Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 2 (2023)

A Machine Learning Approach to Analyze and Predict the Relationship between Sustainable Development Goals with Various Energy

B. Santhosh Kumar[1], Dr. P. Rajesh[2]

[1]Research Scholar, Department of Computer and Information Science, Annamalai University, Annamalainagar, Tamil Nadu, India
Email: santhoshcdm@gmail.com

[2]Assistant Professor, PG Department of Computer Science, Government Arts College, Chidambaram, (Deputed from Department of Computer and Information Science, Annamalai University) Tamilnadu, India
Email: rajeshdatamining@gmail.com

Abstract

Comprehensive worldwide information on sustainable energy is crucial for monitoring advancements, establishing objectives, and forming well-informed policy choices to address climate change and foster sustainable growth. Machine learning, a subset of artificial intelligence (AI), concentrates on creating algorithms and models that empower computers to acquire knowledge and generate forecasts or choices rooted in data. This field harnesses statistical techniques and computational approaches to enhance a computer's proficiency in a particular task. This paper considers global data on sustainable energy between 2000 to 2020. The machine learning approaches are used to analyze and predict the dataset using linear regression, multilayer perceptron, SMOreg, M5P, random forest, and REP tree**.** Numerical illustrations are provided to prove the proposed results with test statistics or accuracy parameters.

Keywords: Machine learning, global data on sustainable energy, decision tree, correlation coefficient, and test statistics.

1. Introduction and Literature Review

Global data related to research on sustainable energy encompasses a wide range of information and statistics, including research funding, publications, technological advancements, innovation trends, and the influence of such research on addressing climate change and achieving sustainable development objectives, all on a global scale. Access to this data is essential for monitoring the advancements and effects of sustainable energy research efforts worldwide.

Machine learning, a subset of artificial intelligence (AI), is dedicated to crafting algorithms and models that empower computers to acquire knowledge from data, enabling them to make predictions and decisions. It achieves this by leveraging statistical techniques and computational methods, enabling computers to enhance their task-specific performance as they encounter more data, all without requiring explicit programming. This versatile field is applied across diverse domains such as natural language processing, computer vision, recommendation systems, and data analysis, serving as a cornerstone for task automation, prediction, and knowledge extraction from extensive datasets. Data mining finds extensive practical use in diverse domains, such as customer segmentation, fraud detection, recommendation systems, market basket analysis, predictive maintenance, and scientific research.

It serves as a crucial instrument for deriving valuable insights from the immense volumes of data generated across various fields.

Employing advanced data analysis methods to scrutinize the primary socioeconomic factors influencing the digital skills of the Spanish population. This analysis helps us ascertain if there are any training requirements concerning their digital proficiency, which, in turn, can positively impact the enhancement of the country's sustainable development [1].

A novel machine learning-based methodology has been developed to connect project-based funding to the Sustainable Development Goals. This approach offers initial assessments of aid contributions from both DAC and non-DAC donors toward achieving these goals. Additionally, it demonstrates the potential for conducting similar analyses at the recipient level and with different types of textual databases, such as private sector reports. This opens up a wide range of possibilities for policy analysis. The methodology, as outlined in this working paper, employs semantic analysis of the textual descriptions of each project contained within the Creditor Reporting System (CRS) [2].

The aim of this research is to explore the alignment between mining operations and the Sustainable Development Goals (SDGs), drawing from insights in the scientific literature and observations of a specific mining activity. To achieve this, we conducted searches in the Web of Science and Scopus databases using keywords "sustainable development goals" and "mining." Additionally, we utilized the Coordination of Improvement of Higher Education Personnel (CAPES) Portal to identify papers linking each SDG with mining. Furthermore, we conducted site visits to three crushed stone mining facilities situated in Monsenhor Gil, Piauí, Brazil, to closely observe the diabase mineral production process. Our literature review revealed numerous avenues through which the mining sector can contribute to the attainment of the 17 SDGs. These include the promotion of employment (SDG 8), the reduction of poverty (SDG 1), and the alleviation of hunger (SDG 2), among others. The site visits underscored that adherence to SDGs related to job promotion, income improvement for local communities, and environmental infrastructure is a common theme in mining activities, particularly those situated in remote and economically disadvantaged regions. It becomes evident when considering regions that have seen economic decline following the cessation of mining operations. The departure of these visited industries would have a substantial negative impact on local residents who depend on them for their livelihoods. However, we observed certain areas of concern during our visits. These include a notable gender imbalance among the workforce, which raises issues related to SDG 5 concerning gender equality. The inherent nature of the mining activity should not justify such gender-related disparities, which may be influenced by entrenched stereotypes dictating which roles are perceived as suitable for men or women. Additionally, there appears to be an absence of investments in inclusive education for both employees and local residents. Furthermore, these industries do not appear to undertake substantial efforts to restore or rehabilitate degraded areas, potentially undermining the pursuit of SDG 13, which pertains to combating climate change and its associated consequences [3].

The Sustainable Development Goals (SDGs) and the Paris Agreement on Climate Change emphasize the need for substantial changes in all nations, necessitating collaborative efforts from governments, civil society, the scientific community, and businesses. However, there is currently no shared understanding of how to put the 17 SDGs into practical action. Building on prior work by The World in 2050 initiative, we introduce six distinct SDG Transformations, which serve as modular building blocks for achieving the SDGs: (i). Education, Gender, and Inequality, (ii). Health, Well-being, and Demography (iii). Energy Decarbonization and Sustainable Industry (iv). Sustainable Food, Land, Water, and Oceans (v). Sustainable Cities and Communities (vi). Digital Revolution for Sustainable Development. Each Transformation outlines key areas for investment and regulatory challenges, urging specific government entities to collaborate with businesses and civil society to take action. These Transformations can be implemented within existing government structures while acknowledging the interconnectedness among the 17 SDGs. Additionally, we present an action plan for the scientific community to provide the necessary knowledge for designing, executing, and monitoring these SDG Transformations [4].

Our research examined the state of academic inquiry into mining, sustainability, and sustainable development by conducting a systematic review aligned with the Sustainable Development Goals (SDGs). We employed the ISI

Web of Science™ Core Collection database, a widely recognized repository of academic journals, encompassing data from 1945 to 2016 (for complete years). The systematic review process encompassed five key phases: (i). Search Terms (ii). Organization (iii). Metrics and Relations (iv). Classification (v). Synthesis. Our findings indicate that despite a noticeable increase in publications over recent years addressing the intersection of mining and sustainability, the predominant focus of these works remains centered on the environmental aspects of the UN's sustainability goals. This underscores the need for further practical and academic efforts within the mining sector to address the broader spectrum of SDGs and fill the existing gaps in research [5].

Data mining is a valuable tool for the practice of examining large pre-existing databases to generate previously unknown helpful information; in this paper, the input for the weather data set denotes specific days as a row, attributes denote weather conditions on the given day, and the class indicates whether the conditions are conducive to playing golf. Attributes include Outlook, Temperature, Humidity, Windy, and Boolean Play Golf class variables. All the data are considered for training purpose, and it is used in the seven-classification algorithm likes J48, Random Tree (RT), Decision Stump (DS), Logistic Model Tree (LMT), Hoeffding Tree (HT), Reduce Error Pruning (REP) and Random Forest (RF) are used to measure the accuracy. Out of seven classification algorithms, the Random tree algorithm outperforms other algorithms by yielding an accuracy of 85.714% [6].

The primary aim of this paper is to examine the Sustainable Development Goals (SDGs) through diverse independent metrics in the Indian states of Tamil Nadu, Kerala, and Karnataka. This analysis encompasses the utilization of data mining and statistical methodologies, incorporating three distinct state SDGs indices, to uncover concealed insights. Additionally, numerical examples are presented to validate the findings put forth in this study [7].

The authors aim to enrich the understanding of the United Nations Sustainable Development Goals (SDGs) by emphasizing the interplay between science and technology. Their approach involves using SDG classifications from scientific publications to construct a machine learning (ML) model for categorizing the relevance of SDGs in patent documents, which serve as indicators of technology development. This ML model was applied to categorize a subset of patent families registered with the European Patent Office (EPO). The analysis sheds light on the extent to which SDGs are addressed in patents. Additionally, a case study was conducted to explore the ML model's capabilities in identifying the SDG orientation of patents. In alignment with global sustainability objectives, these findings have the potential to enhance the identification of science and technology-related artifacts. Furthermore, they provide insights for aligning research and development efforts, patenting strategies, and the measurement and management of their contributions to the realization of SDGs [8].

Presently, spatial information is quantified for remote sensing data, outlining existing spatial machine learning approaches found in the literature and highlighting prospects for advancing these spatial methodologies. Additionally, we outline a basic set of criteria essential for gauging Sustainable Development Goals (SDGs) using satellite imagery data [9].

Data mining is discovering hiding information that efficiently utilizes the prediction by stochastic sensing concept. This paper proposes an efficient assessment of groundwater level, rainfall, population, food grains, and enterprises dataset by adopting stochastic modeling and data mining approaches. Firstly, the novel data assimilation analysis is proposed to predict the groundwater level effectively. Experimental results are done, and the various expected groundwater level estimations indicate the sternness of the approach [10] and [11].

The input for the chronic disease data denotes a specific location as a row; attributes denote topics, questions, data values, low confidence limit, and high confidence limit. All the data are considered for training and testing using five classification algorithms. In this paper, the authors present the various analysis and accuracy of five different decision tree algorithms; the M5P decision tree approach is the best algorithm to build the model compared with other decision tree approaches [12].

2. Backgrounds and Methodologies

Data mining is the practice of uncovering patterns, trends, relationships, or valuable insights within extensive datasets, and Decision Trees are a well-known technique within data mining, particularly useful for tasks involving classification and prediction. These trees represent a type of supervised machine learning algorithm effectively utilized in the field of data mining. [13].

2.1 Linear Regression

Linear regression is a statistical technique employed to comprehend and forecast the connection between two variables by discovering the optimal straight line that most effectively aligns with the data points. It aids in ascertaining how alterations in one variable correspond to changes in another, proving valuable for predictions and trend recognition. The core idea of linear regression is to find the best-fitting straight line (also called the "regression line") through a scatterplot of data points. This line represents a linear equation of the form:

$$y = m_x + b \qquad \dots (1)$$

Where y is the dependent variable (the one you want to predict or explain), x is the independent variable (the one you're using to make predictions or explanations), m is the slope of the line, representing how much, y changes for a unit change in x and b is the y-intercept, indicating the value of y when x is 0.

2.2 Multilayer Perception

A Multilayer Perceptron (MLP) is an artificial neural network consisting of multiple layers of interconnected nodes or neurons. It's a fundamental architecture in deep learning and is used for various tasks, including classification, regression, and more complex tasks like image recognition and natural language processing. The architecture of an MLP typically includes three types of layers:

i. **Input Layer:** This layer consists of neurons receiving input data. Each neuron corresponds to a feature in the input data, and the values of these neurons pass through the network.
ii. **Hidden Layers:** These layers come after the input layer and precede the output layer. They are called "hidden" because their activations are not directly observed in the final output.
iii. **Output Layer:** This layer produces the network's final output. The number of neurons in the output layer depends on the problem type.

2.3 SMO

SMO stands for "Sequential Minimal Optimization," an algorithm used for training support vector machines (SVMs), machine learning models commonly used for classification and regression tasks. The SMO algorithm is particularly well-suited for solving the quadratic programming optimization problem that arises during the training of SVMs.

Step 1. Initialization
Step 2. Selection of Two Lagrange Multipliers
Step 3. Optimize the Pair of Lagrange Multipliers
Step 4. Update the Model
Step 5. Convergence Checking
Step 6. Repeat

2.4 M5P

M5P is a machine learning algorithm used for regression tasks. It is an extension of the decision tree-based model called M5, which Ross Quinlan developed. The M5 algorithm combines decision trees and linear regression to create more accurate and flexible regression models. M5P, specifically, stands for M5 Prime. It enhances the original M5 algorithm to improve its predictive performance. M5P uses a tree-based model to divide the data into subsets based on feature values recursively and then fits linear regression models to each of these subsets. The result is a piecewise linear regression model, where different linear regressions are used for other regions of the input feature space.

Steps involved in the M5P

Step 1. Building the initial decision tree (M5 model): Recursive Binary Splitting and Pruning (optional)
Step 2. Linear Regression Model: Leaf Regression Models and Model Parameters
Step 3. Piecewise Linear Regression: Piecewise Prediction
Step 4. Model Evaluation: Training and Testing.

2.5 Random Forest

Random Forest is a popular machine learning ensemble method for classification and regression tasks. It is an extension of decision trees and is known for its high accuracy, robustness, and ability to handle complex datasets. Random Forest is widely used in various domains, including data science, machine learning, and pattern recognition. The main idea behind Random Forest is to create an ensemble (a collection) of decision trees and combine their predictions to make more accurate and stable predictions. The following steps describe what Random Forest works like Bootstrap Aggregating (Bagging), Decision Tree Construction and Voting for Classification, Averaging for Regression. The key advantages of Random Forest to Reduced overfitting, Robustness and Feature Importance. The steps involved in building a Random Forest are as follows:

Step 1. Data Bootstrapping
Step 2. Random Feature Subset Selection
Step 3. Decision Tree Construction
Step 4. Ensemble of Decision Trees
Step 5. Out-of-Bag (OOB) Evaluation
Step 6. Hyperparameter Tuning (optional)

2.6 REP Tree

REP (Repeated Incremental Pruning to Produce Error Reduction) Tree is a machine learning algorithm for classification and regression tasks. A decision tree-based algorithm constructs a decision tree using a combination of incremental pruning and error-reduction techniques. The key steps involved in building a REP Tree are as follows Recursive Binary Splitting, Pruning and Repeated Pruning and Error Reduction. Below are the steps involved in building a REP Tree.

Step 1. Recursive Binary Splitting
Step 2. Pruning
Step 3. Repeated Pruning and Error Reduction
Step 4. Model Evaluation

2.7 Accuracy Metrics

The predictive model's error rate can be evaluated by applying several accuracy metrics in machine learning and statistics. The basic concept of accuracy evaluation in regression analysis is comparing the original target with the predicted one and using metrics like R-squared, MAE, MSE, and RMSE to explain the errors and predictive ability of the model [14]. The R-squared, MSE, MAE, and RMSE are metrics used to evaluate the prediction error rates and model performance in analysis and predictions [15] and [16].

R-squared (Coefficient of determination) represents the coefficient of how well the values fit compared to the original values. The values from 0 to 1 are interpreted as percentages. The higher the value is, the better the model is.

$$R^2 = 1 - \frac{\Sigma(y_i - \hat{y})^2}{\Sigma(y_i - \bar{y})^2} \qquad \dots (2)$$

MAE (Mean absolute error) represents the difference between the original and predicted values extracted by averaging the absolute difference over the data set.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}| \qquad \text{... (3)}$$

RMSE (Root Mean Squared Error) is the error rate by the square root of MSE.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2} \qquad \text{... (4)}$$

Relative Absolute Error (RAE) is a metric used in statistics and data analysis to measure the accuracy of a forecasting or predictive model's predictions. It is particularly useful when dealing with numerical data, such as in regression analysis or time series forecasting.

$$\mathrm{RAE} = \frac{\sum|y_i - \hat{y}_i|}{\sum|y_i - \bar{y}|} \qquad \text{… (5)}$$

Root Relative Squared Error (RRSE) is another metric used in statistics and data analysis to evaluate the accuracy of predictive models, especially in the context of regression analysis or time series forecasting.

$$\mathrm{RRSE} = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}} \qquad \text{… (6)}$$

Equation 3 to 7 are used to find the model accuracy, which is used to find the model performance and error. Where Y_i represents the individual observed (actual) values, $\hat{Y}_i$ represents the corresponding individual predicted values, $\bar{Y}$ represents the mean (average) of the observed values and Σ represents the summation symbol, indicating that you should sum the absolute differences for all data points.

Numerical Illustrations

The corresponding dataset was collected from the open source Kaggle data repository. The global data on the sustainable energy dataset include 21 parameters which have different categories of data like Sustainable Development Goals with Various Energy Entity (name of the country), Year, Access to electricity (% of the population), Access to clean fuels for cooking, Renewable-electricity-generating-capacity-per-capita, Financial flows to developing countries (US $), Renewable energy share in the total final energy consumption (%), Electricity from fossil fuels (TWh), Electricity from nuclear (TWh), Electricity from renewables (TWh), Low-carbon electricity (% electricity), Primary energy consumption per capita (kWh/person), Energy intensity level of primary energy (MJ/$2017 PPP GDP), Value_co2_Emissions_kt_by_country, Renewables (% equivalent direct energy),Density\n (P/Km2), Land Area (Km2), Latitude, Longitude, gdp_growth, gdp_per_capita [17]. A detailed description of the parameters is mentioned in the following Table 1.

Table 1(a). Global data on sustainable energy sample dataset

Entity	**Year**	**Access to electricity (% of population)**	**Access to clean fuels for cooking**	**Renewable - electricity - generating-capacity - per-capita**	**Financial flows to developing countries (US $)**	**Renewable energy share in the total final energy consumption (%)**	**Electricity from fossil fuels (TWh)**	**Electricity from nuclear (TWh)**	**Electricity from renewables (TWh)**

Afghanistan	2000	1.613591	6.2	9.22	20000	44.99	0.16	0	0.31
Afghanistan	2001	4.074574	7.2	8.86	130000	45.6	0.09	0	0.5
Afghanistan	2002	9.409158	8.2	8.47	3950000	37.83	0.13	0	0.56
Afghanistan	2003	14.73851	9.5	8.09	25970000	36.66	0.31	0	0.63

Table 1(b). Global data on sustainable energy sample dataset

Low-carbon electricity (% electricity)	Primary energy consumption per capita (kWh/person)	Energy intensity level of primary energy (MJ/$2017 PPP GDP)	Value_co2_Emissions_kt_by_country	Renewables (% equivalent primary energy)	Density\n (P/Km2)	Land Area (Km2)	Latitude	Longitude	gdp_growth	gdp_per_capita
63.44086	252.0691	1.41	1550	0	60	652230	33.93911	67.70995	11.22971	242.0313
76.19048	304.4209	1.5	1760	0	60	652230	33.93911	67.70995	5.357403	263.7336
78.94737	354.2799	1.53	1770	0	60	652230	33.93911	67.70995	13.82632	359.6932
73.9726	607.8335	1.94	3560	0	60	652230	33.93911	67.70995	3.924984	364.6635

Table 2: Machine Learning Models with Correlation coefficient

ML Approaches	**GDP Per Capita**
Linear Regression	0.7589
Multilayer Perceptron	0.9206
SMOreg	0.7220
M5P	0.9225

Random Forest	0.9832
REP Tree	0.9440

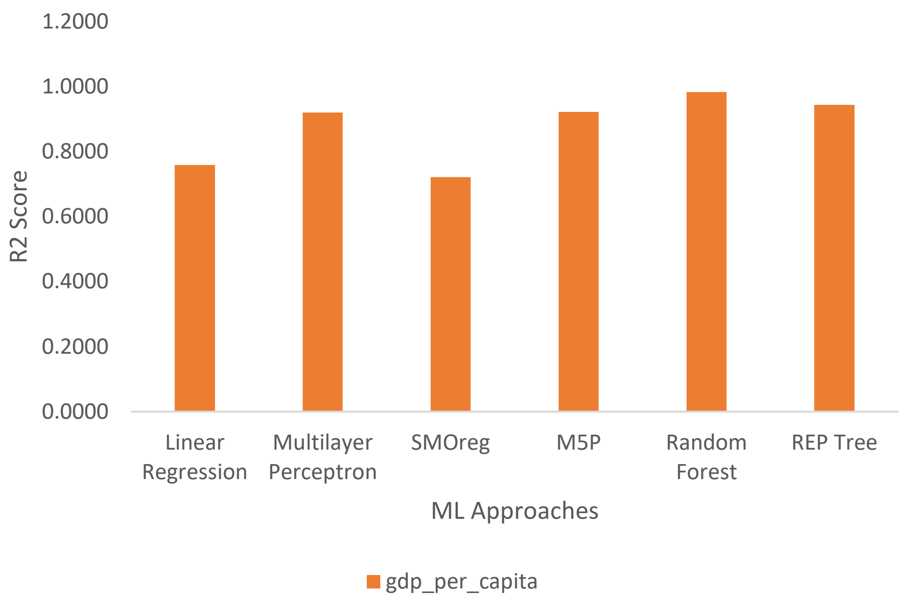

Fig. 1. R2 Score for Machine Learning Approaches

Table 3: Machine Learning Models with Mean Absolute Error

ML Approaches	**GDP Per Capita**
Linear Regression	7712.1782
Multilayer Perceptron	5016.7202
SMOreg	6322.7384
M5P	2631.6264
Random Forest	1648.5053
REP Tree	3004.2536

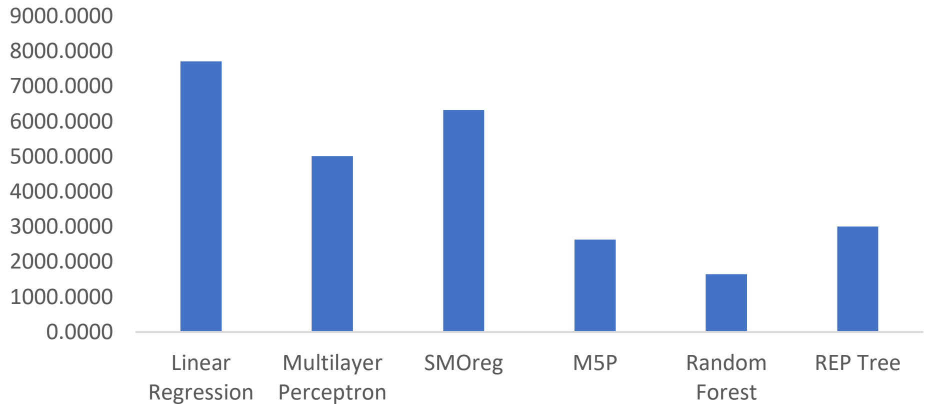

Fig. 2. Machine Learning Models with MAE

Table 4: Machine Learning Models with Root Mean Squared Error

ML Approaches	GDP Per Capita
Linear Regression	12834.2817
Multilayer Perceptron	7858.0937
SMOreg	13942.4033
M5P	7743.2910
Random Forest	3635.4602
REP Tree	6507.6691

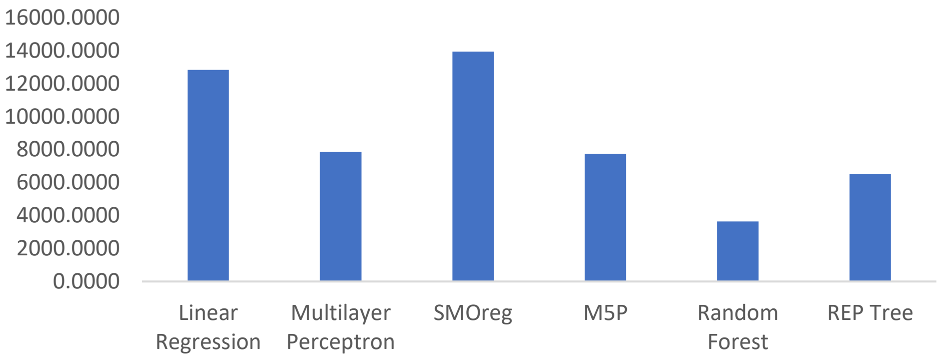

Fig. 3. Machine Learning Models with RMSE

Table 5: Machine Learning Models with Relative Absolute Error (%)

ML Approaches	GDP Per Capita
Linear Regression	55.1533
Multilayer Perceptron	35.8768

SMOreg	45.2168
M5P	18.8199
Random Forest	11.7892
REP Tree	21.4848

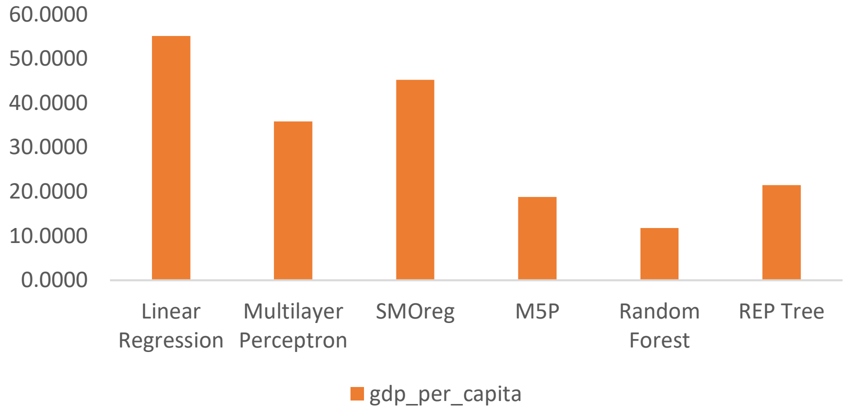

Fig. 4. Machine Learning Models with RAE (%)

Table 6: Machine Learning Models with and Root Relative Squared Error (%)

ML Approaches	GDP Per Capita
Linear Regression	65.1057
Multilayer Perceptron	39.8625
SMOreg	70.7270
M5P	39.2802
Random Forest	18.4420
REP Tree	33.0121

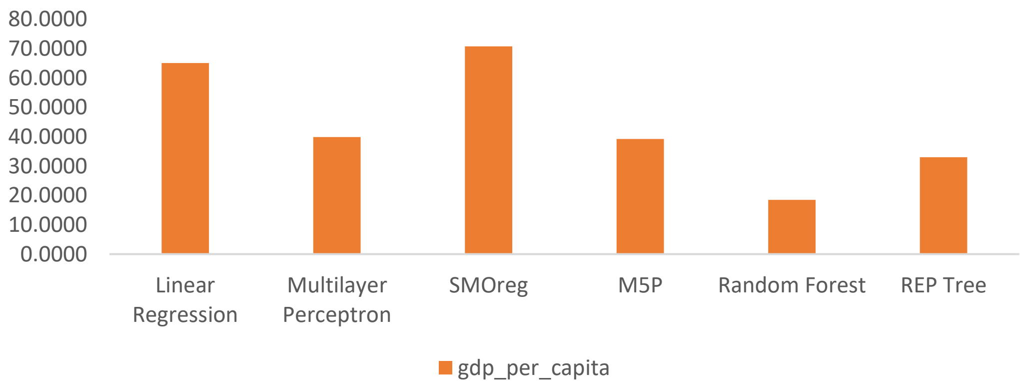

Fig. 5. Machine Learning Models with RRSE (%)

Table 7: Machine Learning Models with Time Taken to Build Model (Seconds)

ML Approaches	GDP Per Capita
Linear Regression	0.0200
Multilayer Perceptron	8.5300
SMOreg	43.7600
M5P	0.3400
Random Forest	1.7400
REP Tree	0.0400

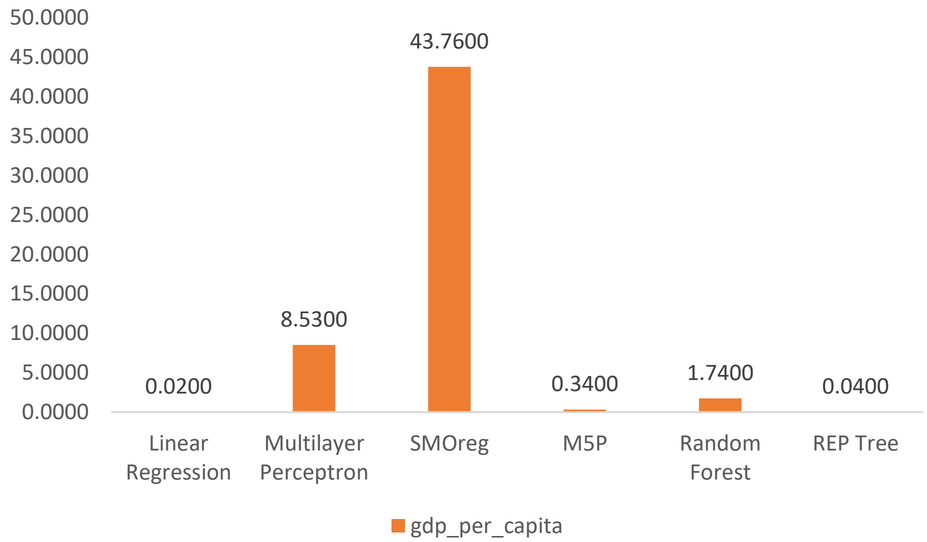

Fig. 6. Machine Learning Models and its Time Taken to Build the Model (Seconds)

3. Result and Discussion

Table 1 explains 21 parameters which include different categories of data like SDG, Meteorology, Types of Land, Chemical fertilizers, Soil Type, and Soil Information. Based on the dataset, it is evident that five different machine learning decision tree approaches are used to find the hidden patterns and which is the best or influencing parameter to decide future predictions. Related results and numerical illustrations are shown between Table 1 to Table 6 and Figure 1 to Figure 4.

They are based on Equation 2, Table 2, and Figure 1, which is used to find the R2 score or correlation coefficient by comparing different parameters. Numerical illustrations suggest that there may be a significant difference from one parameter to another. In thesis research data analysis is fully based on GDP per capita, a measure of a country's economic output per person. In this case, using six different ML approaches returns a robust, strong positive correlation of nearly 0.9.

Further data analysis revealed a gradual improvement in test scores over time. The MAE is used to find model errors using Equations 3 and six machine-learning algorithms will be used in this case. The random forest returns a minimum error for using MAE test statistics. The corresponding results using graphical representation are shown in Table 3 and Figure 2.

The RMSE (root mean square error) measures the difference between predicted and actual values using Equation 4. In this case, the random forest tree returns a minimum error for using RMSE test statistics. SMOreg returns a maximum error. The related numerical illustration is shown in Table 4 and Figure 3.

Relative Absolute Error (RAE) measures accuracy using Equation 5 to compare the difference between predicted and actual values in percentage. In this research, considering Random Forest returns a minimum error. The RAE results are shown in Table 5 and Figure 4. Similar error approaches are reflected in RRSE. Similar numerical illustrations are shown in Table 6 and Figure 5.

Time taken is one of the significant tasks in machine-learning approaches. Based on Table 7 and Figure 6, linear regression, REP tree, M5P, and random forest tree take minimum time to build the model. Finally, multilayer perception and SMOreg take the maximum time to make the model.

4. Conclusion and Further Research

It is essential to consider the limitations of this study. The sample size of each group was relatively small, which could impact the generalizability of the results. Additionally, other variables could influence SDG with energy performance. The findings presented in this study contribute to our understanding that all the parameters return robust positive correlations. In this research, the maximum of the machine learning approaches returns a minimum error with less processing time. Future studies can build upon these, finding the SDG suitable variable for future prediction with increased accuracy using different machine learning and decision tree approaches.

Reference

[1] Hidalgo, A., Gabaly, S., Morales-Alonso, G. and Urueña, A., 2020. The digital divide in light of sustainable development: An approach through advanced machine learning techniques. Technological Forecasting and Social Change, 150, p.119754.

[2] Pincet, A., Okabe, S. and Pawelczyk, M., 2019. Linking Aid to the Sustainable Development Goals–a machine learning approach.

[3] Monteiro, N.B.R., da Silva, E.A. and Neto, J.M.M., 2019. Sustainable development goals in mining. Journal of Cleaner Production, 228, pp.509-520.

[4] Sachs, J.D., Schmidt-Traub, G., Mazzucato, M., Messner, D., Nakicenovic, N. and Rockström, J., 2019. Six transformations to achieve the sustainable development goals. Nature sustainability, 2(9), pp.805-814.

[5] Mesquita, R.F.D., Klein, B., Xavier, A. and Matos, F.R.N., 2017. Mining and the sustainable development goals: a systematic literature review.

[6] Rajesh, P. and Karthikeyan, M., 2017. A comparative study of data mining algorithms for decision tree approaches using the Weka tool. Advances in Natural and Applied Sciences, 11(9), pp.230-243.

[7] Rajesh, P. and Kumar, B.S., 2020. Comparative studies on Sustainable Development Goals (SDG) in India using Data Mining approach. J. Sci, 14(2), pp.91-93.

[8] Hajikhani, A. and Suominen, A., 2022. Mapping the sustainable development goals (SDGs) in science, technology and innovation: application of machine learning in SDG-oriented artefact detection. Scientometrics, pp.1-33.

[9] Holloway, J., Mengersen, K. and Helmstedt, K., 2018. Spatial and machine learning methods of satellite imagery analysis for Sustainable Development Goals. In Proceedings of the 16th Conference of International Association for Official Statistics (IAOS) (pp. 1-14). International Association for Official Statistics (IAOS).

[10] Rajesh, P., Karthikeyan, M. and Arulpavai, R., 2019, December. Data mining approaches to predict the factors that affect the groundwater level using a stochastic model. In AIP Conference Proceedings (Vol. 2177, No. 1). AIP Publishing.

[11] Rajesh, P. and Karthikeyan, M., 2019. Data mining approaches to predict the factors that affect agriculture growth using stochastic models. International Journal of Computer Sciences and Engineering, 7(4), pp.18-23.

[12] Rajesh, P., Karthikeyan, M., Santhosh Kumar, B. and Mohamed Parvees, M.Y., 2019. Comparative study of decision tree approaches in data mining using chronic disease indicators (CDI) data. Journal of Computational and Theoretical Nanoscience, 16(4), pp.1472-1477.

[13] Kohavi, R., & Sahami, M. (1996). Error-based pruning of decision trees. In International Conference on Machine Learning (pp. 278-286).

[14] Akusok, A. (2020). What is Mean Absolute Error (MAE)? Retrieved from https://machinelearningmastery.com/mean-absolute-error-mae-for-machine-learning/

[15] S. M. Hosseini, S. M. Hosseini, and M. R. Mehrabian, "Root mean square error (RMSE): A comprehensive review," International Journal of Applied Mathematics and Statistics, vol. 59, no. 1, pp. 42–49, 2019.

[16] Chi, W. (2020). Relative Absolute Error (RAE) – Definition and Examples. Medium. https://medium.com/@wchi/relative-absolute-error-rae-definition-and-examples-e37a24c1b566

[17] https://www.kaggle.com/datasets/anshtanwar/global-data-on-sustainable-energy