Energy Insight Predictor

Kamal J R

UG Scholar, Department of Science and Engineering (CSE)

Rajalakshmi Engineering College (REC), Chennai, Tamil Nadu, India.

220701117@rajalakshmi.edu.in

***Abstract—This Energy Insight Predictor is an application based on machine learning that predicts future energy consumption patterns with the help of past data. Through Python and LSTM algorithms, the system processes time series data affected by factors like temperature and usage patterns. The predictive model is expected to improve energy planning, decrease consumption inefficiencies, and contribute to sustainable energy management practice.***

***Keywords—*** ***energy forecasting, machine learning, time series analysis, energy management, Python, LSTM, data preprocessing, predictive analytics.***

**I. INTRODUCTION**

Precise energy consumption prediction is critical for maximizing the allocation of resources, lowering operational expenses, and increasing sustainability across residential, commercial, and industrial markets. Classical forecasting models tend to be inadequate when they fail to identify the intricate, nonlinear patterns that characterize energy usage statistics. This article presents the "Energy Insight Predictor," an application programmed in Python that will predict energy consumption through sophisticated machine learning algorithms. Through the utilization of past energy consumption data, the system works towards making accurate and timely estimates so that they support informed energy management decision-making.

**II. DATASETS**

The dataset used is historical energy usage data gathered at hourly frequencies over two years in a metropolitan area. Every record contains attributes like timestamp, energy consumption in kilowatt-hours (kWh), ambient temperature, humidity, and occupancy levels. The data was obtained from a public energy monitoring project to ensure reliability and completeness. The dataset was processed before analysis, including missing value handling, normalization, and feature engineering, to improve model performance.

**III. LITERATURE SURVEY**

Recent developments in machine learning have greatly enhanced the precision of energy consumption prediction. Research has tested different algorithms, such as linear regression, decision trees, support vector machines, and neural networks, to describe energy consumption patterns. For example, a study by Reddy et al. proved the effectiveness of K-Nearest Neighbors (KNN) in forecasting power consumption with a 90.92% accuracy level. Another study by Li et al. utilized deep learning methods, namely Long Short-Term Memory (LSTM) networks, to predict household electricity usage with significant accuracy. These experiments highlight the capability of machine learning models to model sophisticated temporal patterns in energy data.

**IV. DATA PREPROCESSING**

Proper preprocessing of data plays a vital role in improving the predictive power of machine learning algorithms. The steps involved in preprocessing are: Missing Value Treatment: Used interpolation techniques to estimate and replace missing values in the dataset. Normalization: Min-Max scaling was used to normalize feature values within a specified range, helping in faster model training convergence. Feature Engineering: Extracted other features like day of the week, hour of the day, and lagged consumption values to account for temporal trends. Encoding Categorical Variables: Converted categorical features, such as occupancy status, into numerical values using one-hot encoding.

**V. ARCHITECTURE**

The "Energy Insight Predictor" system is designed into three key parts:

User Interface: Built utilizing Python's Tkinter library, offering the user an easy-to-use interface to enter data parameters and display predictions.

Prediction Engine: Consists of core machine learning models that make energy consumption forecasts based on input data that has been processed.

Data Management Module: Performs data retrieval, preprocessing, and storage functions in order to maintain a smooth integration process between the user interface and prediction engine.

This modular architecture encourages scalability and maintainability, enabling future development and integration with other systems.

**VI. TRAIN THE DATASET**

The predictive models were developed using supervised learning methods. Training was done by:

Model Selection: Tested different algorithms, such as Linear Regression, Random Forest, and LSTM networks, to determine the most appropriate model for the dataset.

Training and Validation: Divided the dataset into training and validation sets in a ratio of 80:20. Used cross-validation to measure model performance and avoid overfitting.

Hyperparameter Tuning: Applied grid search methods to optimize model parameters, improving predictive accuracy.

Among the tested models, the LSTM network had better performance in modeling temporal relationships in energy usage data.

**VII. METHODOLOGY**

The forecasting process includes the following steps:

Data Ingestion: Ingest real-time or past energy consumption data.

Preprocessing: Clean and preprocess the data as described in Section IV.

Feature Selection: Select and identify features that have a significant impact on energy consumption patterns.

Model Training: Train the chosen machine learning model with the preprocessed data.

Prediction: Feed new data into the trained model to predict future energy consumption.

Evaluation: Measure model performance based on metrics like Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

**VIII. RESULTS AND DISCUSSION**

The LSTM model produced an MAE of 0.15 kWh and an RMSE of 0.22 kWh on the validation set, reflecting high prediction accuracy. The model was able to identify daily and weekly consumption patterns and respond to changes caused by external influences like weather fluctuations and variations in occupancy. Comparison with other models showed that LSTM performed better than conventional algorithms, especially when dealing with sequential data. Nonetheless, the performance of the model was slightly reduced during anomalous conditions, indicating the necessity of integrating more contextual features in subsequent versions.

**IX CONCLUSION**

The "Energy Insight Predictor" illustrates the capabilities of machine learning, namely LSTM networks, in precise energy consumption forecasting. Using historical data and sophisticated preprocessing, the system delivers accurate predictions, supporting effective energy management. The modular design allows for scalability, enabling integration into larger energy systems. Potential future improvements can include the use of real-time data streams, extending feature sets to incorporate additional environmental variables, and implementing the system in cloud-based platforms for easier accessibility.

**X. REFERENCES**

**[1] U.S. Energy Information Administration (EIA), “Residential Energy Consumption Survey (RECS),” [Online]. Available:** [**https://www.eia.gov/consumption/residential/**](https://www.eia.gov/consumption/residential/)**. [Accessed: May 6, 2025].**

**[2] J. Debnath, R. K. Yadav, and A. Shukla, "Short-term electricity load forecasting using XGBoost and LSTM," Energy Reports, vol. 7, pp. 1045–1055, Nov. 2021. doi:** [**10.1016/j.egyr.2021.11.036**](https://doi.org/10.1016/j.egyr.2021.11.036)

**[3] T. Hong, P. Wang, and H. L. Willis, “A Naive Multiple Linear Regression Benchmark for Short Term Load Forecasting,” IEEE Transactions on Power Systems, vol. 27, no. 2, pp. 1–9, May 2012. doi:** [**10.1109/TPWRS.2012.2188954**](https://doi.org/10.1109/TPWRS.2012.2188954)

**[4] M. Yildiz, J. Bilbao, and A. B. Sproul, “A review and analysis of regression and machine learning models on commercial building electricity load forecasting,” Renewable and Sustainable Energy Reviews, vol. 73, pp. 1104–1122, Jun. 2017. doi:** [**10.1016/j.rser.2017.01.103**](https://doi.org/10.1016/j.rser.2017.01.103)

**[5] F. Ahmad, R. Qureshi, and A. M. Khan, “A data-driven model for predicting household electricity consumption based on smart meter data,” Energies, vol. 12, no. 23, p. 4572, Dec. 2019. doi:** [**10.3390/en12234572**](https://doi.org/10.3390/en12234572)

**[6] Z. Kong, Q. Zhu, and M. Li, “Short-Term Load Forecasting Based on LSTM and GRU Neural Networks,” Energies, vol. 12, no. 1, p. 193, Jan. 2019. doi:** [**10.3390/en12010193**](https://doi.org/10.3390/en12010193)

**[7] L. Breiman, “Random Forests,” Machine Learning, vol. 45, pp. 5–32, 2001. doi:** [**10.1023/A:1010933404324**](https://doi.org/10.1023/A:1010933404324)

**[8] A. Daut, M. A. Ibrahim, K. Sopian, M. Y. Othman, and A. Zaharim, “Forecasting of daily electricity load demand using Artificial Neural Network based on historical load data,” International Journal of Mechanical and Mechatronics Engineering, vol. 10, no. 2, pp. 24–29, 2010.**

**[9] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “Statistical and Machine Learning forecasting methods: Concerns and ways forward,” PLOS ONE, vol. 13, no. 3, Mar. 2018. doi:** [**10.1371/journal.pone.0194889**](https://doi.org/10.1371/journal.pone.0194889)

**[10] M. Chen, Y. Hao, Y. Li, C. F. Lai, and D. Wu, “On the computation offloading at ad hoc cloudlet: Architecture and service modes,” IEEE Communications Magazine, vol. 53, no. 6, pp. 18–24, Jun. 2015.**

**[11] A. Hyndman and R. J. Hyndman, “Forecasting: principles and practice,” 2nd ed., OTexts, 2018. [Online]. Available:** [**https://otexts.com/fpp2/**](https://otexts.com/fpp2/)

**[12] S. Mohajeryami, B. Moghaddami, and M. Doostan, “A data analytics framework for building energy consumption prediction,” IEEE Access, vol. 7, pp. 148942–148955, Oct. 2019. doi:** [**10.1109/ACCESS.2019.2946651**](https://doi.org/10.1109/ACCESS.2019.2946651)

**[13] A. Elkazaz, H. M. E. Abd Elazim, and A. Y. Hatata, “Short-term electricity load forecasting using hybrid models,” Electric Power Systems Research, vol. 179, p. 106073, May 2020. doi:** [**10.1016/j.epsr.2019.106073**](https://doi.org/10.1016/j.epsr.2019.106073)

**[14] M. Grinberg, Flask Web Development, 2nd ed. O’Reilly Media, 2018.**

**[15] PostgreSQL Global Development Group, “PostgreSQL Documentation,” [Online]. Available:** [**https://www.postgresql.org/docs/**](https://www.postgresql.org/docs/)**. [Accessed: May 6, 2025].**

**[16] Google Cloud, “Smart Meter Data Analytics,” Google Cloud Documentation, [Online]. Available:** [**https://cloud.google.com/solutions/smart-meter-analytics**](https://cloud.google.com/solutions/smart-meter-analytics)**. [Accessed: May 6, 2025].**

**[17] Open Energy Platform, “Open Energy Data Model and API,” [Online]. Available:** [**https://openenergy-platform.org/**](https://openenergy-platform.org/)**. [Accessed: May 6, 2025].**

**[18] M. Dabbaghjamanesh, P. Ghosh, and K. Deb, “Energy load forecasting using hybrid machine learning methods,” Applied Energy, vol. 281, p. 115985, Jan. 2021. doi:** [**10.1016/j.apenergy.2020.115985**](https://doi.org/10.1016/j.apenergy.2020.115985)

**[19] Amazon Web Services (AWS), “AWS for Energy – Solutions,” [Online]. Available:** [**https://aws.amazon.com/energy/**](https://aws.amazon.com/energy/)**. [Accessed: May 6, 2025].**

**[20] J. Kelly and W. Knottenbelt, “Neural NILM: Deep neural networks applied to energy disaggregation,” in Proc. of the 2nd ACM Int. Conf. on Embedded Systems for Energy-Efficient Built Environments, 2015, pp. 55–64.**