Journal of Science Engineering Technology and Management Science
Volume 02, Issue 06, June 2025 ISSN: 3049-0952
www.jsetms.com DOI:10.63590/jsetms.2025.v02.i06.pp177-186

MACHINE LEARNING-BASED ENERGY CONSUMPTION FORECASTING IN SMART GRIDS

K. Swayam Prabha[1*], K. Murali Krishna [2], G. Ranjith Kumar[2], I. Laxman[2], D. Arun kumar[2]
[1]Assistant Professor, [2]UG Student, [1,2] Department of Computer Science and Engineering (AIML)
[1,2]Sree Dattha Institute of Engineering and Science, Sheriguda, Hyderabad, Telangana

To Cite this Article
K. Swayam Prabha, K. Murali Krishna, G. Ranjith Kumar, I. Laxman, D. Arun kumar, "Machine Learning-Based Energy Consumption Forecasting In Smart Grids", Journal of Science Engineering Technology and Management Science, Vol. 02, Issue 06, June 2025,pp:177-186, DOI: http://doi.org/10.63590/jsetms.2025.v02.i06.pp177-186

Submitted: 20-04-2025 Accepted: 27-05-2025 Published: 06-06-2025

ABSTRACT
The integration of machine learning-based energy consumption forecasting into smart grids offers significant benefits across various domains. In the utility sector, accurate forecasts enable optimized energy generation, load balancing, and infrastructure planning, resulting in improved operational efficiency and cost reductions. For renewable energy sources, such forecasting enhances the integration of intermittent resources like solar and wind, supporting grid stability and sustainability. In smart buildings and homes, energy forecasts empower users to manage consumption proactively, improve comfort, and reduce electricity bills. Traditional forecasting methods, such as statistical models and time series analysis, often fall short in handling the complexity of smart grid data. These methods typically depend on manual feature selection and struggle to incorporate contextual factors like weather, holidays, and user behavior, leading to reduced forecast accuracy and scalability limitations. To address these challenges, the proposed system employs machine learning algorithms that learn from diverse features—temporal data, environmental conditions, and grid characteristics—to provide more accurate and dynamic forecasts. Regression models are used to uncover complex patterns, while ensemble learning and optimization techniques further boost performance. This approach offers a robust, scalable solution for modern energy management needs.

Keywords: Smart Grids, Energy Forecasting, Machine Learning, Load Balancing, Renewable Integration

1. INTRODUCTION

The integration of machine learning-based energy consumption forecasting in smart grids heralds a transformative era in the energy sector. It encompasses a multifaceted approach that not only revolutionizes utility operations but also extends its benefits to various sectors. At its core, this integration aims to enhance the efficiency and effectiveness of energy generation, distribution, and consumption processes within smart grid infrastructures. By leveraging advanced algorithms and diverse datasets, this paradigm shift promises to address the challenges posed by the increasing complexity and variability of modern energy systems.

In the realm of energy consumption forecasting, traditional methods often fall short in capturing the intricacies of smart grid data. They struggle to adapt to the dynamic nature of energy usage patterns and may overlook crucial contextual factors. This limitation hampers the ability of utilities to optimize energy generation and distribution, leading to inefficiencies and increased costs, inaccurate forecasts hinder the seamless integration of renewable energy sources, impeding progress towards sustainable energy goals. Thus, there is a pressing need for more advanced and data-driven approaches to energy consumption forecasting that can overcome these challenges and provide reliable predictions.

The motivation behind this research stems from the recognition of the pivotal role that accurate energy consumption forecasting plays in shaping the future of energy systems. By improving forecasting accuracy, we can unlock a myriad of benefits, including enhanced grid stability, reduced reliance on fossil fuels, and greater cost savings. Additionally, empowering consumers with actionable insights into their energy usage promotes energy conservation and efficiency. This research seeks to capitalize on the potential of machine learning algorithms to revolutionize energy forecasting and pave the way for a more sustainable and resilient energy ecosystem.

2. LITERATURE SURVEY

Smart Grids (SG) have emerged as a solution to the increasing demand on energy worldwide. The grid refers to the traditional electrical grid that is a collection of transmission lines, substations, and other components that make sure energy is delivered from the power plant to the home or business [1]. The smartness in the SG resides in the two-way communication between the utility and the customers, in addition to the sensing along the lines. The main components of a SG are controls, computers, automations, in addition to other new technologies that are working together to accommodate for the quick increase in the energy demand. The SG has many benefits among which we state: more efficient energy transmission, improving security, reducing peak demand which helps with the decrease of electricity rates, etc. SG are also known by the use of renewable energy sources. The prediction and scheduling are two of the main pillars of efficient Energy Management Systems (EMS). EMSs are very crucial for the well-functioning of the SG. They are responsible for managing the power flux within the SG elements in order to minimize the costs and optimize the quality [2]. The prediction of the energy consumed by different appliances is one of the building blocks of the concept of SGs. The energy consumption can be seen as a nonlinear time series with a number of complex factors [3]. With many renewable energy sources used in the SGs, the energy prediction methods are getting more and more accurate, and hence, the prediction becomes a crucial part in the efficient planning of the entire SG. There are different approaches that are used for the prediction of the energy consumption. The most popular ones use machine learning (ML). Machine learning (ML) is one of the growing technical fields that merge between computer science and statistics. It tackles the issue of building computers that learn through experiences and hence provide more improved algorithms. ML keeps witnessing advances thanks to the new algorithms and the availability of online data, in addition to the accessibility of the computing power [4]. Artificial Neural Networks (ANN) are one of the ML algorithms are widely used in this context. ANNs have seen light in the early 1940s but have not been widely used until lately. They became very popular thanks to the outstanding results they offer. They are very powerful with large datasets which gives the neural network enough data to train the model. In brief, ANNs are inspired by the way the brain processes information. They build an informational processing model that mimics the work of the neurons in the brain [5]. Their ability to learn quickly is what makes ANNs very powerful. This learning is done through an information flow that goes in two directions. Patterns from the training dataset are given to the ANN through the input neurons, then goes through the hidden layers and arrives to the output neurons.

Genetic Algorithms (GA) are considered the best solution for task and operation scheduling. They emerged from the research of Mr. John Holland conducted at the University of Michigan in 1960. However, it took them almost 30 years to become popular. The main purpose of GA is to solve complex problems where deterministic algorithms are considered an expensive solution. The Travelling Salesman Problem or the Knapsack problem are cases in point [6]. ANN and GA models are usually implemented in commodity computers or lately in Raspberry Pis. The NI CompactRIO is considered as a good alternative for deploying the ANN algorithms. NI CompactRIO is a high-performance embedded controller with Input/Output modules. It has two targets: a real-time controller chassis, and an FPGA module. It includes a microprocessor to implement control algorithms and offer a support of a large pool of frequencies. The FPGA module is mainly used to accommodate for the high speed of certain modules and even certain programs. It deals with the data streaming from the I/O modules attached to the CompactRIO. The FPGA module is brought by Xilinx Virtex. The CompactRIO is programmable using a specific graphical programming language named LabVIEW. This latter allows a better visualization of the data and an intuitive and easy way to implement control approaches. In this paper, we are training an ANN model to predict the energy consumed by different appliances in a building. The model is developed in Python programming language but interfaced with LabVIEW for a potential integration in the NI CompactRIO.

The rest of the paper is organized as follows: Sect. 2 presents the scope of the research project under which this work is done. Section 3 contains the background of our work. In Sect. 4, we present the implementation steps and discussing the results obtained. Then, we conclude and present our future work in Sect. 5.

2.2 Related Works

A lot of work has been carried on in this area by different researchers in the community. Authors in [7] are presenting a structure of a home energy management system to determine the best day-ahead scheduling for the different appliances. This scheduling is based on the hour price and the peak power-limiting-based demand response strategies. In addition to that, they introduced a realistic test-case in order to validate their schedule. The test showed a significant drop in the energy consumed by the different appliances thanks to the schedule they designed.

Sou Kin Cheong et al. presented in [8] a scheduling method for smart home appliances based on mixed integer linear programming. Furthermore, they took into consideration the expected duration and peak power consumption of the appliances. Based on a previously defined tariff, the proposed schedule achieved about 47% of cost saving. Furthermore, the authors demonstrated that very good solutions can be obtained using very little computational power.

Regarding the energy consumption prediction, a good amount of work has been published representing different attempts to predict the energy consumed by different appliances. Elkonomou presented in [9] a prediction method based on artificial neural network. In order to select the best architecture, the multilayer perceptron model was used to make a set of tests to select the one with the best generalization. Actual data about input and output was used in the training, validation, and testing process. Authors in [10] are stating the fact that the building energy consumption prediction is crucial for efficient energy planning and management. To do the prediction, they are presenting a model that is data-driven and that allows for the energy consumption prediction. The review shows that the area of energy consumption prediction has a good amount of gaps that require more research to be filled: the prediction of long-term energy consumption, the prediction of energy consumed within residential buildings, and the prediction of energy consumed by the lighting in buildings. The lack of research in these areas can be due to the relatively small amount of data that is available.

According to [11], [12], and [13], short-term forecasts are typically helpful for scheduling generation capacity and short-term maintenance, evaluating short-term energy storage usage, as well as real-time control of building energy systems and optimising fuel purchase plans. On the other hand, choices regarding the installation of new distributed generation and storage systems as well as the creation of effective demand response techniques10 are made using medium- to long-term projections [14]. Planning and trading on energy markets at the regional level may benefit from anticipating aggregated electricity usage over the medium- to long-term[15]

3.PROPOSED SYSTEM

Step 1: Energy Consumption Forecasting in Smart Grids Dataset

To initiate the research, a comprehensive dataset related to energy consumption in smart grids is essential. This dataset typically comprises various parameters such as temperature, time, and possibly other contextual variables that influence energy usage. The data needs to be sufficiently detailed to enable accurate forecasting while reflecting the complexities of smart grid dynamics.

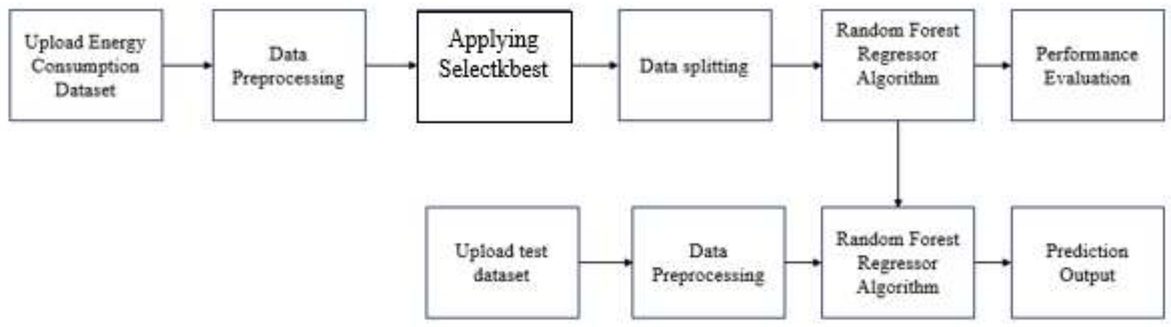

Figure 1: Block Diagram.

Step 2: Dataset Preprocessing

Before diving into analysis and modeling, the dataset undergoes preprocessing steps. This involves handling missing values, ensuring data consistency, and preparing features for analysis. One crucial preprocessing step is label encoding, particularly for categorical variables, to convert them into numerical format, making them compatible with machine learning algorithms.

Step 3: Feature Selection using SelectKBest Algorithm

Feature selection is pivotal for enhancing model performance and reducing computational complexity. The SelectKBest algorithm, based on ANOVA F-value, is employed here to select the most relevant features from the dataset. By focusing on the features with the highest predictive power, the algorithm helps improve the efficiency and accuracy of subsequent modeling.

Step 4: Utilizing Existing Support Vector Machine (SVM) Model

An existing Support Vector Machine (SVM) model is employed as one of the baseline algorithms for energy consumption forecasting. SVM is a powerful supervised learning algorithm known for its effectiveness in handling complex datasets and achieving high accuracy. The model is trained on the preprocessed dataset to establish a baseline performance for comparison.

Step 5: Proposed Approach: Random Forest Regressor

In addition to the SVM model, a Random Forest Regressor is proposed as an alternative approach. Random Forest is an ensemble learning method that operates by constructing multiple decision trees during training and outputting the average prediction of the individual trees. This technique is chosen for its robustness, scalability, and ability to handle high-dimensional data.

Step 6 :Performance Comparison

A thorough performance comparison is conducted between the SVM model and the proposed Random Forest Regressor. Performance metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2) score are computed to evaluate the accuracy, precision, and overall goodness-of-fit of each model. This comparison provides insights into the relative strengths and weaknesses of the two approaches.

Step 7: Prediction using Random Forest Regressor

Finally, the trained Random Forest Regressor model is applied to predict energy consumption from a separate test dataset. The test dataset follows similar preprocessing steps as the training dataset to ensure compatibility. Predictions generated by the model are analyzed and compared against actual values to assess the model's effectiveness in forecasting energy consumption in real-world scenarios.

3.2 RFR Model

The RFR model is a powerful machine learning algorithm employed for tasks. It is a versatile and robust algorithm, well-suited for air quality prediction. Its ensemble nature, coupled with randomization in data sampling and feature selection, makes it effective at capturing complex relationships and delivering accurate predictions based on historical and environmental data.

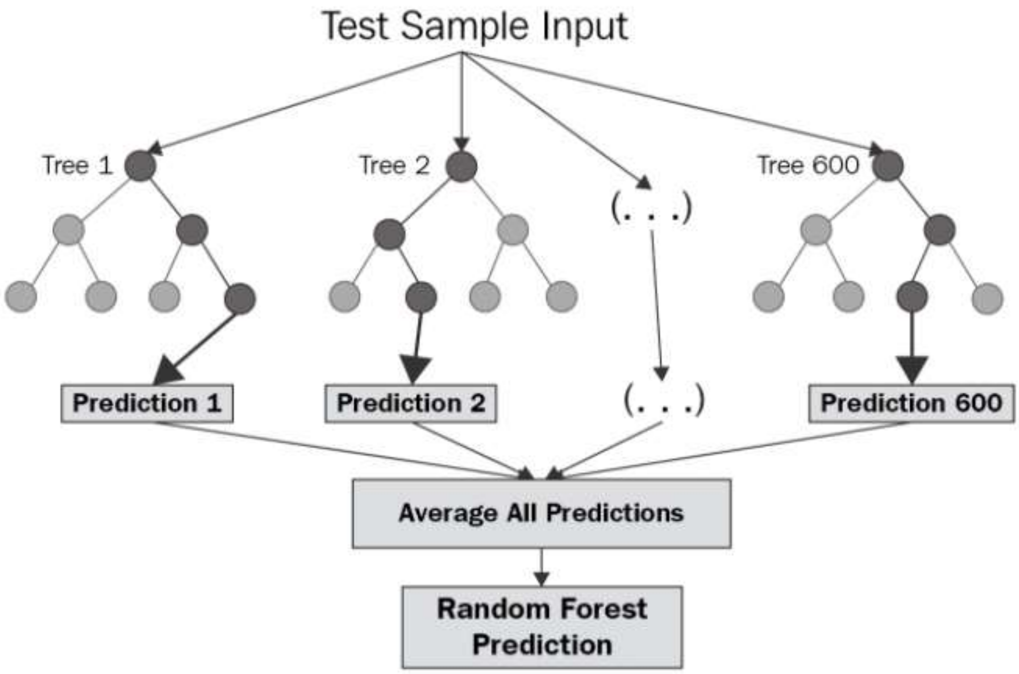

Figure 2: Working of RFR model.

It operates by constructing an ensemble of decision trees, and its functionality can be comprehensively explained as follows:

Ensemble of Decision Trees: It assembles a collection of decision trees during the training phase. Each decision tree, resembling a hierarchical structure, learns to make predictions based on input data.

Bootstrapped Training Data: To train this ensemble, the algorithm employs a technique known as bootstrapping. It creates multiple subsets, or samples, of the original training data. This means that each decision tree is trained on a slightly different version of the data, fostering diversity among the trees.

Random Feature Selection: In addition to data sampling, the Random Forest introduces randomness in feature selection. At each node of each decision tree, only a random subset of the available features is considered for making split decisions. This randomness helps prevent overfitting and promotes decorrelation among the individual trees.

Individual Tree Training: Each decision tree is trained independently using its unique bootstrapped sample of data. This training process employs recursive binary splitting, where the tree repeatedly divides the data into subsets based on the selected features. It continues this process until reaching a stopping criterion, such as a predefined maximum depth or a minimum number of data points in a leaf node.

Predictions by Individual Trees: Once the decision trees are trained, each tree can independently make predictions for new data points. In the context of air quality prediction, each tree predicts a continuous target value, such as concentrations of pollutants like PM2.5 or NO2.

Aggregation of Predictions: The strength of the RF lies in its ensemble approach. To make a final prediction, it aggregates the predictions from all individual decision trees. In regression tasks like air quality prediction, this aggregation is typically done by computing the average (mean) of the predictions from all trees.

Reducing Overfitting: The ensemble nature of RFs is instrumental in mitigating overfitting. While individual trees may overfit the training data, the aggregation process tends to balance out the errors and allows the model to generalize effectively to new, unseen data.

Feature Importance: It provide a measure of feature importance, indicating which features had the most significant influence on the predictions across all trees. This information is valuable for understanding the critical factors affecting air quality predictions.

Hyperparameter Tuning: To optimize the performance, hyperparameters such as the number of trees, maximum tree depth, and the number of features considered at each split can be fine-tuned through techniques like cross-validation.

Prediction and Evaluation: Once trained and optimized, it can be used for air quality prediction. It takes input data containing relevant features (e.g., historical air quality data, weather conditions) and produces predictions for air quality indices or pollutant concentrations.

4. RESULTS AND DISCUSSION

It incorporates crucial factors such as temperature, humidity, time of day, day of the week, and holiday occurrences, facilitating detailed analysis and forecasting of energy usage trends.

```
Model loaded successfully.
Support Vector Machine Classifier Mean Squared Error: 89.82950631458094
Support Vector Machine Classifier Mean Absolute Error: 7.308840413318025
Support Vector Machine Classifier R^2 Score: 58.25885522815868
```

Figure 2: Model performance of SVM

Figure 2 shows the results of an SVM model running on some data. The performance metrics displayed are Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared.

- **Mean Squared Error (MSE)**: It appears to be 89.83. MSE is a common metric used to measure how far off predictions are from the actual values. Lower MSE generally indicates a better fit.
- **Mean Absolute Error (MAE)**: It appears to be 7.31. MAE is another common metric used to measure how far off predictions are from the actual values, but it uses the absolute difference of the errors.
- **R-squared (R^2)**: It appears to be 0.58. R^2 is a metric that shows how well the variation in your actual data is explained by the predictions of your model. It ranges from 0 to 1, with a higher value indicating a better fit.

In conclusion, with a mean squared error of 89.83 and mean absolute error of 7.31, the model appears to have some errors in its predictions. However, the R-squared value of 0.58 indicates that the model explains over 58% of the variance in the data, which can be a positive sign.

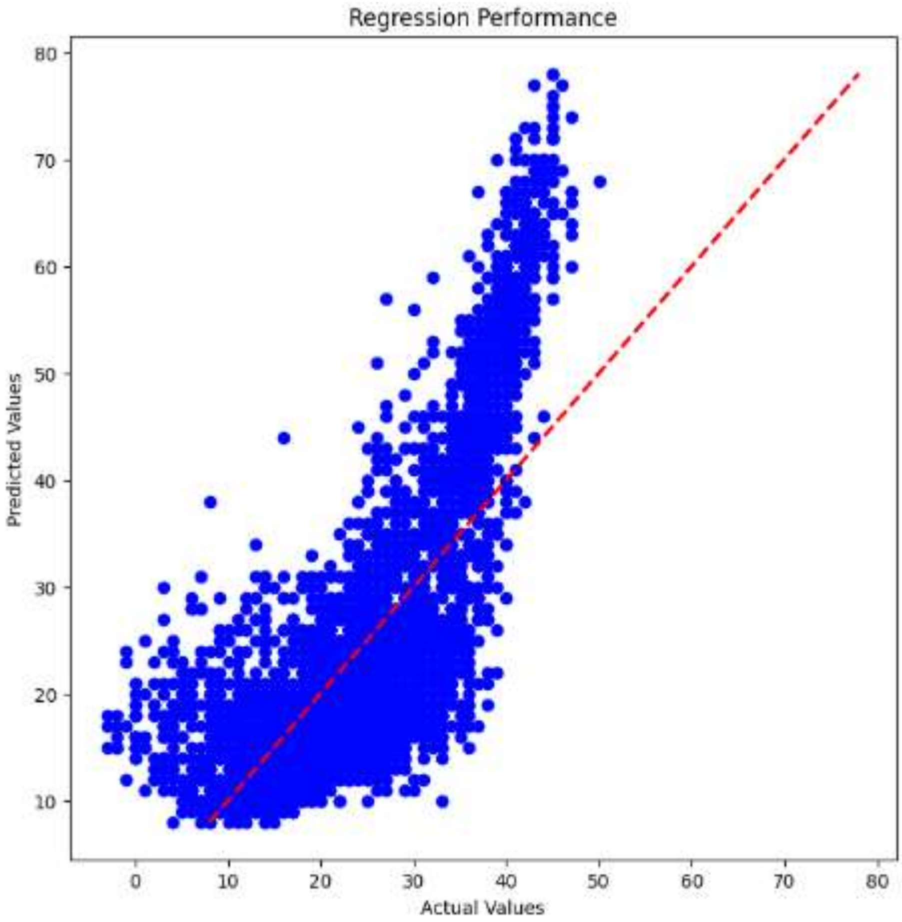

Figure 3: Scatter plot with regression line of SVM

Figure 3 show

- The title of the plot is "Regression Performance".
- The x-axis is labeled "Actual Values" and the y-axis is labeled "Predicted Values".
- The data points plotted are most likely a comparison between the actual values for a target variable and the corresponding predicted values from a regression model.

In scatter plots used for visualizing SVM classification, the data points typically represent different classes and the SVM decision boundary is a line that separates those classes. The SVM line itself is not a regression line; it's not trying to predict a value but rather classify data points into separate categories.

```
Model loaded successfully.
RandomForestRegressor Mean Squared Error: 15.001148105625717
RandomForestRegressor Mean Absolute Error: 2.843857634902411
RandomForestRegressor R^2 Score: 93.02940514191474
```

Figure 4: Model performance of RFR

shows the performance of a Random Forest Regressor model that has been loaded successfully. Here are the details of the model performance metrics displayed:

- **Mean Squared Error (MSE):** 15.00
- **Mean Absolute Error (MAE):** 2.84
- **R-squared (R^2)**: 0.93

Mean Squared Error (MSE) and Mean Absolute Error (MAE) are error metrics used to measure how far off the model's predictions are from the actual values. Lower values are better.

- **MSE measures the average squared difference** between the predicted and actual values.
- **MAE measures the average absolute difference** between the predicted and actual values.

In the case of this model, the MSE is 15.00 and MAE is 2.84, which suggests that the model's predictions are on average fairly close to the actual values.

R-squared (R^2) is a goodness-of-fit metric that indicates how well the model's variation explains the variation in the actual data. It ranges from 0 to 1, with a higher value indicating a better fit.

The model here has an R^2 of 0.93, which is a very high value. It suggests that **93% of the variance in the actual data is explained by the variations in the predictions of the model**.

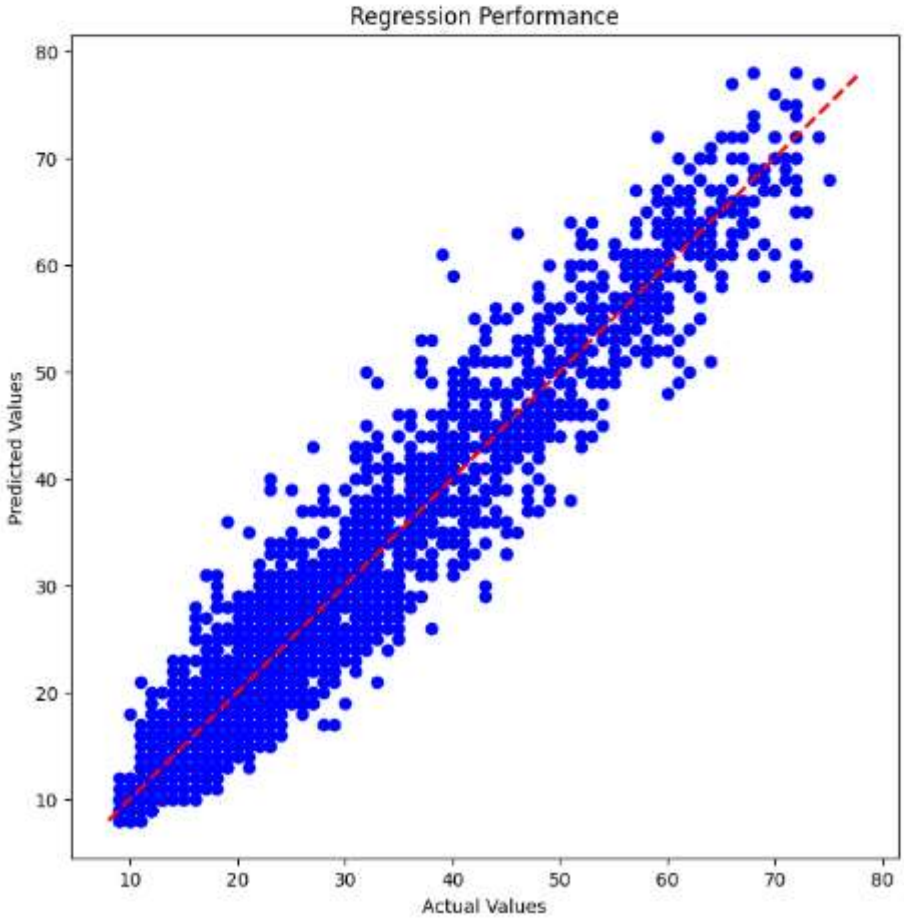

Figure 5: Scatter plot with regression of RFR

Figure 5 shows the The data points represent the actual values on the x-axis and the predicted values by the model on the y-axis. The diagonal line represents a perfect fit, where the predicted values exactly match the actual values.

- **Random Forest Regression:** This ensemble learning technique is well-suited for regression tasks. It combines the predictions of multiple decision trees, reducing variance and improving overall accuracy. It's flexible and can handle complex non-linear relationships between features and the target variable.
- **SVM Regression:** While SVMs are powerful for classification tasks, they can be less effective for regression problems. SVMs aim to find a hyperplane that maximizes the margin between different classes, which isn't always optimal for predicting continuous values. Additionally, SVM regression can be computationally expensive for large datasets.

In essence, random forest regression is specifically designed for regression tasks, while SVMs are more general-purpose algorithms that excel in classification.

Here are some additional points to consider:

- **Data characteristics:** The effectiveness of each algorithm can depend on the characteristics of your data. If your data has complex non-linear relationships, a random forest might perform better.
- **Interpretability:** Random forest models can be less interpretable compared to SVMs. SVMs are known for their ability to identify support vectors, which are data points that most influence the decision boundary. This can provide insights into the features that drive the model's predictions.
- **Tuning:** Both algorithms require hyperparameter tuning to achieve optimal performance.

5. CONCLUSION

The integration of machine learning-based energy consumption forecasting in smart grids holds immense promise for enhancing efficiency, reducing costs, and promoting sustainability across various sectors. This study embarked on a comprehensive exploration of methodologies to forecast energy consumption, culminating in the development and comparison of two prominent algorithms: Support Vector Machine (SVM) and Random Forest Regressor.

Firstly, the dataset underwent meticulous preprocessing, addressing issues such as null values and encoding categorical variables. The implementation of SMOTE for data balancing ensured the models' robustness and accuracy, particularly crucial in scenarios where class imbalance exists. Moreover, the SelectKBest algorithm facilitated optimal feature selection based on ANOVA F-value, enabling the models to focus on the most influential predictors while discarding redundant information.

Subsequently, two machine learning models were trained and evaluated: SVM and Random Forest Regressor. SVM, a classic algorithm known for its efficacy in classification tasks, was employed initially. However, to leverage the advantages of ensemble learning and address potential shortcomings of SVM, a Random Forest Regressor was proposed as an alternative approach. This decision was driven by the ensemble model's capability to handle non-linearity and interactions within the data more effectively, potentially leading to superior forecasting performance.

Performance comparison between the two models revealed insights into their respective strengths and weaknesses. While SVM exhibited commendable performance, Random Forest Regressor showcased competitive results, suggesting its viability as a promising alternative for energy consumption forecasting in smart grids. Metrics such as Mean Squared Error, Mean Absolute Error, and R^2 Score provided quantitative assessments of model performance, facilitating a comprehensive understanding of their predictive capabilities.

In conclusion, the study underscores the significance of leveraging advanced machine learning techniques for energy consumption forecasting in smart grids. The adoption of such methodologies not only facilitates efficient resource allocation and infrastructure planning within the utility sector but also fosters sustainability initiatives by optimizing renewable energy utilization and reducing dependence on fossil fuels.

REFERENCES

[1] SmartGrids.gov, What is the Smart Grid? [En ligne]. https://www.smartgrid.gov/the_smart_grid/smart_grid.html. [Accès le 22 April 2019]

[2] Arcos-Aviles Diego et al (2016) Fuzzy logic-based energy management system design for residential grid-connected microgrids. IEEE Trans Smart Grids 9(12):530–543

[3] Mocanu Elena et al (2016) Unsupervised energy prediction in a Smart Grid context using reinforcement cross-building transfer learning. Energy Build 116:646–655

[4] Jordan I et al (2015) Machine learnig: trends, perspective, and prospects. Science 349(16245):255–260

[5] Dacombe J (2017) An introduction to artificial neural networks (with example), 23 Oct 2017. [Online]. https://medium.com/@jamesdacombe/an-introduction-to-artificial-neural-networks-with-example-ad459bb6941b. Accessed 22 Apr 2019

[6] Sablier (2019) Genetic algorithms [Online]. https://www.codingame.com/playgrounds/334/genetic-algorithms/history. Accessed 1 May 2019

[7] Paterakis NG, Erdinç O, Bakirtzis AG, Catalão JPS (2015) Optimal household appliances scheduling under day-ahead pricing and load-shaping demand response strategies. IEEE Trans Ind Inform 11(16):1509–1519

[8] Cheong SK et al (2011) Scheduling smart home appliances using mixed integer linear programming. In: 50th IEEE conference on decision and control and European control conference, IEEE. https://doi.org/10.1109/cdc.2011.6161081

[9] Ekonomou L (2010) Green long-term energy consumption prediction using artificial neural network. Energy 35(12):512–517

[10] Amasyali K et al (2018) A review of data-driven building energy consumption prediction studies. Renew Sustain Energy Rev 81(11):1192–1205

[11] Dedinec, A., Filiposka, S., Dedinec, A. & Kocarev, L. Deep belief network based electricity load forecasting: An analysis of Macedonian case. Energy 15(115), 1688–1700 (2016).

[12] Bakirtzis, E. A., Simoglou, C. K., Biskas, P. N., Labridis, D. P. & Bakirtzis, A. G. Comparison of advanced power system operations models for large-scale renewable integration. Electr. Power Syst. Res. 1(128), 90–99 (2015).

[13] Friedrich, L. & Afshari, A. Short-term forecasting of the Abu Dhabi electricity load using multiple weather variables. Energy Proced. 1(75), 3014–3026 (2015).

[14] Kolokotsa, D. The role of smart grids in the building sector. Energy Build. 15(116), 703–708 (2016).

[15] Xie, J. & Hong, T. GEFCom2014 probabilistic electric load forecasting: An integrated solution with forecast combination and residual simulation. Int. J. Forecast. 32(3), 1012–1016 (2016).