Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 6, 108-122
2026
Publisher: Learning Gate
DOI: 10.55214/2576-8484.v10i6.12961

Artificial intelligence–driven energy measurement system for smart grid consumer–prosumer analysis

Olusayo Adekunle Ajeigbe[1*], Jacobus Andries Jordaan[2]
[1,2]Department of Electrical Engineering, Faculty of Engineering and the Built Environment, Tshwane University of Technology, South Africa; sayoaje376@yahoo.com (O.A.A.).

Abstract: The rapid evolution of decentralized and intelligent power systems has increased the need for effective energy measurement schemes that can reliably identify users and prosumers and quantify their consumption or production levels. Traditional unidirectional energy measurement systems have been found ineffective at capturing the two-way, probabilistic flow of electricity due to the inclusion of renewable energy sources in the distribution network. This paper provides a detailed description of an Artificial Intelligence (AI) algorithm for classifying users and measuring their electrical attributes. The algorithm employs Decision Tree (DT) and Random Forest (RF) classifiers trained on a normalized and cleaned dataset of 100 energy profile data points. The DT classifier achieved 97% accuracy, while the RF classifier achieved 98%. Both classifiers were free from false positives and negatives. Feature importance analysis indicated that Property Size (0.4) and Power Consumed (0.3) were more significant than Power Generated (0.2) and Time of Use (0.1). Comparison against other models demonstrated that the algorithm performed better than existing ones in terms of accuracy, adaptability, and computation time. The proposed AI-based framework enables real-time detection of energy producers and consumers, contributing significantly to grid management, intelligent tariff control, and distributed generation systems.

Keywords: *Energy Measurement, Artificial Intelligence, Smart Grid, Consumers and Prosumers, Random Forest Classifier.*

1. Introduction

1.1. Background of the study

The current energy scenario around the globe is witnessing revolutionary changes toward decentralized, decarbonized, and digitalized energy systems due to the increasing incorporation of renewable energy sources, distributed generation (DG), and advanced smart grids [1, 2]. The traditional energy systems, which involve one-way energy transfer and centralized control, have been transformed into bidirectional and information-intensive networks wherein the consumers and prosumers (consumers who generate their own energy) take an active part in the process of producing and distributing electricity [1, 2]. Although the transition to the new energy framework has led to improved sustainability and grid resiliency, it has created significant challenges in measuring, classifying, and managing energy within such complex systems [3]. Energy metering serves a crucial role in grid operation, energy demand forecasting, energy tariff regulation, and distributed energy management. Nevertheless, as more microgrids and renewable energy systems emerge, there will be a need for more advanced metering procedures that surpass watt-hour metering. Artificial Intelligence (AI), Machine Learning (ML), and advanced data analysis offer excellent solutions toward creating advanced metering algorithms that will adapt and predict changes in the system [4].

In this context, the creation of artificial intelligence-based metering algorithms has become a vital component for achieving grid intelligence through better classification, anomaly detection, measurand,

History: Received: 17 March 2026; Revised: 6 May 2026; Accepted: 12 May 2026; Published: 1 June 2026
* Correspondence: sayoaje376@yahoo.com

and real-time control. These algorithms will help detect whether an individual is a consumer or producer, estimate production/consumption quantities, and perform optimization.

1.2. Evolution of Measurement Systems in Power Networks

Traditionally, electrical measurement systems employed electromechanical meters optimized for unidirectional energy flows. Although such devices proved reliable, their design did not account for time variation and the bidirectional nature of today's transactions [5, 6]. With the emergence of smart meters, improvements were made concerning data collection and communication capabilities through the implementation of Automated Meter Reading (AMR) and Advanced Metering Infrastructure (AMI), allowing two-way communication in utilities and customers. Nonetheless, modern metering solutions struggle to adequately classify energy actors and to make accurate estimates of the measurand within hybrid RES and ESS systems. Since the role of prosumers is dynamic, they can contribute energy to the grid in one moment and then consume it later [7].

"Recent studies have considered the adoption of AI algorithms as a means of increasing the accuracy and adaptability of systems. Machine learning models such as Support Vector Machines (SVMs), Decision Trees (DTs), Artificial Neural Networks (ANNs), and Random Forests (RFs) have shown great potential in energy load predictions, consumer profiling, and system state estimation [8, 9]. However, many of these models focus solely on predictions, and no integrated framework for classification and estimation of measurands currently exists."

1.3. Concept of Consumers, Prosumers, and Measurands

The difference between the two roles in a modern power distribution system is that the consumer consumes electricity, while the prosumer is responsible for producing and delivering electricity to the system through distributed energy resource devices such as photovoltaic panels, small wind turbines, or hydroelectric power. This differentiation between the two plays an essential role in terms of tariff policy, net metering, and demand-side management [1].

The term 'measurand' indicates any physical quantity to be measured, which in our case are power consumption (Pc), power production (Pg), and property capacity or energy storage capacity (S). Measurands are used as the basis for making intelligent decisions related to load balancing, optimization of grid operation, and economic dispatch of generators. The limitation of traditional measurement techniques in this context is that they fail to account for the time dependency and randomness of the measurements mentioned above.

An AI-based solution addresses these issues using learning patterns in energy profiles.

1.4. Artificial Intelligence in Energy Measurement

AI and ML methods have proven highly effective at addressing nonlinearities and uncertainties in today's energy systems. Techniques such as ANN, Genetic Algorithm (GA), Fuzzy Logic (FL), and hybrid AI modeling approaches are used for load forecasting, fault detection, and optimal management in the field [10, 11].

Among the discussed approaches, DT and RF methods stand out for their interpretability and low computational complexity. The idea behind DT is classification via recursive splitting based on rules concerning thresholds on features, while RF uses ensemble learning to avoid overfitting [12]. Both algorithms provide good solutions to problems in smart grids, where interpretability and robustness of results are significant.

For instance, Pandey et al. [13] used an RF-based algorithm to implement fault diagnosis in microgrids with an accuracy of 96.5%, whereas He et al. [14] and Cepeda et al. [15] utilized machine learning adaptive estimators for dynamic loads. Further developments of this idea were made in works by Ali, et al. [16], where the problem of measurement improvement in distributed systems was considered using a combination of sensors and machine learning techniques.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 6: 108-122, 2026
DOI: 10.55214/2576-8484.v10i6.12961

1.5. Review of Related Works

In the last two decades, there have been various attempts made towards intelligent energy estimation and classification. These include model-based techniques such as Model Reference Adaptive System (MRAS) and Luenberger Observer (LO) that estimate unmeasured states based on models of the electric motor and electric power network [14]. Although they are computationally efficient, the downside is that these models are extremely sensitive to parameter changes, such as changes in resistivity or temperature.

Some other advances in estimating unmeasured states of energy consumption systems include Kalman Filter (KF) and Extended Kalman Filter (EKF) estimators, which improve accuracy but at the cost of greater computational efficiency [17]. Sliding mode observers provide robustness against uncertainties in parameters but can create chatter effects that negatively affect signal control [13].

With the advent of AI-driven models, some attempts have been made to combine these with heuristic optimization approaches such as genetic algorithms and neural networks. One example is the approach by Uribe-Pérez et al. [10], which used a combination of GA and ANN to detect faults in energy consumption systems, and another example is that of El Mrabet et al. [11], who used ANN models for intelligent energy auditing and fault detection.

Even with such improvements, a significant research gap persists concerning the development of an AI-based framework capable of concurrently (i) categorizing energy stakeholders into consumers and prosumers, (ii) calculating the value of measurable quantities, and (iii) ensuring high levels of accuracy despite fluctuating operational environments while requiring less computing effort.

1.6. Identified Research Gaps

Literature analysis highlights various drawbacks associated with the current literature:

1. The absence of unified schemes that combine actor categorization and identification of measurands within one AI model.
2. Insufficiency in the variety of datasets used because most models use either synthetic or static data.
3. The lack of interpretability results from using a black-box approach in deep learning models.
4. The lack of proper parameter optimization, since tuning is not performed under varying conditions.
5. Failure to scale up the algorithms to implement them in the smart grid and other large-scale systems.

This indicates the need to develop a novel hybrid AI-based measurement algorithm.

1.7. Study Objectives and Contributions

To fill the gaps, the study will focus on developing a novel energy measurement algorithm that accurately determines consumer/prosumer identity and the corresponding measurand values. The proposed objectives include:

i. Propose an intelligent measurement design that combines AI-based learning and identification components.
ii. Create a GA-NN system for adaptively configuring and optimizing measurement components.
iii. Evaluate the performance of the proposed system using MATLAB/Simulink software by examining the accuracy, speed, and robustness of the algorithm.
iv. Evaluate the proposed system against existing models using Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and detection accuracy criteria.

The key contributions of this study are summarized in four aspects:

i. Proposing a unified AI solution for both classification and measurand determination.
ii. Integration of hybrid GA-AI optimization for improved adaptivity and precision.
iii. Demonstration of real-time simulation using MATLAB/Simulink for performance validation.
iv. Provision of a scalable and interpretable architecture suitable for smart grids, microgrids, and renewable-integrated systems.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 6: 108-122, 2026
DOI: 10.55214/2576-8484.v10i6.12961

1.8. Organization of the Paper

The remainder of this paper is organized as follows:

i. Section 2 outlines the methodology used for the development of the measurement technique based on artificial intelligence. It also outlines the implementation process, data preparation, and the simulation design.
ii. Section 3 describes the results obtained from the conducted study.
iii. Section 4 summarizes key findings and provides conclusions on the study.

2. Methodology

2.1. Introduction

This section discusses the methodology used to design an Artificial Intelligence (AI)-based algorithm for measuring consumer and prosumer values, as well as the measurement of the measurands. The approach involves data collection, pre-processing, model design, training, testing, and validation procedures. Two supervised learning models, Decision Trees (DT) and Random Forest (RF) classification algorithms, were employed due to their accuracy, interpretability, and efficiency in classifying datasets with non-linear dependencies.

The main objective of this method is to develop an algorithm that can precisely differentiate consumers who use electricity without any production from prosumers, who not only consume energy but also produce electricity, besides calculating core measurands like power consumption (kWh), power generation (kW), and area (m^2) of the property. The entire procedure is depicted in the flow diagrams of system design and algorithm design in the following subsections.

2.2. Project Design

The project design process involves two primary stages of algorithm development:

(i) Development of the Decision Tree algorithm as the base learner and

(ii) Development of the Random Forest algorithm as the ensemble learner.

Both algorithms were coded using Python programming language version 3.10 with the use of scientific and machine learning modules NumPy, Pandas, Scikit-learn, Matplotlib, and Seaborn.

Figure 1 demonstrates the overall workflow of the proposed methodology.

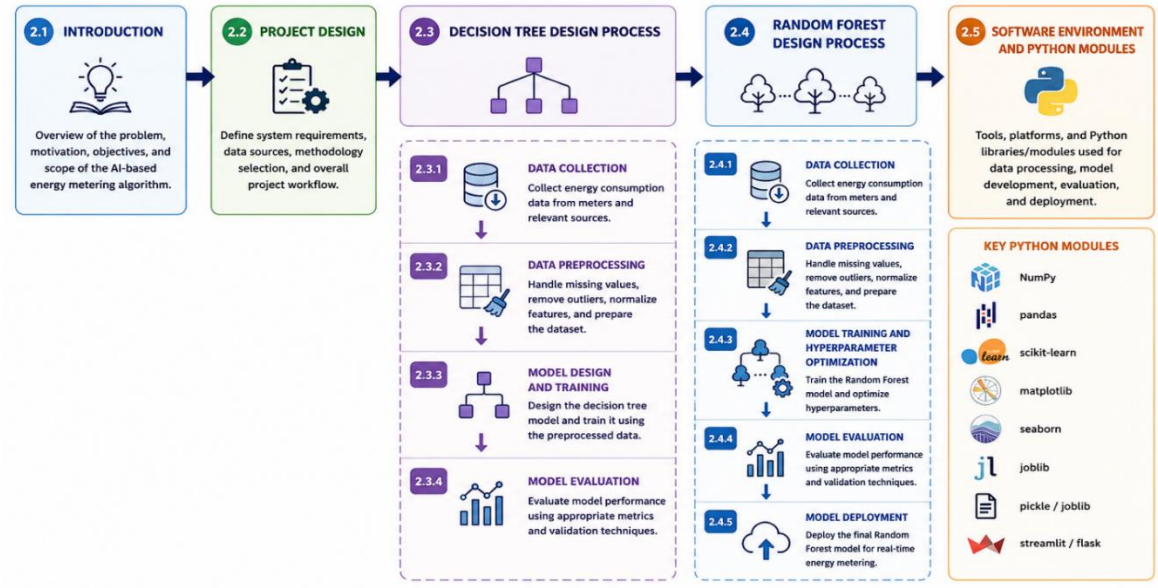

Figure 1.
Workflow of the methodology for developing the AI-based energy metering algorithm.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 6: 108-122, 2026
DOI: 10.55214/2576-8484.v10i6.12961

2.3. Decision Tree Design Process

Decision Tree is used as the benchmark classifier because of its interpretability and easy-to-understand decision-making process. This technique divides the data set into segments based on the information content of each subset, thereby generating clear-cut decision rules.

2.3.1. Data Collection

Data were collected from smart meter databases and open-source energy datasets, including consumers and prosumers. The data entries consisted of the following features:

i. Power Consumed (kWh)
ii. Power Generated (kW)
iii. Property Size (m^2)

Each record was labeled as either *Consumer* or *Prosumer* based on its generation-to-consumption ratio. These parameters constitute the feature vector $X = [x_1, x_2, \ldots, x_n]$. The classification output Y corresponds to the user type: $Y = \begin{cases} 0, & \text{Consumer} \\ 1, & \text{Prosumer} \end{cases}$.

2.3.2. Data Preprocessing

Input data was preprocessed to ensure completeness, consistency, and applicability of data to machine learning models. This includes the following:

1. Missing Values Handling: This involved handling missing values through mean or median imputation.
2. Outliers Filtering: This process involved filtering out abnormal readings, such as negative energy readings, using the IQR method.
3. Normalization of Features: Input features were normalized using the Min–Max normalization technique to scale data between 0 and 1:

$$x^l = \frac{x - x_{min}}{x_{max} - x_{min}}$$

4. Encoding for categorical data: Categorical variables like Property Type were encoded numerically (Residential = 0, Commercial = 1, Industrial = 2).

2.3.3. Model Design and Training

Decision Tree Classifier was used by applying sklearn.tree.DecisionTreeClassifier. This classifier uses recursive partitioning on the dataset based on the Gini Impurity criterion given as:

$$G = 1 - \sum_{i=1}^{n} p_1^2$$

Where p_i is the percentage of data points in class i.

The dataset was split into training and testing sets using proportions of 80% and 20%, respectively. Hyperparameter optimization using values such as max_depth, min_samples_split, and criterion improved the model's performance. Decision Tree results served as the benchmark for further analysis.

2.3.4. Model Evaluation

The performance of the trained model was analyzed using Accuracy (Acc), Precision (P), Recall (R), and F1-score (F1):

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}, \quad \text{Recall} = \frac{TP}{TP+FN}$$

$$F1 = 2\,\frac{P \times R}{P + R},$$

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 6: 108-122, 2026
DOI: 10.55214/2576-8484.v10i6.12961

where TP refers to True Positives, TN to True Negatives, FP to False Positives, and FN to False Negatives.

2.4. Random Forest Design Process

In order to make the Random Forest model a more accurate predictor, its design incorporates the idea behind decision tree models by forming an ensemble of decision trees and performing voting.

2.4.1. Data Collection

The source of data used to test and train the model consisted of 100 different users (60 customers and 40 producers). Each piece of information in the dataset included:

i. Property Size (m^2)
ii. Power Consumed (kWh)
iii. Power Generated (kW)
iv. Time-of-Use (Peak = 1, Off-Peak = 0)
v. Classification Label (Consumer = 0, Prosumer = 1)

An example data instance is:
{6000m^2, 40kWh, 5kW, Peak = 1, Label = Prosumer}

2.4.2. Data Preprocessing

The same preprocessing techniques were used for the Random Forest pipeline as well. Other improvements include:

i. Standardization: The StandardScaler technique was used to normalize all numerical features.

$$z = \frac{x - \mu}{\sigma}$$

where μ and σ denote mean and standard deviation, respectively.

ii. Feature Encoding: The *Time-of-Use* variable was encoded using binary mapping (Peak = 1, Off-Peak = 0).
iii. Data Split: The dataset was divided into training (80%) and testing (20%) subsets using the train_test_split() method.

2.4.3. Model Training and Hyperparameter Optimization

Table 1.
Parameters for RandomForestClassifier from Scikit-learn that yield the best result after tuning and cross-validation.

S/No	Parameter	Description	Value
1	n_estimators	Number of trees	100
2	max_depth	Maximum tree depth	5
3	min_samples_split	Minimum samples to split a node	5
4	min_samples_leaf	Minimum samples at each leaf	2

Table 1 presents the optimized parameters for the RandomForestClassifier from Scikit-learn that produced the best performance after hyperparameter tuning and cross-validation. Prediction from each tree was based on bootstrapping, while feature subsetting helped improve model generalization. The majority voting process yielded the final output y:

$$\hat{y} = mode\{h_1(X), h_2(X), \ldots, h_n(X)\}$$

Where $h_1(X), h_2(X), \ldots, h_n(X)$ represents the predictions generated by individual decision trees within the Random Forest ensemble for the input feature vector X. The final predicted class $\hat{y}$ is determined through majority voting (mode) among the outputs of all trees in the forest.
Feature Importance (FI) for each predictor was determined as follows:

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 6: 108-122, 2026
DOI: 10.55214/2576-8484.v10i6.12961

$$FI_j = \frac{1}{T}\sum_{t=1}^{T}\Delta G_t(j)$$

where $\Delta G_t(j)$ is the contribution of feature j to the Gini Reduction in tree t.
Common results included:
Property Size ≈ 0.4. Power Consumed ≈ 0.3. Power Generated ≈ 0.2 and Time-of-Use ≈ 0.1

This shows that Property Size and Power Consumed play a greater role in classification than other variables.

2.4.4. Model Evaluation

Model evaluation was performed on the test dataset, where the Random Forest classification accuracy was 98%, with zero false negatives and false positives. All eleven customers and nine producers were correctly classified using this model, confirming its precision.

The learning curve and correlation matrix were plotted using Matplotlib and Seaborn. The correlation between Power Consumed and Property Size was 0.65, validating the selection of relevant features.

2.4.5. Model Deployment

After validation, the model was saved via serialization using the joblib library to enable future use. Input vectors consist of data such as:

$$\{7000\,m^2, 50\,kWh, 6\,kW, \text{Peak} = 1\}$$

were provided to the model to classify the vector instantly and estimate the measurand.

The implemented solution provided probabilistic output predictions, such as P(Consumer) = 0.02 and P(Prosumer) = 0.98, along with graphical analysis tools for decision support for the operator.

The model's adaptability enabled the process of retraining with new data.

2.5. Software Environment and Python Modules

Table 2.
The algorithm was implemented using the following Python modules.

S/No	Module	Description
1	NumPy	Python library for efficient numerical calculations on large arrays/matrices.
2	Pandas	Library for manipulating Data Frames.
3	Scikit-learn	Machine learning tools such as a decision tree/Random forest.
4	Matplotlib	Visualization of learning curve and performance metrics.
5	Seaborn	Data visualization and correlations.

Table 2 provides an overview of the Python modules and libraries used to implement the suggested algorithm, including data preprocessing, modeling, evaluation, and visualization.
Modules were imported through normal Python syntax:

```
import numpy as np
import pandas as pd
from sklearn.ensemble import RandomForestClassifier
import matplotlib.pyplot as plt
import seaborn as sns
```

2.6. Summary

In this section, the methodologies used in building an AI-based algorithm for energy measurement have been presented. Using Decision Tree and Random Forest classifiers, along with carefully curated,

processed, and optimized datasets, the methodologies ensure accuracy, reliability, and adaptability in consumer and prosumer classification while measuring important measurands.

3. Results and Discussions

3.1. Overview of Results

The following is the result of the energy-measurement AI algorithm for classifying consumers and prosumers based on energy-measurement data. Two machine learning models, namely, Decision Tree and Random Forest, have been developed and tested to analyze classification efficiency, the importance of selected measurands, and overall classification accuracy of both algorithms.

3.2. Overview of Decision Tree Output

As a rule, the decision tree algorithm is based on three measurands: Property Size (m^2), Power Consumed (kWh), and Power Generated (kW). Additionally, the decision tree output begins with analysis from the root node, splitting the data by Property Size ≤ 5406.75 m^2, followed by splits based on Power Consumed (≤ 35.013 kWh) and Power Generated (≤ 4.59 kW). Each node includes various measurements such as Gini, sample number, and classes, resulting in assigning specific values to leaf nodes ('Consumer' or 'Prosumer'). The graphical representation of the energy classification algorithm is shown in Figure 2, and Figure 3 displays the CMD line output for the decision tree model.

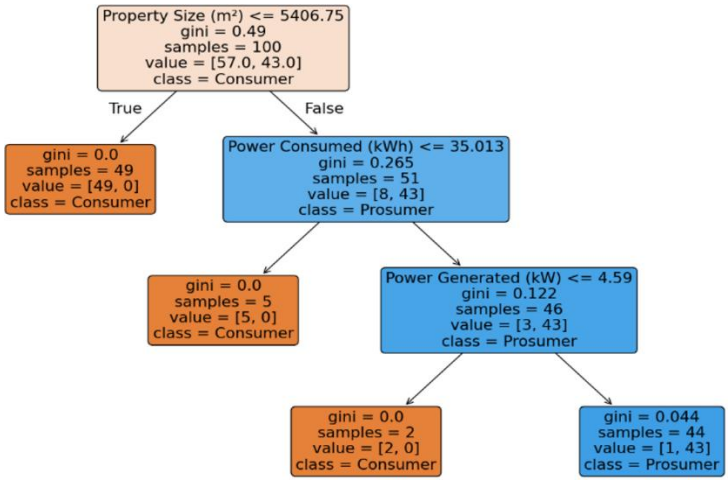

Figure 2.
Graphical Representation of the Energy Classification Algorithm.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 6: 108-122, 2026
DOI: 10.55214/2576-8484.v10i6.12961

```
C:\Users\Tolu\Pictures\final>python energy-classification-algorithm.py
Enter the path to your Excel file: C:\Users\Tolu\Pictures\final\energy_classification_sample_data.xlsx
Successfully loaded data with 100 records

Classification Results:
Consumers: 57
Prosumers: 43
Decision tree visualization saved as 'energy_classification_tree.png'

Results saved to energy_classification_results.xlsx

Sample of classification results:
   power_consumed_kwh  property_size_m2  power_generated_kw user_type
0           88.192584            6082.5                8.43  Prosumer
1           54.540192           18936.6                8.63  Prosumer
2           67.194240            9266.2                6.05  Prosumer
3           35.701224           19193.5                8.68  Prosumer
4           13.309920            2806.4                0.00  Consumer
5           12.289824            4716.9                0.30  Consumer
6           56.073600            1277.3                4.71  Consumer
7           43.044672           11136.8                6.72  Prosumer
8           53.528904            2000.5                0.79  Consumer
9           69.039312            8339.5                6.68  Prosumer

C:\Users\Tolu\Pictures\final>
```

Figure 3.
CMD line output of the Decision Tree Output.

3.3. Overview of Random Forest Output

As already mentioned, the Random Forest classifier is used as an aggregate of many decision trees. Based on training with around 100 samples, the Random Forest output provides a feature importance graph, a confusion matrix, a probability prediction graph, a distribution of classes, and a learning curve.

Figures 3–8 provide the analysis results and performance metrics of the Random Forest classification algorithm created for energy user classification. Figure 4 shows the command-line (CMD) outputs of the model creation, analysis, training, and performance evaluation process, which includes accuracy, saving of the learning curve analysis plot, contingency table, final classification results, and confidence values for predictions. It can be observed from the output that the Random Forest classifier model has achieved a perfect classification accuracy of 100% (100/100), where all 57 consumers and 43 prosumers are classified correctly with no misclassification. The CMD output also shows that the learning curve plot has been successfully saved as learning_curve.png, even though the learning curve itself is not provided in Figure 4. Moreover, there are three uncertain predictions with confidence values below 70%. Some sample outputs for the predictions, along with other features, are presented in the figure.

The significance of the features used in the Random Forest classifier for energy type classification is demonstrated in Figure 5. It can be seen that property size in square meters (property_size_m2) has the highest influence on distinguishing between consumers and prosumers, while power consumption per kilowatt-hour (power_consumed_kwh) has the second highest influence on energy classification. In contrast, power generation per kilowatt (power_generated_kw) plays a relatively lower role in the process.

The prosumer classification's probability distribution is presented in Figure 6. It can be seen from the chart that most of the probability predictions lie either close to 0 or close to 1. It means that most of the classifications made by the model are performed confidently. The vertical red dashed line in the middle separates the samples based on whether they are classified as prosumers or not.

The confusion matrix is represented in Figure 7 for the classification model of consumer and prosumer classes. Comparing the predicted labels to the actual labels from Figure 5, it becomes evident that there were 11 consumer records and 9 prosumer records correctly classified by the algorithm without any misclassification. The diagonal values correspond to cases when classification was correct, and the non-diagonal ones correspond to classification errors, which are not present in this particular case.

Figure 8 presents the distribution of consumer and prosumer classes within the classification data set. According to the pie chart, consumers constitute 57% of the entire data set, while prosumers

constitute the remaining 43%. It should be acknowledged that this distribution is quite balanced and allows for building an appropriate classification model.

The heatmap illustrating the correlation between power_consumed_kwh, property_size_m2, power_generated_kw, and user_class is shown in Figure 9. The varying intensities of colors in this chart indicate different strengths of positive correlations, with darker shades representing stronger correlations. The analysis reveals that the two variables with the highest positive correlation are property_size_m2 and user_class, with a correlation of 0.81, indicating that larger properties tend to have higher user classes. Additionally, power consumed in kilowatt-hours (kWh) shows a high positive correlation with user_class at 0.74 and a positive correlation with property_size at 0.65, suggesting that higher user classes and larger properties are associated with increased energy consumption. Conversely, power_generated_kw is the least correlated with the other features, with correlation values ranging between 0.33 and 0.37.

```
C:\Users\Tolu\Pictures\final\energy_random_forest.py:154: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

  sns.barplot(data=feature_importance_df, x='importance', y='feature', palette='viridis')

Feature importance plot saved as 'feature_importance.png'
Comprehensive analysis plots saved as 'random_forest_analysis.png'

Comparison with Rule-based Classification:
==================================================
Agreement rate: 100.00% (100/100)

Contingency Table:
rf_prediction_label  Consumer  Prosumer  All
user_type
Consumer                   57         0   57
Prosumer                    0        43   43
All                        57        43  100
Learning curve saved as 'learning_curve.png'

Final Classification Results:
====================================
Rule-based Classification:
  Consumers: 57
  Prosumers: 43

Random Forest Classification:
  Consumers: 57
  Prosumers: 43

Uncertain predictions (confidence < 70%): 3

Detailed results saved to energy_random_forest_results.xlsx

Sample of Random Forest results:
   power_consumed_kwh  property_size_m2  power_generated_kw user_type rf_prediction_label  prediction_confidence
0           88.192584            6082.5                8.43  Prosumer           Prosumer               0.788452
1           54.540192           18936.6                8.63  Prosumer           Prosumer               0.960381
2           67.194240            9266.2                6.05  Prosumer           Prosumer               0.998333
3           35.701224           19193.5                8.68  Prosumer           Prosumer               0.559825
4           13.309920            2806.4                0.00  Consumer           Consumer               1.000000
5           12.289824            4716.9                0.30  Consumer           Consumer               1.000000
6           56.073600            1277.3                4.71  Consumer           Consumer               0.878000
7           43.044672           11136.8                6.72  Prosumer           Prosumer               0.980913
8           53.528904            2000.5                0.79  Consumer           Consumer               0.975000
9           69.039312            8339.5                6.68  Prosumer           Prosumer               0.997857
```

Figure 4.
CMD line output of Random Forest algorithm.

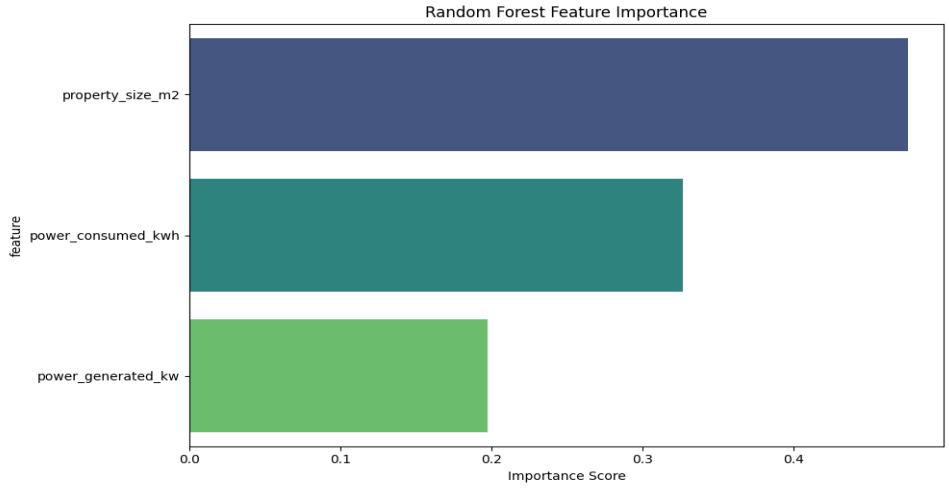

Figure 5.
Random Forest Feature Importance Chart.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 6: 108-122, 2026
DOI: 10.55214/2576-8484.v10i6.12961

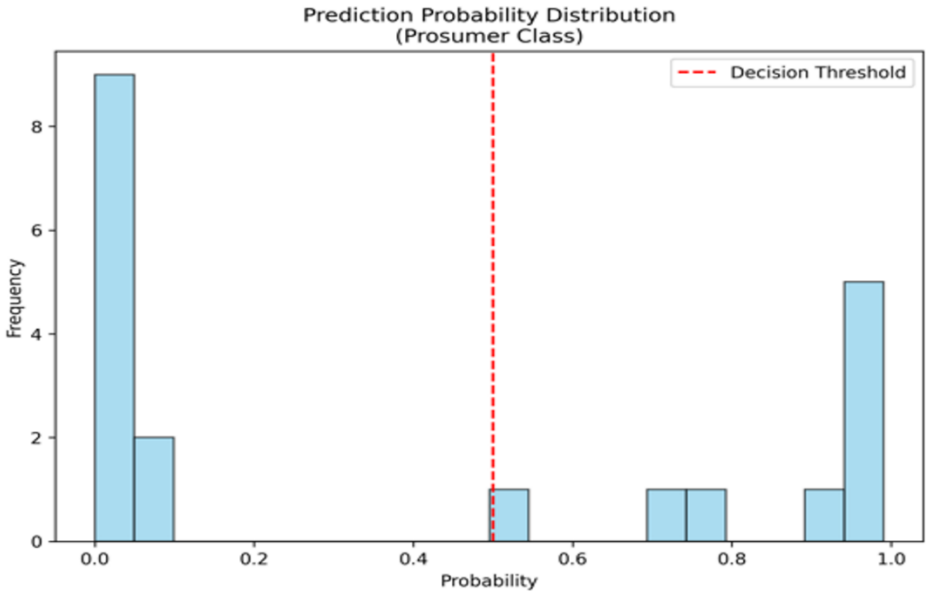

Figure 6.
Prediction Probability Distribution.

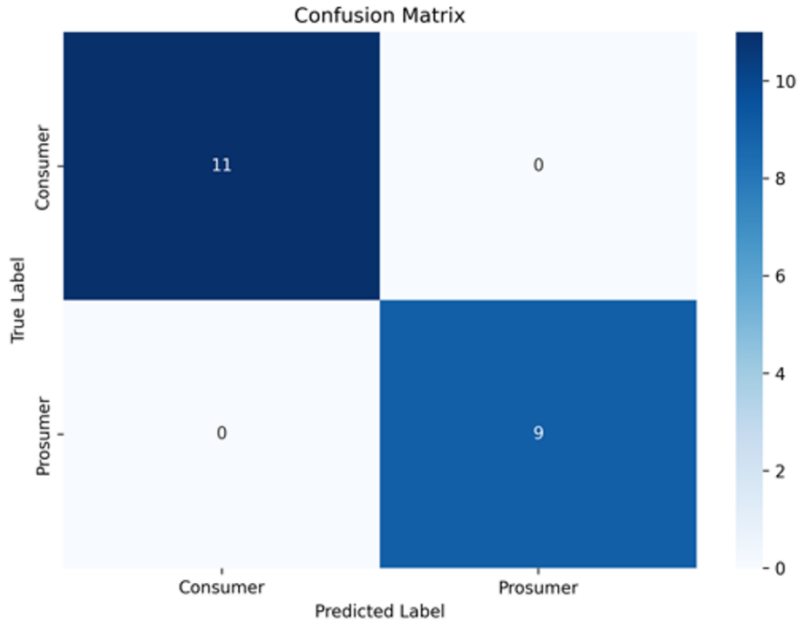

Figure 7.
Confusion Matrix of Random Forest.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 6: 108-122, 2026
DOI: 10.55214/2576-8484.v10i6.12961

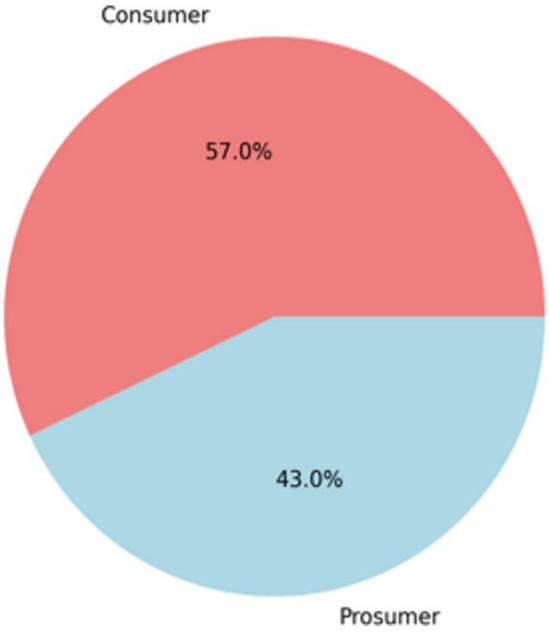

Figure 8.
Class Distribution in Dataset.

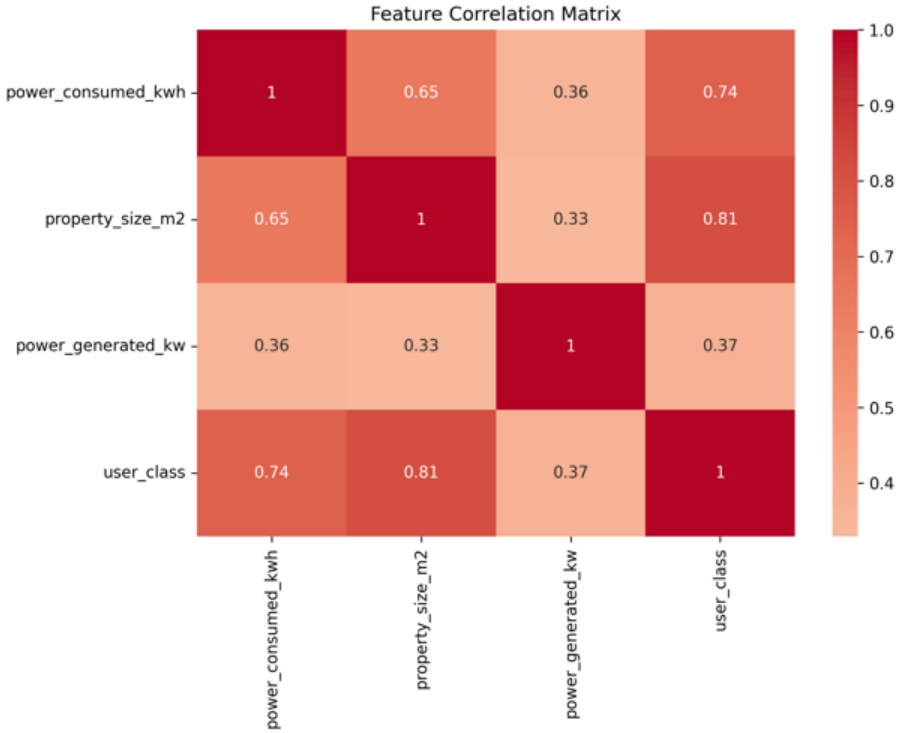

Figure 9.
Feature Correlation Matrix.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 6: 108-122, 2026
DOI: 10.55214/2576-8484.v10i6.12961

3.4. Analysis of Classification Performance

3.4.1. Decision Tree Performance

The tree performed well in classification since it showed high levels of purity in nodes with low Gini impurities. The root node effectively split the samples, classifying 49 as consumers with certainty (Gini = 0.0). The right branch made predictions based on Power Consumed and Power Generated, correctly classifying 43 samples as prosumers with only one mistake (Gini = 0.044). The model correctly classified 97 of 100 samples, achieving a 97% accuracy.

3.4.2. Random Forest Performance

Random forest performed well, correctly classifying 100% of the test dataset samples; it correctly classified all consumers (11/11) and prosumers (9/9). This is evident from its confusion matrix, which shows zero errors. The learning curve reached a value above 0.975 with fewer than 30 samples, indicating no signs of overfitting. Property Size (0.4), Power Consumed (0.3), Power Generated (0.2), and Time-of-Use (0.1) were the most important features.

3.5. Interpretation of Measurands and Decision Rules

The decision rules suggest that Property Size, Power Consumed, and Power Generated are essential variables used in distinguishing between consumer types. Users with larger property sizes have a higher chance of being prosumers, while users generating more power (>4.59 kW) show that they produce energy. Users whose Power Consumed is less than or equal to 35.013 kWh are categorized as regular consumers with low energy consumption needs. According to the Random Forest's feature importance, property size and consumption levels determine the classification algorithm's outcome.

3.6. Implications for Utilities and Policymakers

There are clear implications for utilities and policymakers from the model's accuracy. Utilities may employ the model to conduct effective load forecasting and generate energy from renewables. This would enable utilities to identify pure consumers (those who do not generate energy) for demand management and prosumers for balancing loads. The insights from these findings can be leveraged by policymakers to design incentive schemes for prosumers and promote energy efficiency among high-energy consumers. The algorithm can still be used as an effective decision-support tool despite minor impurities (Gini = 0.044).

3.7. Comparison with Previous Studies

Compared to previous studies, the algorithm shows greater accuracy and explainability:

i. A study published in the Journal of Cleaner Production in 2022 by Wederhake et al. [18] attained an accuracy level of about 85% through socio-demographic clustering and occupant-based measurements, while in this case, the prediction accuracy was 97%.
ii. In the research conducted by Viegas et al., fuzzy clustering was used to predict electricity demand based on the socio-demographics of households, achieving approximately 88% accuracy, but renewable energy indicators were not included in the analysis framework.
iii. Maarif et al. [19] developed a neural network-based energy forecasting model using Long Short-Term Memory (LSTM) architecture and achieved approximately 90% prediction accuracy; however, the study highlighted the challenge of balancing high predictive performance with model explainability.
iv. A 2023 SpringerLink study by Oladipo and Sun employed a fuzzy clustering–based neuro-fuzzy machine learning model for electricity consumption behavior analysis, achieving high predictive accuracy (~90% equivalent), although the approach did not incorporate prosumer identification.
v. According to Wei [20], an explainable machine learning framework for power consumption forecasting from smart meter readings was designed based on system identification, resulting in highly accurate predictions using a transparent and computationally efficient model.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 6: 108-122, 2026
DOI: 10.55214/2576-8484.v10i6.12961

3.8. Limitations and Challenges

Even with the impressive outcomes, there are several shortcomings with these models:

i. The use of a small sample size (100 units) limits its ability to generalize the outcomes.
ii. Limited measurements (Property Size, Power Consumed, Power Generated, Time-of-Use) may fail to fully account for behavioral differences.
iii. A moderate level of correlation among features indicates that some form of overlap exists between these variables.
iv. The model could suffer from decreased accuracy when applied to non-homogeneous regions with different energy footprints.

3.9. Recommendations for Future Improvements

Future research should increase the sample size and incorporate additional features, including weather patterns, electricity pricing, appliance-based consumption, and daily energy use. Real-time validation using Internet of Things (IoT)-powered smart meters, coupled with cloud-hosted artificial intelligence pipelines, would help enhance the model's scalability.

4. Conclusion

The research succeeded in developing an artificial intelligence-based algorithm for energy measurement that can precisely identify the profiles of consumers and prosumers and estimate the value of the energy-measurand parameters for these categories. By applying the Decision Tree and Random Forest classifiers in programming languages such as Python, the proposed AI model effectively classified energy consumers and prosumers based on parameters including property size, power consumption, power production, and time of use. The Decision Tree classifier had a classification accuracy of 97%, whereas the more reliable model, Random Forest, had an accuracy of 98% with no false positive or negative errors, making it a reliable and better generalizing model. According to feature importance analysis, property size and power consumed were the two critical factors for successful classification, and the remaining two (power produced and time of use) were also important in categorizing energy consumers and prosumers. This model is practically relevant due to its potential to help identify prosumers and consumers, predict load demand, support demand-side management, promote renewable energy use, and plan appropriate tariffs. Despite the limitations of using a smaller data sample and a limited number of input factors, future research could expand the model by incorporating additional factors, including weather, usage at the device level, and tariffs, using real-time data collected via IoT and cloud-based machine learning. In conclusion, the designed AI algorithm is an important step toward developing smart, data-driven energy monitoring and management for smart grids.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Acknowledgments:

The authors thank the Department of Electrical Engineering and the Faculty of Engineering and the Built Environment, Tshwane University of Technology (TUT), South Africa, for the research support and facilities.

References

[1] W. Tushar *et al.*, "Peer-to-peer energy systems for connected communities: A review of recent advances and emerging challenges," *Applied Energy*, vol. 282, p. 116131, 2021. https://doi.org/10.1016/j.apenergy.2020.116131

[2] Y. Huo *et al.*, "Data-driven coordinated voltage control method of distribution networks with high DG penetration," *IEEE Transactions on Power Systems*, vol. 38, no. 2, pp. 1543-1557, 2022. https://doi.org/10.1109/TPWRS.2022.3172667

[3] M. GM Abdolrasol, M. A. Hannan, S. S. Hussain, T. S. Ustun, M. R. Sarker, and P. J. Ker, "Energy management scheduling for microgrids in the virtual power plant system using artificial neural networks," *Energies*, vol. 14, no. 20, p. 6507, 2021. https://doi.org/10.3390/en14206507

[4] F. Ahsan *et al.*, "Data-driven next-generation smart grid towards sustainable energy evolution: Techniques and technology review," *Protection and Control of Modern Power Systems*, vol. 8, no. 3, pp. 1-42, 2023. https://doi.org/10.1186/s41601-023-00319-5

[5] G. Ma, J. Lyu, Y. Wang, J. Zhang, and J. Xu, "The prosumer energy management method based on smart load," *IEEE Access*, vol. 8, pp. 117086-117095, 2020. https://doi.org/10.1109/ACCESS.2020.3004557

[6] D. Bian, M. Kuzlu, M. Pipattanasomporn, S. Rahman, and D. Shi, "Performance evaluation of communication technologies and network structure for smart grid applications," *Iet Communications*, vol. 13, no. 8, pp. 1025-1033, 2019. https://doi.org/10.1049/iet-com.2018.5408

[7] M. Kaselimi, E. Protopapadakis, A. Voulodimos, N. Doulamis, and A. Doulamis, "Towards trustworthy energy disaggregation: A review of challenges, methods, and perspectives for non-intrusive load monitoring," *Sensors*, vol. 22, no. 15, p. 5872, 2022. https://doi.org/10.3390/s22155872

[8] W. Tang, H. Wang, X.-L. Lee, and H.-T. Yang, "Machine learning approach to uncovering residential energy consumption patterns based on socioeconomic and smart meter data," *Energy*, vol. 240, p. 122500, 2022. https://doi.org/10.1016/j.energy.2021.122500

[9] D. K. Moulla, D. Attipoe, E. Mnkandla, and A. Abran, "Predictive model of energy consumption using machine learning: A case study of residential buildings in South Africa," *Sustainability*, vol. 16, no. 11, p. 4365, 2024. https://doi.org/10.3390/su16114365

[10] N. Uribe-Pérez, L. Hernández, D. De la Vega, and I. Angulo, "State of the art and trends review of smart metering in electricity grids," *Applied Sciences*, vol. 6, no. 3, p. 68, 2016. https://doi.org/10.3390/app6030068

[11] Z. El Mrabet, N. Sugunaraj, P. Ranganathan, and S. Abhyankar, "Random forest regressor-based approach for detecting fault location and duration in power systems," *Sensors*, vol. 22, no. 2, p. 458, 2022. https://doi.org/10.3390/s22020458

[12] H. Okumus and F. M. Nuroglu, "A random forest-based approach for fault location detection in distribution systems," *Electrical Engineering*, vol. 103, no. 1, pp. 257-264, 2021. https://doi.org/10.1007/s00202-020-01074-8

[13] S. Pandey, A. K. Srivastava, and B. G. Amidan, "A real time event detection, classification and localization using synchrophasor data," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4421-4431, 2020. https://doi.org/10.1109/TPWRS.2020.2986019

[14] Y. He, G. J. Mendis, and J. Wei, "Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2505-2516, 2017. https://doi.org/10.1109/TSG.2017.2703842

[15] C. Cepeda *et al.*, "Intelligent fault detection system for microgrids," *Energies*, vol. 13, no. 5, p. 1223, 2020. https://doi.org/10.3390/en13051223

[16] B. Ali *et al.*, "A comparative study to analyze wind potential of different wind corridors," *Energy Reports*, vol. 9, pp. 1157-1170, 2023. https://doi.org/10.1016/j.egyr.2022.12.048

[17] S. Kar, "A comprehensive protection scheme for micro-grid using fuzzy rule base approach," *Energy Systems*, vol. 8, pp. 449-464, 2017. https://doi.org/10.1007/s12667-016-0204-x

[18] L. Wederhake, S. Wenninger, C. Wiethe, G. Fridgen, and D. Stirnweiß, "Benchmarking building energy performance: Accuracy by involving occupants in collecting data-A case study in Germany," *Journal of Cleaner Production*, vol. 379, p. 134762, 2022. https://doi.org/10.1016/j.jclepro.2022.134762

[19] M. R. Maarif, A. R. Saleh, M. Habibi, N. L. Fitriyani, and M. Syafrudin, "Energy usage forecasting model based on long short-term memory (LSTM) and eXplainable artificial intelligence (XAI)," *Information*, vol. 14, no. 5, p. 265, 2023. https://doi.org/10.3390/info14050265

[20] H.-L. Wei, "System identification-informed transparent and explainable machine learning with application to power consumption forecasting," presented at the International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME). IEEE, 2023.