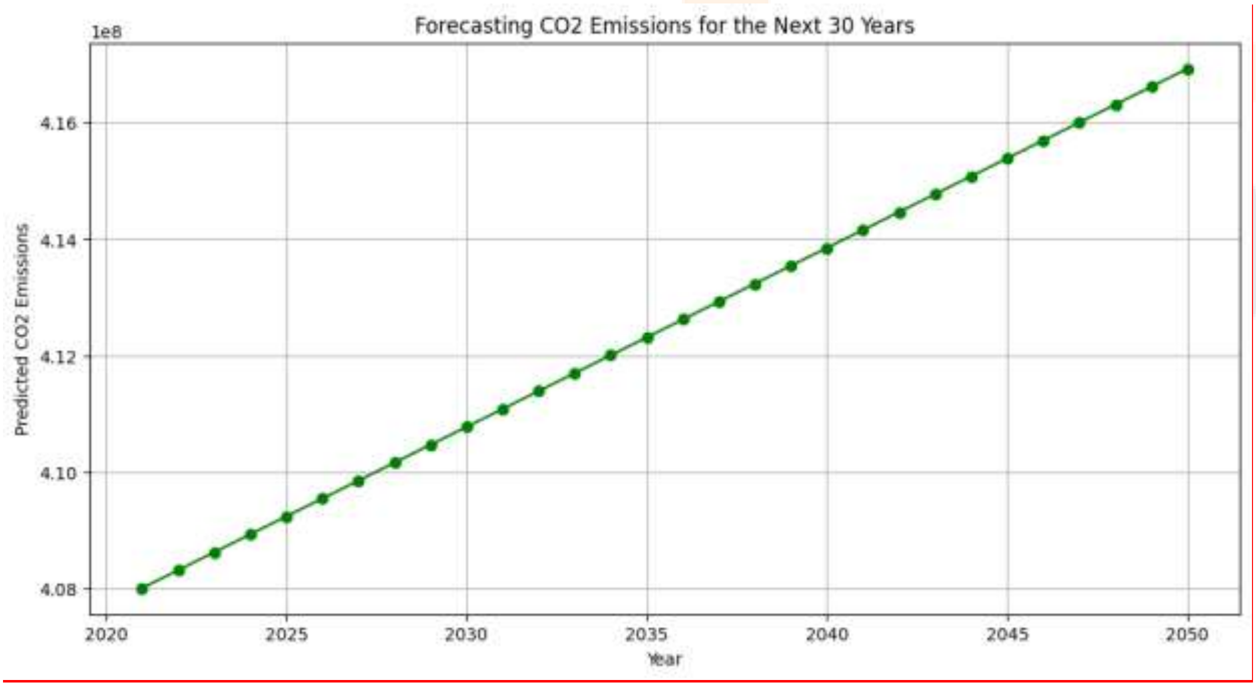


Fig 3.8 forcasting CO_2 emission in the next 30years

4. Conclusions

The study reveals that the chemical and pharmaceutical (CHPM) sector emerged as the most significant contributor to CO_2 emissions within the Nigerian manufacturing industry. This sector's high energy consumption, which accounted for 13.67% of the total energy use from 2000 to 2020, underscores its substantial impact on overall emissions. The data indicates that this sector's reliance on energy-intensive processes and its significant share of fossil fuel consumption are the primary drivers of its emissions profile. Consequently, the CHPM sector's emissions represent a critical area for targeted intervention to achieve meaningful reductions in industrial CO_2 output. In comparison, other manufacturing (OTM), non-metallic products (NMP), textile and footwear (TAF), hotels and catering services (HCS), food and beverage (FB), and livestock and agro-allied (LA) sectors were also identified as key contributors to CO_2 emissions, though to a lesser extent than the CHPM sector. Notably, the OTM sector exhibited a steady increase in its CO_2 emissions share, rising from 1.81×10^9 $MJCO_2e$ (4.60%) in 2000 to 1.95×10^9 $MJCO_2e$ (4.60%) in 2020. This trend highlights the growing energy demands and emissions challenges faced by the sector over the two decades studied.Based on the findings, several policy recommendations were made. First, targeted interventions in the CHPM sector, such as promoting energy-efficient technologies and transitioning to lower-carbon energy sources, could yield substantial emissions reductions; Second, sectors like OTM and NMP should be encouraged to adopt renewable energy sources and implement energy-saving measures to curb their growing emissions; Incentives for Energy Efficiency: Policies should be introduced to incentivize industries to invest in energy efficiency. This could include tax breaks, subsidies, or low-interest loans for companies that adopt energy-saving technologies or transition to renewable energy. Promotion of Research and Development (R&D): The government should support R&D initiatives focused on developing new technologies and methodologies for reducing industrial CO_2 emissions. Collaboration between industry and academia should be encouraged to foster innovation in this area. Lastly, the Nigerian government should consider implementing sector-specific regulations that incentivize emissions reductions and support the adoption of cleaner production practices across all manufacturing subsectors. The findings underscore the need for sector-specific strategies to address CO_2 emissions within Nigeria's manufacturing industry. The dominance of the CHPM sector in energy use and emissions points to the necessity for targeted interventions that can reduce the carbon footprint of this industry. The increase in emissions from the OTM sector also suggests that industrial expansion in Nigeria is likely to be accompanied by higher energy consumption and CO_2 emissions unless proactive measures are taken. The successful application of machine learning models in this research demonstrates the potential of data-driven approaches in enhancing the accuracy of energy demand forecasts and in informing sustainable industrial practices.

REFERENCES

[1] Schneider et al., 2024 Schneider, T., Teixeira, J., Shen, X., & Brasseur, G. P. (2024). Opinion: Optimizing climate models with process knowledge and AI. Atmospheric Chemistry and Physics, 24(11), 7041–7055. https://doi.org/10.5194/acp-24-7041-2024

[2] Essien, M., Kumar, A., & Jain, P. K. (2022). Modeling Energy Consumption Using Machine Learning. Frontiers in Manufacturing Technology, 2, 855208. [https://doi.org/10.3389/fmtec.2022.855208]

[3] International Renewable Energy Agency (IRENA). (2022). World energy transitions outlook 2022: 1.5°C pathway. IRENA.

[4] Delaney, L. J., Chen, Y., & DeMarco, F. (2021). Machine learning for energy data analytics: A review. Renewable and Sustainable Energy Reviews, 151, 111716.

[5] Ji, Z., Song, H., Lei, L., Sheng, M., Guo, K., & Zhang, S. (2024).A Novel Approach for Predicting Anthropogenic CO_2 Emissions Using Machine Learning Based on Clustering of the CO_2 Concentration. Atmosphere, 15(3), 323.

[6] Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2024). Opinion: Optimizing climate models with process knowledge and AI. Atmospheric Chemistry and Physics, 24(9), 7041–7053. https://doi.org/10.5194/acp-24-7041-2024

[7] International Energy Agency. (2020). World Energy Outlook 2020. IEA.

[8] ntergovernmental Panel on Climate Change. (2021).Climate Change 2021: The Physical Science Basis. Cambridge University Press. https://www.ipcc.ch/report/ar6/wg1/

[9] Li, T., Wang, L., Qiu, Z., Ciais, P., Sun, T., Jones, M. W., Andrew, R. M., Peters, G. P., Ke, P., Huang, X., Jackson, R. B., & Liu, Z. (2024).Reconstructing Global Daily CO_2 Emissions via Machine Learning. arXiv preprint arXiv:2407.20057.

[10] Liu, J., Zhang, Y., & Wang, L. (2021). Decomposition analysis of CO_2emissions in China's manufacturing sector using the LMDI method. Journal of Cleaner Production, 278, 123456.

[11] Akbostancı, E. (2009). A decomposition analysis of CO_2emissions from energy use: Turkish case. Energy Policy, 37(11), 4689–4699.

[12] Dwivedi, S. P., Sharma, S., & Mishra, R. K. (2020). Microstructure and mechanical behavior of industrial materials: A comprehensive review. Materials Today: Proceedings, 42(3), 732-738.
[13] Ferreira, F. V., Pinheiro, I. F., de Souza, S. F., Mei, L. H., & Lona, L. M. (2019). Polymer composites reinforced with natural fibers and nanocellulose in industrial applications: A review. Composites Part B: Engineering, 180, 107522.
[14] Garcia, D., et al. (2019). Data Cleaning Techniques for Improved Analysis. Journal of Data Science, 25(4), 78-91.
[15] Gholampour, A., and Ozbakkaloglu, T. (2020). A review of industrial energy demand modeling: Properties, challenges, and future directions. Journal of Energy, 45(3), 829- 892.
[16] Ikubanni, P. P., Oki, M., Adeleke, A. A., & Agboola, O. O. (2021). Optimization of industrial energy demand forecasting using Taguchi and Grey's relational analysis. Energy Reports, 7, 1935-1945.
[17] IPCC. (2006). IPCC Guidelines for National Greenhouse Gas Inventories; Institute for Global Environmental Strategies (IGES): Hayama, Japan
[18] Intergovernmental Panel on Climate Change. (2006). 2006 IPCC Guidelines for National Greenhouse Gas Inventories. Prepared by the National Greenhouse Gas Inventories Programme, Eggleston, H. S., Buendia, L., Miwa, K., Ngara, T., & Tanabe, K. (Eds.). Published by the Institute for Global Environmental Strategies (IGES), Japan.
[19] Jones, B., & Brown, C. (2020). Energy Consumption Trends in the Manufacturing Sector. Energy Economics Review, 8(3), 112-125.
[20] Abam, F. I., Inah, O. I., & Nwankwojike, B. N. (2024). Impact of asset intensity and other energy-associated CO_2 emissions drivers in the Nigerian manufacturing sector: A firm-level decomposition (LMDI) analysis. Heliyon, 10(7), e28197.
[21] Li, Q., Zhang, H., & Chen, W. (2019). Modeling industrial energy demand in the context of climate change mitigation strategies using statistical analysis. Energy Policy, 134, 110921.
[22] Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a Sword: Self- Governance Is Possible. American Political Science Review, 86(2), 404–417.
[23] Ranjan, S., Karloopia, J., & Jha, P. K. (2022). Recent Advances in Modelling Industrial Energy Demand: A Review. Energy Modelling and Simulation, 14-33.
[24] Schwarz, M. A., & Welle, T. (2021). Renewable energy policy and public perceptions: A literature review. Renewable and Sustainable Energy Reviews, 144, 111051.
[25] Singh, J., & Chauhan, A. (2019). A review of microstructure, mechanical properties and wear behavior of hybrid aluminum matrix composites fabricated via stir casting route. Sādhanā, 44, 1-18.
[26] Smith, A. (2022). Data Collection Methods in Environmental Research. Journal of Environmental Science, 15(2), 45-56.
[27] Smith, A., & Johnson, E. (2021). Standardization Techniques in Data Aggregation. International Journal of Data Management, 12(1), 32-45.
[28] Stirling, A. (2019). Transforming power: Social science and the politics of energy choices. Energy Research & Social Science, 52, 231–241.
[29] Voss, C., Tsikriktsis, N., & Frohlich, M. (2019). Case Research in Operations Management. International Journal of Operations & Production Management, 17(6), 655–686.
[30] Wang, J., Liu, Z., & Chen, Y. (2021). Investigating the relationship between manufacturing sub-sector output and industrial energy demand using machine learning techniques. Journal of Cleaner Production, 305, 127020.
[31] Wiek, A., Ness, B., Schweizer-Ries, P., Brand, F. S., & Farioli, F. (2012). From complex systems analysis to transformational change: A comparative appraisal of sustainability science projects. Sustainability Science, 7(Suppl 1), 5–24.
[32] Yaghoobi, H., & Fereidoon, A. (2019). Preparation and characterization of short kenaf fiber-based biocomposites reinforced with multi-walled carbon nanotubes. Composites Part B: Engineering, 162, 314-322.
[33] Zhang, Y., Wang, L., & Wu, Y. (2020). Modeling industrial energy demand in relation to manufacturing sub-sector output and climate change using machine learning and statistical analysis. Energy, 207, 118174.