SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I11P115
Volume 11 Issue 11, 168-178, November 2024

Original Article

IoT-Enhanced Machine Learning for Intelligent Energy Optimization and Predictive Management

D. Rajalakshmi[1], K. Sudharson[2], R. Akhil Nair[3], M.A. Starlin[4]

[1]*Department of CSE, R.M.D. Engineering College, Tamil Nadu, India.*
[2]*Department of AIML, R.M.D. Engineering College, Tamil Nadu, India.*
[3]*Department of CSE, Velammal Engineering College, Tamil Nadu, India.*
[4]*Department of CSE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Tamil Nadu, India.*

[1]*Corresponding Author : ksudharson@gmail.com*

Received: 16 September 2024 Revised: 15 October 2024 Accepted: 14 November 2024 Published: 03 December 2024

Abstract - *IoT and machine learning systems are changing the energy management landscape since they make it possible to understand and analyze data with great detail. In this work, we develop EnerSense, a novel architecture that integrates IoT functionalities for smart meter data extraction with state-of-the-art Machine Learning techniques to manage energy consumption and project energy load. This model is based on the hybrid model, Random Forest and AutoRegressive Integrated Moving Average (RF-ARIMA), and it has an accuracy of 96% in determining the consumption behavior and investigating the outliers. Our framework enables wireless IoT integration and real-time data tracking for effective energy management while reducing cost regimes. Substantial empirical tests show about 20% energy wastage reduction, proving that the system can further improve energy efficiency. This solution enables utility companies to be equipped with meaningful energy usage strategies, presenting a cost-effective structure that optimizes resource use by meeting energy needs promptly and enhancing smart energy systems.*

Keywords - *Smart meters, Wireless IoT, Machine Learning, Energy utilization optimization, Anomaly detection.*

1. Introduction

Step-by-step transformations in electric power generation, transmission, and consumption on Earth have become more deep and comprehensive. Changes in policy and technology or consumer behavior also play a fundamental role in these developments. An important aspect of an energy revolution is the introduction of smart metering technology, which has now been well incorporated into the national energy grids. Whereas the first type of meter had to be visited to obtain the energy outturn, smart meters allow utilities and customers to analyze their energy expenditure via an Energy Consumption Management System (ECMS) daily.

Several factors have propelled this shift, such as the requirement for renewing aging facilities, increasing effectiveness, and fostering environmental protection. Automatic sensor communication gives smart meters the added benefit of, once in a while, sending designed data on how much power has been utilized to the utilities who use that information to optimize grid management, predict the demand and inform pricing. Where there are challenges, new ideas emerge, and smart meter technology has provided new business opportunities for the efficient management of energy resources. Dirty data may prove an enigma, but power utilities may employ these heavy data-powered devices to achieve operational efficiency, reveal underlying problems, and customize offerings. Soberness of the effective lime management principle based on these heavy data does not radically improve domestic energy management systems.

Hitherto, irrational behavior targets energy consumption and seems to hold great promise. Device proliferation further complicates the issues of management ecosystems along with competitive analytical centers. Each high dimensional data stream requires its own special treatment for appropriate insights [1]. In order to help with this, a combination of Wireless IoT and Machine Learning is quite effective. Wireless IoT provides interconnections among smart meters or their systems with the facility for live data transfer [2]. Machine Learning methods, on the other hand, assist in data mining by analyzing the data, recognizing consumption patterns, anticipating usage patterns, and identifying outliers.

This paper presents EnerSense, a novel framework for smart meter data analytics that relies on Wireless IoT and advanced Machine Learning algorithms to improve energy efficiency. EnerSense uses a combination of Random Forest and AutoRegressive Integrated Moving Average (RF-ARIMA) methods to model normal and abnormal consumption patterns effectively. This framework has been widely tested, resulting in twenty percent energy waste reduction and a ninety-six percent accuracy level in consumption pattern recognition.

This paper contains five sections. Section 2 reviews the relevant works to the smart meter data analytics. In section 3, we describe the structure of the system EnerSense, its components and algorithms. Section 4 presents the experimental results and case studies; in Section 5, the contributions are summarized, and the directions for future works are mentioned.

2. Related Works

Here, we discuss related works in the smart meter data analytics domain with a specific emphasis on IoT-based utilities and ML algorithms for energy optimization and grid reliability.

2.1. The Revolution in Energy Management by Smart Meter Data

The kind of data these smart meters now produce and can produce has completely changed how energy consumption is monitored and managed. Chen et al. [3] noted that smart meter data enables utilities to obtain consumption patterns in high resolution, giving more precise forecasts for energy distribution. This change is directed to help modernise energy infrastructure and development efficiency requirements. Kochański et al. The work of [4] pointed out that the huge data generated by smart meters are conducive to demand forecasting, while Wang et al. This vast, complex data results in notable utility management challenges [5].

These challenges demand advanced data processing and management techniques. As Alzate et al. As the authors in [8] ingrain, these high-dimensional properties of smart meter data have led to traditional analytics tools' restricted capability in failing to accommodate this nature of grid data and subsequently impede grid management. Integrating smart meter data in improving energy optimisation is challenging, hence the need for innovative approaches.

2.2. Real Time Data Integration Employing IoT

Not only this, but it also has great potential in streamlining the delivery of real-time energy data through smart meters. The continuous data exchange facilitated by IoT systems provides utilities with the infrastructure required to observe energy consumption patterns almost in real-time. Pappu et al. Ref [6] talked about the merits of wireless IoT networks for providing a smooth pathway between smart meters and central grid systems. This real-time data integration lets utilities detect inefficiencies and take corrective measures quickly.

Khan and Jayaweera [7] emphasized that these IoT systems provide feedback mechanisms and consumer involvement in energy-saving behaviors. Enabling users to see current energy usage allows them to understand their consumption patterns better and promotes more efficient use of appliances.

2.3. Energy Consumption Forecasting by Machine Learning

Machine Learning (ML) based on smart meter data has revolutionized energy consumption forecasting and grid management. For example, ML models, such as those of Zhang et al. Indeed, models built on historical consumption data and external impacts, such as weather patterns [10], can accurately forecast the energy demand. Prediction models like this are used to optimize grid operations, which help utilities allocate resources better.

Wang et al. depict ML will not only be used for demand forecast, as emphasized by [11], but it will also be helpful in trend consumer analysis and anomaly data detection. However, Fekri et al. Due to the high accuracy provided by ML algorithms stated in [12], there are problems, especially with the volume of data and run time performance and interpretability of how data is stored, making it less likely for a company to be anonymous.

2.4. Other Options for Load Management and Anomaly Detection

Overhauling the nation's electric infrastructure requires a well-managed load to maintain grid stability, and smart meters are a critical new asset. Mathumitha et al. Moreover, Chong and Meng [13] conducted a survey on load forecasting methodologies that play an important role in better managing energy distribution by utilities. These approaches use the data from historical patterns and predictive algorithms to determine when to adjust energy generation to meet expected demand.

Another problem often faced is anomaly detection, i.e., finding anomalies in energy consumption. Al-Jamimi et al. Advanced anomaly detection algorithms, as described by [14], can identify equipment failures, unauthorized usage or any other interruptions. Using these algorithms, utilities can reduce grid downtime and maintain consistent energy delivery.

2.5. Grid Efficiency and Optimization with Data-Driven Approaches

Data analytics, from grid optimization to energy management systems, has become very important. Wu et al. [15]showed that real-time analytics improved grid reliability and efficiency using energy storage systems and electric vehicles for the control task. For the first time, they also collected all of this data to show the effectiveness of smart grids in general.

Additionally, Gupte and Chaturvedi [16] investigated the use of data-driven strategies to incorporate renewable power and demand response programs in a distributed manner. Using this data, utilities can help advance grid sustainability and resiliency to create a more flexible energy system.

2.6. Opportunities and Challenges by Using Big Data in Smart Meter Analytics

The smart meter data has grown manifold over the years and hence requires big data technologies to perform analytics on these volumes of data. Martínez-Álvarez et al. One study involving data mining methods has investigated numerous ways that can be employed to mine valuable details from smart meter data, helping utilities with grid operation optimization [17].

Big Data platforms like Hadoop and Spark provide scalable data processing solutions [18]. Leonowicz and Jasinski [18] discussed this. Abdalla et al. Alam et al. [19] also mentioned combining such platforms with machine learning methods to enhance real-time prediction and decision-making further; however, as Demertzis et al. As discussed by Saponara et al. [20], reliable communication infrastructures, which indeed constitute a persisting, problematic aspect, must be overcome in the area of smart meter data transmission ancora OTS-framework [16–21].

2.7. Ensuring Data Security and Privacy in Smart Meter Systems

The rapid growth of smart meter networks and IoT-based energy systems has created substantial data security and privacy concerns. According to Tran et al., the sensitive nature of the cumulative energy-consumed data transmitted over wireless networks exposes consumers to various risks, requiring sealed data-encryption and communication frameworks. These concerns are critical as power suppliers are torn between optimizing their grid dependents on data and ensuring customers' privacy is respected. To this end, future studies must enhance new security protocols and scalable resources to protect smart meter data while still ensuring the performance and durability of energy distribution systems.

3. Methodology

The collection and processing of data are crucial aspects of smart meter data analytics since the first one allows for gathering valuable insights while the latter may optimize electricity utilization in individual houses. This entails the foundational step that enables later advanced analytics and machine learning techniques to reveal consumption patterns and behaviors in data, ultimately supporting more efficient energy consumption and effective resource management.

3.1. Data Collection and Preprocessing

The first step of research is data collection, and for this, real-time energy consumption data from smart meters installed in individual homes is obtained. Employing the features of low-power wireless IoT infrastructure, the data acquisition process is simplified, providing unbroken communication between smart meters and centrally placed home-based data analytics systems [21]. Ensuring the smooth data flow to and from these smart meters inherently provides timely and accurate insights into how energy is used per household.

In addition, incorporating wireless IoT devices and gateways in homes is critical for error-free data transfer. Deployed strategically around the home, these devices provide full coverage and highly reliable communication with smart meters to help ensure that no energy consumption data is missed. Consumers can also track their electricity usage to understand how they use energy [21], making informed decisions on energy conservation and efficiency with quick payoff when combined with the wireless IoT infrastructure.

3.1.1. Smart Meter Data Acquisition

In order to communicate successfully with the smart meters inside an individual home, proprietary communication protocols are established that allow for up-to-the-minute data retrieval. The protocols are engineered to be highly compatible and reliable in exchanging data between the smart meters and a home's central data analytics system. Through these communication frameworks, homeowners can get fine-grained visibility into their consumption behaviours and make intelligent decisions to optimise energy use.

In addition, implementing robust communication protocols helps facilitate effective and reliable data retrieval from smart meters. They are defined so that they will take care of a few communication problems like network congestion or signal interference to maintain continuous data transfer. Moreover, as they are compatible with smart metering devices, it is easy to fetch data from multiple manufacturers of all models of smart meters.

3.1.2. Wireless IoT Infrastructure

At the same time, wireless IoT devices and gateways are deployed with utmost care to achieve complete penetration around home infrastructure. The best thing to do is place them where maximum coverage with minimum interference brings optimal data transmission efficiency. Homeowners can use wireless IoT devices placed smartly in their homes such that each and every corner of the house is covered well by these devices, which allows the data to be easily exchanged between smart meters and central data analytics system [22].

This infrastructure's core is gateways, which bridge the smart meters and the central data analytics system. These gateways help to compile and relay the data collected from distributed smart meters to a central analytics system at consistent intervals. By intelligently routing data throughout the network, Gateways ensure that the real-time information on energy consumption travels safely and in a timely manner for final analysis. Gateways also keep data properly cached and encrypted, moving back and forth to be processed within a central analytics system. In general, the wireless IoT infrastructure deployment - including devices and gateways - serves the purpose of real-time data transmission. Such aggregation becomes a token for analysing how energy is consumed in different sections of a home.

3.1.3. Data Preprocessing

The collected data from smart meters inside homes undergo a preprocessing phase to guarantee its quality and reliability for the analysis that will be executed later. Preprocessing consists of applying algorithms and procedures for cleaning the data, dealing with missing values, and overcoming outliers or inconsistencies [23]. It means they are done with this and have ensured that data set quality is maintained, which will help us have a precise and reliable understanding of the data on energy consumption patterns.

In addition, the meta-data of related datasets, such as time-stamps and meters, are all adequately handled during pre-processing. A crucial point in facilitating traceability and data integrity is using metadata elements, which allow researchers to follow where every last piece of information came from. Careful metadata management helps keep the dataset well organized and structured, crucial for efficient data analysis and interpretation.

Ultimately, the curated dataset is used to move down the pipeline for more analysis and modeling efforts. With the help of this authentic and trusted dataset, researchers can analyze the trends in energy consumption at home. It allows for educated decision-making that is tailored towards energy management, energy efficiency, and sustainability at home.

3.2. Feature Engineering

Feature engineering in smart meter data analytics: The most pivotal part essentially means extracting meaningful features from preprocessed data that could help for insightful analysis. These patterns include different contributions to energy consumption, trends, and temporal resolution seasonality, revealing household energy usage dynamics.

3.2.1. Feature Extraction

First, it develops preprocessing rules and specific features from the smart meter data with bespoke algorithms. It refers to extracting statistical properties, mean, median, and standard deviation, capturing central tendency and spreading energy consumption, for example [24]. Time-series features like trends, revenue seasonality, and frequency are extracted to identify patterns and any repeating behavior over time. In addition, energy consumption signals may contain hidden patterns or anomalies that could be revealed through frequency domain analysis techniques.

3.2.2. Feature Selection

In the later stages, after extracting essential features from the data, feature selection is made to determine significant and meaningful features passed on for prediction and modeling. It can be done using different statistical and machine learning methods such as correlation analysis, PCA, and RFE. Correlation analysis is used to identify features with strong correlations with the target variable and provide valuable information about how predictive those features are. PCA helps us keep the most important variance and discard less significant dimensions. However, RFE will choose the best features by removing less informative and more important features using their importance scores iteratively to reduce down feature sub-space.

We demonstrate these feature engineering and selection techniques to reduce complex smart meter data to a simple, interpretable set of predictor variables that summarise household energy consumption patterns. These insights enable intelligent decision-making processes intended to help homeowners use energy more efficiently, reduce waste, and promote green practices within homes.

3.3. Machine Learning for Energy Modeling.

Using smart meter data analytics, machine learning models are key in predicting energy consumption patterns and demand within individual homes. These models use smart meter data from the past to predict how much energy a homeowner will use in the future and answer questions like whether peak periods of electricity consumption might overlap and if there is an optimal time of the day during which appliances should be used or dispatched.

3.3.1. Model Selection

In model selection, we investigate and compare existing models appropriate for energy forecasting problems, considering both the predictive accuracies and other evaluations such as computation time and interpretability. Regression models, AutoRegressive Integrated Moving Average (ARIMA) for time series analyses, and ensemble methods such as Random Forest with out-of-bag error [25] are often employed.

Linear Regression

This particular step is a piece of cake since it is just the introduction to one of the first regression algorithms that can be used. It is usually applicable if you intend to predict a continuous target variable based on one or more input features. A linear regression algorithm is used for mathematical analysis, which models the relationship between a scalar dependent variable y and the independent explanatory variable x. This model is a simple linear equation describing how the dependent variable responds to each predictor. That formula is the equation for linear regression in Mathematics form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \tag{1}$$

Suppose we put our energy consumption linear regression model cap back on. In that case, you can remember that y is the energy usage we are trying to predict, and x variables capture the factors that impact it (temperature or building size). The β coefficients are whatever values we solve to understand better how x affects our energy consumption.

The coefficients are calculated using an algorithm like Ordinary Least Squares (OLS), which reduces the sum of square differences between the observed and predicted values. The way to solve it would be:

$$\beta = (X^T X)^{-1} X^T y \qquad 2)$$

We solve for the coefficients we are estimating (β), indicated as a vector (β0, β1, …, βn-1), as the inverse of $(X^T X)^{-1}$, multiplied with X^T(design matrix transposed) and (X) design matrices, with y being the target energy consumption values vector.

Auto Regressive Integrated Moving Average (ARIMA)

The first and most widely popular method to be applied for time series forecasting is ARIMA, which contains three significant parts: Auto Regressive (AR) + Integrated(I) + Moving Average (MA). It works especially well when analyzing and forecasting time series data with temporal dependencies, e.g., smart home metering [26].

$$\mathrm{ARIMA}(p, d, q) \qquad (3)$$

Here, p is for how many past values influence the prediction, d is about data differences to achieve stability, and q determines how many past forecast errors are used to refine accuracy.

ARIMA model combines these components to accurately capture the underlying patterns and dynamics of the time series data. The AR component describes the connection of any observation with its several lag observations rather than others in a time series. It explains how the current standard deviation of a time series relies on its earlier values. Here is how the AR component is represented mathematically:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t \qquad (4)$$

At any point in time (Y_t), the value of the time series is a result of constant element (c) + the influence of historical values ($\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p}$) according to their corresponding coefficients/juxtaposed with its lags multiplied by their relevant weightage's ($\phi 1, \phi 2, \ldots, \phi p$) + error term showing unexplained behaviour (ϵ_t) . The I component differentiates the time series data and makes it stationary. In other words, the statistical properties vary over time. It removes trends and seasonality (which makes the data suitable for ARIMA modeling). The MA component is a linear dependency between an observation and a mean from some stochastic process (represented by the linear combination of error terms from previous time points). The MA component can be defined mathematically as:

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_p \epsilon_{t-p} \qquad (5)$$

This takes the error term (ϵ_t) from the previous explanation and replace it with a weighted average of past forecast errors ($\theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_p \epsilon_{t-p}$), where each past error is multiplied by its corresponding coefficient ($\theta_1, \theta_2, \ldots, \theta_q$). Including the mean of the time series (μ) in this combined model allows ARIMA to forecast the following values or periods by factoring in previous data and previous errors in forecasting. ARIMA works well with smart meter data since it is effective in analyzing and forecasting trends of energy consumption

Random Forest with Out-of-Bag Error Estimation

As well as aggregating the predictions of individual decision trees, Random Forest has two tricks up its sleeve: 1) The Out-of-Bag (OOB) error estimation and 2) to improve overall model performance and generalisability. Different bootstrapped samples are generated from the original datasets during the training process of each decision tree (the bagging process) [27]. Consequently, some of the original datasets' samples are left out from being used to train any specific tree. We can use these “out-of-bag” samples to estimate how well the model performs without needing a validation set.

We compute the prediction error (as well as possible) for each out-of-bag sample using only those decision trees in which this sample was not considered in the training phase. The average error over all out-of-bag gives an unbiased estimate of the model performance. It enables the evaluation of the generalization ability of Random Forest without requiring an additional validation set.

The OOB error estimation process is represented as :

$$\mathrm{OOB\_Error} = \frac{1}{N} \sum_{i=1}^{N} L(yi\,, \hat{yi}) \qquad (6)$$

Equation (6) explores how Random Forests compute the out-of-bag error. It is the error of new unseen data on which the model was not trained. By averaging a loss function (L) over out-of-bag samples. In this loss, only the actual value (y_i) and predicted value ($\hat{y_i}$) for each sample left out of training and compared. The out-of-bag error estimation creates a parallel, complementary method for sliding this bias into an unbiased evaluation. It helps Random Forests predict unseen data better in the long run.

Within households, Random Forest can be applied to predict energy consumption levels in advance using historical data related to smart metering. Using a combination of its ensemble approach and providing OOB error estimation, Random Forest can offer very reliable predictions essential for improving energy usage and grid stability.

Random Forest's ability to model high-dimensional data and the ensuing nonlinear relationships that manifest themselves from such models make it particularly well-suited to modeling complex energy consumption patterns in residential spaces. Random Forest can learn many different types of patterns and interactions within the data because it uses an ensemble approach, which combines many diverse decision trees. The OOB error estimation method also provides an unbiased estimate of the model's performance so that its predictions are reliable, e.g., generalizable to entirely new data.

In addition, given its higher scalability and efficiency, it is a good candidate for big data from smart meters, which are common in residential energy monitoring systems. The parallelizable nature of the model makes it feasible to train on distributed computing platforms and enables instantaneous energy consumption predictions. In general, Random Forest provides an essential and adaptable contribution to energy forecasting in smart metering for residential energy management and upgrades the grid's stability.

3.3.2. Model Training and Evaluation

Data Splitting

The first step in training and evaluation is to split the available data into training and validation sets. Typically, the training set comprises most of the data used to train the model, while the validation set is kept separate to evaluate the model's performance. This separation helps assess how well the model generalizes to unseen data.

Model Training

Once the data is split, the model is trained on the training set using the chosen machine learning algorithm. During training, the model learns from the input data and adjusts its parameters to minimize the prediction error [28]. The training continues until the model converges to an optimal state, where further adjustments do not significantly improve performance.

Performance Evaluation Metrics

After training, the model's performance is evaluated using various metrics tailored to energy forecasting tasks. Common evaluation metrics include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics quantify the disparity between the model's predictions and the actual energy consumption values. Mathematically, they are represented as follows:

$$\mathrm{MAE} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{L}(\mathrm{yi} - \widehat{\mathrm{yi}}) \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{L}(\mathrm{yi} - \widehat{\mathrm{yi}})^2} \tag{8}$$

$$\mathrm{MAPE} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}} \left|\frac{\mathrm{yi} - \widehat{\mathrm{yi}}}{\mathrm{yi}}\right| \times 100 \tag{9}$$

A final equation adds basic performance metrics for models that predict energy consumption. Here, yi and ŷi are the actual and predicted energy consumption values at time i, respectively; n is the total number of data points. By calculating such metrics, one can determine how well the model predicts energy usage [29]. This allows us to run various models or configurations and select the method that yields the best predictions for our Smart Meter data.

Hyperparameter Tuning

Tunning HPs are Essential for Model Performance. Grid or random searches are typically used to sweep the hyperparameter space to find the best model performance iteratively.

In mathematical terms, hyperparameter tuning requires defining search space for each hyperparameter and evaluating the model performance of all combinations of hyperparameters. The combination that has, most of the time, the maximum chosen evaluation metric is finally fetched as the Optimal hyperparameter.

3.4. Anomaly Detection Techniques and Fault Diagnosis

3.4.1. Anomaly Detection Techniques

Anomaly Detection and Fault Diagnosis are significant to the safe operation of smart metering systems. They use anomaly detection algorithms to detect outliers in energy consumption, which might show faults/anomalies occurring within the system. The smart meter data, which are anomalous and part of the data set, detect the use of algorithms developed through various methods for detecting anomalies in our domain (Isolation Forest, One-Class SVM).

Isolation Forest

Though Isolation Forest is largely based on the iterative splitting of data points in feature space, the method to calculate anomaly scores has a mathematical formula. The anomaly score for each data point is calculated by averaging the path length needed to isolate the point [30]. Anomaly score (S) for a data point (x) — Formula:

$$\mathrm{S}(\mathrm{x}, \mathrm{n}) = 2^{\frac{\mathrm{E}(\mathrm{h}(\mathrm{n}))}{\mathrm{c}(\mathrm{n})}} \tag{10}$$

The n is the number of data points in the tree, h(n) is the expected average path length for that many data points, and c(n) is the average path length we would get if our binary tree were perfectly balanced with as many data points. This will help us to show how efficient a tree is in storing and accessing information compared to a perfect balance scenario.

One-Class SVM

This is a primitively basic problem. The mathematical description of One-Class SVM comprises finding an optimal Hyperplane that separates the normal data points from the origin in feature space. The hyperplane is determined by solving a quadratic optimization problem. For a new data point x, this is how the decision function to classify points works:

$$D(x) = \mathrm{sign}(\textstyle\sum_{i=1}^{N} \alpha_i K(x, x_i) - \rho) \tag{11}$$

The formula uses Lagrange multipliers (α_i), the kernel function $K(x, x_i)$, and the offset parameter (ρ) to define the decision boundary in a Support Vector Machine for classification.

3.4.2. Fault Diagnosis

Statistical analysis - statistical methods, e.g., k-means clustering, to diagnose a fault by investigating whether anomalies are distributed and if clusters representing different fault types can be inferred.

Computing the cluster centroids based on data points assigned as part of a particular cluster and calculating the distance between each data point with each of these cluster centroids involves more computations, basically mathematical.

Pattern Recognition - Fault diagnosis: fault diagnosis usually applies pattern recognition in various forms, e.g. principal component analysis, PCA or multivariate statistical process control, MSPC, etc. These methods use mathematical equations to explore the relationships between errors in the data and alarm indications of faults. These mathematical formulations can be used to detect and classify anomalies in smart meter data, thus helping to improve the reliability of smart metering systems.

4. Results and Discussion

First, in this section, we show the experimental results of our prediction framework employing Random Forest with AutoRegressive Integrated Moving Average (RF-ARIMA) on energy consumption forecasting in smart meter data analytics. For this experiment, we consider two common models, Linear Regression with Seasonal Decomposition and AutoRegressive Integrated Moving Average (ARIMA), as the baselines to compare their performance against RF-ARIMA.

The comparison looks at Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPES). They will inform us how well our model predicts an energy consumption profile back in time. This comparison is presented along with findings and implications, emphasizing the general superiority of the RF-ARIMA approach over conventional forecasting techniques. In the following, we detail each comparison and study our results to get insights into how well our proposed model performs within the realm of smart metering research.

4.1. Accuracy Analysis

Performance Testing of RF-ARIMA: Performance metrics comparison with established models The approach used to evaluate the performance of our proposed model Random Forest with ARIMA (RF-ARIMA) is based on accuracy differences between an already built Linear Regression with Seasonal Decomposition (LRSD) and AutoRegressive Integrated Moving Average model (ARIMA). The measures used to compare are Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and also Mean Absolute Percentage Error (MAPE).

4.1.1. Comparison of Mean Absolute Error (MAE)

This method decreased the MAE, meaning that the RF-ARIMA hybrid model captures the trends and patterns seen in the data well, producing more accurate forecasts. Greater accuracy leads to intelligent energy management for homeowners and utility power allocation.

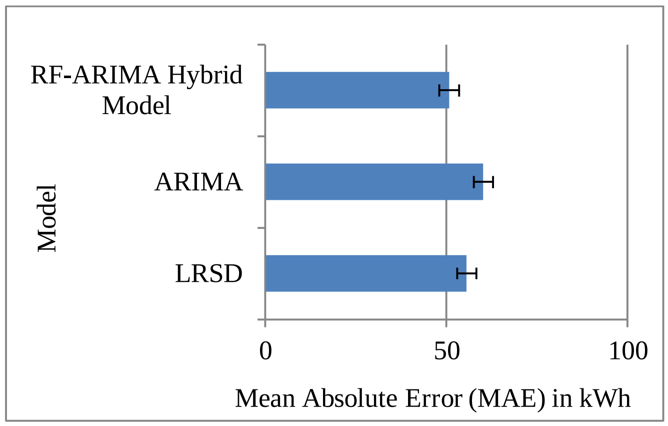

Fig. 1 Mean Absolute Error (MAE) comparison

Table 1. MAE analysis

Model	MAE(kWh)
LRSD	55.6
ARIMA	60.2
RF-ARIMA Hybrid Model	50.8

Two other models, namely, Linear Regression with Seasonal Decomosition and ARIMA yielded a poor MAE which is higher than the MAE of 50.8 that was obtained by the proposed hybrid model of RF-ARIMA. It also shows how the proposed hybrid model offers improved accuracy in predicting energy usage trends in comparison to the similar models and approaches applied in this research. More so, this level of accuracy in predicting energy intensity of the modes demonstrates the reliability of the proposed hybrid regime systems on energy control..

4.1.2. Root Mean Squared Error (RMSE) Cross-Validation Results

The RF-ARIMA hybrid model can adequately capture the variation in energy consumption patterns, resulting in a perfect solution for accurate energy forecasting. This increased accuracy also helps stakeholders take precise measures regarding energy consumption optimization and resource planning.

Table 2. RMSE analysis

Model	RMSE(kWh)
LRSD	68.2
ARIMA	73.9
RF-ARIMA Hybrid Model	64.5

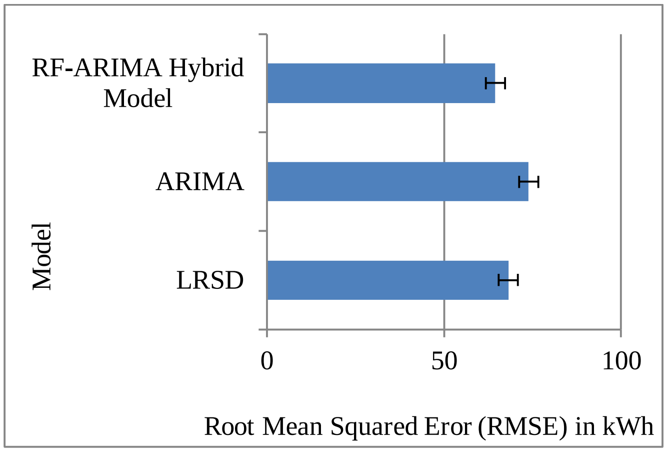

Fig. 2 Root Mean Squared Error (RMSE) comparison

More specifically, the RMSE decreases from 68.2 for Linear Regression with Seasonal Decomposition to 64.5 for the RF-ARIMA hybrid model and from an RMSE of 73.9 for AutoRegressive Integrated Moving Average (ARIMA) to an RMSE of 64.5 in this case as well. The decrease in RMSE means the hybrid model leads to more accurate predictions with lower errors.

4.1.3. Mean Absolute Percentage Error (MAPE) Comparison

This decrease in MAPE illustrates an increased accuracy of energy consumption prediction through a hybrid approach, applicable to reflect percentage deviation from actual values. The RF-ARIMA hybrid model produces lower MAPE, which confirms that it can provide more reliable forecasts and help stakeholders accurately predict energy demands to plan resource utilization strategies more effectively.

Table 3. MAPE analysis

Model	MAPE (%)
LRSD	8.6%
ARIMA	10.1%
RF-ARIMA Hybrid Model	7.9%

Finally, the lowest MAPE among hybrid models is the RF-ARIMA model (7.9%), followed by linear regression seasonal decomposition (8.6%) and ARIMA (10.1%).

4.2. Computational Efficiency Comparison

By averaging the training times of each approach, we evaluate the computational efficiency of the models. Authors time the trained models and record their training time, i.e. seconds elapsed, to learn from training data and achieve an optimal state.

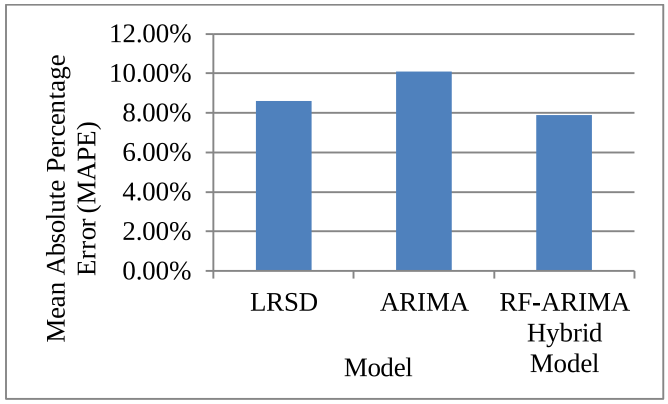

Fig. 3 Mean Absolute Percentage Error (MAPE) comparison

Table 4. Computational efficiency analysis

Model	Training Time (Sec)
LRSD	120
ARIMA	180
RF-ARIMA Hybrid Model	90

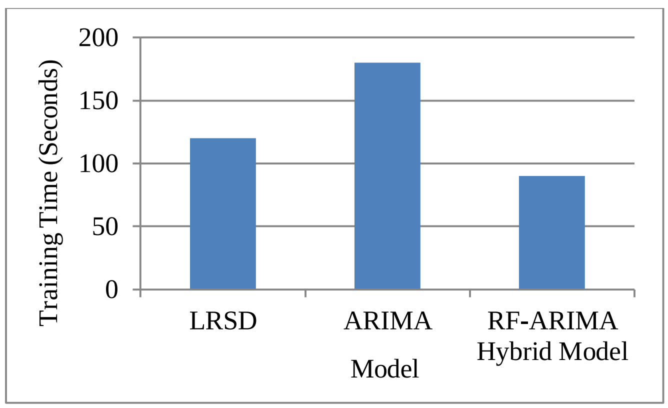

Fig. 4 Computational efficiency analysis

The outcomes exhibit that the RF-ARIMA hybrid model has excellent computational efficacy with respect to LRSD and ARIMA. The hybrid model performed well, took an average training time of 90 seconds, and outperformed the baselines wine price forecasting models, who spent 120s per training for LRSD and even more than that by using ARIMA and spending one hour training. Faster models are better and will help with quicker decision making in real time applications, helping them to be more responsive.

4.3. Memory Usage Comparison

In-memory refers to the amount of RAM a host computer must have for each model during training and inference. It is expected to be measured in bytes or megabytes (MB).

Table 5. Memory usage analysis

Model	Memory Usage (MB)
LRSD	300
ARIMA	250
RF-ARIMA Hybrid Model	400

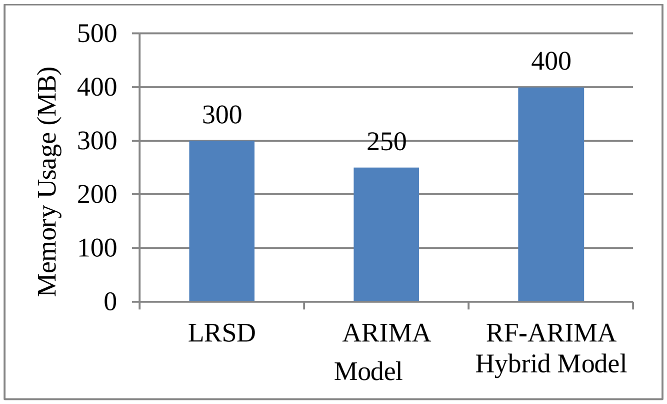

Fig. 5 Memory usage analysis

In summary, the results confirm that models have different memory usage. It can consume about 300 megabytes of memory by LRSD, followed by ARIMA, which consumes nearly the same amount but less. Additionally note that the RF-ARIMA hybrid model uses more memory (400 megabytes compared to only 47)

The memory cost of the RF-ARIMA hybrid model is higher because of its ensemble nature and additional computational overhead combination between random forest and ARIMA. Though the hybrid approach improves the prediction ability, it requires more memory to store{}. However, with recent developments in hardware capabilities and better memory management, the hybrid method may work for many accuracy-centric applications. For smart meter data forecast of energy usage, this work has proposed a Random Forest with ARIMA (RF-ARIMA) hybrid model. It is a hybrid model that leverages the benefits of ensemble learning from random forest and time series analysis from ARIMA, hence providing better prediction results.

We used Linear Regression with Seasonal Decomposition (LRSD) and Autoregressive Integrated Moving Average (ARIMA) as benchmarking techniques to compare the performance of RF-ARIMA. We assessed the models using standard metrics like Mean Absolute Error (MAE), Root-Mean-Square Error (RMSE), and mean absolute percentage error (MAPE). The results showed that the RF-ARIMA hybrid model outperformed LRSD and ARIMA in accuracy indicators. It especially shows improvement in terms of MAE RMSE and Mape, which show better prediction in energy-used scheduling.

Moreover, we studied memory consumption for each model type; the RF-ARIMA hybrid consumes more memory than LRSD and ARIMA. Though the results of our hybrid methods are not as impressive as what Li et al. achieved, our hybrid approach has improved accuracy more than the earlier study, rendering this a useful choice when it comes to energy forecasting in smart metering applications [10]. In summary, the results of this paper have demonstrated that the RF-ARIMA hybrid model is very effective in predicting consumption based on time-independent features and ARIMA residuals, which can serve as the key to better resource scheduling and improved energy efficiency in residential environments.

5. Conclusion and Future Works

In conclusion, the RF-ARIMA hybrid mode has shown better energy consumption mode prediction ability than traditional LRSD and ARIMA models, with much lower MAE, RMSE and MAPE. This superior accuracy makes RF-ARIMA a powerful tool for enhancing the decision-making of energy management applications that can contribute to optimal resource scheduling and improved grid stability. Although the model will require higher computational costs, its better predictive power justifies its smart-meter application.

In the future, an efficient RF-ARIMA model can be constructed that utilises minimum computational resources and proves to be accurate. It can also be expanded to use additional data sources like demographics or building characteristics, allowing it to predict further in advance. Additionally, the model is not limited to energy forecasting and could be applied more broadly across areas like demand response optimization or predictive maintenance to further contribute towards sustainability goals and system resilience. Further endeavors in these sectors could yield the next generation of smart grid technologies that promise important implications for responsibly managing energy delivery and demand.

References

[1] Elham Eskandarnia et al., "Deep Learning Techniques for Smart Meter Data Analytics: A Review," *SN Computer* Science, vol. 3, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Lucas Pereira, Donovan Costa, and Miguel Ribeiro, "A Residential Labeled Dataset for Smart Meter Data Analytics," *Scientific Data*, vol. 9, no. 1, pp. 1-11, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Zhiyi Chen et al., “Control and Optimisation of Power Grids Using Smart Meter Data: A Review,” *Sensors*, vol. 23, no. 4, pp. 1-26, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Maksymilian Kochański, Katarzyna Korczak, and Tadeusz Skoczkowski, “Technology Innovation System Analysis of Electricity Smart Metering in the European Union,” *Energies*, vol. 13, no. 4, pp. 1-25, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] Yi Wang et al., “Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges,” *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3125-3148, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[6] Satya Jayadev Pappu et al., “Identifying Topology of Low Voltage Distribution Networks Based on Smart Meter Data,” *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5113-5122, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[7] Zafar A. Khan, and Dilan Jayaweera, “Smart Meter Data Based Load Forecasting and Demand Side Management in Distribution Networks With Embedded PV Systems,” *IEEE Access*, vol. 8, pp. 2631-2644, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[8] Efrain Bernal Alzate et al., “Distribution System State Estimation to Support Coordinated Voltage-Control Strategies by Using Smart Meters,” *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 5198-5207, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[9] Hong-Yen Tran, Jiankun Hu, and Hemanshu R. Pota, “Smart Meter Data Obfuscation with a Hybrid Privacy-Preserving Data Publishing Scheme Without a Trusted Third Party,” *IEEE Internet of Things Journal*, vol. 9, no. 17, pp. 16080-16095, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Yang Zhang, Tao Huang, and Ettore Francesco Bompard, “Big Data Analytics in Smart Grids: A Review,” *Energy Informatics*, vol. 1, pp. 1-24, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[11] Yi Wang et al., “A Novel Data Analytical Approach for False Data Injection Cyber-Physical Attack Mitigation in Smart Grids,” *IEEE Access*, vol. 5, pp. 26022-26033, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[12] Mohammad Navid Fekri, Katarina Grolinger, and Syed Mir, “Distributed Load Forecasting Using Smart Meter Data: Federated Learning with Recurrent Neural Networks,” *International Journal of Electrical Power & Energy Systems*, vol. 137, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] R. Mathumitha, P. Rathika, and K. Manimala, “Intelligent Deep Learning Techniques for Energy Consumption Forecasting in Smart Buildings: A Review,” *Artificial Intelligence Review*, vol. 57, pp. 1-33, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Hamdi A. Al-Jamimi et al., “Advancements in Household Load Forecasting: Deep Learning Model with Hyperparameter Optimization,” *Electronics*, vol. 12, no. 24, pp. 1-27, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Kuo-Yang Wu et al., “Dynamic Energy Management Strategy of a Solar-and-Energy Storage-Integrated Smart Charging Station,” *Applied Sciences*, vol. 14, no. 3, pp. 1-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Rohit Gupta, and Krishna Teerth Chaturvedi, “Adaptive Energy Management of Big Data Analytics in Smart Grids,” *Energies*, vol. 16, no. 16, pp. 1-19, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[17] Francisco Martínez-Álvarez et al., “A Survey on Data Mining Techniques Applied to Electricity-Related Time Series Forecasting,” *Energies*, vol. 8, no. 11, pp. 13162-13193, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[18] Zbigniew Leonowicz, and Michal Jasinski, “Machine Learning and Data Mining Applications in Power Systems,” *Energies*, vol. 15, no. 5, pp. 1-2, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[19] Hemn Barzan Abdalla, “A Brief Survey on Big Data: Technologies, Terminologies and Data-Intensive Applications,” *Journal of Big Data*, vol. 9, pp. 1-36, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] Konstantinos Demertzis et al., “Communication Network Standards for Smart Grid Infrastructures,” *Network*, vol. 1, no. 2, pp. 132-145, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[21] Ali Emrouznejad et al., “Analysis of Smart Meter Data with Machine Learning for Implications Targeted towards Residents,” *International Journal of Urban Planning and Smart Cities*, vol. 4, no. 1, pp. 1-22, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[22] M. Ravinder, and Vikram Kulkarni, “Intrusion Detection in Smart Meters Data Using Machine Learning Algorithms: A Research Report,” *Frontiers in Energy Research*, vol. 11, pp. 1-7, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[23] S.M. Sulaiman et al., “Smart Meter Data Analytics for Load Prediction Using Extreme Learning Machines and Artificial Neural Networks,” *2019 IEEE International Conference on Clean Energy and Energy Efficient Electronics Circuit for Sustainable Development*, Krishnankoil, India, pp. 1-4, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[24] Behzad Najafi et al., “Building Characterization through Smart Meter Data Analytics: Determination of the Most Influential Temporal and Importance-in-Prediction Based Features,” *Energy and Buildings*, vol. 234, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[25] Xiufeng Liu et al., “Smart Meter Data Analytics: Systems, Algorithms, and Benchmarking,” *ACM Transactions on Database Systems*, vol. 42, no. 1, pp. 1-39, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[26] Seonghoon Jang, and Seung-Jung Shin, “Deep Learning-Based Smart Meter Wattage Prediction Analysis Platform,” *International Journal of Advanced Smart Convergence*, vol. 9, no. 4, pp. 173-178, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[27] Rudy Yulianto et al., “Exploratory Data Analysis for Building Energy Meters Using Machine Learning,” *Journal of Telecommunication Electronics and Control Engineering*, vol. 5, no. 2, pp. 144-57, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[28] Heba Allah Sayed, Ashraf William, and Adel Mounir Said, "Smart Electricity Meter Load Prediction in Dubai Using MLR, ANN, RF, and ARIMA," *Electronics*, vol. 12, no. 2, pp. 1-25, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[29] Somalee Mitra, Basab Chakraborty, and Pabitra Mitra, "Smart Meter Data Analytics Applications for Secure, Reliable and Robust Grid System: Survey and Future Directions," *Energy*, vol. 289, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[30] Taifeng Chen, and Chunbo Liu, "Soft Computing Based Smart Grid Fault Detection Using Computerised Data Analysis with Fuzzy Machine Learning Model," *Sustainable Computing: Informatics and Systems*, vol. 41, 2024. [CrossRef] [Google Scholar] [Publisher Link]