

HOUSE HOLD ELECTRICITY CONSUMPTION PREDICTION USING ML

¹ K.Suresh Babu, ² Gyara Pavan, ³ Devara Swathi, ⁴ Gunti Bhanu Prasad, ⁵ Malga Akshaya

¹Assistant Professor, ^{2,3,4,5}Students

Department of ECE

Siddhartha Institute of Technology & Sciences, Narapally

k.sureshbabu@siddhartha.org.in, 24tq1a0418@siddhartha.co.in, 24tq1a0411@siddhartha.co.in,

24tq1a0417@siddhartha.co.in, 24tq1a0429@siddhartha.co.in

ABSTRACT

Household electricity consumption has grown steadily with the wider use of air conditioners, refrigerators, water heaters, washing machines, and electronic devices. Most households learn how much energy they have used only when the monthly bill arrives, by which time it is too late to change their behaviour for that billing period. Distribution companies also need reliable estimates of future demand at the level of feeders and transformers in order to plan supply and avoid overloading. This paper presents a household electricity consumption prediction system that uses machine learning to forecast energy use over the coming hours, days, and billing period.

The system takes its data from two sources. The first is a public household power consumption dataset recorded at one-minute resolution, containing active power, reactive power, voltage, current, and sub-metering readings for different groups of appliances. The second is a small smart-metering setup built by the project team using an ESP32 microcontroller and a non-invasive current transformer sensor with a voltage sensing module, which measures the load of a real household and sends readings over Wi-Fi to a server. Weather data such as outdoor temperature and humidity and calendar information such as weekends and holidays are added because they strongly affect domestic load.

The raw readings are cleaned to remove missing intervals and measurement errors, aggregated to hourly and daily totals, and converted into features describing recent usage, time of day, day of week, season, and temperature. Several regression models are trained and compared, including Linear Regression, Decision Tree, Random Forest, Gradient Boosting, XGBoost, and a Long Short-Term Memory network. The models are evaluated with a time-ordered split, and performance is measured using RMSE, MAE, MAPE, and the coefficient of determination. The results show that gradient boosting and LSTM models capture daily and weekly patterns well, with lagged consumption and outdoor temperature as the most important inputs.

The predictions are presented on a web dashboard that shows real-time power, daily consumption, the forecast for the next day, and an estimate of the monthly bill based on the applicable tariff slabs. The dashboard also displays how consumption is divided among appliance groups and highlights unusual usage, such as a sudden rise in night-time load, which may indicate a faulty appliance or equipment left on unintentionally.

The system is useful for households that wish to manage their electricity use and budget, and it also illustrates how aggregated forecasts can support demand planning

by distribution companies. Future work includes appliance-level load disaggregation, integration with rooftop solar generation data, and demand-response features that suggest shifting flexible loads to periods of lower tariff.

I. INTRODUCTION

Electricity has become central to daily life, and the pattern of domestic consumption has changed considerably over the past two decades. Households now operate many appliances with large and variable loads, such as air conditioners, geysers, induction cooktops, and electric vehicle chargers. Consumption changes from hour to hour depending on the activities of the family, from day to day depending on the weather, and across the year depending on the season. For most consumers, however, this changing pattern remains invisible, because the only information they receive is a single number on a monthly bill.

Lack of visibility into consumption has practical consequences. A household cannot easily tell which activities contribute most to its bill or whether a sudden increase is due to weather, a change in routine, or a faulty appliance. Telescopic tariff structures, in which the rate per unit increases once consumption crosses certain slabs, mean that a small increase in usage near a slab boundary can raise the bill noticeably. On the supply side, distribution companies must forecast demand to schedule power purchase and to plan the capacity of transformers and feeders. Errors in these forecasts lead to either costly short-term purchases or supply interruptions during peak periods.

Machine learning offers an effective way to forecast electricity consumption because domestic load follows repeating patterns that are strongly linked to time, weather, and recent history. Regression models can learn these relationships from historical meter readings and produce short-term forecasts with useful accuracy. From an electronics perspective, reliable forecasting also depends on accurate measurement. Current transformer sensors, voltage sensing circuits, and analog-to-digital conversion must be designed and calibrated carefully so that the calculated power reflects the true load, and the sampling and communication scheme must deliver readings consistently to the server.

This paper describes a household electricity consumption prediction system that combines a public high-resolution dataset with readings from a low-cost smart meter built on an ESP32. The system cleans and aggregates the data, creates time, weather, and lag features, trains and compares several regression and deep learning models, and presents forecasts and bill estimates through a dashboard. The aim is to give households a clear picture of their current and expected consumption so that they can control their costs, and to show how such forecasts can support better planning of electricity supply.

II. LITERATURE SURVEY

Forecasting electricity demand is a well-established field, but most early work concentrated on aggregate load at the level of a utility, city, or region rather than individual households. Classical approaches used multiple linear regression with

temperature and calendar variables, as well as time-series models such as ARIMA and exponential smoothing, and these methods performed well for aggregate load because the combined demand of many consumers is smooth and highly regular. Studies that applied the same methods to single households found much larger errors, since individual consumption is strongly affected by the unpredictable behaviour of a few people. A kettle switched on or an air conditioner started at an unusual time produces sharp spikes that cannot be anticipated from past averages. Researchers therefore concluded that household forecasting requires methods that can capture nonlinear relationships and that performance should be judged with an understanding of the inherent randomness at this level.

The spread of smart meters made detailed household data available and led to extensive research on machine learning methods for residential load forecasting. Support vector regression, decision trees, random forests, and gradient boosting machines were trained on hourly or half-hourly consumption together with temperature, humidity, time of day, day of week, and holiday indicators. Comparative studies generally reported that ensemble methods outperformed linear regression and single trees, and that lagged consumption values from the previous hour, the same hour on the previous day, and the same hour in the previous week were among the strongest predictors. Some studies clustered households by their typical daily profile before training separate models for each group and found that this improved accuracy. Others noted that feature engineering based on domain knowledge, such as heating and cooling degree hours, often contributed more to accuracy than the choice of algorithm.

Deep learning approaches have received considerable attention in recent years. Recurrent neural networks, especially Long Short-Term Memory networks and gated recurrent units, were applied to sequences of past consumption to predict future load, and several studies reported lower errors than classical and ensemble methods for short horizons. Convolutional layers were combined with recurrent layers to extract local patterns from load sequences, and attention mechanisms were used to weight the most relevant past time steps. However, researchers also observed that deep models require large amounts of training data and careful tuning, can overfit on data from a single household, and are harder to interpret than tree-based models. Several authors recommended that deep models be compared against strong and well-tuned baselines, since reported improvements were sometimes the result of weak baselines rather than genuine advantages of the more complex architecture.

The measurement side of household energy monitoring has also been studied extensively. Low-cost energy monitors based on current transformer sensors, voltage transformers or resistive dividers, and microcontrollers have been developed to measure real power, apparent power, power factor, and energy. Studies described the importance of sampling voltage and current simultaneously at a rate high enough to capture the waveform, calibrating the sensors against a reference meter, and correcting phase error introduced by the transformers. Wireless transmission of readings using Wi-Fi, ZigBee, or Bluetooth was compared in terms of reliability and power use. Related research on non-intrusive load monitoring showed that appliance-level consumption can be estimated from the aggregate signal by recognising the

characteristic power signatures of individual devices, providing more detailed feedback to consumers without installing a sensor on every appliance.

Work on consumer feedback has shown that information about consumption can lead to meaningful savings when it is timely, specific, and easy to understand. Studies of in-home displays and web portals reported reductions in consumption when households could see real-time usage, comparisons with previous periods, and projections of their bill. Feedback that identified the largest contributors to consumption or highlighted unusual usage was found to be more effective than simple totals. Researchers also emphasised that forecasts presented to consumers should be expressed in familiar terms such as units and rupees, and linked to the tariff structure they actually pay. The present system combines these strands of work by integrating a calibrated low-cost meter, a public high-resolution dataset, comparison of machine learning forecasting models, and a dashboard that translates forecasts into bill estimates and usage alerts.

III. SYSTEM ANALYSIS

The system is analysed as a pipeline that begins with electrical measurement and ends with forecasts and bill estimates for the user. The smart meter node uses a current transformer clamped around the phase conductor and a voltage sensing module connected through an isolation transformer, both sampled by the ESP32 analog-to-digital converter at a rate sufficient to capture the mains waveform. Real power, RMS voltage, RMS current, and power factor are calculated on the node and averaged over one-minute intervals before being sent to the server over Wi-Fi. The server also loads the public household dataset, which provides a long history for model training. Weather data for the location and a calendar of weekends and holidays are merged with the readings. The preprocessing module removes invalid readings, fills short gaps, and aggregates the data to hourly and daily totals. Features describing recent consumption, time, weather, and calendar are created, and several models are trained and compared. The selected model produces forecasts that the dashboard converts into bill estimates according to the tariff slabs.

Existing System

At present, most households receive information about their electricity consumption only through the monthly or bi-monthly bill issued by the distribution company. The bill states the total units consumed and the amount payable but gives no breakdown by day, time of day, or appliance. Conventional electromechanical and basic electronic meters display a cumulative reading that consumers rarely check, and even when they do, converting the reading into a daily pattern or a bill projection requires manual calculation. Some smart meters and utility portals provide daily or hourly consumption, but they generally show past usage without forecasts, and the data is often delayed by a day or more. Commercial energy monitors are available but are relatively costly and rarely offer predictive features. Distribution companies forecast demand at an aggregate level using historical load curves and weather, but these forecasts are not shared with individual consumers. As a result, households cannot

anticipate high bills, identify wasteful usage, or plan their consumption around tariff slabs.

Disadvantages of Existing System

- Consumption information is available only once per billing period.
- Bills show total units without any breakdown by time or appliance.
- Cumulative meter readings are difficult for consumers to interpret.
- Utility portals, where available, show past usage without forecasts.
- Households cannot anticipate crossing a higher tariff slab.
- Faulty appliances and unintended loads go unnoticed for long periods.
- Commercial energy monitors are costly and lack predictive features.
- Weather effects on consumption are not made visible to the user.
- Aggregate utility forecasts are not shared with individual consumers.
- No historical record is maintained for analysis by the household.

Proposed System

The proposed system forecasts household electricity consumption using machine learning and presents the results through a simple dashboard. A low-cost smart meter built with an ESP32, a non-invasive current transformer, and a voltage sensing module measures real power, voltage, current, and power factor, averages them over one-minute intervals, and transmits them over Wi-Fi to a Python server. A public household consumption dataset with one-minute resolution provides a long training history. Outdoor temperature, humidity, and a calendar of holidays and weekends are added as external inputs. The data is cleaned, aggregated to hourly and daily levels, and converted into lag, rolling average, time, calendar, and weather features. Linear Regression, Decision Tree, Random Forest, Gradient Boosting, XGBoost, and LSTM models are trained and compared using RMSE, MAE, MAPE, and R-squared. The selected model forecasts consumption for the next hours, the next day, and the rest of the billing period, and the dashboard converts these forecasts into bill estimates and flags unusual usage.

Advantages of Proposed System

- Provides forecasts of consumption for the next hours, day, and billing period.
- Estimates the monthly bill using the actual tariff slab structure.
- Uses an inexpensive, calibrated ESP32-based smart meter for live data.
- Includes weather and calendar effects that strongly influence domestic load.
- Compares classical, ensemble, and deep learning models to select the best.
- Shows real-time power, voltage, and power factor on the dashboard.
- Highlights unusual consumption that may indicate faulty or forgotten appliances.
- Displays the share of consumption from different appliance groups.
- Helps households plan usage to avoid crossing higher tariff slabs.
- Maintains a consumption history for long-term analysis and retraining.

IV. METHODOLOGY

The methodology begins with the design and calibration of the smart meter. The current transformer output is converted to a voltage with a burden resistor and biased to the middle of the ADC range, and the mains voltage is stepped down and similarly biased. Voltage and current are sampled together, real power is computed as the average of instantaneous products, and calibration constants are set by comparison with a reference meter under known resistive loads. Readings are collected for several weeks and combined with the public dataset. The data is cleaned by removing zero and out-of-range values, interpolating short gaps, and excluding days with long outages. Hourly and daily totals are computed, and features are created, including consumption lags of one hour, twenty-four hours, and one week, rolling means, hour, weekday, month, holiday flag, temperature, and cooling degree hours. The data is split in time order, models are trained and tuned with time-series cross-validation, and RMSE, MAE, MAPE, and R-squared are calculated on the held-out period. The best model is saved and connected to the dashboard.

SYSTEM ARCHITECTURE

The architecture of the system is organised into layers that move data from the electrical supply to the user. The Metering Layer consists of the current transformer, voltage sensing module, and ESP32 that sample the waveforms and compute power and energy. The Communication Layer sends one-minute readings to the server over Wi-Fi using lightweight messaging. The Data Storage Layer holds live meter readings, the public training dataset, weather records, and the holiday calendar. The Preprocessing Layer removes invalid readings, fills gaps, and aggregates data to hourly and daily levels. The Feature Engineering Layer creates lag, rolling, time, calendar, and weather features. The Model Training Layer fits regression, ensemble, and LSTM models, and the Evaluation Layer compares their errors and selects the best one. The Forecast and Billing Layer produces short-term and billing-period forecasts and converts them into cost using tariff slabs, and the Dashboard Layer displays live power, consumption history, forecasts, bill estimates, and usage alerts.

House hold electricity consumption prediction using ML

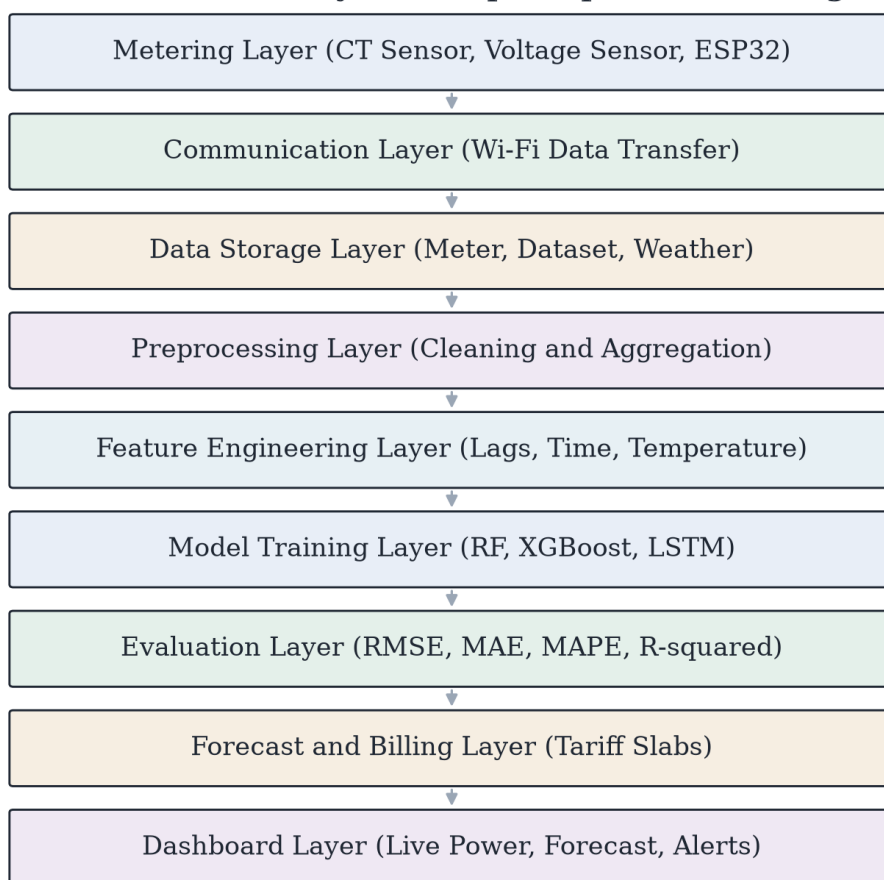
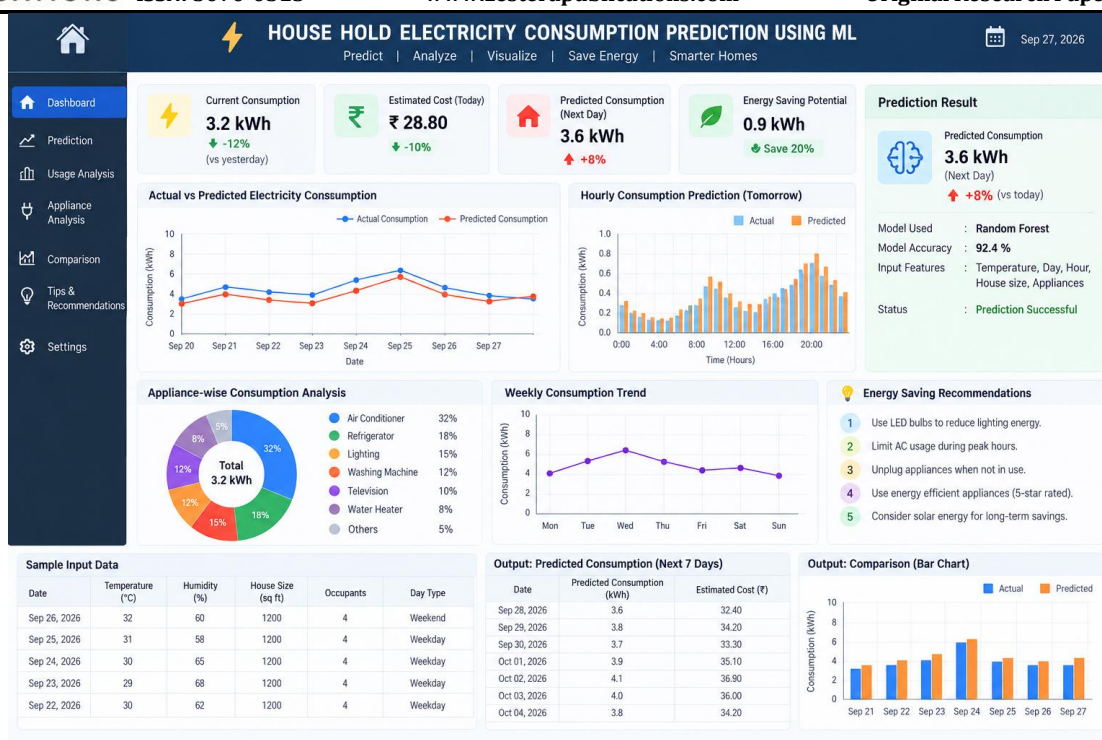


Fig. 1 System Architecture

V. RESULT AND OUTPUT



VI. CONCLUSION

This paper has presented a household electricity consumption prediction system that combines a low-cost smart meter, a public high-resolution dataset, and machine learning forecasting models. The system gives households a continuous view of their consumption and an estimate of how much energy they are likely to use in the coming hours, days, and billing period, replacing the single monthly figure that most consumers currently rely on.

The comparison of models showed that ensemble methods such as Gradient Boosting and XGBoost, as well as the LSTM network, captured daily and weekly consumption patterns more accurately than linear regression and single decision trees. Lagged consumption values and outdoor temperature were the most influential inputs, confirming that recent behaviour and weather together explain much of the variation in domestic load. The use of a time-ordered evaluation ensured that the reported accuracy reflects genuine forecasting ability.

The measurement hardware played an important role in the usefulness of the system. Careful biasing, simultaneous sampling of voltage and current, and calibration against a reference meter allowed the ESP32-based meter to report real power and power factor with reasonable accuracy at very low cost. Translating forecasts into bill estimates based on tariff slabs, and highlighting unusual consumption such as increased night-time load, makes the output directly meaningful to the household.

Future work can extend the system with non-intrusive load monitoring to estimate appliance-level consumption from the aggregate signal, integration of rooftop solar generation and net metering, and suggestions for shifting flexible loads such as

washing machines and water heaters to periods of lower demand. Aggregating anonymised forecasts from many households could also help distribution companies plan transformer loading and improve demand forecasting at the feeder level.

REFERENCES

- [1] Kumar, R. D., Prudhviraaj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In *Handbook of Artificial Intelligence and Wearables* (pp. 145-158). CRC Press.
- [2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In *The International Conference on Artificial Intelligence and Smart Environment* (pp. 557-564). Cham: Springer Nature Switzerland.
- [3] Sv satyakrishna, shirisha rang, bhargavi nalacheruve. (2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm
- [4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.
- [5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in *Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT)*, Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.
- [6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in *Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment*, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.
- [7] R. D. Kumar, V. N. S. Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in *Blockchain for Smart Cities*, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.
- [8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.
- [9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," *Journal of Machine and Computing*, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.
- [10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in *Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE)*, Apr. 2025.
- [11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," *SN Computer Science*, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.