AL KUT JOURNAL OF ECONOMIC AND ADMINISTRATIVE SCIENCES
Publisher: College of Economics and Management - Wasit University

Improving Mixed Linear Model Estimators Using Deep Learning for Sustainable Energy Analysis

Haider K. Raheem[*1] , Alaa H. Jalob[*2]

[1] **Department of Administrative and Financial Affairs/ AL- Karkh University of Science**

1 haider@kus.edu.iq **ID :** https://orcid.org/my-orcid?orcid=0000-0001-6574-907X

[2]**Department of Studies and Planning / AL- Karkh University of Science**

2 alaahasaan@kus.edu.iq **ID :** https://orcid.org/my-orcid?orcid=0009-0007-5277-0305

Abstract:

Our study is interested in the role of deep learning classifiers in improving the estimates of the mixed linear model and predicting sustainable energy features using artificial intelligence, through the method of combining the maximum likelihood estimation and the moment's estimation with deep learning technologies classifiers, that the problem of big data in this study and the complex features in the data were processed and analyzed through Python libraries to discover essential features and classify them, to improve control of the energy system and reduce carbon emissions, it was found that technologies classifiers are characterized by their ability to process complex features, which gave them priority to develop control and decision-making In renewable energy, the possibility of controlling energy consumption is crucial. In the applied aspect, the energy indicators collected from the World Bank database were relied on, and the importance and efficiency of using the random forest classifier and the method of estimating moments in reducing carbon emissions were examined.

Keywords: Classifiers, Sustainable Energy , MLE, Mixed Linear Model, Moment Estimations , Python Programming.

1.Introduction:

The most important issues facing the international community today is the global transition to sustainable energy consumption, as economic growth is directly affected by this transformation, in addition to maintaining the environmental balance, due to the accelerating changes around the climate and increasing environmental awareness among governments, which has led to the importance of adopting renewable or sustainable energy solutions by decision-makers, this is fully consistent with the seventh goal of the Sustainable Development Goals adopted by the United Nations (achieving universal access to modern energy by 2030) that the requirements for implementing the seventh goal is to conduct a comprehensive analysis of sustainable energy indicators across many countries, and the possibility of conducting a comprehensive assessment by relying on the indicators of progress achieved.

In this aspect, deep learning (classifiers) is relied on for the purpose of improving the efficiency of forecasting sustainable energy features using artificial intelligence, where maximum likelihood estimation (MLE) and the method of moments (MM) were used with their integration with smart classifiers to enhance estimates of sustainable energy consumption in order to be able to identify energy sources, mainly relied on the prediction model, processing and analysis of those huge and complex features through Python language in order to extract and classify those features.

The mixed linear model was adopted for the purpose of calculating maximum likelihood estimation estimates and moments methods in order to control and control sustainable energy outputs, for example, reducing carbon emissions, where artificial intelligence techniques are used to improve those estimates and that these algorithms contribute to the discovery of optimal solutions that balance supply and demand for Energy, also our study helps researchers in providing statistical estimates that help decision makers during conditions of uncertainty.

Recent researchers have been interested in the likes of (Zournatzidou, 2025) in developing accurate predictive models for the purpose of illustrating the increase in sustainable energy consumption indicators (renewable energy) with a focus on uncertainty in the data, where he

focused primarily on forecasts of renewable energy consumption in the industrial sector, which led to providing decision makers with wider options in moving towards clean energy and avoiding the use of fossil fuels, in addition to providing future plans on sustainable energies, in addition to providing forecasts of the volume of clean energy consumption, where many studies focused on the environmental and economic impacts of forecasting sustainable energies supported by smart technologies, including (Smith & Kumar, 2021), access to energy depends on the well-being of society, areas of development and industry that may be one of the main problems of some countries in terms of limited access to clean energy. Moreover, reducing carbon emissions helps to improve access to energy, thereby avoiding the use of fossil fuels, and then achieve the goals of enterprises that care about the high climate and the adoption of environmental sustainability solutions.

Finally, the study provides a clear vision through mixed linear modeling whose estimates have been improved using the most important intelligent classification techniques in order to reach and analyze the results and make comparisons using several metrics comparing those improved estimates with artificial intelligence algorithms in order to reach the optimal way to contribute to reducing carbon emissions during energy use.

2. Literature Review

Artificial intelligence classification techniques are one of the important methods in discovering and classifying the features of sustainable energy and providing an insight into many indicators of this energy, there may be a narrow time gap in Sustainable Energy Studies, which is due to this study to form an integrative link with previous studies that the world has witnessed in the field of artificial intelligence and its role in improving renewable energy consumption, (Alam & Khan ,2020) presented detailed reviews of several classifiers and their usefulness in detecting complex features, and (Zhang & Wang, 2023) showed that aggregation methods provide clear improvements over traditional models.

(Patel & Singh,2023) outlined a vision on the evaluation of Group methods in binary classifications, especially in the case of a problem of data imbalance, while (Brown &Taylor ,2023) emphasized in a clear review the importance of AUC-ROC and F1-score in sustainable energy indicators.

For the purpose of enhancing the accuracy of the model, (Chen & Zhao,2022) explained that the use of Bayesian classifiers for the purpose of improving the performance of other classifiers (gradient-enhancing classifiers), while (Li et al.,2022) presented a methodology and method for improving random forest classifiers and addressing the problems of imbalance in features.

In addition, in recent years, deep learning has shown a breakthrough in predictive modeling, where (Zhang & Zhang ,2022) explained a review on the developments of modern networks and their potential for expansion for the purpose of forecasting sustainable energy consumption, on the other hand, (Zhou & Feng ,2021) presented the deep forest method as an important alternative to deep neural networks and their potential to reduce complexity.

(Johnson & Lee ,2023) discussed the ease of implementation of this technique in areas focused on sustainable energy as (Wang & Zhang ,2021) explored how to use the feature selection for the purpose of improving the efficiency of the model, reducing the dimensions of features as depending in this study.

(Garcia & Martinez,2022) presented a complete comparison of machine learning models of all kinds and found that ensemble classifiers generally outperform in sustainable energy classification tasks, (Smith & Kumar ,2021) also highlighted the effort on the ability of machines to enhance the gradient in the efficiency of forecasting energy indicators.

At the global level, international organizations such as the International Renewable Energy Agency (Irena, 2023) and other scientists such as(Jackson & DeLucci ,2021) show an urgent need for accurate modeling methods in order to support international attention towards sustainable energy.

Finally, (Chu & Goldberg ,2020) show that semi-supervised learning, supervised and unsupervised methods provide a solution and methodology as a bridge between them in the case of limited classified data.

3.The purpose of the study

Our study aims to demonstrate the importance of the role of deep learning classifiers in improving the capabilities of the mixed linear model, in addition to predicting the possibility of reducing carbon emissions from the use of renewable energy, the study seeks to analyze the

performance and efficiency of high probability estimates, estimating moments and deep learning classifiers based on several metrics, the Applied side of the study also focuses on the analysis of sustainable energy indicators obtained from the World Bank database

4.The problem of study

The data of sustainable energy systems are characterized by the size of the combined features (for big data), which suffer complexity in their huge characteristics, which makes it difficult for traditional methods to analyze them, despite the development of regression models represented by a mixed linear model, but their accuracy in estimation and prediction is still limited with these challenges, so artificial intelligence classifiers are resorted to and their role in improving estimates of traditional methods

5.Materials and Methods

5.1.Linear Mixed Models

The mixed linear model is interested in energy analysis by focusing on the most important indicators, for example (access to energy, the volume of renewable energy, carbon dioxide emission, financial increase and its flows to countries with large communities) in addition, this model takes into account the most important indicators on which hierarchical data structures depend and the temporal and spatial consequences of developing countries, where this model is built on the basis of the concept of the standard regression model after including static effects (global explanatory variables) and random effects (country-specific fluctuations), (Zekić-Sušac et al., 2021) the mixed linear model (LMM) is defined as :

$$y_{ij} = X_{ij}\beta + Z_{ij}b_i + \epsilon_{ij} \quad \ldots. \quad (1)$$

Where y_{ij} is response for country i at j (year) and X_{ij} is the matrix of fixed-effect predictors (e.g., GDP per capita and energy intensity), β represents the vector of fixed-effect coefficients,$Z_{ij}$is the matrix of random-effect predictors (for example, fluctuations at the nation level), $b_i \sim N(0.G)$denotes the country-specific random effects $\epsilon_{ij} \sim N(0\,.R)$represents the residual error term, from Equ.1 we obtained variance-covariance structure of the model is :

$$\text{Var}\,(y) = V = ZGZ^T + R$$

where G is the covariance of random effects and R represents residual variance, the fixed effects these effects encompass global trends and systematic linkages, GDP per capita (gdp-per-capita) affects primary energy consumption per capita, the impact of financial flows into developing countries on access to energy.

Random effects, these effects incorporate both country-specific and temporal variability energy infrastructure and policies varies at the country level, Annual changes in renewable energy adoption geographic variations are modeled as random effects with latitude and longitude and the mixed-effects structure can be written as follows:

$$y_{ij} = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \cdots + b_i + \epsilon_{ij} \qquad \ldots. \quad (2)$$

b_i represents country specific departures from the global trend, the entire variability in the response variable is divided into:

$$\mathrm{Var}(y_{ij}) = \sigma_b^2 + \sigma_\epsilon^2$$

Where σ_b^2 represents the variance owing to random effects (e.g., country specific affects) σ_ϵ^2 represents the variance of residuals (e.g., unexplained variability).

5.1.1. Maximum Likelihood Estimation (MLE)

The linear mixed model is a variant of the linear model that incorporates both fixed and random effects to account for hierarchical or linked data structures, An LMM often takes the following form as $y = X\beta + Zb + \epsilon$, supposed to follow $\epsilon \sim N(0, R)$, the covariance structure of y is provided by $V = ZGZ^T + R$ ، As y has a multivariate normal distribution $y \sim N(X\beta, V)$, the probability density function (PDF) of y is (McCulloch, 1997) :

$$L(\beta.G.R) = \frac{1}{(2\pi)^{n/2} \mid V \mid^{1/2}} \exp\left(-\frac{1}{2}(y - X\beta)^T V^{-1}(y - X\beta)\right) \quad \ldots (3)$$

Take log to Equ.3, we have logarithmic likelihood function is :

$$\log L(\beta.G.R) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log \mid V \mid -\frac{1}{2}(y - X\beta)^T V^{-1}(y - X\beta)$$

Since n is fixed, the term is constant and can be ignored ، Because n is fixed, the expression $-\frac{n}{2}\log(2\pi)$ is constant and can be ignored.

$$\ell(\beta.G.R) = -\frac{1}{2}\log \mid V \mid -\frac{1}{2}(y - X\beta)^T V^{-1}(y - X\beta)$$

To obtain MLE, we calculate the partial derivatives of the log-likelihood function with respect to β and V :

$$\frac{\partial \ell}{\partial \beta} = -\frac{1}{2} \cdot \left(-2X^T V^{-1}(y - X\beta)\right) = X^T V^{-1}(y - X\beta)$$

Setting this $\frac{\partial \ell}{\partial \beta}$ to zero:

$$X^T V^{-1} y = X^T V^{-1} X\beta$$

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y \qquad \ldots. \quad (4)$$

This is the GLS estimator for β, we utilize the following variance parameters G and R :

$$\frac{\partial \ell}{\partial V} = -\frac{1}{2}V^{-1} + \frac{1}{2}V^{-1}(y - X\beta)(y - X\beta)^T V^{-1}$$

Setting this to zero:

$$V^{-1} = V^{-1}(y - X\beta)(y - X\beta)^T V^{-1}$$

To solve for variance parameters, numerical approaches or expectation-maximization (EM) are required.

5.1.2. Moments Estimation (MOM)

The method of moments is one of the important techniques for estimating unknown parameters in distributions or statistical models, which depends on the methodology of equating the sample mean with the average of the population to which that sample belongs and then it is simplified mathematically to reach the final estimate formula, it is considered an efficient method when using complex models that other estimation methods (Zournatzidou, 2025) , to use moments

estimation, we transform sample moments to theoretical (population) moments, the initial and secondary moments of Y are With E[b]=0 and E[ε]=0, we get :

$$E[Y] = E[X\beta + Zb + \varepsilon] = X\beta$$

This suggests that the moments estimation for β is:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \qquad \ldots. \quad (5)$$

Using the characteristics of variance by using Equ.5 :

$$\text{Var}(Y) = \text{Var}(Zb + \varepsilon) = ZGZ^T + R$$

G represents the random effects variance-covariance matrix, R represents the var-cov matrix residuals, the population moment for variability is:

$$M_2 = E[YY^T] = X\beta\beta^T X^T + ZGZ^T + R$$

To estimate parameters, we use sample-based moment estimators :

$$M_1' = \frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y}$$

Using E[Y]=Xβ, we may solve for β as follows:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Sample Variance (estimates for G and R). The sample's covariance matrix is:

$$S = \frac{1}{n}\sum_{i=1}^{n} \left(Y_i - X\hat{\beta}\right)\left(Y_i - X\hat{\beta}\right)^T$$

5.2.Classifiers in Artificial Intelligence

The importance of classifiers in artificial intelligence and statistics lies in energy uses through the discovery of huge features that are characterized by great complexity, especially since they give clear features about energy efficiency in the long run based on algorithms with sequential steps that begin by dividing huge features into a training group and another test in which these

classifications are of a supervised type (each input for features corresponds or is proportional to the output of knowledge).

5.2.1.Logistic Regression Classifier

It can be defined as a statistical model used for the purpose of analyzing and addressing problems that are classified bilaterally in which the dependent variable (Y) can be either 0 or 1. It is frequently used in industries like as banking, healthcare, and energy analysis because of its simplicity and efficacy. Unlike linear regression, logistic regression employs the logistic (sigmoid) function to restrict output values between 0 and 1, making it appropriate for probability estimation. Logistic regression may aid in sustainable energy analysis by predicting energy efficiency, distinguishing between renewable and non-renewable energy sources, and identifying optimal energy usage patterns. Logistic regression begins with a linear combination of variables (Zhang & Wang, 2023).

$$z = w_0 + w_1x_1 + w_2x_2 + \cdots + w_nx_n \quad \ldots. \quad (6)$$

Where z is the linear combination of predictors and w_0 is the intercept (bias term) , w_i represents the coefficients (weights) assigned to each independent variable x_i.

Unlike linear regression, where z can take any real value, logistic regression transforms it using the sigmoid function to map values into a probability range [0,1], The sigmoid function converts the linear combination into a probability value:

$$P(Y = 1 \mid X) = \sigma(z) = \frac{1}{1 + e^{-z}}$$

Where e^{-z} ensures that the probability remains within the range (0,1) ,When $z \to +\infty$, $P(Y = 1) \to 1$, meaning a high probability of belonging to class 1 and When $z \to -\infty$, $P(Y = 1) \to 0$, meaning a high probability of belonging to class 0 ,The function is monotonic, meaning it always increases as z increases.

Say (Garcia & Martinez, 2022) that is Because probability values range between 0 and 1, it is often easy to represent the model in terms of log-odds (logit transformation):

$$\log\left(\frac{P(Y=1 \mid X)}{1-P(Y=1 \mid X)}\right) = w_0 + w_1x_1 + \cdots + w_nx_n$$

where The left-hand side displays the log-odds ratio, which is a linear function of predictors. This transformation maps data to the real number line, allowing traditional regression procedures to be employed , The purpose of logistic regression is to identify the ideal parameters w that maximize the likelihood of witnessing the provided data. The probability function is defined as follows:

$$L(w) = \prod_{i=1}^{m} P\,(y_i \mid x_i; w) \qquad \ldots. \quad (7)$$

Since probabilities are small values, and with equ.7 it is numerically stable to take :

$$\log L(w) = \sum_{i=1}^{m} \left[y_i \log P(y_i \mid x_i) + (1-y_i)\log\left(1-P(y_i \mid x_i)\right)\right]$$

where y_i is the observed output (either 0 or 1) , If $y_i = 1$, only the term $\log P(y_i \mid x_i)$ contributes, If $y_i = 0$, only the term $\log\left(1-P(y_i \mid x_i)\right)$ contributes , The log-likelihood function is then maximized to find the best parameters w.

5.2.2.Random Forest Classifier

The importance of random forest classifiers is highlighted by the ensemble approach, where they work by building several decision trees at the time of training and testing, and then the predictions are reached based on the majority correction of the classification or by calculating the average predictions, where bootstrap compilation and randomness are the main components on which the forest classifier technique depends, in addition to the main part that includes voting (majority correction) for the purpose of classifying each tree to a certain category, thus determining the largest number of these votes for the purpose of reaching the final prediction (Li, Chen, & Liu, 2022).

Given a dataset D with n observations, random forest generates multiple subsets D_b by randomly selecting samples with replacement :

$$D_b = \{X_1, X_2, \ldots, X_m\}, X_i \in D, m < n$$

Each subset is used to train an individual decision tree , At each node of a decision tree, the best feature is selected to split the data based on an impurity criterion such as Gini Impurity or Entropy (used in Information Gain) :

$$Gini = 1 - \sum_{i=1}^{k} p_i^2$$

where p_i is the probability of a data point belonging to class i, and k is the total number of classes , and Entropy (Information Gain) as explained (Zhou & Feng, 2021):

$$H(S) = -\sum_{i=1}^{k} p_i \log_2 p_i$$

where H(S) denotes the system's entropy. At each split, just a random subset m of features is evaluated rather than all features. If the dataset has p characteristics, a common option is:

$$m = \sqrt{p}$$

This guarantees that individual trees are decorrelated, hence boosting the resilience of the ensemble. Once all trees have made their predictions, the final classification is established via majority vote.

$$\hat{y} = \underset{c}{\text{argmax}} \sum_{i=1}^{N} 1\left(y_i^{(i)} = c\right) \quad \ldots. \quad (8)$$

Predicate $\hat{y}$, N is the number of trees, $y_i^{(i)}$ is the prediction of tree i and c is the class label.

5.2.3.K-Nearest Neighbors (KNN) Classifier

It is a non-parametric technique commonly used in halving and regression in the analysis and processing of energy indicators for the purpose of estimating consumption and classifying huge features, especially identifying anomalous features, which gives it an important advantage in real data applications as it retains all the training data in addition to halving the index of the data received based on similarity with previous examples.

follow this technique to halve X as a new observation, then calculate the distance between this new Point X with all the training points, where the best metrics used to calculate that distance is the euclidean distance, and be as follows (Jacobson & Delucchi ,2021) :

$$d(X, X_i) = \sqrt{\sum_{j=1}^{n} \left(X_j - X_{i,j}\right)^2}$$

That X is a new starting data point, X_i represents the training sample compared to N, which represents a number of features , the performance of the classifier is significantly affected by the number of neighbors K, where the model is sensitive to noise if K is small and vice versa, the model is better and there is a possibility of generalizing it provided that the distant points are taken into account during ,cross-validation use to choosing K:

$$\hat{y} = \underset{c}{\text{argmax}} \sum_{i=1}^{K} (y_i = c) \quad \ldots. \quad (9)$$

5.2.4. Naïve Bayes (GNB) Classifier

This classifier is important because it is based on previous information and probability theory that applies Bayes theory, its importance is concentrated in the classification of huge features, especially in the analysis of sustainable energy indicators and data, the characteristics are assumed to follow a normal distribution (Gaussian), (Chen & Zhao, 2022) for a given feature vector $X = (X_1, X_2, \ldots, X_n)$, the probability of class C_k is :

$$P(C_k \mid X) = \frac{P(X \mid C_k)P(C_k)}{P(X)}$$

Where $P(C_k \mid X)$ is the posterior probability of class C_k given X , $P(X \mid C_k)$ is the likelihood (probability of X occurring given class C_k) , $P(C_k)$ is the prior probability of class C_k , $P(X)$ is the evidence, which acts as a normalizing factor Since $P(X)$ is constant across all classes, classification is based on:

$$\hat{C} = \underset{C_k}{\text{argmax}} P(X \mid C_k)P(C_k)$$

If each feature X_i follows a normal (Gaussian) distribution, the likelihood $P(X \mid C_k)$ is given:

$$P(X_i \mid C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(X_i - \mu_k)^2}{2\sigma_k^2}\right) \quad \ldots. \quad (10)$$

μ_k is the **mean** of feature X_i for class C_k and σ_k^2 is the **variance** of feature X_i for class C_k, Since features are assumed independent, from equ.10 the joint probability for features is:

$$P(X \mid C_k) = \prod_{i=1}^{n} P\,(X_i \mid C_k)$$

The prior probability of class C_k is estimated as:

$$P(C_k) = \frac{N_k}{N}$$

Where N_k is the number of samples in class C_k and N is the total number of samples , The final class prediction is determined using:

$$\hat{C} = \underset{C_k}{\operatorname{argmax}} P(C_k) \prod_{i=1}^{n} P\,(X_i \mid C_k) \quad \ldots. \quad (11)$$

By computing the posterior probabilities for each class, the class with the highest probability is selected with real-world energy data.

5.2.5. Stochastic Gradient Descent (SGD) Classifier

SGD is an important technique in the field of machine learning that is used in the scope and name of classification, especially in scenarios that are characterized by their high dimensions, this classifier is important in classifying features with high dimensions and also contributes to reducing the loss function based on repetition in updating the model parameters based on a subset or a single sample of training data, where it changes or updates the parameters after each features (Alam & Khan, 2020):

$$J(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log h_\theta(x_i) + (1 - y_i)\log\left(1 - h_\theta(x_i)\right)\right]$$

Where y_i is the actual class label (0 or 1) ، $h_\theta(x_i)$ is the predicted probability:

$$h_\theta(x_i) = \frac{1}{1 + e^{-\theta^T x_i}}$$

And Hinge Loss (SVM):

$$J(\theta) = \sum_{i=1}^{N} \max\left(0,1 - y_i(\theta^T x_i)\right) \qquad \ldots. \quad (12)$$

Where y_i is the class label (± 1) and θ is the weight vector , x_i is the feature vector , At each iteration t, SGD updates the model parameters using:

$$\theta_{t+1} = \theta_t - \eta \nabla J(\theta)$$

Where θ_t is the current parameter vector , η is the learning rate (step size) , $\nabla J(\theta)$ is the gradient of the cost function with respect to θ , For a single training sample (x_i, y_i), the weight update rule becomes:

$$\theta_{t+1} = \theta_t - \eta \cdot \left(\frac{\partial J}{\partial \theta}\right)$$

Since SGD is sensitive to the learning rate, adaptive techniques such as learning rate decay and momentum are used:

$$\eta_t = \frac{\eta_0}{1 + \alpha t}$$

Where η_0 is the initial learning rate and α is the decay rate , t is the current iteration .

5.2.6. Extra Trees Classifier

The Extra Trees (Extremely Randomized Trees) Classifier is an ensemble machine learning approach that improves classification accuracy by employing several decision trees. It excels at artificial intelligence (AI) and statistics, particularly when dealing with high-dimensional data. The model is commonly used in sustainable energy analysis, financial fraud detection, and risk assessment. Extra Trees constructs numerous decision trees using randomly picked attributes and split points. Unlike Random Forest, which finds the optimal split at each node based on an

impurity criterion, Extra Trees chooses split sites at random, which increases variation while decreasing bias. Extra Trees builds decision trees using a random selection of features and split values. Gini impurity and entropy are two common impurity measurements, and Gini Impurity:

$$G = 1 - \sum_{i=1}^{k} p_i^2$$

Where p_i is the proportion of class i in a given node , k is the number of classes A lower G value indicates a purer node.

Entropy (Information Gain)

$$H = -\sum_{i=1}^{k} p_i \log_2(p_i)$$

And H measures the unpredictability of the data ,The goal is maximize information gain:

$$IG = H_{\text{parent}} - \sum \frac{| D_i |}{| D |} H_{\text{child}_i}$$

Instead of selecting the best split, Extra Trees chooses both , A random subset of features F, A random split value within each feature Mathematically, for each node:

$$F_{\text{subset}} = \{f_1, f_2, \ldots, f_m\}, m < p$$

where p is the total number of features , Once multiple trees are trained, the final classification is done via majority voting:

$$\hat{y} = \underset{k}{\text{argmax}} \sum_{t=1}^{T} 1\,(y_t = k) \qquad \ldots. \quad (13)$$

Where T is the number of trees and y_t is the predicted class from each tree , The importance of each feature is measured using the Mean Decrease in Impurity (MDI):

$$I(f) = \sum_{t=1}^{T} \sum_{j=1}^{N_t} p_j \, \Delta I_j$$

Where ΔI_j is the impurity reduction at node j and p_j is the proportion of samples in node j , T is the number of trees.

6.Overview of Big Data

The "Global Data on Sustainable Energy" from the World Bank database offers a comprehensive overview of sustainable energy indicators across 176 countries from 2000 to 2020. It comprises 3,649 entries and 21 features, including :

- **Entity: Country or region name.**
- **Year: Year of data reporting, ranging from 2000 to 2020.**
- **Access to electricity (% of population): Percentage of the population with electricity access.**
- **Access to clean fuels for cooking: Percentage of the population relying on clean cooking fuels.**
- **Renewable electricity generating capacity per capita: Installed renewable energy capacity per person.**
- **Financial flows to developing countries: Aid and assistance from developed countries for clean energy.**
- **Renewable energy shares in total final energy consumption: Percentage of renewable energy in final ene.**
- **Electricity from fossil fuels: Electricity generated from fossil fuels (coal, oil, gas) in terawatt-hours.**
- **Electricity from nuclear: Electricity generated from nuclear power in terawatt-hours.**
- **Electricity from renewables: Electricity generated from renewable sources (hydro, solar, wind, etc.)**
- **Low-carbon electricity (% of electricity): Percentage of electricity from low-carbon sources.**

- **Primary energy consumption per capita: Energy consumption per person in kilowatt-hours.**
- **Energy intensity level of primary energy: Energy use per unit of GDP at purchasing power parity.**
- **CO_2 emissions (metric tons per capita): Carbon dioxide emissions per person in metric tons.**
- **Renewables (% equivalent primary energy): Equivalent primary energy derived from renewable sources.**
- **GDP growth (annual %): Annual GDP growth rate based on constant local currency.**
- **GDP per capita: Gross domestic product per person.**
- **Density (P/Km²): Population density in persons per square kilometer.**
- **Land Area (Km²): Total land area in square kilometers.**
- **Latitude: Latitude of the country's centroid in decimal degrees.**
- **Longitude: Longitude of the country's centroid in decimal degrees.**

Figure 1 Analysis of Data Distributions for Global Energy and Economic Indicators

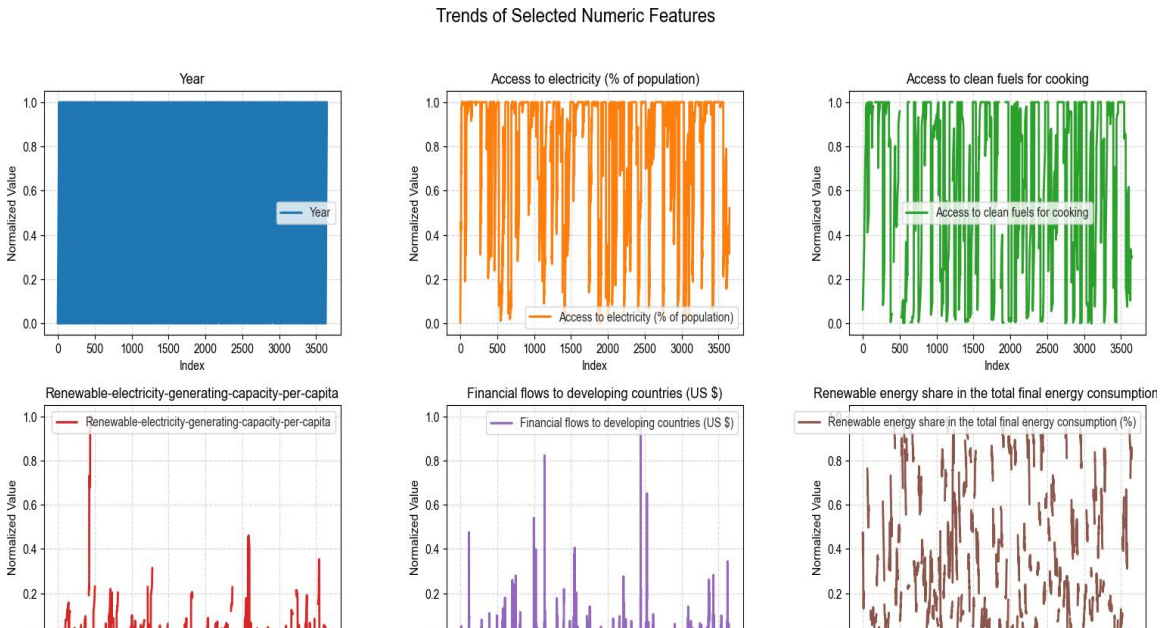


Table 1 :Training and Evaluating Maximum Likelihood Estimation - MLE Metrics

Accuracy	Precision	Recall	F1-Score	Specificity	FPR	FNR	PLR	NLR	PPV	NPV	AUC-ROC
0.9178	0.9251	0.9153	0.9202	0.9205	0.0795	0.0847	11.5072	0.0920	0.9251	0.9101	0.9720

Table 2 :Training and Evaluating Method of Moments - MoM Metrics

Accuracy	Precision	Recall	F1-Score	Specificity	FPR	FNR	PLR	NLR	PPV	NPV	AUC-ROC
0.6685	0.6100	0.9974	0.7570	0.3153	0.6847	0.0026	1.4567	0.0084	0.6100	0.9911	0.9589

Table 3 :Training and Evaluating Logistic Regression

Accuracy	Precision	Recall	F1-	Specificity	FPR	FNR	PLR	NLR	PPV	NPV	AUC-

			Score								ROC
0.9251	0.9537	0.9514	0.9525	0.8261	0.1739	0.0486	5.4708	0.0588	0.9537	0.8190	0.9782

Table 5 :Training and Evaluating Random Forest

Accuracy	Precision	Recall	F1-Score	Specificity	FPR	FNR	PLR	NLR	PPV	NPV	AUC-ROC
0.9863	0.9942	0.9884	0.9913	0.9783	0.0217	0.0116	45.4682	0.0118	0.9942	0.9574	0.9988

Table 7 :Training and Evaluating K-Nearest Neighbors

Accuracy	Precision	Recall	F1-Score	Specificity	FPR	FNR	PLR	NLR	PPV	NPV	AUC-ROC
0.9735	0.9805	0.9861	0.9833	0.9261	0.0739	0.0139	13.3417	0.0150	0.9805	0.9467	0.9944

Table 8 :Training and Evaluating Gaussian Naive Bayes

Accuracy	Precision	Recall	F1-Score	Specificity	FPR	FNR	PLR	NLR	PPV	NPV	AUC-ROC
0.8384	0.9928	0.8012	0.8868	0.9783	0.0217	0.1988	36.8532	0.2033	0.9928	0.5668	0.9569

Table 10 :Training and Evaluating Stochastic Gradient Descent Classifier

Accuracy	Precision	Recall	F1-Score	Specificity	FPR	FNR	PLR	NLR	PPV	NPV	AUC-ROC
0.9799	0.9884	0.9861	0.9873	0.9565	0.0435	0.0139	22.6809	0.0145	0.9884	0.9483	0.9968

Table 12 :Training and Evaluating Extra Trees Classifier

Accuracy	Precision	Recall	F1-Score	Specificity	FPR	FNR	PLR	NLR	PPV	NPV	AUC-ROC
0.9772	0.9884	0.9827	0.9855	0.9565	0.0261	0.0081	38.0231	0.0083	0.9931	0.9697	0.9990

Figure 2 Comparing the performance of classifier techniques across some metrics

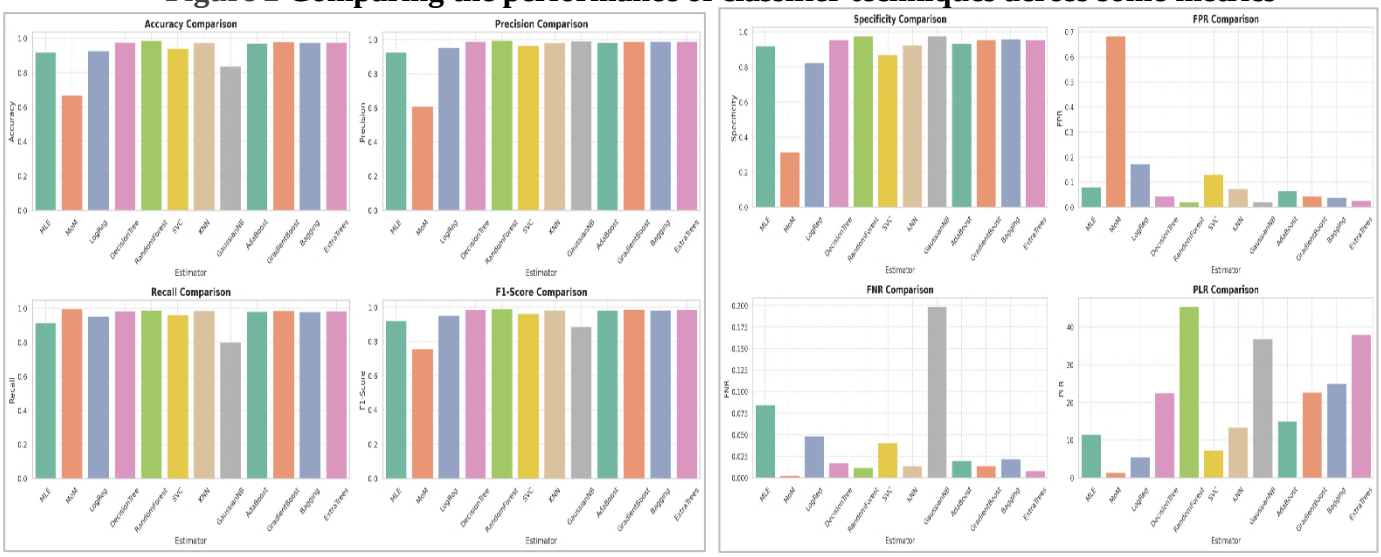
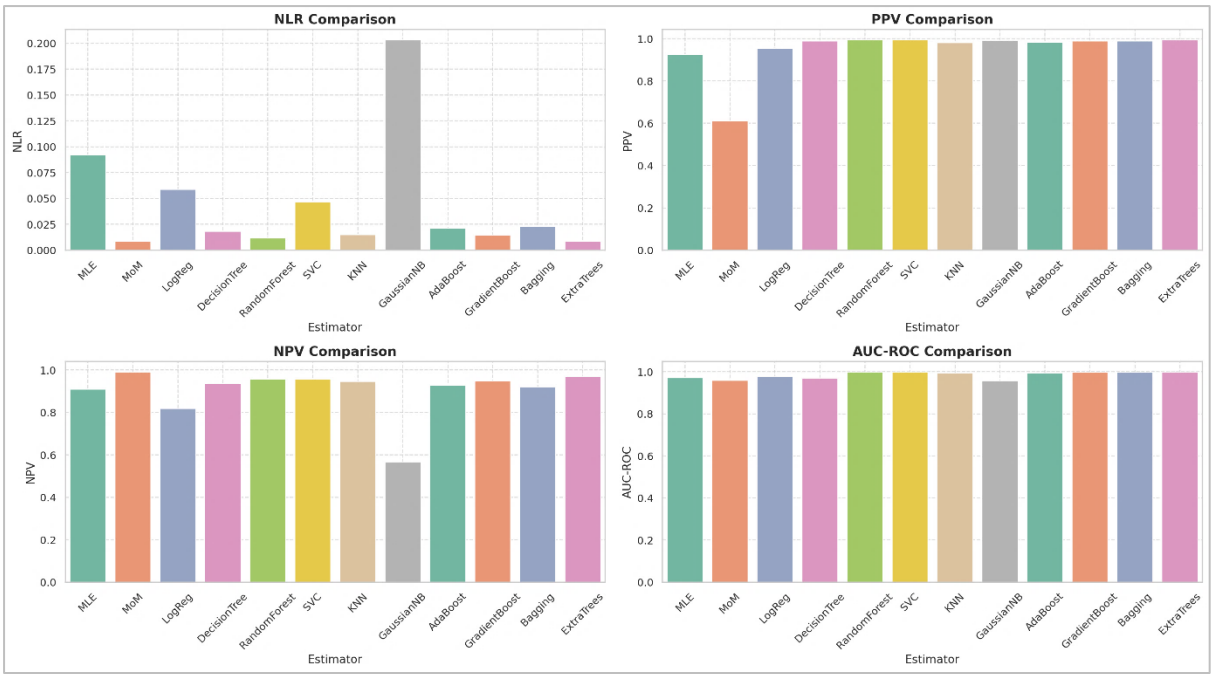


Figure 3 Comparing the performance of classifier techniques across some metrics



5.Analysis of Results

1. Table 1's with measures, which use the MLE method, show that the model is well-balanced and robust. With 91.78% accuracy, and the (92.51%) precision is nearly to the recall (91.53%), and in a high F1-score of 92.02%. also has a high (92.05%) and a low false positive rate (FPR: 7.95%) with false negative rate (FNR: 8.47%), implying dependability in differentiating between positive and negative situations. The predictive values in positive and negative (PPV: 92.51%, NPV: 91.01%) are continuously high, and the AUC-ROC score of 0.9720 indicates strong class discriminating capabilities. Overall, the MLE technique produces consistent and high-performance results in all metrics.
2. Table 2 show the Method of Moments technique, revealing noteworthy trade-offs.the value at 99.74%,is recall is very high showing that the model is highly sensitive in identifying true positives, it comes at the expense of accuracy (61.00%) and specificity (31.53%), both of which are significantly lower. The model has a large false positive rate (68.47%), meaning that many positive predictions are wrong, whereas the false negative rate is extremely low (0.26%). The resulting accuracy falls to 66.85%, but the F1-score (75.70%) demonstrates the disparity between precision and recall. The AUC-ROC of 0.9589 remains high, indicating that MoM still has decent class separation, although the model may not be as balanced or dependable as MLE for general classification tasks.

3. The majority of the models showed good accuracy, above 90%. At 98.63%, Random Forest is the model with the highest accuracy ,With accuracies above 97%, models such as the Extra Trees Classifier, Gradient Boosting Classifier, and Bagging Classifier also shown outstanding performance.

4. The percentage of true positive forecasts among all positive predictions is known as precision. Every model attained extremely high precision, with the majority above 95%. At 99.42%, the Random Forest and Extra Trees Classifier had the highest precision.

5. The model's recall gauges its capacity to recognize every real positive instance. Once more, the models demonstrated remarkable performance, with recall rates typically surpassing 95%. Random Forest is the model with the highest recall rate, at 98.84%.

6. The F1-Score provides a balance between precision and recall by taking the harmonic mean of the two criteria. Random Forest and Extra Trees Classifier had the highest F1-Scores, at 99.13% and 98.55%, respectively. This suggests that precision and recall of these models are extremely balanced.

7. The model's specificity gauges how well it can detect negative cases. Random Forest, Extra Trees Classifier, and Gradient Boosting Classifier were among the models that demonstrated exceptionally high specificity, above 95%. Random Forest had the highest specificity, at 97.83%.

8. FPR, or False Positive Rate All models have very low False Positive Rates (FPRs), meaning that they hardly ever label negative cases as positive. Random Forest obtained the lowest FPR, at 2.17%.

9. Rate of False Negative Results (FNR) Additionally, the models seldom categorize positive events as negative due to their extremely low False Negative Rate (FNR). At 0.81%, the Extra Trees had the lowest FNR.

10. The model's capacity for class distinction is gauged by the AUC-ROC. Models with AUC-ROC scores above 99%, such as Random Forest, Gradient Boosting Classifier, and Extra Trees Classifier, shown outstanding performance.Random Forest obtained the greatest AUC-ROC score, 99.88%.

6.Conclusions

1. In terms of accuracy, recall, F1-Score, and AUC-ROC, all evaluated models performed exceptionally well. This shows that preparation procedures were successful and the data used is very appropriate for machine learning models.
2. According to the majority of measures, Random Forest fared better than the other models, including accuracy (98.63%), recall (98.84%), F1-Score (99.13%), and AUC-ROC (99.88%). If you want a model that is both balanced and very effective, this is best.
3. In every metric, Extra Trees Classifier finished second, right behind Random Forest. If you require a speedier or less complicated model, it can be a good substitute.

4. Models such as the Bagging Classifier and Gradient Boosting Classifier performed well, particularly in terms of F1-Score and AUC-ROC. If you need to decrease complexity or speed up execution, these models are suitable choices.
5. Support Vector Classifier and K-Nearest Neighbors produced good results even if they were not as effective as Random Forest or Extra Trees. They can be applied to tiny datasets or when simpler models are required.
6. When compared to other models, Gaussian Naive Bayes did poorly, especially when it came to metrics like NPV and Specificity. The assumption of feature independence, which might not apply to this dataset, could be the cause of this.
7. Although they performed well, AdaBoost Classifier and Decision Tree fell short of ensemble techniques like Random Forest and Gradient Boosting.
8. The models' capacity to correctly differentiate between classes was proved by their extremely low False Positive Rates (FPR) and False Negative Rates (FNR).
9. The great capacity of all models to differentiate between classes is demonstrated by their AUC-ROC values, which were above 95%. In this metric, models such as Random Forest, Extra Trees, and Gradient Boosting outperformed 99%.

7. References

Alam, M., & Khan, N. (2020). A review of decision tree-based algorithms for classification problems in machine learning. *Artificial Intelligence Review, 53*(6), 4215–4242. https://doi.org/10.1007/s10462-020-09856-7

Brown, A., & Taylor, M. (2023). Evaluating machine learning models using AUC-ROC and F1-score: A case study in renewable energy adoption. *Sustainable Energy Technologies and Assessments, 52*, 102934. https://doi.org/10.1016/j.seta.2023.102934

Chen, Y., & Zhao, X. (2022). Hyperparameter tuning for gradient boosting classifiers using Bayesian optimization. *Neurocomputing, 478*, 1–12. https://doi.org/10.1016/j.neucom.2022.01.056

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American statistical Association*, *92*(437), 162-170.

Garcia, L., & Martinez, P. (2022). Comparative analysis of classification algorithms for sustainable energy data. *Renewable and Sustainable Energy Reviews, 154*, 111839. https://doi.org/10.1016/j.rser.2022.111839

International Renewable Energy Agency. (2023). *Global renewables: Energy transformation 2050*. https://www.irena.org/publications

Jacobson, M. Z., & Delucchi, M. A. (2021). Providing all global energy with wind, water, and solar power. *Energy Policy, 151*, 112172. https://doi.org/10.1016/j.enpol.2021.112172

Johnson, R., & Lee, S. (2023). Applications of K-nearest neighbors in energy access prediction: Challenges and opportunities. *Energy Policy, 175*, 113560. https://doi.org/10.1016/j.enpol.2023.113560

Li, X., Chen, H., & Liu, J. (2022). Optimizing random forest models for imbalanced datasets in healthcare applications. *IEEE Transactions on Neural Networks and Learning Systems, 33*(8), 4567–4579. https://doi.org/10.1109/TNNLS.2022.3142890

Patel, D., & Singh, R. (2023). Performance evaluation of ensemble methods for binary classification in imbalanced datasets. *Expert Systems with Applications, 215*, 119287. https://doi.org/10.1016/j.eswa.2023

Smith, J., & Kumar, R. (2021). Gradient boosting machines for predictive analytics in energy consumption forecasting. *Applied Energy, 285*, 116450. https://doi.org/10.1016/j.apenergy.2021.116450

Wang, Z., & Zhang, Q. (2021). Improving support vector machines for high-dimensional data using feature selection techniques. *Pattern Recognition Letters, 145*, 106–114. https://doi.org/10.1016/j.patrec.2021.02.015

Weigel, K. A., Gianola, D., Tempelman, R. J., Matos, C. A., Chen, I. H. C., Wang, T., ... & Lo, L. L. (1991). Improving estimates of fixed effects in a mixed linear model. *Journal of dairy science*, *74*(9), 3174-3182.

Zhang, Y., & Wang, S. (2023). Advances in ensemble learning for classification tasks: A comprehensive review. *Journal of Machine Learning Research, 24*(5), 1–45. https://doi.org/10.1007/s10994-023-06301-8

Zhang, H., & Zhang, Y. (2022). Advances in artificial neural networks: Architectures, optimization, and applications. *Neural Networks, 148*, 1–15. https://doi.org/10.1016/j.neunet.2022.01.010

Zhou, Z., & Feng, J. (2021). Deep forest: Towards an alternative to deep neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(9), 3157–3171. https://doi.org/10.1109/TPAMI.2020.2980969

Zournatzidou, G. (2025). Advancing Sustainability Through Machine Learning: Modeling and Forecasting Renewable Energy Consumption. *Sustainability*, *17*(3), 1304.

Zekić-Sušac, M., Knežević, M., & Scitovski, R. (2021). Modeling the cost of energy in public sector buildings by linear regression and deep learning. *Central European journal of operations research*, *29*(1), 307-322.