2025 Volume: 01 Issue: 07 www.ijrpa.com ISSN 2456-9995

Research Article

International Journal Research Publication Analysis

Page: 01-13

PREDICTIVE MODEL FOR THE ANALYSIS OF CARBON EMISSIONS WITH MACHINE LEARNING

***Falodun Olugbenga Abiola, Omowole Fadesakin, Ayileka Ojo Samson, Ogunlade Adedayo, Aloba Tosin Olugbenga, Ige Samuel Adeniyi**

Electrical and Electronics Engineering, Rufus Giwa Polytechnic, Owo, Nigeria.

Article Received: 28 October 2025

Article Revised: 17 November 2025

Published on: 08 December 2025

***Corresponding Author: Falodun Olugbenga Abiola**

Electrical and Electronics Engineering, Rufus Giwa Polytechnic, Owo, Nigeria. DOI: https://doi-doi.org/101555/ijrpa.1691

ABSTRACT

This paper examines telecommunication base station (TBS) carbon remediation with the development of a versatile modeling framework along with analysis and forecasting of CO_2 emissions based on energy use data. Emissions of diverse forms of energy were calculated, considering their emissions factor (EF). The source EF was multiplied with region specific carbon intensity factors. An exploratory analysis was done using Python machine-learning packages, including regression and Random Forest regression to identify emission drivers and provide time-series forecasts for emissions. The analysis reveal there's a strong dependence on emissions on energy source, with renewables proposed as having substantial reduction potential. There were also higher emissions during the weekdays for operational load. Overall, the research highlights the influence of moving to renewables and patterns of operating TBS for lowering the burden on the environment, and intends to provide practical context for policy makers and network managers looking to achieve sustainability goals.

KEYWORDS: Carbon Emissions, Machine Learning, Predictive Modeling, Renewable Energy Integration, Telecommunication Base Station.

INTRODUCTION

It has become difficult to envision a world without smartphones, making telecommunication devices the most important development of the 21^{st} century. The rapid adoption of these devices worldwide has significantly affected daily life and economic growth, accelerating the spread of information, encouraging competition, and increasing access to various services. The expansion has also contributed to global GDP [1]. However, the broader effects of telecommunication devices, particularly regarding climate change and carbon dioxide (CO_2) emissions, are not as widely understood[2].For much of the 20th century, CO_2 emissions and greenhouse gases have gotten limited political and policy attention. However, it has become evident that the greenhouse effect plays a key role in warming the planet with CO_2 responsible for over half of the intensified greenhouse effect and predicted to remain a major influence [3].

Despite these early warnings, rising global populations and economic growth have led to an increase in CO_2 emissions and to limit the most severe effects of climate change, CO_2 emissions must drastically decrease this decade and reach net zero by 2050 [4]. In this context, the role of technology, especially Information and Communication Technology (ICT), in combating climate change is increasingly highlighted. Most research tends to focus on one of these effects, and although ICT's potential to reduce CO_2 emissions is well-regarded, evidence of its overall macro-level impact remains scarce.

The research project will explore the effects and predict the anticipated future carbon emissions of Telecommunication Base Stations (TBS). This research will model energy consumption for TBS and calculate CO2 emissions based on energy data and emission factors. Later, a Python machine learning model will be designed to test energy scenarios of TBS, and use regression and time series modelling techniques to predict emissions into the future. The results will be utilized to provide recommendations for best practices for reducing the carbon footprint of TBS.

Carbon Footprint and Emissions

A carbon footprint is the total greenhouse gas emissions which is associated with a product. Calculating the exact footprint has been a daunting task due to the extensive data required and the difficulty in controlling natural sources of carbon dioxide [5]. This has been a major challenge as there is no consistent framework for assessing greenhouse gas emissions from these devices. This consistency is essential for manufacturers, operators, and users to effectively understand, manage, and reduce their environmental impact. Despite significant progress in reporting emissions, the data remains inconsistent and unaudited. This makes it difficult to verify improvements as a standardized set of accounting principles for greenhouse gases is needed to ensure accurate emissions [6]. Teehan [8], employed a process-sum life cycle (LCA) approach to calculate the embodied GHG emissions for 11 ICT products. These include various types of desktops, laptops, tablets, mobile devices, server and network switch. Using linear regression models, they identified and simplified relationship between product mass and GHG emissions and develop a more refined model that includes display, battery and circuit board for higher accuracy.

A Washington Post report shows that carbon dioxide and other greenhouse gases are already altering the Earth's climate, with more severe disruptions expected. Also, a study from the American Consumer Institute Center for Citizen Research examined how widespread broadband access in the U.S. could help to reduce emissions [2]. Belkhir [9] aimed to assess the global carbon footprint of the ICT industry, focusing on its contributions to greenhouse gas (GHG) emissions, because ICT has been considered here for enabling efficiencies that reduce emissions in other sectors but has received little scrutiny as a direct contributor to GHG emissions. Despins [10], explored the potential of green communication systems and green ICT to mitigate global greenhouse emissions through green communication research and international efforts to standardize methods for accurately measuring ICT's carbon reduction potential. These standardized metrics are crucial for assessing the economic viability of green ICT within the low-carbon economy.

Predictive Modeling of DateTime Datasets

Predictive modeling of data sets is increasingly becoming more important in applications ranging from energy forecasting and financial market analysis to weather forecasting and telecommunication. Datetime data sets naturally carry a unique set of issues due to their temporal nature, including trends, seasonality, autocorrelation, and irregular intervals. A variety of statistical and machine learning techniques have therefore been developed and applied to model and forecast such data effectively. Selection of an appropriate model is often dependent on the nature of temporal features, data quality, and even forecast horizon [11]. Autoregressive Integrated Moving Average (ARIMA) model is one of the longest-standing and most widely utilized methods of time series forecasting. ARIMA as well as its seasonal variant SARIMA are applied extensively in time series forecasting due to the fact that both trend and seasonality can be described in terms of differencing and lag variables [12].

In order to model more complex patterns, especially for non-linear and noisy data, machine learning models such as Random Forests (RF) and Gradient Boosting Machines (GBM) have gained popularity. These models do not have stationarity or linearity assumptions and are capable of capturing interaction among multiple features, such as time-derived features such as day of the week, hour, or lag values. Random Forests, being ensemble decision tree-based models, are highly robust against overfitting and are capable of handling high-dimensional feature sets. However, their accuracy often heavily depends on being tightly integrated with feature engineering, especially when temporal relationships have to be included manually using lagged variables [13].

There are many machine learning methods used to take advantage of datetime modeling with different strengths and weaknesses. K-Nearest Neighbors (KNN) is a non-parametric method that assumes similarity based on past observations but is useful on small cyclical data, computationally expensive on large amounts of observations, and not useful when trends shift. Support Vector Regression (SVR) can be used to handle noisy and/or high-dimensional data, it uses kernel functions to create a non-linear treatment for predictions. However, the model itself is tricky to use because it requires good feature engineering to properly adjust for temporal dependencies. Deep learning approaches like Long Short Term Memory (LSTM) networks rely on deep learning concepts to find long-term structure in sequential data [14], [15], [16], [17], [18]. The absence of a holistic and scalable modeling framework that incorporates regionally specific emission factors, machine learning-based driver analysis, and time-series forecasting in relation to telecommunication base stations (TBS), represents a significant void within actionable solutions for the reduction of carbon footprints as global networks continue to grow.

METHODOLOGY

Python

Setup

The dataset which contained the daily energy consumption data was imported into a *pandas DataFrame* for analysis. The main fields were; site id, the date and time for each measurement period start and end timestamp and the power consumption various metrics, average, minimum, and maximum metrics in kilowatts.

Feature Engineering

Two key features were extracted from the raw data to improve the analysis. The first is **Daily Energy Consumption** in kWh, which was calculated as the difference of the Consumers Load and the average power multiplied by 24 hours. This feature converts power readings to total daily energy usage. The second feature is **Day Type Classification**, a binary variable distinguishing the two types of days, weekends and weekdays.

Carbon Emission Modeling

The research commenced with the modeling of power consumption of a transmission base station in Nigeria, data on the energy consumption of TBS would include parameters based on load demand and uptime hours, and environmental conditions, using the historical power consumption data.

The Linear Emissions Factor (LEF) model is a basic mathematical model which estimates emissions as the sum of energy consumed from each energy supply multiplied by the respective emissions factor. For example, if a TBS uses energy from diesel generators, energy from grid electricity and energy from renewable sources, the emissions formula would be:

$$CO_2 = \sum_{i=1}^{n} E_i \times EF_i \quad (3.1)$$

Where:

E_i *is the energy consumed from source*

EF_i *is the emission factor of that source*

Renewable energy sources will also typically have emissions factors close to zero due to their carbon light footprint. This linear format allows emissions to be attributed to the actual energy sources, which can then be used to prioritize mitigation strategies.

The model will need two main types of data:

The energy consumption records should be a complete record of all consumption of all sources, including the diesel consumption logs, electricity meter readings and the solar generation logs. In order to estimate emissions from the different energy sources, you will need to apply the relevant emission factors (EF) to the different source and region specific usage, with Nigeria's national average EF for grid electricity in central Nigeria being around 0.7 kg CO_2/kWh. You may also be able to source the relevant regional grid emission factors from local agencies such as the Nigerian Electricity Regulatory Commission (NERC) or international databases such as the IPCC or IEA whose output values vary according to grid mixes. Carbon emissions are calculated by applying constants carbon intensity factors to four energy sources. Information used for this calculation would be gotten from [19].

Carbon Emission Factor for Coal

From [19], taking Coal as 100 kgCO_2 per mmBtu, we convert it to gCO_2 per kWh.

Mass conversion is; $1kg\ CO_2 = 1000g\ CO_2$ (3.2)

Hence, $100kg\ CO_2 = 100 \times 1000 = 100{,}000g\ CO_2$ (3.3)

Energy Conversion is; $1mmBtu = 1{,}000{,}000Btu$ (3.4)

Hence, $1kWh = 3{,}412Btu$

(3.5)

$1mmBtu = \frac{1{,}000{,}000}{3{,}412} \approx 293.07kWh$ (3.6)

So **Thermal Energy Basis** (which assumes the output is thermal energy),

$\frac{gCO_2}{kWh} = \frac{100{,}000g\ CO_2}{293.07kWh} \approx \frac{341.2g\ CO_2}{kWh}$ (3.7)

Hence, for **Electrical Energy Basis** with an efficiency of 40%, efficiency '**η**'= **0.4**

$\frac{gCO_2}{kWh} = \frac{341.2}{0.4} \approx \frac{853g\ CO_2}{kWh}$ (3.8)

Hence, the carbon emission factor for coal (853 gCO_2/kWh)

Carbon Emission Factor for Grid

First, we would make some key assumptions;

Fuel mix and emission factors:

a. This will comprise of coal which is around 100 kg CO_2/mmBtu, natural gas around 50 kg CO_2/mmBtu, all having weighted average emission factor of 65 kg CO_2/mmBtu.

1. Average thermal efficiency: This has to be about 40% for fossil plants, accounting for aging infrastructure.
2. Share of fossil generation of 85% of total electricity.

   Then we would calculate emission intensity for fossil generation.

$\frac{gCO_2}{kWh_{fossil} = F_{avg} \times 1000 \times \frac{3.412}{\eta}}$ (3.9)

$\frac{gCO_2}{kWh_{fossil} = 65 \times 1000 \times 8.53}$ (3.10)

$\frac{gCO_2}{kWh_{fossil} = 554}\frac{gCO_2}{kWh}$ (3.11)

Then we would calculate grid average, which includes non-fossil sources

Now, fossil generator is about 85% of emission = 554gCO_2/kWh, non-fossil share is 15% carbon emission intensity which is 0 gCO_2/kWh.

Hence, Grid average= (554 X 0.85) + (0 X 0.15) = 470.9 gCO_2/kWh, which is roughly 471 gCO_2/kWh

Carbon Emission Factor for Natural Gas

Taking $\eta = 40$, $\frac{gCO_2}{kWh = 53.06 \times 1000 \times 8.53} = 452.6 gCO_2/kWh$. (3.12)

Hence, the carbon emission factor for natural gas is 453 gCO_2/kWh

Carbon Emission Factor for Renewable Energy

All sources of electricity produce some GHG emissions over their lifecycle, but renewable energy would have a lot less emissions than a fossil fuel-fired power plant. One study estimates renewable energy sources emit roughly 50g or less of CO2 emissions per kWh over their lifecycle versus a coal which emits about 1000 g CO2/kWh and natural gas with 475 g CO2/kWh. The production of solar panels does require a considerable amount of energy; however, studies found the energy consumed in the production process is offset, depending on module type, in about two years of operation [20].

This prompted us to use a 50 gCO_2/kWh emission factor for renewable energy.

Emissions (in $kgCO_2$) for all scenarios are calculated according to the equation:

$$\frac{Daily\ Energy\ (kWh) \times Carbon\ Intensity\ (gCO_2/kWh)}{100} \quad (3.13)$$

The resulting columns will be appended to the DataFrame.

Machine Learning Integration

Regression Models for Understanding Drivers of Emission

This Regression analysis is the best way to determine the relationships between the variables driving CO_2 emissions. This would be expressed as:

$$CO_2 = \beta_0 + \beta_1(Diesel\ Energy) + \beta_2(Grid\ Energy) + \beta_3(Renewable\ Energy) + \beta_4(Load) \quad (3.14)$$

Where $\beta_1\beta_2\beta_3$ would show the marginal impact of different energy sources on emissions, β_4 would reveal how network traffic affects power demand.

Nonlinear relationships would need machine learning algorithms such as Random forest that can account for the interactions variables. Our regression will provide output that identifies the most productive levers to address emissions.

Implementation

Python Implementation

The subsequent step would be to develop a model in Python for analyzing the collected data and estimating CO_2 emissions. The code will have input values for power consumption of the TBS, the energy mix ratio, and emission factors for each energy source. The emissions would be calculated using the formula:

$$CO_2 emissions = (Energy\ from\ fossil\ fuel \times fossil\ fuel\ emission\ factor) + (Energy\ from\ Grid \times Grid\ emission\ factor) \quad (3.16)$$

- Where the energy imports would be expressed in kilowatt-hours or liters of the fossil fuel source.

The validity of the model could come in the form of backtesting the forecasts to historical data and comparing the estimated TBS emission reduction from renewable pilot projects compared to the actual TBS emission reduction from pilot projects.

Visualization

The visualization will start with Daily Energy Consumption vs Carbon Emissions to form a baseline, and then will look at emissions performance across several energy types, providing ranges and anomalies. Ultimately, we should better understand emissions related to the telecommunication base station as we develop a boxplot comparison for weekdays and weekends. From this, we will be able to discuss takeaways like total monthly energy consumption, average daily CO2 emissions, peak emission days, and the potential for savings based on the use of renewables. Next, predictive modeling would be done, using the most relevant input variables through feature importance while optimizing the model size, eliminating irrelevant features, and finally ensuring that the measures align with the modeled application and domain knowledge. Finally, there will be a side-by-side comparison of an energy consumption forecast model and the forecast made against data points and if mismatches occur a regression model will be built to efforts to improve the forecast.

RESULTS AND DISCUSSION

Exploratory Data Analysis

After the data was loaded onto the python platform for coding, the data was cleaned and the date columns were converted to datetime for easy identification. Then the average daily energy consumption was calculated and the calculated carbon intensity was loaded onto the platform which was then used to calculate the emissions for different scenarios.

Daily Energy Consumption and Carbon Emissions

Here, we illustrate the daily energy consumption (with blue bars) alongside CO_2 emissions (with an orange line) over 30 days. There isn't much of a change with energy consumption, only minor fluctuations, but with CO_2 emissions, there is significant variation with sharp peaks and drops. This indicates that CO_2 emissions are influenced by daily energy consumption, but may also depend on other factors.

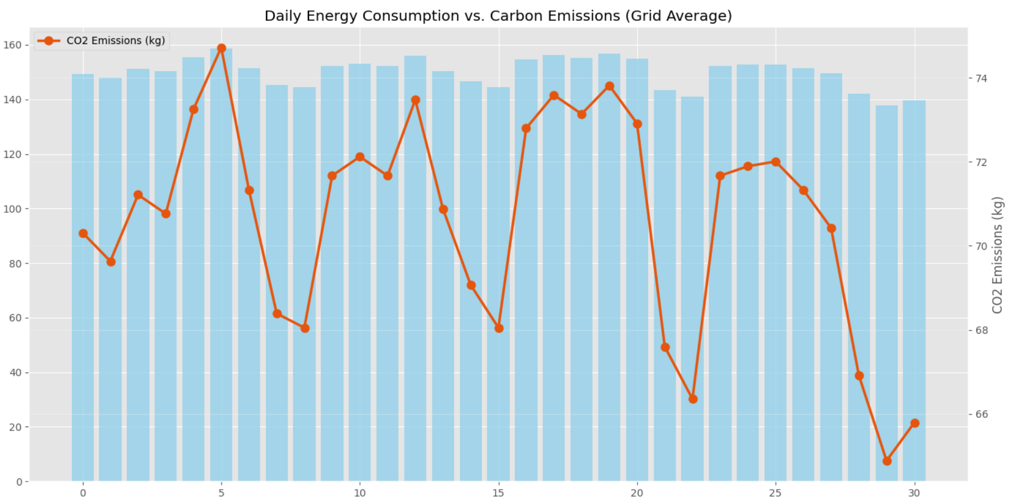

Figure 1: Daily Energy Consumption vs Carbon Emissions. (Grid Average)

Carbon Emission Comparison by Energy Source

Here, we compare the CO2 (carbon dioxide) emissions by the energy source over time in kilograms. It clearly shows that coal is the most carbon-intensive source followed by the grid average, then natural gas, and lastly renewable energy with almost no pollution. This data demonstrates the environmental benefit of moving from fossil fuels to renewable energy.

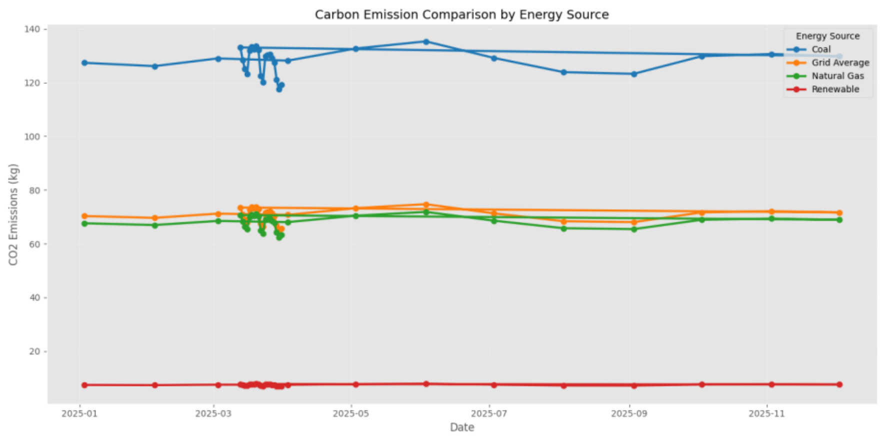

Figure 2: Carbon Emission Comparison by Energy Source.

Carbon Emission Range

We then got the range of carbon emissions over time, presenting what appeared to be a typical "band" of output with an average line and also highlighted specific "anomalies" in which emissions were relatively high spikes. The point being, emissions are highly predictable and appear to lie within expected boundaries or are of a predictable nature. However, an anomaly represents an event that is either significant or an unexpected event.

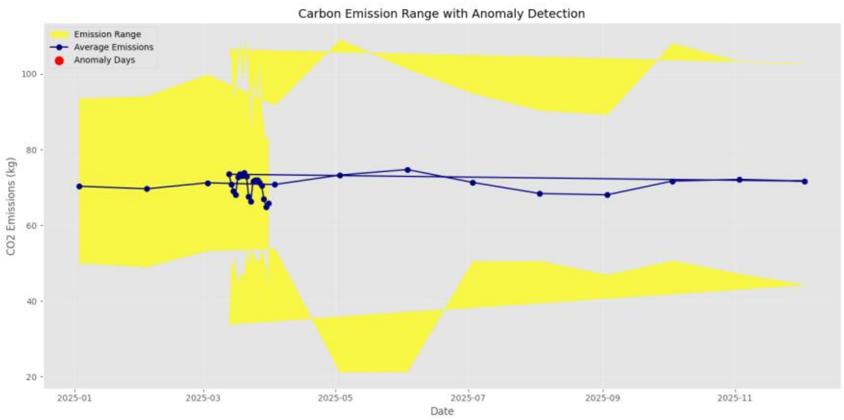

Figure 3: Carbon Emission Range with Anomaly Detection.

Comparing Carbon Emissions (Weekdays with Weekends)

Emissions are always higher on weekdays (71.7kg) than on weekends (67.8kg). This suggests that job-related commercial and industrial activity during a normal work week is a major contributor to emissions, and essentially zeroes during a weekend time period of reduced economic operations.

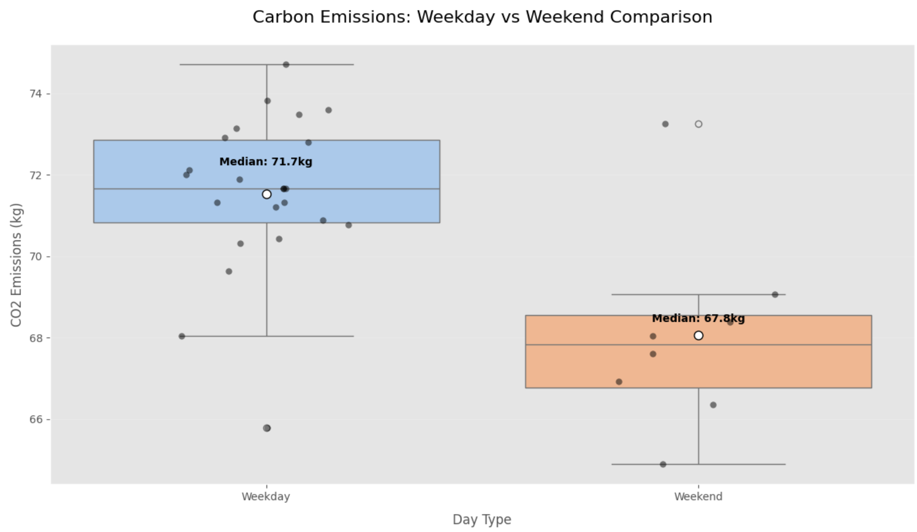

Figure 4: Carbon Emissions: Weekday vs Weekend Comparison.

Potential for Renewable Integration

While the grid consumption shown will produce high daily CO2 emissions, there is a considerable potential for both switching to renewable energy and all its associated costs and savings by opting for renewable energy and avoiding those CO2 emissions. With the right mix of renewables, there is the substantial potential to save almost 2000 kg/month of CO2 emissions but also will likely result in considerable cost savings.

Predictive Modeling

We first attempted to create and test a predictive model of target variable based on selected features of the dataset. A Random Forest Regressor, a machine learning algorithm, was chosen because of its strength and accuracy in dealing with non-linear relationships and complex data patterns. The energy consumption forecast model which performed well in train data (low error, high accuracy), but performed okay in unseen test data (error and negative R^2).

The model appears sufficiently good in the validation dataset since it has a MAE of 1.26 kWh and R^2 of .88, which explains .88 of the variability, and on the training dataset the MAE is 5.24 and the R^2 is -0.26. The confusion in wording is what the graph is showing, even though it is very limited data, that appears to show a comparison of forecasted energy consumption to actual measured energy consumption over several days at the end of March in 2025. I think it is reasonable to assume that the prediction is probably not very good, especially in terms of previous poor test model metrics (Test R^2: -0.26). If the likelihood of my forecasts were not too different from actual consumption, it could very well have been due to luck or a clear simple trend during that time, but, because the test R^2 was negative and the data explains nothing when predictions were made for new data, the model seems to fail on that task systematically.

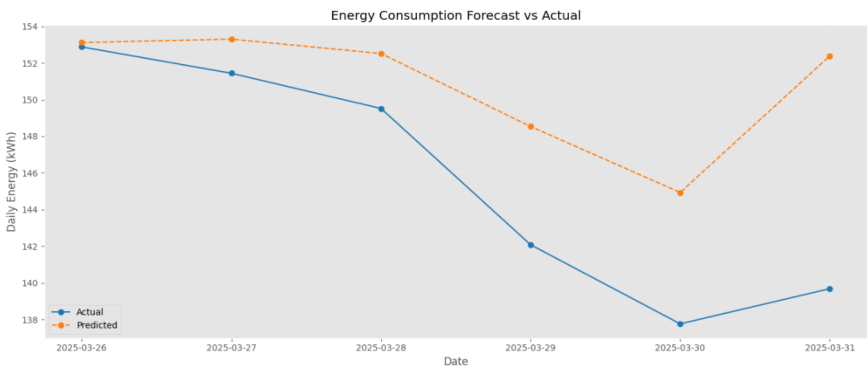

Figure 5: Energy Consumption Forecast vs Actual Forecast.

CONCLUSION

This paper has focused on analyzing the carbon emissions of telecommunication base stations (TBS) through detailed energy consumption modeling and the application of predictive machine learning techniques. Using regression analysis and time series forecasting, the study evaluated the relationship between energy sources and emissions, identifying coal as the most carbon-intensive and renewables as the least. The models, particularly Random Forest regression, were employed to capture non-linear patterns in the data, while anomaly detection highlighted irregular spikes in emissions. Although limitations in dataset size affected prediction accuracy, evidenced by a negative R^2 in test results, the analysis confirmed that emissions follow predictable patterns and that renewable energy integration could significantly reduce the TBS carbon footprint. Overall, the study demonstrates the effectiveness of combining statistical and machine learning models for emissions analysis while stressing the importance of improved datasets to enhance predictive reliability.

REFERENCES

1. K. Bahia, P. Castells, and X. Pedros, "The impact of mobile technology on economic growth: global insights from 2000-2017 developments," *30th Eur. Reg. ITS Conf.*, 2019, [Online]. Available: https://econpapers.repec.org/RePEc:zbw:itse19:205164
2. J. P. Fuhr and S. B. Pociask, "Broadband Services : Economic and Environmental Benefits," *Am. Consum. Inst.*, pp. 1–50, 2007.
3. T. Dietz, R. L. Shwom, and C. T. Whitley, "Climate change and society," *Annu. Rev. Sociol.*, vol. 46, no. May 2022, pp. 135–158, 2020, doi: 10.1146/annurev-soc-121919-054614.
4. J. Yu, Y. M. Tang, K. Y. Chau, R. Nazar, S. Ali, and W. Iqbal, "Role of Solar-Based Renewable Energy in Mitigating CO2 Emissions: Evidence from Quantile-on-Quantile Estimation," *Educ. Res.*, no. May, pp. 1–20, 2022.
5. M. Sloma, "Carbon footprint of electronic devices," *Electron Technol. Conf. 2013*, vol. 8902, p. 890225, 2013, doi: 10.1117/12.2030271.
6. E. Sutherland, "Climate Change: The Contribution of Telecommunications," *SSRN Electron. J.*, no. 76, 2011, doi: 10.2139/ssrn.1529685.
7. J. Malmodin, A. S. Moberg, D. Lundén, G. Finnveden, and N. Lövehagen, "Greenhouse gas emissions and operational electricity use in the ICT and entertainment & Media sectors," *J. Ind. Ecol.*, vol. 14, no. 5, pp. 770–790, 2010, doi: 10.1111/j.1530-9290.2010.00278.x.
8. P. Teehan and M. Kandlikar, "Comparing embodied greenhouse gas emissions of modern computing and electronics products," *Environ. Sci. Technol.*, vol. 47, no. 9, pp. 3997–4003, 2013, doi: 10.1021/es303012r.

9. L. Belkhir and A. Elmeligi, “Assessing ICT global emissions footprint: Trends to 2040 & recommendations,” *J. Clean. Prod.*, vol. 177, pp. 448–463, 2018, doi: 10.1016/j.jclepro.2017.12.239.
10. C. Despins *et al.*, “Leveraging green communications for carbon emission reductions: Techniques, testbeds, and emerging carbon footprint standards,” *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 101–109, 2011, doi: 10.1109/MCOM.2011.5978422.
11. L. Saeeda and P. Křemen, “Temporal knowledge extraction for dataset discovery,” *CEUR Workshop Proc.*, vol. 1927, 2017.
12. K. Kalpakis, D. Gada, and V. Puttagunta, “Distance measures for effective clustering of ARIMA time-series,” *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 273–280, 2001, doi: 10.1109/icdm.2001.989529.
13. A. Joly, “Exploiting Random Projections and Sparsity with Random Forests and Gradient Boosting Methods,” 2016.
14. N. F. Rajani, B. Krause, W. Yin, T. Niu, R. Socher, and C. Xiong, “Explaining and Improving Model Behavior with k Nearest Neighbor Representations,” 2020, [Online]. Available: http://arxiv.org/abs/2010.09030
15. H. Muthiah, U. Sa, and A. Efendi, “Support Vector Regression (SVR) Model for Seasonal Time Series Data,” *Proc. Second Asia Pacific Int. Conf. Ind. Eng. Oper. Manag.*, no. September 14-16, 2021, pp. 3191–3200, 2021.
16. F. Zhang and L. J. O’Donnell, “Support vector regression,” *Mach. Learn. Methods Appl. to Brain Disord.*, vol. 11, no. 10, pp. 123–140, 2019, doi: 10.1016/B978-0-12-815739-8.00007-9.
17. A. Sherstinsky, “Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network,” *Phys. D Nonlinear Phenom.*, vol. 404, no. March, pp. 1–43, 2020, doi: 10.1016/j.physd.2019.132306.
18. T. Chai and R. R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature,” *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, 2014, doi: 10.5194/gmd-7-1247-2014.
19. M. K. Albertini and R. F. de Mello, “Emission Factors for Greenhouse Gas Inventories,” *Adv. Data Anal. Classif.*, vol. 7, no. 4, pp. 435–464, 2013, doi: 10.1007/s11634-013-0145-3.
20. S. Tierney and L. Bird, “Setting the Record Straight About Renewable Energy | World Resources Institute,” 2020. [Online]. Available: https://www.wri.org/blog/2020/05/setting-record-straight-about-renewable-energy