"Power Consumption and Monitoring Recommendation System"

A PROJECT REPORT

Submitted by

Mohit Kumar Shrivastava (12311713)
Satyam Vishwakarma (12308720)

in partial fulfillment for the award of the degree of

Bachelor of Technology
IN
Computer Science and Engineering

Lovely Professional University, Punjab

Lovely Professional University, Punjab

BONAFIDE CERTIFICATE

Certified that this project report **"Power Consumption Monitoring and Recommendation System"** is the bonafide work of " Mohit Kumar Shrivastava (12311713) and Satyam Vishwakarma (12308720) **"** who carried out the project work under my supervision.

SIGNATURE

SUPERVISOR
Dr.Vishu and Premananda Sahu

Department of Machine Learning, CSE,
Lovely Professional University, Punjab

SIGNATURE
HEAD OF THE DEPARTMENT

TABLE OF CONTENTS

POWER CONSUMPTION MONITORING AND RECOMMENDATION SYSTEM

[1] Mohit Kumar Shrivastava, [2]Satyam Vishwakarma

1 School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab
mohitshrivastav13p@gmail.com
2 School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab
satyamv927@gmail.com

Abstract-In order to facilitate anomaly detection and predictive modelling, this project offers a thorough examination of building power consumption data using machine learning approaches. To comprehend consumption patterns, the dataset—which includes timestamped energy usage measurements for HVAC, lighting, electronics, and environmental conditions—is first pre-processed and visualized. Standard regression measures, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2), are used to assess the effectiveness of a Random Forest Regressor, which is used to estimate total power usage based on certain variables. For comparative analysis, a Linear Regression model is also trained, and a boxplot is used to display the prediction errors.
An Isolation Forest is used for unsupervised anomaly detection in order to find anomalous consumption patterns. In order to help identify any energy inefficiencies or flaws, anomalies are identified and displayed in a time-series plot. Lastly, the system offers practical suggestions for lowering energy consumption, with a focus on customers who account for the top 5% of total consumption. This method provides a useful framework for facility management systems or smart building energy monitoring, forecasting, and optimization.

Keywords-: Power Monitoring, Smart Metering, IoT, Machine Learning, Consumption Optimization

I.INTRODUCTION

Understanding and controlling power use is crucial for cutting expenses, boosting operational efficiency, and improving sustainability in today's energy-conscious environment. Large amounts of power usage data can now be examined to identify consumption trends, forecast future demand, and identify anomalous activity that can point to inefficiencies or problems thanks to the expanding availability of smart energy meters and Internet of devices.
The goal of this research is to apply machine learning techniques to analyze power usage data at the building level. Timestamped records of energy use in a variety of categories, including kitchen appliances, electronics, lights, HVAC systems, and environmental variables like humidity and temperature, are included in the collection. In order to facilitate temporal analysis, the software starts with data ingestion and preprocessing, reading the dataset from an Excel sheet and transforming the timestamp column into a datetime format. After that, the features are chosen according to how well they relate to the goal variable, Total_Consumption_kWh.The main model used to forecast the overall energy usage using the chosen input features is a Random Forest Regressor. The model's performance is verified by separating the data into training and testing sets. The accuracy and dependability of the regression model are evaluated using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) Score. Additionally, a Linear Regression model is trained for comparative analysis. The code uses the Isolation Forest technique to detect anomalies in addition to the prediction task. This kind of unsupervised learning finds data points that differ greatly from the norm, which may be a symptom of malfunctioning systems, incorrect data entry, or unusual usage patterns. A complete classification report, precision, recall, F1-score, and other classification metrics are used to assess the identified anomalies once they have been transformed into binary labels. This technique aids in identifying possible problems for more research, even when the ground truth for anomalies is not externally established.The programming incorporates visualizations to facilitate improved comprehension and decision-making. Normal and abnormal consumption trends throughout time are shown in a time-series line plot. Facility managers or analysts can more easily spot energy usage anomalies or spikes thanks to this visual representation. Additionally, the tool targets the top 5% of users in the dataset with recommendations for energy efficiency. These recommendations include easy yet efficient ways to save money and energy, such as turning to energy-efficient lighting, lowering HVAC use during peak hours, and unplugging electronics.
All things considered, this project exemplifies a comprehensive method of energy data analytics by combining anomaly detection, prediction, and

interpretability through visual aids and useful insights. It demonstrates how cutting-edge machine learning methods can be used to address practical energy management icommercial, industrial, and residential.

II. EXISTING WORK

The 1980s saw the initial development of the NILM approach [2]. The process of determining the power consumption of each individual appliance based on the total power demand recorded by a single smart meter that monitors several appliances is called energy disaggregation, or NILM in the literature [3]. Vitality consumption is the sum of the electricity used by all of the appliances in a building. As a result, each appliance's electricity use needs to be noted. Equation (1) states that the total active energy of all the appliances at time t can equal the aggregated energy of N appliances.

$$y(t) = \sum_{i=1}^{N} y_i(t) + e(t)$$

The energy accumulated at time t is represented by $y(t)$, the number of appliances is represented by N, the fraction of energy consumption at time t by appliance i is represented by $yi(t)$, and noise or undesired energy is represented by e(O). Three categories can be used for the examination of energy data: algorithms that are 1) supervised, 2) unsupervised, and 3) evolutionary. To create a model, supervised learning techniques need labelled samples of every appliance. Neural networks were initially used to identify the energy usage of household appliances in order to suggest the application of machine learning techniques in NILM [4]. Then, for energy disaggregation, backpropagation, learning vector quantisation, recurrent neural networks, and sparse auto-encoders [5–7] were suggested. Deep neural networks outperform Hidden Markov Models, despite the challenging training process. Recently, it has been suggested that malfunctioning appliances can be identified by using DAEs and the Long Short-Term Memory (LSTM) architecture [8]. An additional field in which neural networks for NILM are the Convolutional Neural Network (CNN) [10] and the Restricted Boltzmann Machine (RBM) [9, 10]. Unsupervised methods are less expensive and more dependable for real-time NILM than supervised methods as they don't require pre-training. Three subgroups of unsupervised NILM techniques can be distinguished: 1) Unsupervised methods that create appliance models through unlabelledtraining. They frequently use HMMs and generate appliance models either automatically or manually throughout the training phase [11]; for instance, the energy usage of every household device can be measured using a single hardware configuration. This hardware configuration uses certain machine learning techniques, such as Combinatorial Optimisation (CO) and Factorial Hidden Markov Model (FHMM), to break down the generated device-specific values by combining household energy readings. Following transmission to a cloud database, these values are shown to the user via a visual interface resembling a dashboard [12]. 2) Unsupervised methods that create appliance models using labelled data, then utilise those models to break down building energy. These methods necessitate gathering appliance data. Appliance models are created using this data and subsequently installed in new structures. This group includes the majority of deep learning-based NILM techniques, and 3) Unsupervised approaches don't require training prior to energy disaggregation. These Energy can be broken down using methods that don't require measured data or prior knowledge. Some NILM investigations have also made use of the evolutionary and genetic algorithms. The primary purposes of the genetic algorithm are to optimise the current parameters in fuzzy systems and to find patterns and characteristics in the energy profiles of appliances [13, 14, 15]. For NILM, the differential evolution algorithm is a straightforward and reliable metaheuristic method that makes use of genetic algorithm operators like crossover, mutation, and selection [16].

III. METHODS AND TECHNIQUES

The Power Consumption Monitoring and Recommendation System collects data from appliances in real time using smart meters or Internet of Things-based sensors, then sends the data using wireless technologies like Wi-Fi or Zigbee. To find usage trends, predict loads, and spot abnormalities, this data is processed using machine learning and data analytics tools like Python and TensorFlow and stored on cloud platforms like AWS or Firebase.

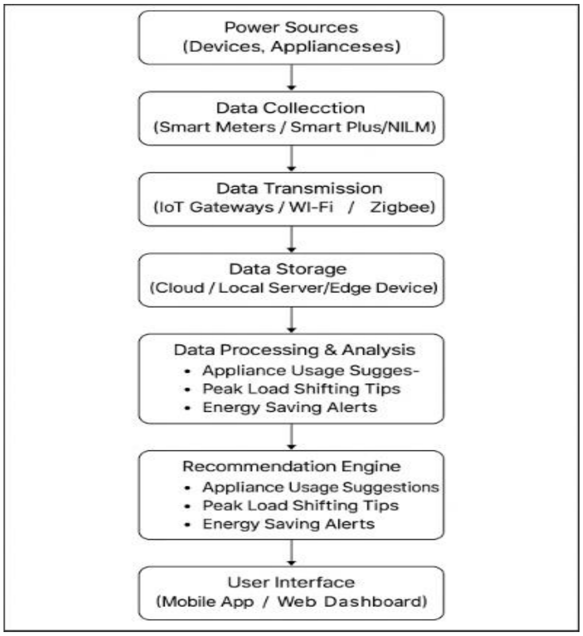


Fig.1. Workflow Model

The above workflow diagram has expressed in following manner:

A. Feature engineering and data preprocessing

Data formatting and cleaning: Time-based power usage records are included in the dataset. Converting timestamp values into a common datetime format is a crucial step in time-series analysis and pattern recognition.

Selection of Features: Important factors including lighting, electronics, temperature, humidity, HVAC use, and power rates are chosen as features. These characteristics are thought to have a significant impact on the overall power consumption.

B. Regression Models for Supervised Learning

Definition of Random Forest Regression: An ensemble learning technique that increases accuracy and robustness by constructing several decision trees The project's goal was to forecast overall energy usage by utilizing a variety of factors. It works particularly well for managing non-linear interactions. A statistical technique that uses a straight line to represent the relationship between a dependent variable and one or independent.The project's goal is to serve as a baseline model for comparison with more intricate models, such as Random Forest. It aids in determining whether less complex models can perform comparably.

C. Measures of Evaluation (for Regression)

MAE, or mean absolute error calculates the average size of a series of predictions' errors without taking into account their direction. The average of the squares of the mistakes is calculated via the Mean Squared Error (MSE). Compared to MAE, it penalizes greater errors more. The percentage of the dependent variable's variance that can be predicted from the independent variables is indicated by the R-squared (R^2) score. A better fit is indicated by a score nearer 1.

D. Metrics for Classification Evaluation

Following the identification of anomalies, classification measures are used to evaluate their performance:

Precision: Indicates the proportion of expected anomalies that really occurred.

Recall: Indicates the proportion of real anomalies that were accurately found.

The F1 Score, which provides a fair assessment in cases of class imbalance, is the harmonic mean of precision and recall.

Classification Report: A thorough analysis of each class's support, F1 score, precision, and recall (normal vs. anomalous).

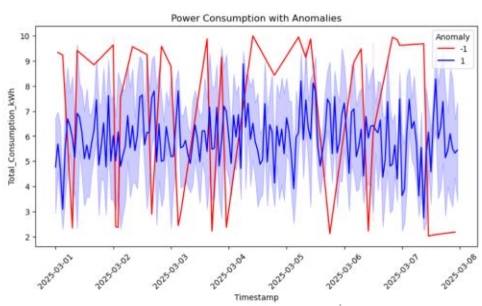
E. Techniques for Data Visualization

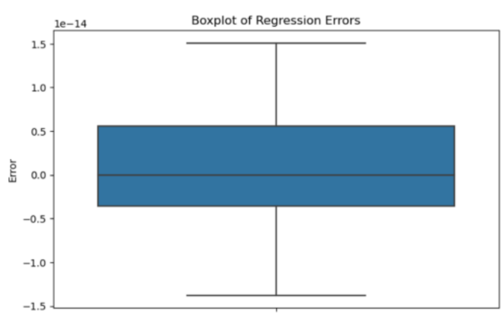
Time series data is represented by line plots, which highlight anomalies found and demonstrate how power use changes over time. Boxplots are a useful tool for understanding model performance and identifying outliers by visualizing the distribution and spread of regression mistakes.

The provided code uses a variety of data science and machine learning approaches to offer a thorough approach to power consumption data analysis and modeling. For efficient time-series handling, the dataset is first pre-processed by transforming timestamp values into datetime format. The goal variable, which is total power consumption, is predicted by key energy-influencing factors, including HVAC, lighting, electronics, kitchen appliance utilization, temperature, humidity, and tariff rates. Random Forest Regression and Linear Regression are the two supervised machine learning models used. An ensemble learning technique called Random Forest is used because it is reliable and can represent intricate, non-linear relationships by averaging predictions from several decision trees, which lowers overfitting and increases accuracy.

IV.EVALUATION METRICS AND RESULT OF THE PROPOSED STUDY

The efficiency of the machine learning models utilized for both regression and anomaly detection tasks was evaluated in the proposed study using a variety of evaluation criteria. Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) were the three primary evaluation metrics used for the regression task, which involved predicting total energy consumption based on a variety of features, including HVAC, lighting, electronics, kitchen usage, temperature, humidity, and tariff rate. An intuitive sense of how far the model is off on average in its predictions is provided by MAE, which calculates the average size of the errors in a collection of forecasts without taking into account their direction. The Isolation Forest method was used to find anomalies. By choosing a feature at random and then choosing a split value at random from the feature's maximum and lowest values, this unsupervised learning method isolates anomalies. After transforming the anomalies (-1 for outliers and 1 for normal) into binary labels, the model outputs were assessed using the following classification metrics: F1-Score, Precision, and Recall. A balanced assessment is provided by the F1-score, which provides a harmonic mean of precision and recall. Precision measures the proportion of discovered anomalies that were truly true anomalies, while recall evaluates the proportion of true anomalies that were effectively detected.Overall, the suggested study combines efficient anomaly detection with high-performing regression modeling to provide a solid method for forecasting and examining power usage trends.





V.CONCLUSION

In both residential and business settings, the Power Consumption Monitoring and Recommendation System is essential for advancing sustainability and energy efficiency. This solution gives consumers the ability to make educated judgements to cut down on needless power consumption by continuously tracking energy usage trends and detecting high-consumption gadgets. Proactive behaviour modification is made possible by the combination of intelligent data and tailored recommendations, which lowers costs and the carbon footprint. All things considered, a system like this helps both individual users and more general environmental and energy conservation objectives.

REFERENCES

[1] Cominola, A., Giuliani, M., Piga, D., Castelletti, A., & RizzoliA. E. (2017). A hybrid signature-based iterative disaggregation algorithm for non-intrusive load monitoring. Applied energy, 185, 331-344.

[2] Hart, G. W. (1992). Nonintrusive appliance load monitoring. Proceedings of the IEEE, 80(12), 1870-1891

[3] Piccialli, V., & Sudoso, A. M. (2021). Improving non-intrusive load disaggregation through an attentionbased deep neural network. Energies, 14(4), 847.

[4] Kolter, J. Z. N. A., Batra, S., & Ng, A. (2010). Energy disaggregation via discriminative sparse coding. Advances in neural information processing systems, 23, 1153-1161

[5] Yang, H. T., Chang, H. H., & Lin, C. L. (2007, April). Design a neural network for features selection in nonintrusive monitoring of industrial electrical loads. In 2007 11th International Conference on Computer Supported Cooperative Work in Design (pp. 1022-1027). IEEE.

[6] Lin, Y. H., & Tsai, M. S. (2010, May). A novel feature extraction method for the development of nonintrusive load monitoring system based on BP-ANN. In 2010 International Symposium on Computer, Communication, Control and Automation (3CA) (Vol. 2, pp. 215-218). IEEE.

[7] Kelly, J., & Knottenbelt, W. (2015, November). Neural nilm: Deep neural networks applied to energy disaggregation. In Proceedings of the 2nd ACM international conference on embedded systems for energy-efficient built environments (pp. 55-64).

[8] Mocanu, E., Nguyen, P. H., Gibescu, M., & Kling, W. L. (2016). Deep learning for estimating building energy consumption. Sustainable Energy, Grids and Networks, 6, 91-99.

[9] Mocanu, D. C., Mocanu, E., Nguyen, P. H., Gibescu, M., & Liotta, A. (2016, October). Big IoT data mining for real-time energy disaggregation in buildings. In 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 003765-003769). IEEE.

[10] Zhang, C., Zhong, M., Wang, Z., Goddard, N., & Sutton, C. (2018, April). Sequence-to-point learning with neural networks for non-intrusive load monitoring. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 32, No. 1).

[11] Deshpande, R., Hire, S., & Mohammed, Z. A. (2022). Smart Energy Management System Using Non-intrusive Load Monitoring. SN Computer Science, 3(2), 1-11.

[12] Kim, H., Marwah, M., Arlitt, M., Lyon, G., & Han, J. (2011, April). Unsupervised

disaggregation of low frequency power measurements. In Proceedings of the 2011 SIAM international conference on data mining (pp. 747-758). Society for Industrial and Applied Mathematics.

[13] Egarter, D., & Elmenreich, W. (2013, July). Evonilm: Evolutionary appliance detection for miscellaneous household appliances. In Proceedings of the 15th annual conference companion on Genetic and evolutionary computation (pp. 1537-1544).

[14] Trung, K. N., Dekneuvel, E., Nicolle, B., Zammit, O., Van, C. N., & Jacquemod, G. (2014, June). Event detection and disaggregation algorithms for nialm system. In Proceedings of 2nd International Non-Intrusive Load Monitoring (NILM) Workshop.

[15] Hassan, T. (2012). Bi-level characterization of manual setup residential non-intrusive demand disaggregation using enhanced differential evolution. In Proc. 1st Int. Workshop Non-Intrusive Load Monitoring.