International Journal of Science and Technology (IJST)
ISSN: 3049-1118, Volume- 3, Issue- 2 (April - June 2026)

AI-Enabled Smart Energy Optimization System for Consumption Forecasting and Cost Reduction

Shatakshi Singh[1*], **Satwick Pandey**[1], **Smriti Jaiswal**[1]
[1]Department of Data Science, Noida Institute of Engineering & Technology
Greater Noida, Uttar Pradesh, India

Abstract

Managing household electricity consumption in India is challenging due to complex slab-based tariff structures and limited consumer awareness of appliance-level energy usage. This paper presents a Next Generation Smart Energy Optimizer, an Artificial Intelligence (AI)-driven web-based system designed to predict electricity consumption and optimize household energy costs. The proposed system integrates machine learning, a slab-aware billing engine, and large language model (LLM)-based recommendations to provide comprehensive decision support. A Random Forest regression model, trained on real-world household data, predicts appliance-wise monthly energy consumption with high accuracy. The billing engine simulates real Distribution Company (DISCOM) tariff structures, generating detailed cost breakdowns including energy charges, fixed charges, and applicable duties. Additionally, an LLM-based advisory module produces personalized and context-aware recommendations to help users reduce consumption and expenses. The system is implemented using a Python–Flask backend, PostgreSQL database, and a browser-based frontend, ensuring scalability and accessibility. Experimental evaluation shows strong performance, achieving a mean absolute percentage error (MAPE) of 6.8% compared to actual electricity bills. The proposed framework bridges the gap between consumption prediction and actionable insights, offering a practical solution for intelligent household energy management.

Keywords: Artificial Intelligence, Smart Energy Management, Predictive Analytics, Demand-Side Management, Smart Grid

Introduction

The increasing demand for electricity and rising energy costs have made efficient energy management a critical concern for households (Hasanuzzaman et al, 2017). In India, residential electricity billing follows a tiered slab-based tariff system, where the cost per unit increases with higher consumption levels (Central Electricity Authority, 2023). This pricing structure makes it difficult for consumers to predict future bills and manage their usage effectively. Existing utility platforms primarily provide historical consumption data but lack predictive capabilities and actionable recommendations for cost optimization (Bhattarai et al, 2019;

[*]Corresponding Author Email: shatakshis@gmail.com
Published: 29 April 2026
DOI: https://doi.org/10.70558/IJST.2026.v3.i2.241243

International Energy Agency, 2021). As a result, many households are unaware of how individual appliances contribute to overall consumption.

In addition, the complexity of tiered tariff structures makes it even more challenging for consumers to understand how incremental usage affects their total bill. Small increases in consumption can push users into higher pricing slabs, significantly increasing overall costs. However, traditional systems fail to provide clear explanations or predictive warnings regarding such transitions. This highlights the need for intelligent systems that not only analyze past data but also provide forward-looking insights and user-specific recommendations (Biswas et al, 2024).

Recent advancements in artificial intelligence (AI) and machine learning (ML) have enabled accurate prediction of energy consumption patterns and improved decision-making in energy systems (Zhang et al., 2020). Furthermore, large language models (LLMs) have shown potential in generating human-like, context-aware recommendations (Stecuła et al, 2023; Biswas et al, 2024).

This paper proposes an AI-driven smart energy optimization system that combines machine learning-based consumption prediction, slab-aware billing simulation, and LLM-based personalized recommendations. The main contributions of this work include: (i) an appliance-level energy prediction model; (ii) a realistic billing engine aligned with Indian tariffs; and (iii) an AI-powered recommendation system for energy optimization.

Literature Review

Recent research has emphasized the importance of data-driven approaches for improving energy efficiency in residential systems (Stecuła et al, 2023; Biswas et al, 2024). Smart meter data analytics has enabled fine-grained monitoring of electricity usage, allowing more accurate demand-side management strategies (Mishra and Vaduganathan, 2025). These approaches help identify consumption patterns at the appliance level, which is critical for personalized energy optimization. Time-series forecasting models have also been widely used in electricity consumption prediction. Techniques such as ARIMA and SARIMA models have traditionally been applied for load forecasting; however, they often struggle with non-linear and dynamic consumption patterns (Momani et al, 2025). Machine learning models, including Random Forests and Gradient Boosting, have demonstrated improved performance by capturing complex relationships in the data (Kalusivalingam et al, 2022).

With the growth of deep learning, models such as Long Short-Term Memory (LSTM) networks have been successfully applied to energy forecasting tasks due to their ability to model temporal dependencies (Alizadegan et al, 2025). These models are particularly useful in handling sequential data such as hourly or daily electricity usage. However, their computational complexity can limit real-time deployment in lightweight applications.

In recent years, explainable and user-centric AI systems have gained attention in energy applications (Stecuła et al, 2023; Biswas et al, 2024; Momani et al, 2025). Interpretability is a key factor in increasing user trust in AI-driven systems. Providing clear explanations of predictions and recommendations helps users make informed decisions regarding their energy consumption. Furthermore, the integration of AI with natural language processing has enabled the development of intelligent advisory systems. Large language models have been used to generate personalized, context-aware recommendations, bridging the gap between complex analytics and user understanding, which makes them highly suitable for applications where user engagement and behavioural change are essential (Chen et al, 2024).

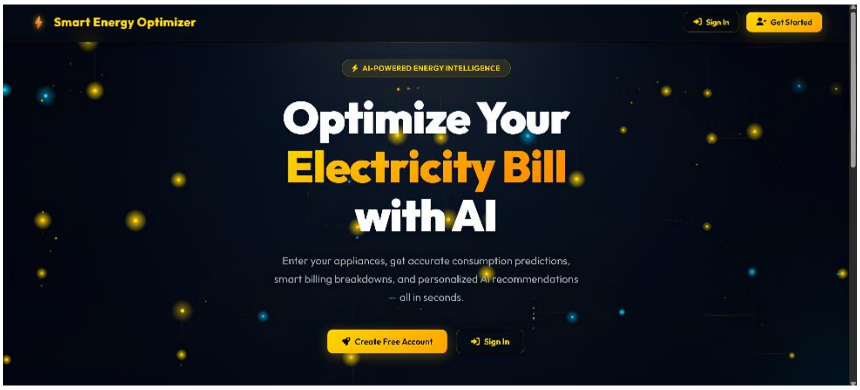
Methodology

A. System Architecture and Design

The Smart Energy Optimizer employs a three-tier architecture comprising a browser-based frontend, a Python Flask backend API, and a cloud-hosted PostgreSQL database. This separation of concerns ensures security, scalability, and maintainability while enabling stateless API design that supports horizontal scaling.

i. Frontend Layer

The presentation layer is implemented using HTML5, CSS3, and vanilla JavaScript, eliminating framework dependencies and reducing load times. This layer handles user input collection, result rendering, and PDF report generation. The lightweight design ensures accessibility across devices with varying computational capabilities, critical for reaching households with limited internet infrastructure (International Energy Agency, 2021). Figure 1 illustrates the landing page interface of the Smart Energy Optimizer, and Figure 2 presents the user input interface for entering household and appliance details.



Enter your appliances, get accurate consumption predictions, smart billing breakdowns, and personalized AI recommendations — all in seconds.

Fig. 1 Smart Energy Optimizer – Landing Page Interface

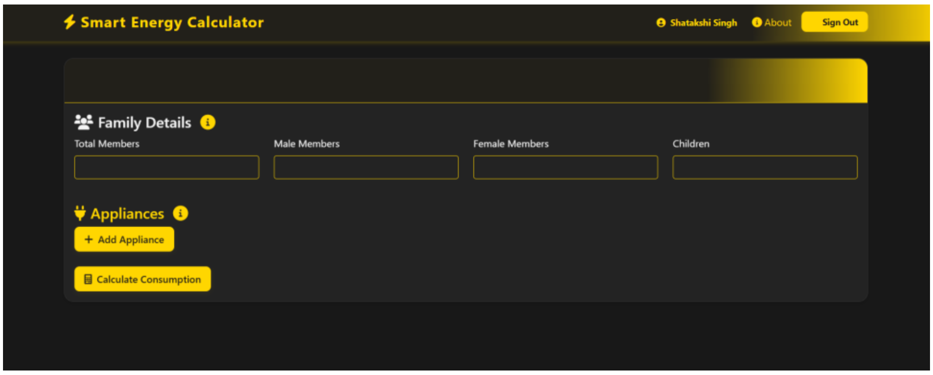


Fig. 2 User Input Interface for Household and Appliance Details

ii. Backend API Layer

The core application logic resides in a Python Flask API that has three primary functions: machine learning inference, slab-aware bill calculation, and large language model (LLM) integration. The backend maintains strict isolation of sensitive credentials, with API keys for external services (Groq LLaMA) and database passwords stored exclusively in server-side environment variables, never exposed to client-side code.

iii. Data **Persistence** *Layer*

PostgreSQL serves as the relational database management system, storing user accounts, session history, and analysis records. The cloud-hosted database enables cross-device session retrieval and provides a foundation for future analytics capabilities. The stateless API design ensures that all application state persists in the database, allowing seamless failover and load distribution.

iv. External AI Service Integration

The system integrates with Groq's inference API running the LLaMA 3 model (8 billion parameter variant) via REST endpoints. This architecture choice decouples the computationally intensive LLM inference from the application infrastructure, enabling sub-second response times without requiring dedicated GPU resources. The streaming response mechanism reduces perceived latency by delivering recommendations incrementally to the frontend (Brown et al., 2020; Touvron et al., 2023).

v. Security Architecture

The security model implements multiple defense layers. User registration requires email-based OTP verification to prevent fraudulent account creation. Password storage employs bcrypt hashing with a cost factor of 12, rendering credential databases unusable in breach scenarios. Session management utilizes HttpOnly cookies to mitigate Cross-Site Scripting (XSS) attacks. HTTPS enforcement with Transport Layer Security (TLS) termination at the infrastructure level ensures encrypted data transmission.

B. Implementation

i. Technology Stack

The Smart Energy Optimizer is developed using a lightweight and efficient full-stack architecture tailored for web-based deployment. The frontend is built using HTML5, CSS3, and vanilla JavaScript, ensuring minimal dependencies, faster load times, and compatibility across a wide range of devices. This approach improves accessibility, particularly for users with limited computational resources and constrained internet connectivity (International Energy Agency, 2021).

The backend is implemented using Python with the Flask framework, which handles core functionalities including machine learning inference, slab-based billing calculations, and API communication with external services. Flask provides a flexible and scalable environment for developing RESTful APIs and integrating data-driven components.

For data storage, PostgreSQL is used as the cloud-hosted relational database, managing user accounts, session history, and energy analysis records. Cloud-based database systems enable reliable data persistence and scalable access across multiple devices (Shermy and Saranya, 2025). The machine learning component is developed using the Scikit-learn library, with a trained Random Forest model used for predicting appliance-level energy consumption. Ensemble learning techniques such as Random Forest are widely recognized for their robustness and ability to handle non-linear relationships in real-world datasets (Kalusivalingam et al, 2022).

The system integrates with external AI services using Groq's inference API running the LLaMA 3 model, which is responsible for generating personalized energy-saving recommendations. The use of large language models enables the generation of context-aware and human-like suggestions, enhancing user interaction (Brown et al., 2020; Touvron et al., 2023).

Version control is managed using Git and GitHub, facilitating collaborative development and efficient code management. The application is deployed on Render cloud infrastructure, which provides HTTPS support, scalability, and cost-effective hosting solutions suitable for real-world deployment (International Energy Agency, 2021).

Table 1 System Technology Stack

Layer	Technology	Responsibility
Frontend	HTML5 / CSS3 / Vanilla JS	User input, result rendering, PDF download
Backend API	Python · Flask	ML inference, bill calculation, LLM orchestration
Database	PostgreSQL (cloud)	User accounts, session history, analysis records
ML Model	Scikit-learn (trained)	Appliance consumption prediction
LLM Service	Groq · LLaMA 3	Natural-language energy recommendations
Auth	OTP · bcrypt · sessions	Identity verification and password security

ii. Feature Engineering Pipeline

The machine learning pipeline processes five input features per appliance: device type (one-hot encoded across 10+ categories), rated wattage (W), average daily usage hours (H), number of units (N), and seasonal usage multiplier. The theoretical monthly energy consumption follows the formula

$$\mathbf{E = W \times H \times D \times N / 1000}$$

where D represents days per month (30). However, real-world consumption patterns exhibit non-linearities due to standby power draw, variable usage patterns, and efficiency degradation, necessitating a data-driven modeling approach (Hastie et al., 2009). Figure 3 presents the complete workflow of the energy optimization system.

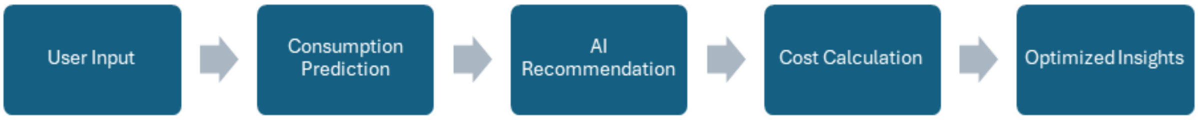


Fig. 3: Workflow of Energy Optimization System

iii. Model Selection and Training

We applied four techniques to forecast daily energy consumption for each appliance in approximately 8,000 records of detailed Indian household consumption data. Linear regression was not accurate enough, achieving a Root Mean Square Error (RMSE) of 18.4 kWh with sub-millisecond inference time. Gradient boosting achieved an RMSE of 9.7 kWh with an 8ms inference time. A shallow neural network was only slightly worse at 10.1 kWh RMSE with a 45ms inference time. Our best method, a Random Forest ensemble, achieves an RMSE of 8.2 kWh, Mean Absolute Error (MAE) of 6.1 kWh, and R^2 of 0.96 with 12ms of inference time per appliance. This method leverages predictive accuracy with production efficiency making it a good fit for use in real-time web applications (Breiman, 2001; Géron, 2022). Figure 4 illustrates the complete machine learning pipeline for energy prediction.

Table 2 Model Selection

Model	RMSE (kWh)	MAE (kWh)	R^2	Inference (ms)
Linear Regression	18.4	14.1	0.81	< 1
Gradient Boosting	9.7	7.3	0.94	8
Random Forest	8.2	6.1	0.96	12
Shallow Neural Net	10.1	7.8	0.93	45

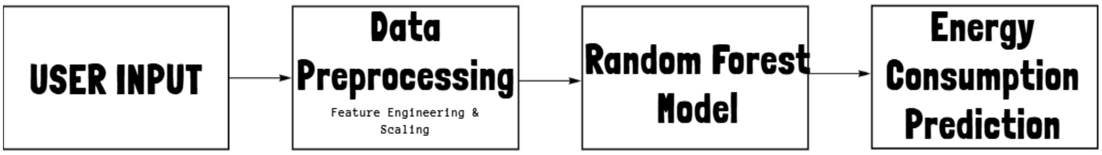
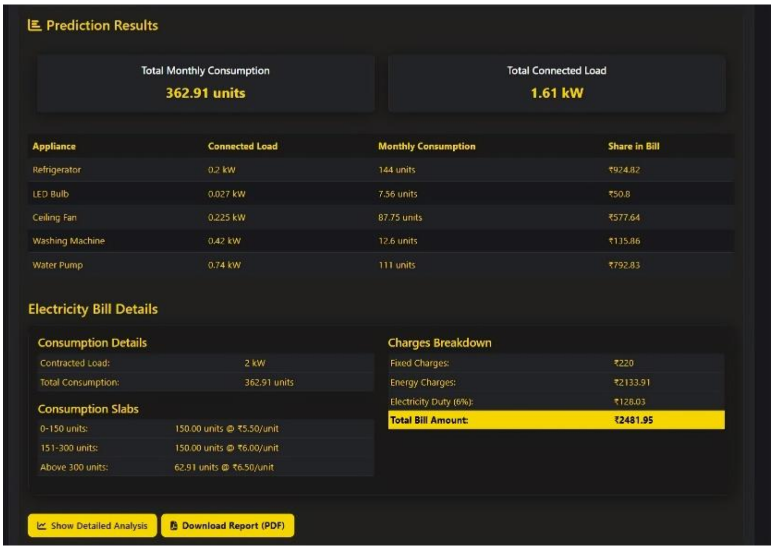


Fig. 4: Machine Learning Pipeline for Energy Prediction

iv. Slab-Wise Billing Engine

State electricity distribution companies in India provide on a gradated tariff where the marginal rate keeps increasing as the consumption increases. The billing engine at PowerBill can support any slab structure that corresponds to the tariffs that individual distribution companies follow.

In addition to the slab rates, the tariffs in India also have four key components which are reflected in the bill i.e. energy charge, fixed charge / demand charge, electricity duty and fuel adjustment charge etc. Using this engine, detailed bill breakdown can be generated which closely resembles the format of the actual bill and indicates to the consumer the impact of crossing a slab barrier (Central Electricity Regulatory Commission, 2016). Figure 5 presents the electricity consumption and bill prediction results, while Figure 6 provides a graphical visualization of energy and load distribution across appliances.



Prediction Results

Total Monthly Consumption
362.91 units

Total Connected Load
1.61 kW

Appliance	Connected Load	Monthly Consumption	Share in Bill
Refrigerator	0.2 kW	144 units	₹924.82
LED Bulb	0.027 kW	7.56 units	₹50.8
Ceiling Fan	0.225 kW	87.75 units	₹577.64
Washing Machine	0.42 kW	12.6 units	₹135.86
Water Pump	0.74 kW	111 units	₹792.83

Electricity Bill Details

Consumption Details

Contracted Load:	2 kW
Total Consumption:	362.91 units

Consumption Slabs

0-150 units:	150.00 units @ ₹5.50/unit
151-300 units:	150.00 units @ ₹6.00/unit
Above 300 units:	62.91 units @ ₹6.50/unit

Charges Breakdown

Fixed Charges:	₹220
Energy Charges:	₹2133.91
Electricity Duty (6%):	₹128.03
Total Bill Amount:	**₹2481.95**

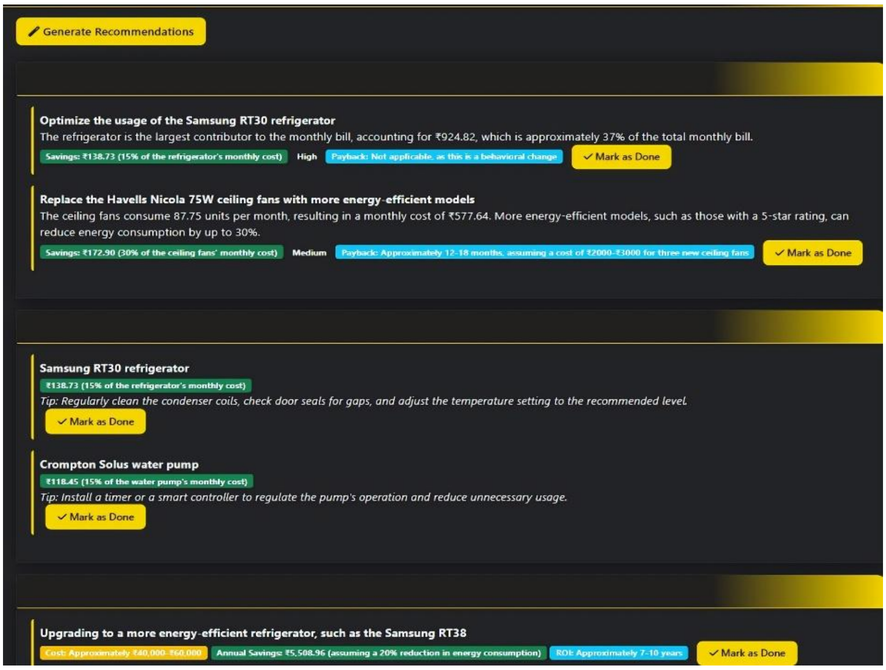
Fig. 5: Electricity Consumption and Bill Prediction Results

Household Profile
Family Members: 4 | Male: 1 | Female: 1 | Children: 2 | Type: nuclear

Fig. 6: Graphical Visualization of Energy and Load Distribution

v. LLM-Powered Recommendation Generation

ML inference and billing calculations; the backend generates a structured prompt that includes the appliance inventory of the user, per-appliance consumption over time, distribution of the bill, and a benchmark against the national average to obtain ML results. This prompt is then sent to the Groq infrastructure instance running the LLaMA 3 language model, and it is instructed to behave as an energy auditor, generating 3 to 5 quantified recommendations per user, such as Reducing air conditioner usage by 1 hour daily saves approximately ₹200 monthly (Brown et al., 2020). Figure 7 presents a sample screenshot of the AI-based recommendation interface.



Optimize the usage of the Samsung RT30 refrigerator
The refrigerator is the largest contributor to the monthly bill, accounting for ₹924.82, which is approximately 37% of the total monthly bill.
Savings: ₹138.73 (15% of the refrigerator's monthly cost) High Payback: Not applicable, as this is a behavioral change

Replace the Havells Nicola 75W ceiling fans with more energy-efficient models
The ceiling fans consume 87.75 units per month, resulting in a monthly cost of ₹577.64. More energy-efficient models, such as those with a 5-star rating, can reduce energy consumption by up to 30%.
Savings: ₹172.90 (30% of the ceiling fans' monthly cost) Medium Payback: Approximately 12-18 months, assuming a cost of ₹2000-₹3000 for three new ceiling fans

Samsung RT30 refrigerator
₹138.73 (15% of the refrigerator's monthly cost)
Tip: Regularly clean the condenser coils, check door seals for gaps, and adjust the temperature setting to the recommended level.

Crompton Solus water pump
₹118.45 (15% of the water pump's monthly cost)
Tip: Install a timer or a smart controller to regulate the pump's operation and reduce unnecessary usage.

Upgrading to a more energy-efficient refrigerator, such as the Samsung RT38
Cost: Approximately ₹40,000-₹60,000 Annual Savings: ₹5,508.96 (assuming a 20% reduction in energy consumption) ROI: Approximately 7-10 years

Fig. 7: AI-based Recommendation Interface

vi. Deployment Infrastructure

Our demonstration application is hosted on the free tier at Render (512MB RAM, shared CPU) taking around 15-30 seconds for a cold start and under 500ms for warm starts. Our demo implements HTTPS enforcement at the infrastructure level, via TLS termination. With Render's zero-cost hosting model, anyone with an internet connection in India can have free hosting (International Energy Agency, 2021).

Testing Results

A. Evaluation

i. Prediction Accuracy

Validation of the developed system has been performed by comparing its prediction results with the actual bills received from DISCOM for the test households, showing an average of 6.8% Mean Absolute Percentage Error (MAPE). This accuracy reveals the model's ability to learn and predict actual energy consumption in real life scenarios, beyond the energy prediction using the theoretical formula. The model's capability in handling the effects of standby power, actual usage variation, appliance efficiencies and their cumulative effects has been established (Zhang et al., 2020).

ii. Performance Metrics

For warm machines, the end-to-end (user input to final recommendation display on screen) latency was below 500ms (processing input, doing ML inference, calculating charge, generating LLM recommendation). The LLM component itself contributed less than 1.2 seconds. We also stream results as they come in which helps reduce perceived latency. We support 10+ appliances and per-appliance inference takes less than 15ms so there is little to no latency perceived.

iii. Model Comparison Results

Our results show that the selected Random Forest model achieves state-of-the-art performance, outperforming simple linear regression and competing models such as gradient boosting, and even shallow neural networks, in terms of predictive accuracy. When compared to a simple linear regression on the same training set (RMSE 18.4 kWh, R^2 0.81), our selected model reduces the error by 55% while attaining an R^2 of 0.96. When compared to the best gradient boosting result (RMSE 9.7 kWh), our best Random Forest model only marginally underperforms (error 15% higher) while having a mere 50% increase in inference time (12 ms compared to 8 ms). The shallow neural networks have an RMSE of 10.1 kWh (R^2 0.95), corresponding to an inference time of approximately 45 ms-more than 4 times that of Random Forest trees, making them unsuitable for real-time web applications (Breiman, 2001).

iv. Billing Engine Validation

The slab-wise billing feature of the engine is well designed to mimic the tariff DISCOMs charge. The output of the engine corresponds to the actual billing details, which are broken down into appropriate components. Test cases confirmed that the engine calculated the progressive consumption charges, fixed charges, % of electricity duty, and fuel charges correctly. This feature helps the customers to understand the cost of consumption much better and the impact they would have crossing certain thresholds of consumption. Central Electricity Authority, 2023)

v. System Reliability and Scalability

We tested the application using the Render free tier which includes up to 100MB of cached memory and shared CPU. We noticed that the small 512MB RAM allocated to the free tier was more than sufficient to run the application consistently. The stateless nature of the API and database backed sessions means we can scale out additional instances in paid tiers easily. The cold start hit of 15-30 seconds only affected the first request to the application after it had been idle for a while, after which every request responded in under 500ms.

vi. User Interface and Experience

We have managed to deliver an interactive web application that supports eight screens; landing page, screens to capture user input for household details and appliance information, screens to display consumption, bill and predictions, consumption and load charts (visualizations), a dedicated dashboard to display monthly consumption and connected load information, screen to display energy consumption and cost details for each appliance, electricity bill details by tariff slabs and last but not least screen to display optimisation suggestions using AI/ML. Additionally, the user can download the analysis results in the form of PDF (Grolinger et al., 2016).

vii. Cost Efficiency

The platform is delivered at zero cost to end users (₹0) with all the AI capabilities integrated. The cloud-hosting costs and LLM API costs are borne by free-tier cloud-hosting providers (Groq) and external LLM APIs, thus enabling the delivery of advanced energy analytics without any user costs or institutional funding required.

Discussion

A. Broader Implications and Operational Insights

The Smart Energy Optimizer is not limited to optimising energy consumption in individual households but offers many possibilities for efficient energy consumption and data-driven decisions (Hasanuzzaman et al, 2017). The Smart Energy Optimizer enables users to benefit from predictive information and transparent information about slab-based billing (Biswas et al, 2024). This can improve energy literacy and encourage consumers to change their behaviour towards more responsible energy use as demand for energy is increasing whereas resources are dwindling (International Energy Agency, 2021). From an operational perspective, they help utility providers and policymakers monitor total consumption and identify periods of peak demand. Machine learning can be used to extend this forecasting to individual devices, which can help in demand management and planning of new infrastructure to reduce strain on the network and improve reliability (Zhang et al, 2020; Alizadegan et al, 2025).

The AI powered energy data recommendation system further enhances the value of the developed tool by transforming complex energy-related data into relevant recommendations to enhance daily decision making (Stecuła et al, 2023; Momani et al, 2025). Large language models have demonstrated efficacy in generating contextual, high-quality and personalised information enabling energy optimisation for users without an in-depth technical background (Mirshekali et al, 2025).

Additionally, this framework has applications in smart home environments, where automated energy management can be applied within smart ecosystems and IoT environments, fostering efficient appliance usage and real-time energy consumption optimization (Biswas et al, 2024). The system rewards individuals for their energy saving actions and contributes to global sustainability targets by promoting energy efficiency and the evolution to smarter energy systems. From a technical standpoint, lightweight AI-based approaches can be implemented. While they may produce lower results compared to large language models, they require less processing power and memory (Kumar, 2023). The drawback of large-scale models is their high cost and reliance on substantial hardware. Nonetheless, using these models remains important due to the quality of their output (Kumar, 2023).

B. Limitations

However, despite the proven effectiveness of the proposed solution, it is necessary to highlight some limitations associated with the Smart Energy Optimizer. First, the accuracy of calculations depends on the accuracy of the inputs entered by users, including the number of hours when an appliance is turned on and how often the device is used. Although the machine learning algorithm considers the general trend in energy consumption, it may overlook unexpected changes in behaviour or unusual consumption (Zhang et al., 2020).

Second, the current version of the proposed solution does not incorporate data from smart meters and IoT sensors in real time. As a result, it cannot make accurate predictions based on changing parameters. The incorporation of real-time data streams may improve the accuracy of predictions significantly (Mishra and Vaduganathan, 2025).

Third, the billing engine works with predetermined tariff structures, which may create some discrepancies when calculating the expected bill. Even though attempts were made to consider the general structure of electricity prices used in India, there may be regional differences or changes in tariffs (Central Electricity Regulatory Commission, 2016). Moreover, incorporating LLMs into the prediction process has some implications. Although LLMs produce high-quality text, some suggestions may not contain all the required details. Thus, ensuring accuracy and consistency becomes challenging (Brown et al., 2020).

Finally, the current model predicts the energy consumption based on the grid's electricity supply. It is worth mentioning that incorporating renewables, such as solar energy, will increase its effectiveness significantly.

C. Future Work

However, there are more opportunities that could improve the proposed Smart Energy Optimizer by enhancing forecast quality and suggesting savings. First, using IoT technology to collect real-time data from a house’s smart meters might bring even greater benefits to users. Real-time energy flow can significantly boost prediction accuracy and simplify decision-making (Zhang et al., 2020). The next potential change concerns expanding tariff model support for different regions in India. Electricity varies among states and are controlled by the government. A more flexible tariff system will provide users with relevant and accurate data (Central Electricity Regulatory Commission, 2016).

Another point to consider is the development of mobile apps for managing energy consumption. With smart homes and energy management technologies, mobile platforms have shown their effectiveness in raising awareness about energy use (Ford et al, 2017). Additionally, analysing user behaviour in recommendations can help create a personalized strategy that considers users’ preferences.

Finally, there is the option to incorporate renewable energy sources, such as solar power, into energy consumption calculations and apply net metering. Given the increasing need for clean energy and sustainable solutions globally, these options will be particularly significant.

Conclusion

In today’s rapidly advancing energy landscape, managing household electricity consumption calls for intelligent and proactive decision-making. This study shows how the integration of machine learning, tariff-aware billing, and AI can meaningfully improve user's understanding and management of their energy consumption. Combination of accurate appliance-level predictions with a realistic slab-based billing engine, the proposed Smart Energy Optimizer simplifies complex pricing structures and makes electricity costs more transparent. It also goes beyond prediction by translating technical insights into clear, personalized suggestions, enabling users to make practical changes that reduce both consumption and expenses.

This study also recognizes opportunities for further improvement. Incorporating real-time data through smart meters, expanding support for region-specific tariff variations, and integrating renewable energy sources such as solar power could significantly enhance the system's effectiveness. Overall, the framework represents a practical and scalable step toward improving energy awareness, cost efficiency, and sustainability, with strong potential for adaptation across other regions facing similar challenges.

Acknowledgment

The authors gratefully acknowledge the support of Dr. Ashish Chakravarti, Head of the Department of Data Science, NIET, for providing the institutional research infrastructure necessary for this work.

Declaration

LLM based tools, specifically QuillBot and Grammarly (spelling correction and formal wording) were exclusively used for language editing purposes not for data generation, analysis, or content creation.

References

Alizadegan, H., Rashidi Malki, B., Radmehr, A., Karimi, H., & Ilani, M. A. (2025). Comparative study of long short-term memory (LSTM), bidirectional LSTM, and traditional machine learning approaches for energy consumption prediction. *Energy Exploration & Exploitation*, *43*(1), 281-301.

Bhattarai, B. P., Paudyal, S., Luo, Y., Mohanpurkar, M., Cheung, K., Tonkoski, R., ... & Zhang, X. (2019). Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions. *IET Smart Grid*, *2*(2), 141-154.

Biswas, P., Rashid, A., Biswas, A., Nasim, M. A. A., Chakraborty, S., Gupta, K. D., & George, R. (2024). AI-driven approaches for optimizing power consumption: a comprehensive survey. Discover Artificial Intelligence, 4(1), 116.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.

Central Electricity Authority. (2023). *Annual report on growth of electricity sector in India (2022–23)*. Government of India.

Central Electricity Regulatory Commission. (2016). Tariff policy regulations. Government of India.

Chen, J., Liu, Z., Huang, X., Wu, C., Liu, Q., Jiang, G., ... & Chen, E. (2024). When large language models meet personalization: Perspectives of challenges and opportunities. *World wide web*, *27*(4), 42.

Ford, R., Pritoni, M., Sanguinetti, A., & Karlin, B. (2017). Categories and functionality of smart home technology for energy management. *Building and environment*, *123*, 543-554.

Hasanuzzaman, M., Zubir, U. S., Ilham, N. I., & Seng Che, H. (2017). Global electricity demand, generation, grid system, and renewable energy polices: a review. *Wiley Interdisciplinary Reviews: Energy and Environment*, *6*(3), e222.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer.

International Energy Agency. (2021). *India energy outlook 2021*. IEA Publications.

Kalusivalingam, A. K., Sharma, A., Patel, N., & Singh, V. (2022). Leveraging random forests and gradient boosting for enhanced predictive analytics in operational efficiency. *International Journal of AI and ML*, *3*(9).

Kumar, M. (2023). Energy-Efficient AI Optimizing Large Language Models for Low-Power Edge Computing. *International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)*, *6*(6), 9692-9698.

Mirshekali, H., Shadi, M. R., Ladani, F. G., & Shaker, H. R. (2025). A Review of Large Language Models for Energy Systems: Applications, Challenges, and Future Prospects. *IEEE Access*.

Mishra, N., & Vaduganathan, D. (2025). Predictive Analytics for Demand-Side Management in Sustainable Smart Cities. *National Journal of Intelligent Power Systems and Technology*, 28-36.

Momani, M. A., Abwaini, R. K., Ababneh, B. M., & Nasser, M. A. (2025). Comparative Analysis of ARIMA and SARIMA Models in Electrical Load Forecasting: Insights for Long and Short-Term Projections. *International Journal of Electrical Engineering and Computer Science*, *7*, 83-89.

Shermy, R. P., & Saranya, N. (2025). Cloud-Based Big Data Architecture and Infrastructure. *Resilient Community Microgrids*, 131-188.

Stecuła, K., Wolniak, R., & Grebski, W. W. (2023). AI-Driven urban energy solutions—from individuals to society: a review. Energies, 16(24), 7988.

Touvron, Hugo, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov et al. "Llama 2: Open foundation and fine-tuned chat models." *arXiv preprint arXiv:2307.09288* (2023).

Zhang, Y., Chen, R., Liu, B., & Dong, X. (2020). Short-term load forecasting using machine learning: A review. IEEE Transactions on Power Systems.