Tec Empresarial
P-ISSN: 1659-2395; E-ISSN: 1659-3359

AI BASED POWER CONSUMPTION ANALYSIS AND PREDICTION FOR ELECTRICAL APPLIANCES.

R.Breesha[1],D.Shofia priyadharshini[2],M.Sheriff [3],Hemalatha G[4],Deepalakshmi L[4],Monika.M[4]

[1,2,3]Assistant Professor,Department of Electronics and Communication Engineering,Vel Tech High tech Dr.Rangarjan Dr.Sakunthala Engineering College,Chennai,India

[4]UG Scholar,Department of Electronics and Communication Engineering,Vel Tech High tech Dr.Rangarjan Dr.Sakunthala Engineering College,Chennai,India.

Abstract — Electricity consumption has been a subject of extensive research in the field of computer architecture for many years. The primary aim of this study is to provide useful guidelines to the machine learning community, enabling them to use and develop energy estimation techniques for machine learning algorithms. Various ensemble models, such as Linear Regression, Random Forest Regression, and LSTM, have been utilized to predict electricity consumption and achieve accurate results. Additionally, this project introduces the latest software tools that support electricity estimation principles, along with two use cases that reinforce the investigation of energy usage in machine learning. Ultimately, this study predicts future energy usage, which is instrumental in enabling the grid to accurately provide energy by leveraging smart meters that provide insights into appliance usage patterns. This information helps users determine when they require more or less energy, making it immensely valuable.

Keyword — LSTM, Linear Regression, Random forest regression, Electricity estimation, Future energy

I. INTRODUCTION

Energy use is on the rise due to the increased use of domestic and industrial applications, such as cars, generators, mobile devices, and household appliances. The growth of smart meter infrastructure (SMI) has made it possible to incorporate active energy systems into intelligent meters, allowing for energy usage modeling and prediction [1]. The use of electrical equipment and consumer behavior is having a significant impact on the electricity market, leading power grid administrators to develop new ways of managing energy demand effectively [2].

Smart residential buildings provide remote control of electronic devices through mobile applications, but the sensors required to support this feature consume a lot of energy. An ensemble regression model combining linear prediction and SVR prediction methods has been developed to increase electricity prediction efficiency. Accurateenergy demand forecasting is crucial to reducing energy waste caused by equipment misuse [3].

Various factors can affect energy consumption, such as climate conditions, building materials, and sub-level structures for heating, lighting, and ventilation [4]. Consumers can also affect energy consumption by changing the load of appliances or the number of occupants in the building [5]. Accurately predicting energy consumption is challenging and requires historical data from 2006 to 2010 to deploy resources effectively [6].

II. RELATED WORKS

Corgnati et al. (2013) used the input (regressor vari-ables) and output variables (response). Based on this data, system parameters will be estimated and thus, a mathematical model could be generated. Several previous studies have analysed the data-driven machine learning approach.

Fu et al. (2015)proposed using one of ML algorithms which is Support Vector Machine (SVM) to predict the load at a building's system-level (air conditioning, lighting, power, and others) based on weather predictions and hourly electricity load input. Overall, SVM method managed to predict the total electricity loadwith root mean square error (RMSE) of 15.2% and mean bias error (MBE)of 7.7%.

Findings byValgaev et al. (2016)proposed a power demand forecast using kNearest Neighbour (k-NN) model at a smart building as part of the Smart City Demo Aspern (SCDA) project. The k-NN forecasting method was approached using a set of historical observations (daily loadcurves) and their successors. The k-NN method is good at classifying data but limited in forecasting future value as it only identifies similar in-stances in large feature space. Therefore, it must be complemented with temporal information identification whereby the prediction will be made for the next 24 h during workdays.

Five methods of machine learning techniques were used for short-term load forecasting by El Khantach et al. (El Khantach et al., 2019)with an initial decomposition of the historical data done periodically into time series of each hour of the day, which finally constituted 24-timeseries that represented every past hour. The five machine learning methods used are multi-layer perceptron (MLP), support vector machine(SVM), radial basis function (RBF) regressor, REPTree, and Gaussian process. The experimentation was done based on data derived from the Moroccan electrical load data. The result showed that MLP method came out as the most accurate with MAPE percentage of 0.96 while SVM came second and although far from the result of MLP, it was still better than the rest. Although the prediction of energy consumption usually uses a classification-based machine learning method, prediction could also be made based on the regression 12 12 method as studied by Gonzalez-Brioneset al (2019).

The research constructed a predictive model by asing the historical data set using Linear Regression (LR), Support Vector Regression (SVR), Random Forest (RF), Decision Tree (DT) and kNearest Neighbour (k- NN). The parameters of the research used one day-before electricity consumption (kWh) as an additional attribute. The results showed that LR and SVR models had the best performance with 85.7%accuracy.

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.

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 13, 13 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'

III. METHODOLOGY

To create a supervised machine learning model that generates continuous values, historical data is utilized during the training process. The model is designed to predict the likelihood of a new power consumption value based on various contextual information, including background data.

Proposed Work

We assessed the effectiveness of our proposed method using a publicly available dataset of individual household power consumption from the UCI machine learning repository. This dataset includes information on electric power usage between 2006 and 2010, and consists of 1048575 rows and 9 columns. We trained each supervised regressor model on the training set, utilizing all of the features, and then applied them to the entire test set to evaluate their performance. To analyze performance over time, we employed three different machine learning algorithms: logistic regression (LR), random forest (RF), and LSTM, utilizing the scikit-learn implementation of each.

SYSTEM ARCHITECTURE

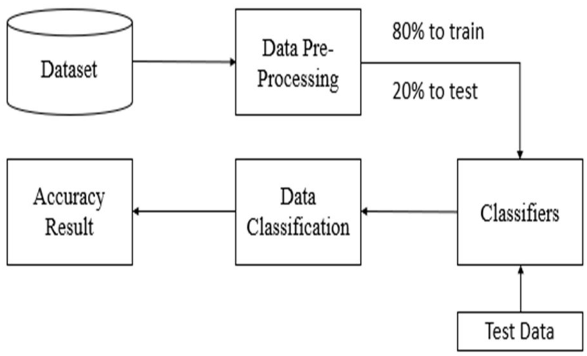


Fig:1 System architecture of the modal.

Modules

A. Data Collection
B. Data Pre-Processing
C. Data Splitting
D. Evaluation Model

A. DATA COLLECTION: Data used in this paper is a set of meter data having several appliances.

- Dataset is a multivariate time series dataset that describes the electric consumption for a household
- Global_active_power , global_reactive power,voltage ,global_intensity, ST, ET etc are the the attributes used
- The dataset is split randomly into two parts : training set , test set.
- Mean Absolute Errors (MAE), Root Mean Squared Errors (RMSE) are calculated for the datasets.
- The algorithm which gives the lowest errors is considered to be accurate.

One way to approach machine learning problems is to begin with a large amount of data, ideally consisting of many examples or observations. It's especially useful if this data is already labeled, meaning that you already know the correct answer for each example in the dataset. This type of data is known as labeled data, and it can be extremely valuable for training machine learning models.

CC BY NC

Table 1: Description of dataset

*	Attributes	UNITS	Remarks
1	Date	dd/mm/yyyy	
2	Time	hh/mm/ss	The time is given in hours, minutes, and seconds (hh/mm/ss), where the hour values range from 0 to 23, and minute values range from 1 to 60
3	Global Active Power (GAP)	Kilowatts	Total average active power of the household for each minute.
4	Global Reactive Power (GRP)	Kilowatts	Total average reactive power of the household for each minute
5	Voltage (V)	Volts	Total average voltage for each minute.
6	Global Intensity (GI	Amperes	Total average current intensity for each minute
7	Sub-metering 1 (S1)	Watt-hours	Active energy related to kitchen, including dishwasher, oven and microwave.
8	Sub-metering 2 (S2	Watt-hours	Active energy related to laundry room, including washing machine, tumble-drier and refrigerator
9	Sub-metering 3 (S3)	Watt-hours	Active energy related to electric water-heater and air-conditioner.

Table 1: Description of dataset

B. DATA PRE-PROCESSING: To prepare your chosen data for analysis, you will need to perform several common pre-processing steps. These steps include formatting, cleaning, and sampling the data.Formatting involves transforming the data into a format that is suitable for analysis. For example, you may need to convert data from a relational database into a flat file or from a proprietary file format into a more accessible format, such as a text file.

Cleaning the data involves identifying and addressing missing or incomplete data instances. You may need to remove data instances that are incomplete or do not contain the necessary information for your analysis. Additionally, you may need to remove sensitive information from the data to protect privacy.

Sampling the data involves selecting a smaller representative sample from the entire dataset. This can be useful for exploring and prototyping solutions before analyzing the entire dataset, which can be computationally expensive and time-consuming. By selecting a smaller sample, you can reduce running times and computational requirements.

•

C. DATA SPLITTING: After preprocessing the data, missing values were removed and the dataset was split into training and testing sets using an 80:20 ratio. Specifically, 80% of the data was used for training, while the remaining 20% was reserved for testing.

D. EVALUATION MODEL: Model Evaluation is a crucial aspect of the model development process, which aids in identifying the best model that accurately represents the data and its performance in the future. It is not recommended to evaluate model performance using the same data used for training since it may result in overfitted and overly optimistic models. There are two methods for evaluating models in data science, namely Hold-Out and Cross-Validation, both of which utilize a test set (not used in training) to evaluate model performance and avoid overfitting.

The performance of each classification model is assessed based on its average, and the results are visually presented through graphs that represent the classified data. Accuracy is defined as the

CC BY NC

percentage of correct predictions for the test data and is calculated by dividing the number of correct predictions by the total number of predictions.

During training, we only need to fit our input data (x_train) and output/label data (y_train). Mini-batch learning with a batch size of 128 and 5 epochs is used for this training. Additionally, a callback named "checkpoint" is included to save the model locally after every epoch if its accuracy has improved compared to the previous epoch. Testing in order to assess the performance of our model, it is necessary to make predictions about the sentiment expressed in the x_test data, and then compare these predictions with the expected output (y_test data). The accuracy of the model can then be calculated by dividing the number of correct predictions by the total amount of data. Our evaluation shows that the model achieved a loss of 0.0092.

Hardware and software specification

a. *Hardware:*

1 GB RAM
80 GB Hard Disk
Intel Processor
LAN

b. *Software :*

Windows OS
Python GUI or Anaconda Navigator

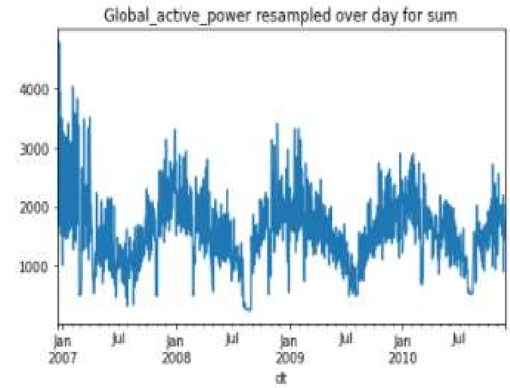
c. *System Requirement:*

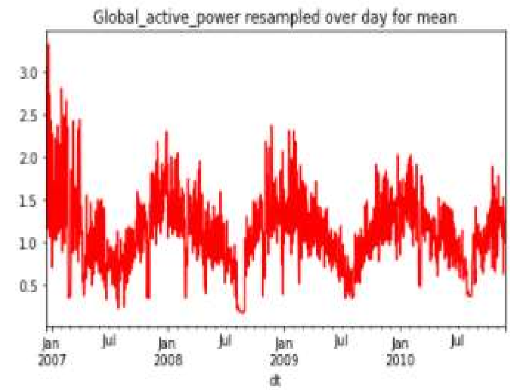
Operating System**:** Windows 7 Ultimate 32 bit / Windows XP

IV. RESULTS

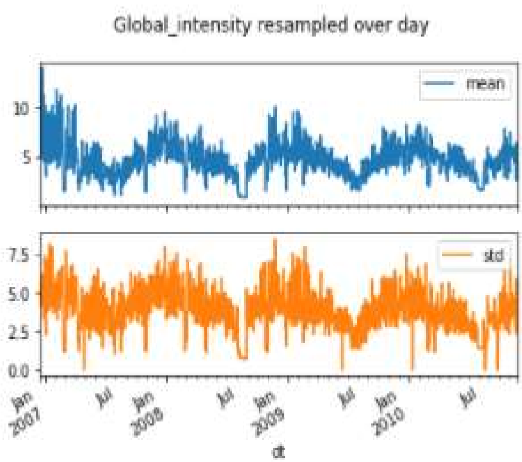
Data mining refers to the process of extracting knowledge from existing data. In the banking and finance industry, it serves as a valuable tool for discovering useful insights from both operational and historical data, which can help facilitate better decision-making. This field is interdisciplinary in nature, combining elements of statistics, database technology, information science, machine learning, and visualization..

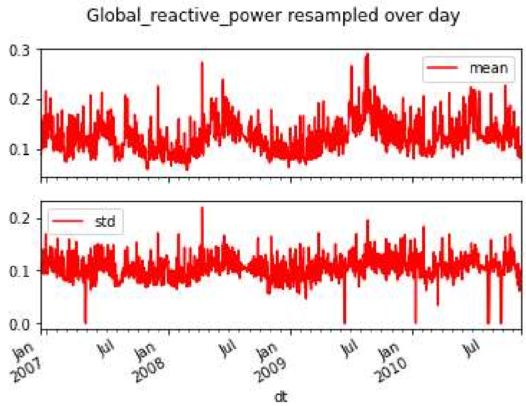
Visualization



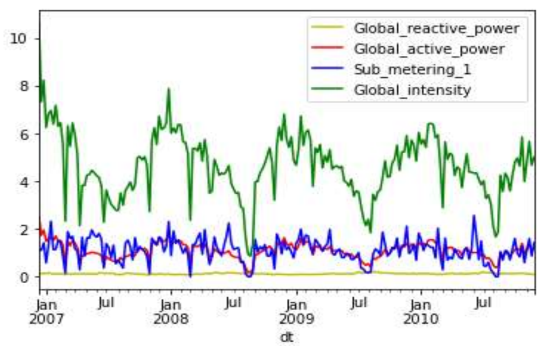


The data mining process typically involves a series of steps, such as data selection, data integration, data transformation, data mining itself, pattern evaluation, and knowledge presentation

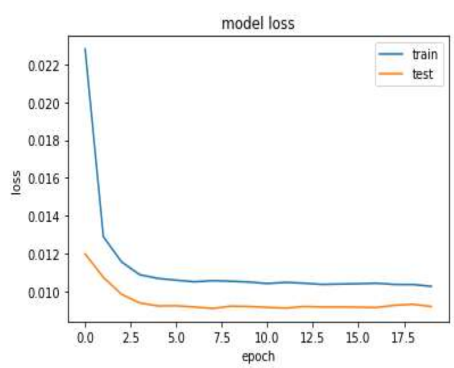




The resampling of the data into decision trees are given accordingly. Each data are classified according to the Data variables like Global _reactive_power, Global_active_power, Sub_meterings, Global_intensity.



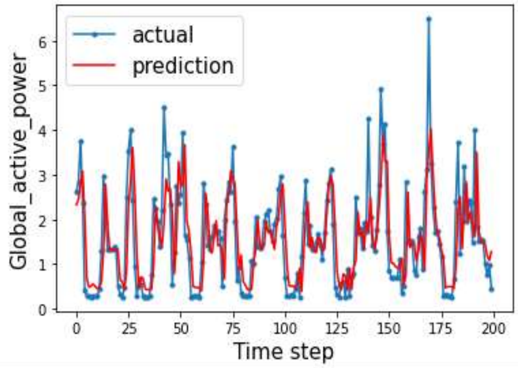
LSTM Result:



Test RMSE: 0.617

A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. Here the LSTM result of the modal is graphed.

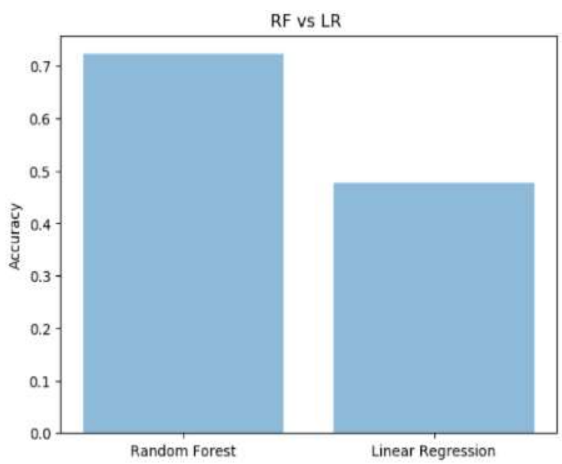
Future Prediction Result :



Prediction Score:

ALGORITHM	OUTPUT (R2 Score)
Prediction Result with Linear Regression	0.475
Prediction Result with Random Forest Regression	0.722

Comparative Result Graph



V. CONCLUSION

In recent years, Machine Learning (ML) techniques have made significant contributions to the development of prediction models for power consumption. These models have greatly improved the accuracy, robustness, and precision, as well as the generalization ability of traditional time series forecasting methods. By utilizing historical data, we can predict future power consumption. In this study, we applied linear regression and random forest regression to the electric power consumption data from a single household. Our results showed an accuracy of 72%. Furthermore, we utilized LSTM for future prediction.

VI. REFERENCES

1. S. Corgnati, T. Bednar, Y. Jang, H. Yoshino, F. M, D. S, D. N, S. M, C. Ghiaus, C. A, T. N, F. V Total Energy Use in Buildings. Analysis and Evaluation Methods, Final Report Annex 53. Statistical Analysis and Prediction Methods (2013)

2. Y. Fu, Z. Li, H. Zhang, P. Xu "Using support vector machine to predict next day electricity load of public buildings with sub-metering devices" Procedia Eng., 121 (2015), pp. 1016-1022

3. O. Valgaev, F. Kupzog, H. Schmeck "Low-voltage power demand forecasting using K-nearest neighbors approach" IEEE Innovative Smart Grid Technologies - Asia (ISGT - Asia) (2016), pp. 1019-1024

4. A. El Khantach, M. Hamlich, N.E. Belbounaguia Short-term load forecasting using machine learning and periodicity decomposition AIMS Energy, 7 (3) (2019), pp. 382-394

5. A. González-Briones, G. Hernández, J.M. Corchado, S. Omatu, M.S. Mohamad Machine learning models for electricity consumption forecasting: a review IEEE Virtual-Ledgers-Tecnologías DLT/Blockchain y Cripto-IOT (2019)

6. 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

7. 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

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

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

10. Q. Li, C. Peng, M. Chen, F. Chen, W. Kang, J. M. Guerrero, et al., "Networked and distributed control method with optimal power dispatch for islanded microgrids," IEEE Transactions on Industrial Electronics, vol. 64, pp. 493-504, 2016.

11. W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu, "Short-term residential load forecasting based on resident behaviour learning," IEEE Transactions on Power Systems, vol. 33, pp.1087-1088, 2017

12. Prince Waqas Khan,Young Cheol Byun, Saang Joon Lee(2020).Machine Machine Learning-Based Approach to Predict Energy Consumption of Renewable and Nonrenewable Power Sources 2020,13,4870.

13. A. Ahmad, M. Hassan, M. Abdullah, H. Rahman, F. Hussein, Abdullah, et al., "A review on applications of ANN and SVM for building electrical energy consumption forecasting, "Renewable and Sustainable Energy Reviews, vol. 33, pp. 102-109, 2014.

14. M. Zhang and C. Bai, "Exploring the influencing factors and decoupling state of residential energy. consumption in Shandong," Journal of cleaner production, vol. 194, pp. 253- 262, 2018 .