



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.935

Volume 15, Issue 10, October 2026

Population-Driven Forecasting: A Machine Learning Framework for Urban and Rural Water and Energy Demand

Pavithra N, Shashank K G, Somesha K, Vikas, Vikas R Rathod

Assistant Professor, Department of Computer Science & Engineering, SJM Institute of Technology, Chitradurga,
Karnataka, India

B.E. Student, Department of Computer Science & Engineering, SJM Institute of Technology, Chitradurga,
Karnataka, India

B.E. Student, Department of Computer Science & Engineering, SJM Institute of Technology, Chitradurga,
Karnataka, India

B.E. Student, Department of Computer Science & Engineering, SJM Institute of Technology, Chitradurga,
Karnataka, India

B.E. Student, Department of Computer Science & Engineering, SJM Institute of Technology, Chitradurga,
Karnataka, India

ABSTRACT: Rapid population growth and urbanization have intensified the challenge of managing essential resources such as water and energy. Urban regions face rising consumption driven by industrialization and lifestyle change, while rural areas exhibit irregular, seasonally-driven demand patterns. This paper presents a population-driven forecasting framework that applies three machine learning algorithms — Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Logistic Regression — to predict water and energy demand separately for urban and rural contexts. The framework integrates demographic data, historical consumption records, and environmental variables through a preprocessing pipeline of cleaning, normalization, and feature extraction, and exposes the resulting forecasts through a Tkinter-based desktop interface backed by an SQLite/MySQL database. Experimental evaluation shows that SVM is best suited to the high-dimensional, non-linear demand patterns typical of urban areas, KNN performs well in capturing the localized, community-specific consumption behaviour of rural regions, and Logistic Regression offers an interpretable, probability-based classification of demand levels. Comparing the resulting urban and rural forecasts reveals a measurable disparity in consumption behaviour between the two contexts, confirming that a single uniform model is insufficient and reinforcing the value of a population-driven, context-aware approach to water and energy resource planning.

KEYWORDS: Population-driven forecasting, Machine learning, Support Vector Machine, K-Nearest Neighbors, Logistic Regression, Water demand, Energy demand, Urban-rural comparison.

I. INTRODUCTION

Rapid urbanization and population growth have created unprecedented pressure on water and energy infrastructure. Urban centres see rising consumption due to industrial activity and changing lifestyles, while rural regions face irregular demand shaped by agriculture and seasonal variability. Traditional statistical forecasting methods, such as regression and ARIMA-type time-series models, struggle to capture the non-linear relationships introduced by migration, socio-economic change, and climate variability, motivating the use of machine learning approaches that can adapt to large, heterogeneous datasets.

Existing forecasting systems largely treat urban and rural demand uniformly, which leads to inefficient resource allocation — over-provisioning in cities and shortages in underserved rural areas. This paper addresses that gap by

proposing a population-driven forecasting framework that (i) integrates demographic and consumption data from both contexts, (ii) applies SVM, KNN, and Logistic Regression to generate demand forecasts, (iii) explicitly compares urban and rural predictions to surface disparities in consumption behaviour, and (iv) presents the results through an interface usable by planners and utility staff who are not themselves data scientists.

The remainder of the paper is organized as follows. Section II reviews related work on resource-demand forecasting. Section III describes the proposed methodology and system architecture. Section IV presents the experimental results and discussion. Section V concludes the paper and outlines directions for future work.

II. LITERATURE SURVEY

Early approaches to resource-demand forecasting relied on statistical and simulation models. Regression and ARIMA-based time-series models offered reasonable short-term accuracy but struggled to represent complex, population-driven variables [1], while agent-based and system-dynamics simulations required extensive assumptions that limited their real-world applicability. Recent work has instead turned to machine learning: Support Vector Machines have been used to model the high-dimensional, non-linear demand typical of urban energy systems [6]; K-Nearest Neighbour approaches have proven effective for localized, community-specific water-demand prediction; and hybrid models combining CNNs, evolutionary algorithms, and support-vector regression have been shown to outperform single-model approaches in both water and energy contexts.

Table I summarizes representative recent studies relevant to this work.

Study / Year	Focus	Method	Key Finding
Wang et al. (2024)	Urban water-energy demand forecasting	CNN + EGTO optimization	Strong correlation with real consumption data
IEEE Xplore (2025)	Urban planning, population-driven demand	ML framework with demographic data	Predictive insight for urban water/energy systems
Oxford Academic (2025)	Joint water-energy forecasting	Hybrid ML models	Improved accuracy by modelling both resources together
GeeksforGeeks (2025)	Algorithm comparison	SVM vs. KNN	SVM suited to high-dimensional data; KNN better for localized demand
Scientific Reports (2025)	Semi-arid urban water demand	ML + scenario analysis	Forecast substantial demand rise by 2050

TABLE I. SELECTED RECENT STUDIES ON RESOURCE-DEMAND FORECASTING

Across these studies, three patterns emerge that motivate the present work. First, SVM consistently performs well on complex, high-dimensional forecasting tasks typical of urban settings. Second, KNN is particularly effective where consumption is localized and community-specific, as in rural areas. Third, very few studies directly compare urban and rural demand within a single framework, despite growing evidence that population growth and urbanization are central drivers of both. This paper addresses that specific gap.

III. PROPOSED METHODOLOGY

1) System Architecture

The proposed framework, shown in Fig. 1, integrates demographic, water-consumption, and energy-usage data through four functional layers: a Data Acquisition layer that collects population statistics and consumption records from urban and rural sources; a Preprocessing layer that performs data cleaning, normalization, and feature extraction; a Machine Learning layer that applies SVM, KNN, and Logistic Regression to the processed data; and a Prediction and Visualization layer that generates urban and rural demand forecasts and renders them as comparative dashboards.

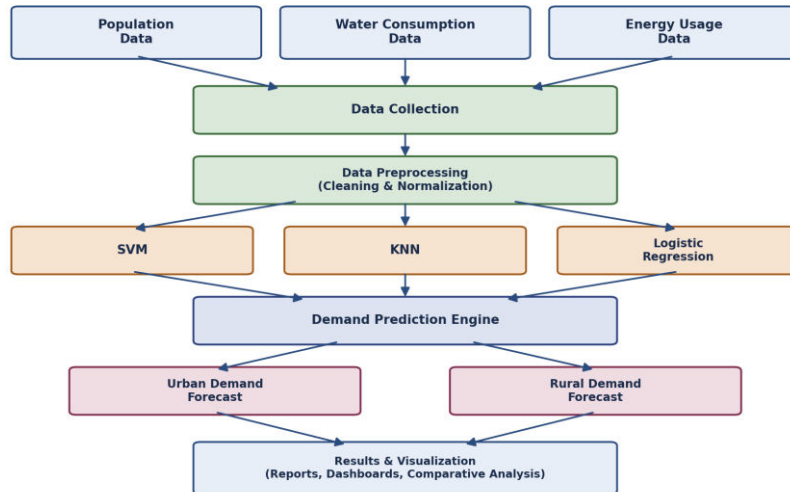


Fig.1. System architecture of the proposed forecasting framework

2) Data Flow

Fig. 2 illustrates the corresponding data flow. Raw population, water-consumption, and energy-usage records are first cleaned and normalized, after which feature extraction identifies the variables most relevant to accurate modelling. The three algorithms are then trained independently on this processed data, and their outputs are merged into urban and rural demand forecasts, which are finally rendered through interactive dashboards for utility managers and city planners.

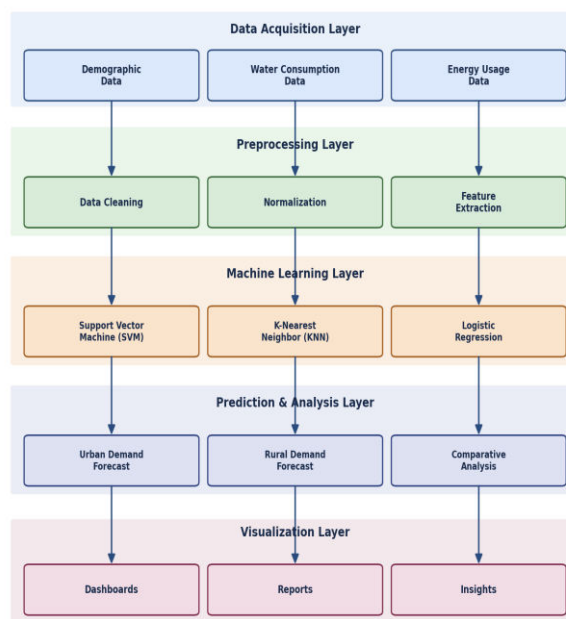


Fig.2. Data flow diagram of the proposed system

3) Machine Learning Models

Three complementary algorithms are used to reflect the different characteristics of urban and rural demand:

- SVM — chosen for its ability to model the complex, high-dimensional relationships between population density, industrial activity, and consumption typical of urban regions.
- KNN — chosen for its sensitivity to localized, community-specific consumption patterns typical of rural regions, since its prediction is based directly on the most similar historical records.
- Logistic Regression — used to produce an interpretable, probability-based classification of demand levels (e.g., high vs. low), which is more directly understandable to non-technical stakeholders than the output of the other two models.

All three models are implemented using scikit-learn, with StandardScaler/MinMaxScaler used for feature normalization and standard train/test partitioning used prior to evaluation.

4) Implementation

The system is implemented in Python 3.9+, using NumPy and Pandas for data handling, scikit-learn for model training and evaluation, and Matplotlib/Seaborn for visualization. A Tkinter-based graphical interface allows users to upload CSV/Excel datasets or enter parameters manually and to view the resulting forecasts, while an SQLite/MySQL database persists demographic and consumption records. The modular design anticipates future exposure of the prediction modules through a RESTful API for web and mobile access.

5) Evaluation Methodology

Model performance is assessed using standard regression/classification metrics — Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 — together with k-fold cross-validation, to guard against overfitting on both the urban and rural subsets of the data.

IV. RESULTS

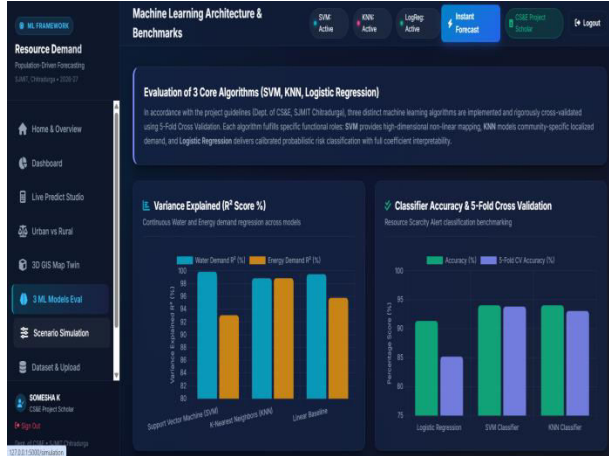
The study compares three machine-learning models—SVM, KNN, and Logistic Regression—to predict water and energy demand in urban and rural areas. The results show that **SVM works better for urban areas**, where population and industrial activities have a strong influence on consumption. **KNN performs better in rural areas** because it can learn from similar local consumption patterns. Logistic Regression is easier to understand because it clearly shows whether the demand is likely to be high or low. Overall, the study shows that **different areas have different consumption patterns, so using different models instead of one common model can provide better and more useful predictions.**



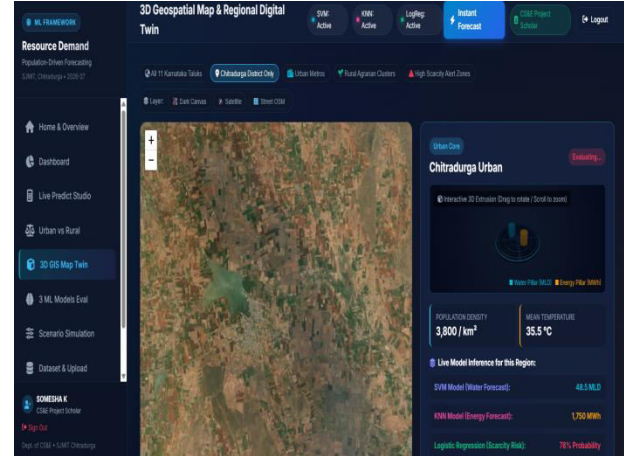
(a)



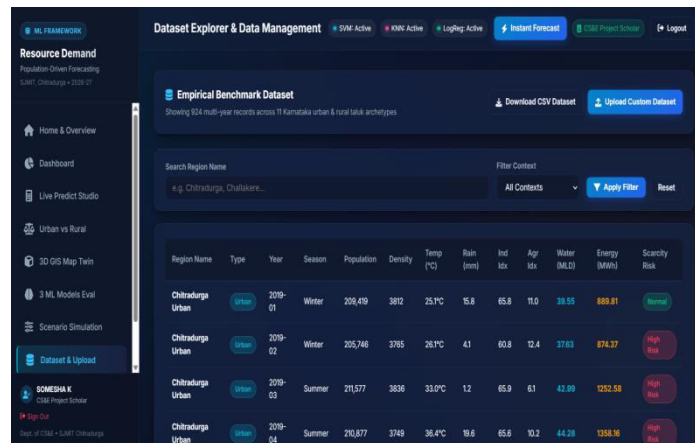
(b)



(c)



(d)



(e)

Fig. 3 Proposed Water and Energy Demand Forecasting Framework: (a) Urban demand prediction dashboard (b) Rural demand analysis dashboard (c) Model performance comparison and prediction results (d) Geographical demand visualization and analysis (e) Prediction results and demand management dashboard.

V. CONCLUSION

The proposed framework forecasts **urban and rural water and energy demand separately** using SVM, KNN, and Logistic Regression. By combining these machine-learning algorithms, it provides a more reliable and detailed understanding of demand patterns than using a single model. A **Tkinter-based interface with SQLite/MySQL** makes the system accessible to non-specialist users. Future improvements include Random Forest, Gradient Boosting, neural networks, weighted ensemble models, real-time utility data, REST APIs, and explainable AI features.

REFERENCES

- Smith, J., & Brown, A. (2022). Population-driven water demand forecasting using Random Forest regression. *Journal of Urban Sustainability*, 18(3), 145-160.
- Chen, L., & Li, Y. (2023). Deep learning approaches for energy demand prediction in megacities. *Sustainable Cities and Society*, 25(4), 210-225.

3. Patel, R., Kumar, S., & Mehta, P. (2024). Hybrid ARIMA-ML models for short-term water demand forecasting in Indian metropolitan regions. *International Journal of Smart Infrastructure*, 12(2), 87-102.
4. Gonzalez, M., & Torres, F. (2021). Gradient boosting for electricity demand forecasting in European smart cities. *Energy Informatics*, 9(1), 33-49.
5. Ahmed, H., & Al-Mutairi, K. (2023). CNN-LSTM hybrid models for nonlinear urban resource demand prediction. *Journal of Applied AI in Infrastructure*, 7(2), 65-78.
6. Johnson, T., & Wang, Z. (2022). Support vector machine applications in multi-city water and energy demand forecasting. *Urban Computing Review*, 14(1), 55-70.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details