Predicting Solar Energy Consumption in a Commercial building using Machine Learning

Divya Rani[1], Harsh Kumar[2], Kunal Singh[3] , Ajay Solanki[4]

[1, 2, 3] Undergraduate students , [4] Assistant Professor
Department of Mechanical Engineering , Delhi Technological University , New Delhi , India

---***---

Abstract - *This research was aimed at investigating different modeling approaches to predict solar power consumption in the building industry in India. LSTM neural networks, and linear regression , two machine learning models were used for evaluating efficiency. Namely, to create models, the analysis of historical records, weather forecasts, and building-specific datasets was performed. The demographic of the region, location, calendar factors, and historical consumption trends were extracted to feature consumption. LSTM networks can well catch complex temporality patterns while having low interpretability, while the linear regression model can assume causation but have reduced accuracy. As shown, the LSTM model was more successful over fitting data with 97.12% accuracy. However, the study shows that both approaches are possible and provide valuable insights into India urban expansion and sustainable energy solutions. The research brings to light the importance of using advanced predictive modeling solutions to address the energy problem and a greener environment in India.*

***Key Words*: Solar Energy , Energy Consumption , Commercial building , Machine learning , Predictive Modeling**

1.INTRODUCTION

India's building industry faces major challenges in regulating energy use and reducing environmental impacts due to rapid urbanization and population growth. The sector is characterized by a high carbon footprint and requires immediate measures to reduce these negative impacts by increasing energy efficiency and use of renewable energy sources. One of the possible ways to tackle these problems is accurate estimation of solar power usage using predictive modeling approaches. This study examines linear regression and long-short-term memory neural network effectives in predictive modeling approaches of solar power consumption in the Indian building sector. It is necessary to examine the extracted key features, such as the demography and geography, and calendar, and the history of the building's energy consumption, which helps to understand the time and environmental patterns of energy use. To do this, the following is used: primary measurement data, historical weather data, forecast weather data, and building-specific datasets .

Thus, this study aims to explain the applicability and productivity of LSTM neural networks and linear regression for prescriptive energy following India's urban development. Research presents considerable conclusions for usage of energy forecasting methods creatively tailored for the building industry in India, enabling policymakers and stakeholders to address energy-related concerns and support ecological sustainability..

2. LITERATURE REVIEW

"Predicting Energy Consumption in Residential Buildings Using Advanced Machine Learning Algorithms" by Dinmohammadi, F., Han, Y., & Shafiee (2023)[1]

The article provides a review of predicting building energy consumption with a focus on machine learning methods. Several machine learning algorithms that are currently widely used in predicting building energy consumption and performance, such as artificial neural networks, support vector machines and Gaussian-based regressions, and clustering, are introduced in this paper. The review underscores the need for accurate energy prediction for efficient decision making on energy consumption and intelligent renovation of buildings. The study recommends a framework to select the best machine learning method for various cases and building usage. Finally, the conclusion of the review argues the current hurdles associated with machine learning on energy prediction and their limitations and possible research avenues for improvement using machine learning in predicting and benchmarking energy.

"Energy consumption prediction by using machine learning for smart building: Case study in Malaysia" by Mel Keytingan , M. Shapi et al. (2021)[2]

This study explored the use of machine learning for the development of a predictive model for energy consumption in smart buildings. The model training, data analysis, and model performance evaluation all took place in the Microsoft Azure Machine Learning Studio. The paper compared models based on three predictive techniques: k-nearest neighbor, support vector machine, and artificial neural network. The results demonstrated the advantage of cloud-based predictive model development and emphasized the importance of data preprocessing in accounting for the heterogeneity of real-life data. The recommendations for further research include utilizing more powerful computational resources, collecting more relevant data, and investigating hybrid and ensemble models to improve predictive accuracy.

"Machine learning for estimation of building energy consumption and performance: a review" by Saleh Seyedzadeh, Farzad Pour Rahimian, Ivan Glesk, and Marc Roper (2018) [3]

The subject of this work is a detailed analysis of implementing machine learning approaches used to efficient analysis of building energy usage. This paper provides an overview of the advantages and constraints of this approach and tracks potential future developments, applications, and other modifications that could further enhance the model's performance. The authors explain why machine learning is suitable for building energy analysis, describe machine learning models, assess how these models are used in the building sector, provide multiple case studies, demonstrate how to choose the right machine learning model, and identify current limitations and ways to improve research. This study will also outline how machine learning can influence building energy forecasts and benchmarking.

"Predicting Daily Mean Solar Power Using Machine Learning Regression Techniques" by Faizan Jawaid and Khurum Nazir Junejo (2016) [4]

The prerequisites of the article related to the importance of accurate sun irradiance forecasting for multiple solar power applications include the prediction of daily mean solar power with regression techniques of machine learning. The results presentation of the study is to establish the comparison of artificial neural networks and azimuth and zenith parameters modification against various regression methods. The publication is based on the extensive analysis of prediction methods and the significance of some meteorological variables on the solar power prediction. The primary objective is the enhancement of solar power mean forecasting for practical application.

3. METHODOLOGY

The methodology of this study includes several stages, including dataset selection, data exploration, data preprocessing, model development , model evaluation, and Results . These stages are illustrated in Figure 1 and described in detail below.

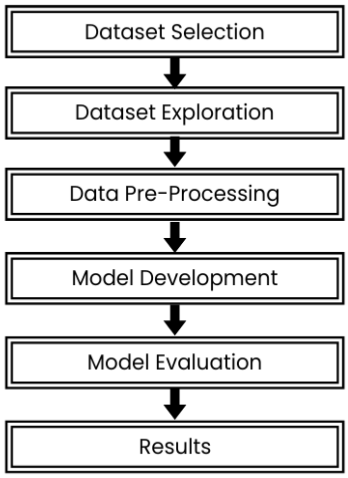

Fig 1. *The Methodology of the Research*

3.1 Dataset Selection : Numerous publicly available databases and online resources provide a wealth of data on solar energy consumption across various buildings with unique characteristics. The dataset from a business building in Tamil Nadu, India, which was collected every day from December 2017 to May 2019, is the focus of this study. This dataset provides a comprehensive insight into the parameters impacting solar energy utilization in commercial establishments. It is enhanced with details such as hourly solar power usage, floor assignments, staff presence, holiday schedules, and daily external temperatures.

3.2 Dataset Exploration : Hundreds of records were collected in the dataset on a daily basis, and an initial exploration was conducted to understand its findings. Analysis of the visualization trends provided insights as follows:

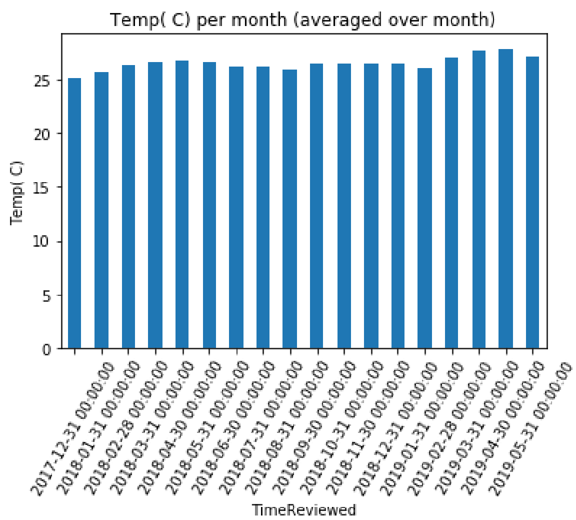

Fig 2 *. Temp (C) per month*
(Averaged over month)

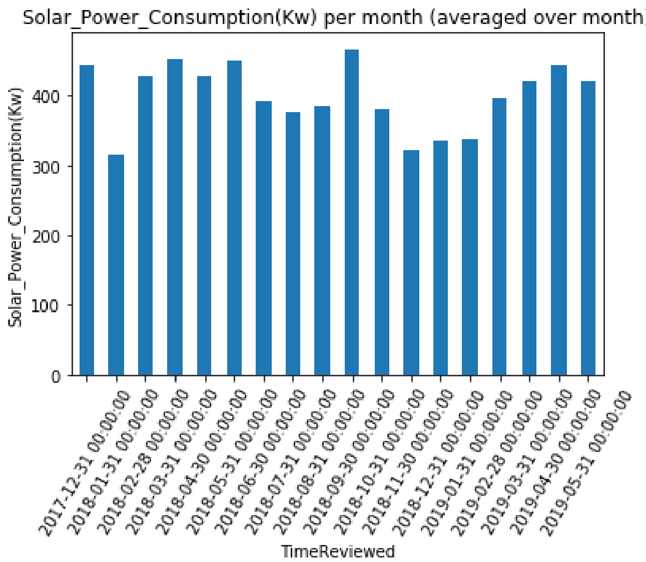

Fig 3. *Solar power Consumption (Kw) per month (averaged over month)*

3.3 Dataset Preprocessing : Extensive preparation activities were conducted to improve the dataset's utility and reliability before it was usable for analysis. Data-cleaning and imputation steps were undertaken to address errors, inconsistencies, and missing data. Additionally, categorical variables were encoded to make them compatible with analytical models; feature selection was used to identify the most relevant variables for this study. In addition, standardization was used to ensure uniformity among the different features. This helped to enhance the dataset, making it more appropriate for use to generate more authentic and valid results about solar power use trends in commercial buildings.

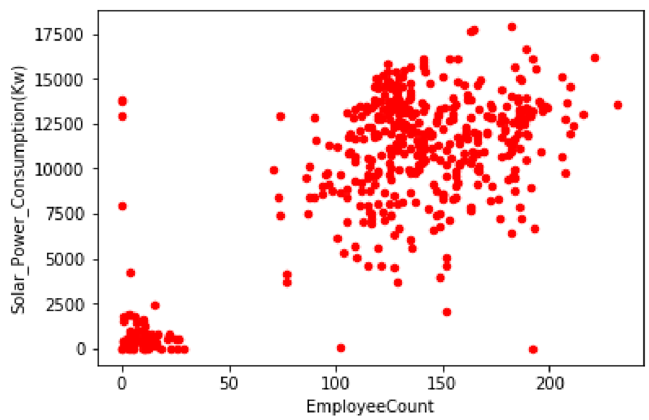

Fig 4 . *Correlation between number of employees and Solar energy*

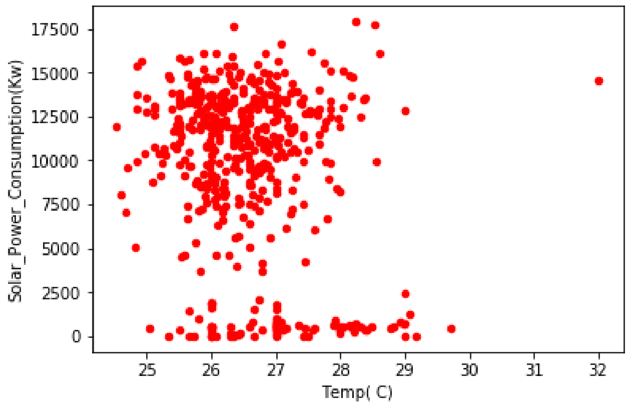

Fig 5. *Correlation between Temperature and Solar energy*

3.4. Model development : A linear regression model was developed earlier for simple insights into solar energy consumption. Then, using LSTM networks positively captured complex temporal patterns and significantly increased forecast precision. This technique was used to gain a deeper understanding of the energy required in commercial buildings, especially solar energy.

3.5. Model Evaluation : The accuracy of the models would be measured by comparing how well the original prediction compares to the different predictions.

4. ML MODELS

4.1 Linear Regression

A linear regression is a statistical technique used to analyze the effect of independent factors on the dependent variable, before which one needs to understand and predict. Temperature, the number of employees, weekdays, and day types, are examples of independent variables relative to solar power consumption. This study utilizes linear regression to investigate trends in energy usage.

The accuracy of the model can be assessed based on how well the original prediction stands up to various predictions. Typically, it is done by dividing a dataset into different parts for testing and training the model . Training enables the model to comprehend the complexities of the interactions between independent and dependent variables. The coefficient of determination indicates how well the model can explain the variance in solar power use. Consequently, linear regression provides a solid basis for making informed energy management and strategy formulation decisions.

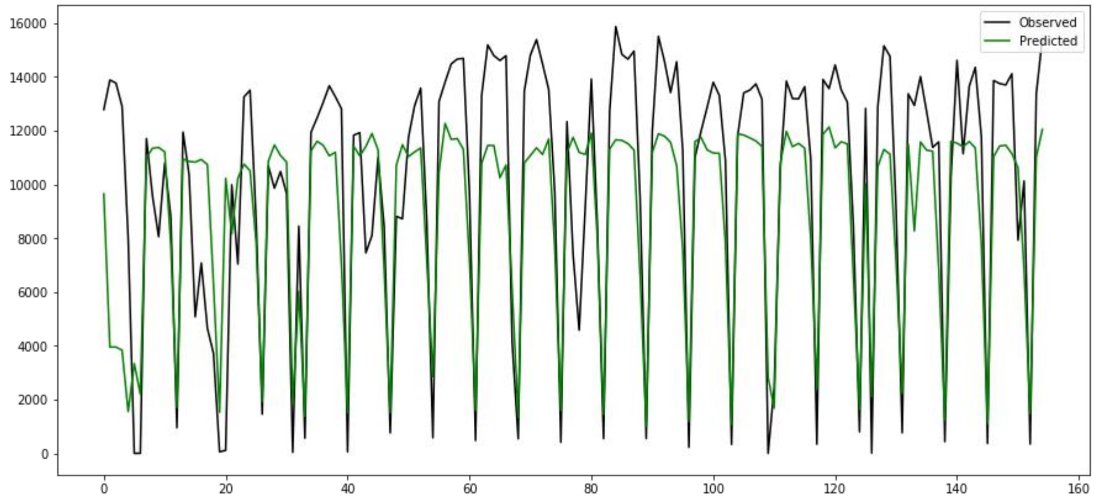

Fig. 6. *Comparing Observed and Predicted Solar Power Consumption using Linear Regression*

The model in this study explains around 63.70% of the variance in solar usage with the independent variables chosen. The R2 score is 0.6370. To enhance forecasting performance, the study aims to incorporate more complex Long Short-Term Memory (LSTM) models , which are especially good at capturing complex temporal patterns.

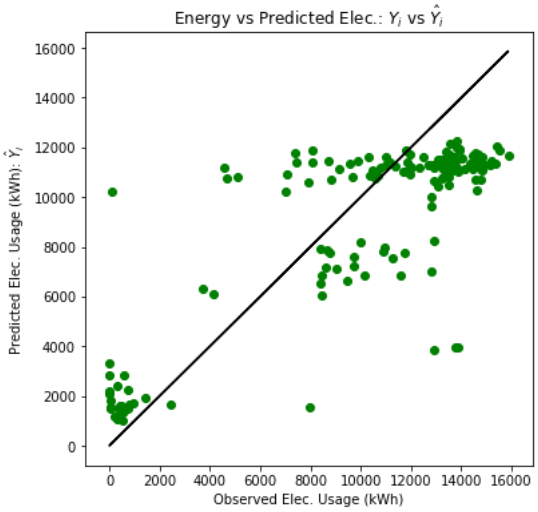

Fig. 7 *Scatter Plot: Comparing Observed vs. Predicted Solar Power Consumption using Linear Regression*

4.2 Long Short Term Memory (LSTM)

Earlier, a particular kind of neural network, long-short-term memory , was utilized for common sense prediction: the future utilization of solar power based on preceding data. Data on day type, manpower count, and temperature was collected and organized using pandas, a data-manipulation tool that simplifies analysis. More thorough data is necessary to understand consumption patterns entirely. Therefore, the capability to differentiate between weekdays, weekends, and public holidays was included . The generator was split into “training” and “test” subsets to provide historical data to the model for training accounting and unobserved data for performance testing . The target variable , solar power consumption, and input data, temperature, and day type were separated. The data was reshaped into a format appropriate for sequence-based modeling to prepare it for LSTM.

An LSTM model was then developed and taught using the training data to recognize various trends and correlations between the input factors and solar usage. This study’s model’s efficiency was measured to be test accuracy of 0.9712. This means that the model can forecast solar use with just a 97.12 percent accuracy. This degree of accuracy is crucial to consider the excellent forecasting accuracy of the LSTM model and to guide energy administration and policy.

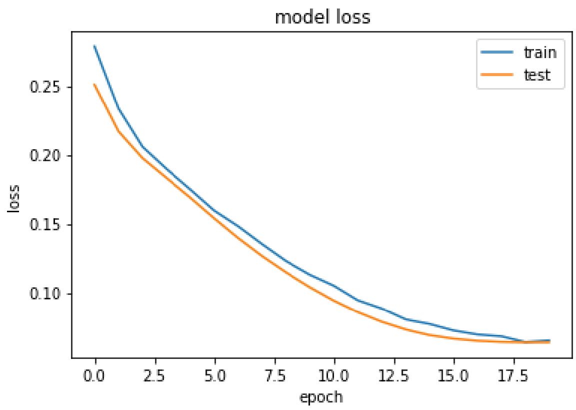

Fig.8 *Visualizing Model Loss Over Epochs*

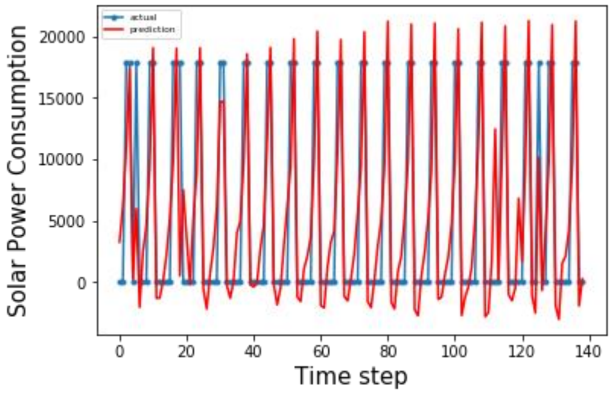

Fig.9 *Comparison of Actual vs. Predicted Solar Power Consumption Over Time using LSTM*

5. RESULTS

The linear regression analysis has provided useful findings regarding the consumption of energy. A majority of the variability in solar power usage, about 63.70%, is explained by the model, as shown by the R2 score of 0.6370. That would indicate a reliable level of predictability, which would enable the professionals to make informed decisions and preparations in energy management.

In particular, when it involves predicting solar power usage, the LSTM model has proven to be very effective at detecting complex sequential patterns. The assessment illustrates the outstanding predictive force of the LSTM model since the test accuracy is 97.72%.

6. CONCLUSION

In conclusion, the present study results affirm the significance of LSTM and linear regression models in evaluating and predicting energy use trends, particularly those related to solar power consumption. Indeed, the interplay of such variables as temperature oscillations, the number of employees, day categorization, and solar power consumption levels has been one of the most clarified by both models. Ultimately, the achievement and the potential of forecast methodologies in guiding energy management solutions are alluring, evidenced by the high R^2 indicator and the acceptable level of precision manifested by the LSTM model.

Therefore, forecasting of solar power consumption in India is crucial in view of the rapidly evolving country's energy supply and demand dynamics and the government's commitment to renewable energy. This research will enable relevant stakeholders to make informed decisions on energy distribution, grid improvements, and future investment in infrastructure. Moreover, utilizing LSTM modeling in combination with linear regression eases the process of proper resource allocation and the need to consume non-renewable finite sources of energy as well as move sustainable developments further. The present report identifies actionable outcomes that are instrumental in driving radical innovation in renewable energy in India.

REFERENCES

[1] Dinmohammadi, F., Han, Y., & Shafiee, M. (2023, April 27) "Predicting Energy Consumption in Residential Buildings Using Advanced Machine Learning Algorithms" doi.org:10.3390/en16093748

[2] Shapi, M. K. M., Ramli, N. A., & Awalin, L. J. (2021, March 1). "Energy consumption prediction by using machine learning for smart building: Case study in Malaysia" doi.org:10.1016/j.dibe.2020.100037

[3] Seyedzadeh, S., Rahimian, F. P., Glesk, I., & Roper, M. (2018, October 2). "Machine learning for estimation of building energy consumption and performance: a review" doi.org:10.1186/s40327-018-0064-7

[4] Jawaid, F., & NazirJunejo, K. (2016, August 1). "Predicting daily mean solar power using machine learning regression techniques" doi.org:10.1109/intech.2016.7845051

[5] Mosavi, A., & Bahmani, A. (2019, March 10). "Energy Consumption Prediction Using Machine Learning; A Review" doi.org:10.20944/preprints201903.0131.v1