Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Predicting Energy Consumption Patterns with Advanced Machine Learning Techniques for Sustainable Urban Development

Syed Ali Reza[1], Md Sakibul Hasan[2], Mohammad Hamid Hasan Amjad[3], Md Shafikul Islam[4], Md Masud Karim Rabbi[5], Arat Hossain[6], Md Shihab Sadik Shovon[7] and Tanaya Jakir[8]

[1]*Department of Data Analytics, University of the Potomac (UOTP), Washington, USA*
[2]*Information Technology Management, St Francis College.*
[3]*Master of Science in Engineering Management, College of Graduate and Professional Studies, Trine University.*
[4]*Assistant professor, Department of Computer Science and Engineering, Uttara University*
[5]*Master's in Business Administration, International American University*
[6]*Information Technology Management, St Francis College.*
[7]*Master of Science in Engineering Management, College of Graduate and Professional Studies, Trine University.*
[8]*Master's in business Analytics, Trine University*
Corresponding Author: Syed Ali Reza, E-mail: syed.alireza@student.potomac.edu

| ABSTRACT

As urbanization continues to expand and evolve in the USA, power demand has increased manifold, and with it has arisen significant environmental problems in the form of increased greenhouse gas emissions and loss of resources. In this paper, we explore how future machine-learning techniques could predict power consumption in U.S. cities. The central aim of this research is to develop advanced machine learning models with the potential to effectively predict energy consumption in cities. This involves not only identifying the key variables behind energy consumption but also selecting and fine-tuning machine learning algorithms that are most capable of understanding the dynamics of urban energy intricacies. This study focuses on the energy consumption patterns in the large cities of the United States, recognizing the diversity of challenges and opportunities presented by different geographic and demographic situations. The dataset used in this research project offered a comprehensive view of energy consumption across various fields of household, commercial, and industrial consumption, giving a holistic view of energy dynamics within cities. It integrated data collected from smart meters that offer granular electricity consumption patterns at the level of individual households and businesses with weather reports that detail ambient conditions governing energy demand, such as temperature and humidity fluctuations. Government energy records add historical context and policy information, further enhancing the dataset and enabling close analysis of trends and patterns in energy consumption. The next phase was to select and train three distinct machine models to explore the energy consumption dataset, namely, Logistic Regression, Random Forest, and XG-Boost algorithms. Random Forest outperformed Logistic Regression and XG-Boost slightly in terms of accuracy and other evaluation metrics. However, all models exhibit relatively low accuracy, suggesting the need for further tuning, feature engineering, or alternative models to improve predictions. In major cities in the U.S. such as Los Angeles, Chicago, and New York, smart power forecasting based on AI is revolutionizing power distribution and power planning in cities. By utilizing advanced machine learning models, these cities can process vast amounts of information and predict power usage with high accuracy. The incorporation of artificial intelligence (AI) in urban power planning has been a defining feature of modern-day power management in the USA. Major cities such as Los Angeles, Chicago, and New York are increasingly adopting AI-powered power forecasting technologies to rationalize power distribution. The integration of machine learning insights in U.S. government-driven green construction is instrumental in driving sustainable construction in infrastructure. By utilizing data-driven approaches, policymakers are in a position to identify the optimal design methods and low-power technologies with high performance in buildings.

| KEYWORDS

Energy Consumption, Machine Learning, Sustainable Urban Development, Smart Cities, Predictive Analytics

| ARTICLE INFORMATION

ACCEPTED: 20 February 2025 **PUBLISHED:** 02 March 2025 **DOI:** 10.32996/jcsts.2025.7.1.20

I. Introduction

Background and Context

Barua et al. (2025), reported that urbanization is occurring at an unprecedented rate in the USA, and it is estimated that by 2050, nearly 68% of the global population will be residing in urban areas. The phenomenon brings with it a myriad of challenges, not the least of which is the question of energy consumption, which has been rising steadily in urban areas as a result of increasing population density, economic activities, and technological advancement. Anonna et al. (2023), found that the impact of this rising energy demand is profound, given that urban areas are drivers of carbon emissions and environmental degradation. Indeed, the International Energy Agency (IEA) notes that cities are the source of approximately 70% of global energy-related CO2 emissions. Against this backdrop, there is a requirement to develop effective strategies that not only mitigate the environmental impacts of urban energy consumption but also ensure a stable and efficient energy supply to meet the needs of urban populations (Al Mukaddim et al.,2024).

According to Ahmed et al. (2025), accurate energy demand forecasting is essential for sustainable cities in the USA. Forecasts enable city planners and utilities to anticipate demand variability, optimize resource allocation, and schedule energy conservation. However, traditional energy forecasting models are often founded on past trends and simplifying assumptions that cannot keep pace with the complexity and dynamics of modern urban energy demand. Chowdhury et al. (2024), argued that as urban energy demand patterns become more influenced by such dynamics as climate change, technological change, and socio-economic processes, there is an increasing requirement for more sophisticated modeling methods that can adapt to these dynamics and make more precise forecasts.

Problem Statement

Hasan et al. (2024a), postulated that the limitations of conventional forecasting methods present a significant challenge to effective energy management in cities. The models typically utilize linear regression or time-series analysis that may not completely represent the complexity of variables underlying energy consumption, including weather, demographics, economic activity, and technology adoption rates. The result is that these traditional approaches typically provide poor predictions, leading to inefficiencies in energy consumption and distribution. In addition, they may fail to identify emerging trends, including increasing penetration of renewable sources of energy and the growing market for electric vehicles, which demand a more nuanced understanding of energy dynamics (Gazi et al., 2024).

This lack of traditional models highlights the urgent need for data-driven intelligence that can harness the power of advanced analytics and machine learning. By applying machine learning algorithms, it is possible to handle enormous amounts of data from disparate sources, detect hidden patterns, and generate predictive intelligence that can inform decision-making (Hossain et al., 2025). These models can learn from the dynamic nature of urban energy consumption and hence provide a more accurate representation of current and future consumption patterns. Thus, there is an urgent need for research that explores the application of machine learning in energy consumption forecasting, which can eventually help achieve optimized energy distribution, enhanced efficiency, and sustainability in urban settings (Reza et al., 2024).

Research Objective

The central aim of this research is to develop advanced machine learning models with the potential to effectively predict energy consumption in cities. This involves not only identifying the key variables behind energy consumption but also selecting and fine-tuning machine learning algorithms that are most capable of understanding the dynamics of urban energy intricacies. The research aims to develop practical conclusions that can assist policymakers and city planners in developing measures to enhance sustainability and simplify energy management policies.

Using machine learning algorithms, this project seeks to analyze historical energy consumption data alongside auxiliary datasets such as weather, economic indicators, and demographic information. The goal is to create predictive models that can accurately forecast energy demand in various urban areas so that stakeholders can make informed decisions regarding energy supply and infrastructure development. In so doing, this project seeks to contribute to the broader debate of sustainable urban development and energy management with a focus on the importance of integrating advanced analytical methods into policy-making and urban planning processes.

Scope and Relevance

This study focuses on the energy consumption patterns in the large cities of the United States, recognizing the diversity of challenges and opportunities presented by different geographic and demographic situations. By focusing on the U.S., the research aims to present findings with immediate applicability to urban planners, policymakers, and energy providers operating in this environment. The application of machine learning algorithms is particularly suitable because of increasing access to large datasets and advances in computing capability that allow investigation into complex patterns of energy consumption.

Furthermore, the importance of this research extends beyond intellectual curiosity; it underlines the need for innovative solutions to the challenges posed by rising energy demands in cities. As cities grapple with the dual imperatives of procuring energy and achieving sustainability goals, the findings of this research can inform the development of viable energy management policies that align with broader environmental and economic goals. Lastly, by harnessing the power of state-of-the-art machine learning algorithms, this research seeks to contribute to a more sustainable urban future, one where energy consumption is optimized, resources are used efficiently, and environmental footprints are minimized.

II. Literature Review

Urban Energy Consumption Trends in the USA

Sumon et al. (2024a), examined how urban energy consumption in the United States has evolved dramatically over the past few decades, with several trends emerging due to demographic change, technological advancement, and growing concern for environmental sustainability. Population density, economic activity, and climate are some of the key drivers of urban energy demand, and these variables interact in complex ways to influence consumption. Sumsuzoha et al. (2024), asserted that as urban areas densify, energy demand typically increases, driven by residential, commercial, and industrial demand. High population density tends to generate higher energy consumption for heating, cooling, lighting, and transportation, as more individuals reside in smaller geographic areas. Not only does this urban concentration increase energy demand, but it also presents unique challenges for energy distribution and efficiency.

Second, economic activity is one of the most important determinants of energy consumption patterns. Cities are often the epicenters of economic growth, innovation, and employment generation, and these may reflect greater energy needs for commercial and industrial applications. As businesses expand and new industries emerge, the demands for energy to power operations, production, and transportation expand likewise (Abbas et al., 2025). Lastly, the shift toward a service-oriented economy in the majority of U.S. cities has implications for energy use, as services are likely to embody different energy inputs than manufacturing. This change demands a nuanced picture of how economic conditions influence energy use patterns, particularly in cities where economic activity is most concentrated (Bansod, 2025).

As per Cui et al. (2025), climate is another significant factor affecting energy use in cities. Temperature, humidity, and seasonal weather patterns directly impact heating and cooling requirements, and hence total energy use. For instance, cities in warmer climates experience greater summer demand for air conditioning, while colder cities experience greater winter demand for heating. The rising occurrence of extreme weather events with climate change introduces additional complexity to these dynamics, as cities experience unexpected peaks in energy demand during heat waves or winter storms. It is thus necessary to appreciate the link between climate and energy use to be able to develop adequate energy planning solutions in cities (Shil et al., 2024).

Ekeh et al. (2025), ascertained that Infrastructure is also a key determinant of energy consumption patterns. The design and efficiency of buildings, transportation, and energy distribution systems can all significantly influence energy consumption in cities. For example, poorly insulated buildings can lead to higher heating and cooling costs, while inefficient transportation systems can result in more fossil fuel consumption. Investment in new energy-efficient infrastructure, like smart grids and renewable energy installations, can address some of these problems by raising the efficiency of energy use and reducing overall consumption. Ghallabi et al. (2025), indicated that as cities continue to grow and evolve, these underlying factors and their influence on energy trends remain a vital area of research, providing crucial lessons for policymakers and urban planners looking to enhance sustainability.

Classical vs. Machine Learning-Based Energy Forecasting

Gera et al. (2025), established that forecasting of energy consumption has traditionally relied on statistical models such as autoregressive integrated moving averages (ARIMA) and linear regression. These models have been the cornerstone of energy demand forecasting for many years due to their simplicity and ease of application. However, they have significant limitations that render them ineffective in capturing the dynamics of urban energy consumption patterns. One of the key limitations of traditional statistical models is that they rely on linear relationships. Most of the determinants of energy consumption, such as economic growth, weather, and demographic change, have non-linear relationships that cannot be effectively captured by these models. Traditional approaches, therefore, can yield oversimplified forecasts that fail to capture the true dynamics of energy consumption (Madabathula et al., 2025).

Furthermore, traditional forecasting models typically depend on past trends with the expectation that past patterns will continue without accounting for abrupt changes in technology, policy, or behavior. This expectation can be particularly problematic when it comes to urban energy consumption, where the adoption of electric vehicles, the deployment of renewable energy sources, and the implementation of energy efficiency standards can radically alter consumption patterns (Marella & Palakurti, 2025). Thus, the models can produce forecasts that become outdated very quickly, leading to inefficiencies in energy planning and management. In contrast, machine learning-based approaches possess several strengths that render them more appropriate for handling complex energy consumption data (Mohammed et al., 2025)

Ogunseye et al. (2025), articulated that Machine learning algorithms, including decision trees, support vector machines, and neural networks, possess the capacity to automatically identify patterns and relationships in large datasets without being programmed to do so explicitly. This adaptability allows machine learning models to capture non-linear relationships and interactions between variables, making them more precise in forecasting energy consumption. Through learning from large datasets that comprise various factors influencing energy demand—such as weather, economic indicators, and demographic data—machine learning models can identify intricate patterns that might elude traditional approaches (Palaniappan et al., 2025)

Furthermore, machine learning models are capable of learning and improving over time as new data becomes available, and are therefore particularly well-suited to dynamic systems like urban energy systems. This allows for adjustments to be made in real-time to forecasts based on the latest information, enabling energy providers and urban planners to respond more effectively to changing conditions (Robi & George, 2025). Machine learning methods can also handle high-dimensional datasets and are therefore capable of processing a range of different data sources, like smart meter readings and social media trends, to improve the accuracy of energy forecasts. As more cities adopt the use of smart technologies and the Internet of Things (IoT), the quantity and range of data available to be analyzed will only continue to grow further, again pointing to the relevance of machine learning for energy forecasting (Sahar et al., 2025)

Applications of Machine Learning in Energy Management

According to Satpathy et al. (2025), machine learning for energy management has garnered significant attention in recent years, delivering innovative solutions to forecast peak demand, optimize resource allocation, and incorporate renewable resources. Peak energy demand prediction is one of the primary applications of machine learning, which plays a crucial role in ensuring the reliability of energy supply systems. By having accurate predictions of peak demand, utilities can allocate resources more effectively, prevent shortages, and reduce the need for costly infrastructure development. Vasenin (2025), established that Machine learning algorithms can analyze historical consumption patterns alongside external drivers, such as weather and economic indicators, to predict when and where peak demand is most likely to occur. Based on these insights, energy suppliers can optimize supply chains and demand response initiatives, incentivizing users to modify consumption during peak periods, which leads to a more balanced and efficient energy grid.

Furthermore, machine learning can help optimize resource allocation through better load forecasting and operation planning. Utilities can utilize predictive models to forecast periods of high demand and adjust their generation plans accordingly, such that the supply of energy is aligned with consumption patterns. Such optimization is particularly important in regions with intermittent renewable energy resources, such as solar and wind, where generation capacity can change in line with environmental conditions. By incorporating machine learning into operational systems, energy suppliers can enhance their ability to balance supply and demand, reduce their reliance on fossil fuels, and reduce greenhouse gas emissions.

Sumon et al. (2024b), argued that another crucial application of machine learning in energy management is in integrating renewable energy sources into the existing grids. As the energy mix shifts in favor of cleaner sources, the problem of their intermittency becomes increasingly urgent. Machine learning algorithms can forecast renewable energy output from variables such as weather and historical production data, enabling grid operators to better anticipate fluctuations in supply. Furthermore, machine learning can contribute to the development of smart grids that use real-time data to optimize energy distribution and consumption, enabling easy integration of different energy sources. Reza et al. (2024), reported that by adopting AI-based solutions, cities can render themselves more resilient to the impacts of climate change and build more sustainable energy systems that are aligned with general environmental goals.

Research Gaps

Hasan (2024), argued that despite the positive advances in machine learning applications for energy management, there exist several research gaps that require further exploration. One of the gaps is the lack of adaptive machine-learning models for urban energy prediction. While existing models have been effective in other contexts, many of them fail to capture the unique dynamics and complexities of urban energy demand. Cities continue to evolve, and hence there is an urgent need for research studies that develop and test adaptive models with the ability to learn from the rapidly evolving energy demand under the influence of technological, environmental, and societal dynamics. This involves the incorporation of real-time data streams and the use of advanced algorithms with the ability to modify predictions dynamically based on emerging trends (Hossain et al., 2025b)

Additionally, the need for real-time, high-accuracy predictive models is paramount in the instance of urban energy management. Traditional forecasting methods rely on static datasets and may not be able to capture the real-time changes in energy consumption spurred by the incidence of extreme weather events or sudden changes in consumer behavior (Gazi et al., 2025). By emphasizing research into the development of machine learning models that can provide real-time insights, researchers can significantly enhance the ability of energy providers and urban planners to make informed decisions and respond to evolving energy trends promptly. This avenue of research is in line with the present momentum toward smart city movements, where real-time data and analytics play a key role in optimizing urban resource management (Chowdhury et al., 2024).

III. Data Collection and Exploration

Dataset Overview

The dataset used in this research project offered a comprehensive view of energy consumption across various fields of household, commercial, and industrial consumption, giving a holistic view of energy dynamics within cities. It integrated data collected from smart meters that offer granular electricity consumption patterns at the level of individual households and businesses with weather reports that detail ambient conditions governing energy demand, such as temperature and humidity fluctuations. Government energy records add historical context and policy information, further enhancing the dataset and enabling close analysis of trends and patterns in energy consumption. This integrated approach not only enables a better understanding of energy consumption but also predictive modeling for energy planning and sustainability.

Data Preprocessing

The implemented code script establishes a standard data preprocessing pipeline using Python libraries including scikit-learn and pandas. It begins with the importation of modules for data manipulation, model selection, preprocessing, and pipeline. The script then proceeds to handle missing values by identifying columns with NaNs and imputing them using Simple Imputer with a mean strategy for numeric features and most frequently for categorical features. It then encodes categorical features, specifically "Type of Renewable System," using Label Encoder. Feature engineering is performed to create a new feature "Energy Potential" from pre-existing features. Outliers in numeric features are handled by capping them at the 5th and 95th percentiles. The script then separates the features and target variable (which is "Energy Consumption Category") and splits the data into train and test sets. It then scales the numeric features using Standard Scaler and prints the head of the preprocessed train and test data frames, marking the end of the preprocessing steps before model training.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a fundamental step in the process of data analysis in which datasets are systematically investigated and graphed to identify hidden trends, patterns, and anomalies without a predetermined hypothesis. Statistical techniques and graphical presentations, such as histograms, scatter plots, and box plots, are applied in EDA to summarize the major features of the data and allow the analyst to gain insights into the structure and distribution of the data. In research, EDA is a fundamental step in guiding future analyses, informing model selections, and determining potential variable relations. It helps researchers identify problems with the quality of the data, such as missing observations and outliers, and constructs hypotheses that are later testable, ultimately increasing the validity and soundness of the research results.

a) Pairwise Relationship of Numerical Features

The implemented code script produces a pair plot visualization of numerical features in a Pandas data frame named df with Seaborn. It begins with a list named numerical features with names of columns representing specific measurements such as solar irradiation, temperature, wind speed, humidity, cloud cover, pressure in the atmosphere, sunshine hours, and system efficiency. The function sns. The pair plot is called the Data Frame subset with only the specified numerical features. diag_kind="KDE" is included to display a kernel density estimate on the diagonal for every feature distribution, and corner=True makes the plot basic with only the lower triangular structure of the pairs. Finally, plt. Subtitle adds a super-title on the entire graph with added information, and plt.show() shows the resulting pair plot with the ability to visually explore correlations and distribution among selected numerical features.

Output:

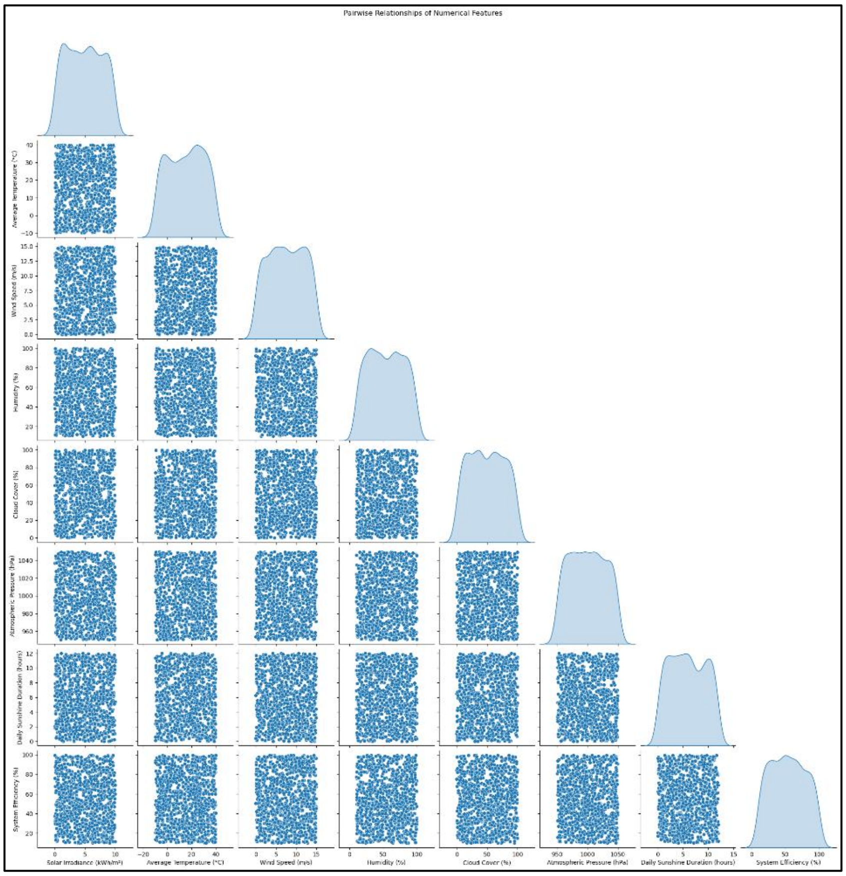

Figure 1: Pairwise Relationship of Numerical Features

The provided figure is a pair plot plotting the pairwise relations among a variety of numerical features in the dataset, and in turn, giving a global overview of their distribution and potential correlations. Each of the pairs in the matrix is a scatterplot representing the connection between two variables, with the diagonal representing the kernel density estimates, and these represent the distribution of every feature on a standalone basis. Interestingly, a variety of variables exhibit distinct patterns that would represent potential correlations; e.g., increased energy usage is associated with increased temperatures, as in the pairs where the energy usage and temperature features have a rising trend. In addition, the density plots reveal the majority of features are skewed to the right with greater values being relatively scarce. The scatted points in a number of the relations represent variability and could warrant investigation. Overall, such visualization is a preliminary step in the process of identifying significant relations and guiding future model-building and areas warranting exploration of the relations in usage.

b) Distribution of Renewable System Types

This Python script generated a count plot using the Seaborn library to visualize the distribution of a categorical feature named "Type of Renewable System" within a Pandas Data Frame, presumably named df. The script initializes a Matplotlib figure with a specified size of 8x4 inches. It then employs sns. Counterplot to calculate and display the frequency of each category, using the Viridis color palette for distinct visual representation. The plot is enhanced with a title, "Distribution of Renewable System Types," and labeled axes for clarity. Specifically, the x-axis, representing the categories of renewable systems, is titled "Type of Renewable System," and the y-axis, indicating the counts, is labeled "Count." To improve readability, the x-axis tick labels are rotated 45 degrees. Finally, plt.show() renders the generated count plot, providing a clear visual summary of the categorical feature's distribution.

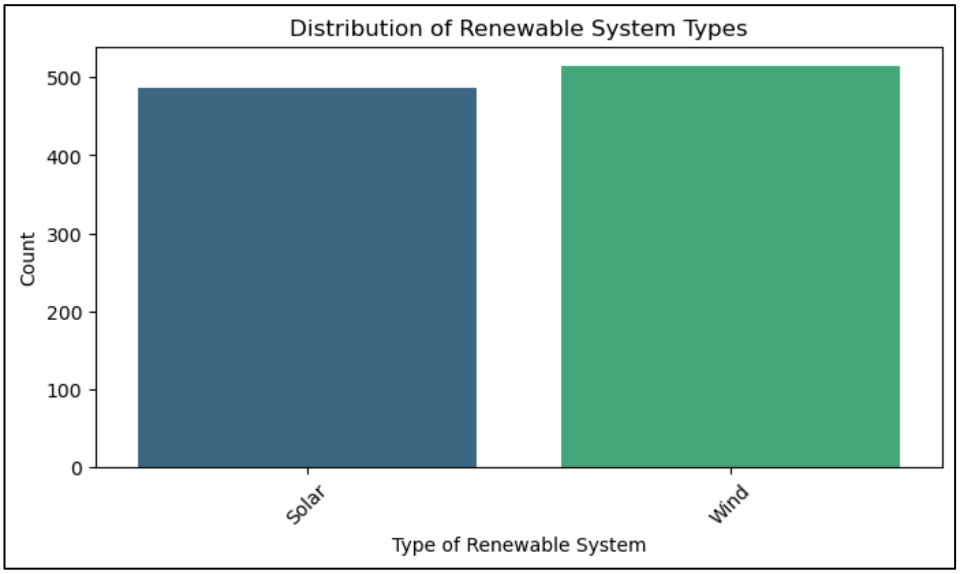

Figure 2: Distribution of Renewable System Types

The chart represents the distribution of system types in the renewable systems, comparing wind and solar installations in the database. It reveals a near-equal distribution in the number of installations of the two system types, with installations in the solar-type totaling approximately 500 and wind installations narrowly above the figure, also at approximately 500. The balanced distribution reflects the high degree of interest and investment in the two renewable resources, a reflection of how crucial they are in the power market. The fact that there is no significant margin in the installations reflects the technologies being adopted at a relatively equal pace, a reflection of a diversification strategy in the generation of renewable power. The information is significant in guiding policymakers and power planners on the need to invest in infrastructure and resources in wind and solar systems to spur a future with a power source.

c) Energy Consumption Category

This Python program uses the Seaborn and Matplotlib modules to plot the distribution of the target variable, "Energy Consumption Category," in a Pandas Data Frame variable df. It initializes a figure in Matplotlib with a size of 6x4 inches. The method sns. countplot is called to get a count graph with the frequency of each category in the target variable. The cool palette is called to color the bars in the graph. The graph is decorated with a title, "Distribution of Energy Consumption Categories," and with axis titles: "Energy Consumption Category" on the x-axis and "Count" on the y-axis. Finally, plt.show() shows the resulting count graph, presenting a graphical overview of the distribution of the energy consumption categories.

Output:

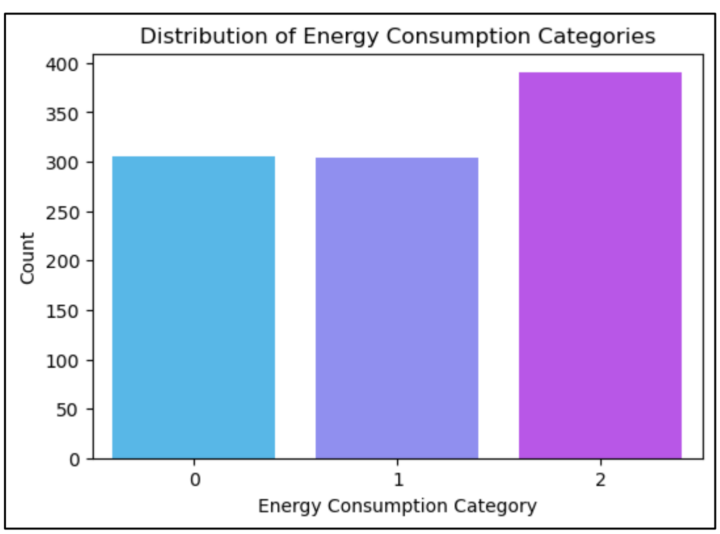

Figure 3: Energy Consumption Category

The chart illustrates the distribution of the categories of consumption, labeled as category 0, category 1, and category 2, with the frequency counts in every category. Category 2 is the largest category with approximately 400 counts, implying a high percentage of the dataset is in category 2. Category 0 and Category 1 are close in counts with approximately 300 in every category, with a relatively even distribution but with a considerably lower presence in category 2. The distribution reveals a potential trend in consumption patterns where category 2 would represent a high usage level or a special demographic characteristic that is significant and warrants investigation. The visualization helps communicate differences in the categories and is informative in helping researchers and policymakers discern consumption trends and guide interventions accordingly.

d) Displays Solar Irradiance

The Python script makes a Seaborn and Matplotlib scatter plot of the relationship between "Solar Irradiance" and "System Efficiency" with points differentiated by "Energy Consumption Category." It makes a figure in Matplotlib with an 8x5-inch dimension. The sns. The scatterplot function creates the scatter plot with "Solar Irradiance" as the x-axis, "System Efficiency" as the y-axis, and "Energy Consumption Category" used in coloring (hue) points. The title is "Solar Irradiance vs. System Efficiency," and labels are provided for axes. It has a legend to interpret coloring by energy consumption category. Finally, plt.show() displays the generated scatter plot to analyze the relationship between system efficiency and solar irradiance given different energy consumption categories.

Output:

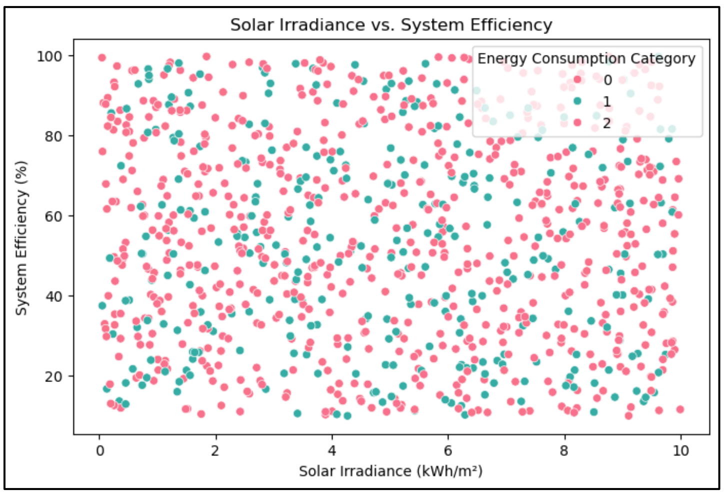

Figure 4: Displays Solar Irradiance

The graph is a scatter plot of system efficiency (in percentage points) and solar irradiance (in kWh/m^2) with points of different colors coding three categories of energy consumption: 0, 1, and 2. The points indicate a broad spread of system efficiency across a broad range of solar irradiance, indicating that although there are highly efficient systems even at low irradiance, there are also systems with poorer performance. There is no apparent linear relationship between system efficiency and solar irradiance, indicating other performance-influence factors aside from irradiance. The color differentiation of points allows a visual observation to be made regarding how each category of energy consumption performs given different levels of solar irradiance and could determine specific trends and deviations in each category. The data could be an important contribution to optimizing solar power systems and determining how consumption patterns affect efficiency outcomes under different conditions.

e) Histogram of Numerical Features

The Python Code script program makes histograms on a list of numerical features in a Pandas DataFrame named df. It operates on columns in the numerical features list and uses the .hist() method to get a grid of histograms. figsize is assigned a figure dimension of 15x10 inches and bins=20 is included to have every histogram with a total of 20 bins. edgecolor='black' puts a black edge around the bars in the histograms and color='skyblue' is included to specify the filling color. Plt. The subtitle puts a title on the figure above the grid of subplots. Plt.show() is invoked on the resulting histograms to visually explore the distribution in every numerical feature.

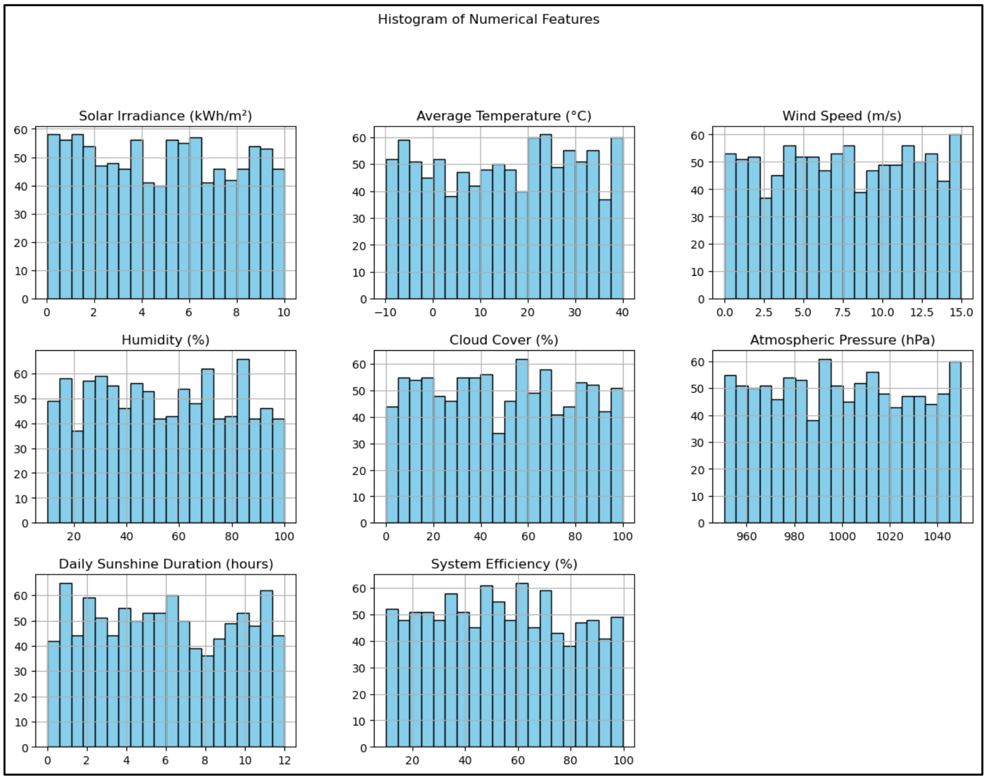

Figure 5: Histogram of Numerical Features

The chart illustrates a series of histograms depicting the distribution of a variety of numerical features related to solar power, e.g., solar irradiation, mean temperature, wind speed, humidity, cloud cover, pressure, daily sunshine hours, and system efficiency. Each histogram portrays information on the frequency within a specified range for each feature. Solar irradiation, e.g., shows a concentration in the value around 5-10 kWh/m², depicting favorable conditions for harnessing solar power. Mean temperature is quite uniformly distributed, while wind speed is skewed in favor of low wind speeds. The histograms on humidity and cloud cover are mostly in intermediate ranges, and these could impact solar efficiency. Atmospheric pressure is quite uniform, while daily sunshine hours are in a concentration of around 5-6 hours. The system efficiency histogram shows a wide variety with a cluster of observations around high efficiency, depicting efficient solar power systems. The visualization in its entirety facilitates understanding the properties and variability in these crucial variables influencing the generation and performance of solar power.

f) Energy Potential vs. Sunshine Duration

The code script in the Python program explores the relationship between a "calculated" "Energy Potential" feature and "Daily Sunshine Duration," grouped by "Energy Consumption Category," in a scatter plot. It begins with the generation of the "Energy Potential" feature as a multiplication of "Solar Irradiance" and "System Efficiency" and inserts it in a new "Energy Potential" column in the Data Frame df. It creates a Matplotlib figure with a dimension of 8x5 inches. The sns. The scatterplot function produces the scatter plot with "Energy Potential" on the x-axis, "Daily Sunshine Duration" on the y-axis, and "Energy Consumption Category" used to differentiate points based on hue. The figure is labeled with a title, "Energy Potential vs. Sunshine Duration," and appropriate axis titles. It includes a legend to differentiate based on the energy consumption category. Finally, plt.show() displays the resulting scatter plot, to observe how energy potential and sunshine duration are related to differing energy consumption categories.

Output:

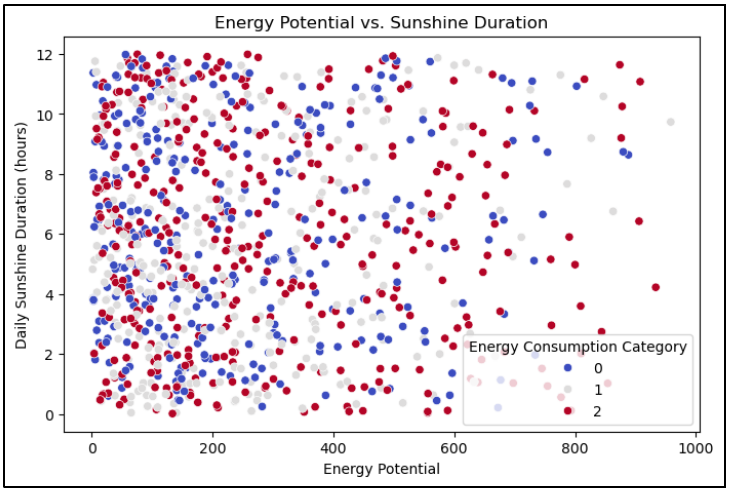

Figure 6: Energy Potential vs. Sunshine Duration

The scatter plot illustrates how daily sunshine duration is related to energy potential (measured in a not specified unit), with points coded following categories of energy consumption: 0, 1, and 2. The graph illustrates a wide distribution in energy potential ranging from 0 to nearly 1000, with daily sunshine duration having a cluster in the area of 0-12 hours. Interestingly, there is a general upward trend where increased energy potential is related to increased sunshine durations, but the scatter reveals significant variability, implying other factors could play a role in influencing energy potential alongside sunshine hours. Category 2 and Category 1 are coded with distinct colors, with Category 2 having a cluster at high energy potential, implying a potential connection with high consumption categories and high potential in producing energy. The visualization helps understand how daily sunshine duration could impact energy potential and guide future research in optimizing energy systems based on these factors.

IV. Methodology

Feature Engineering

In the feature engineering step, we targeted identifying and creating significant predictive features with high potential to influence the usage patterns in power. The key among these are temperature, usage timing, type of structure, and occupation. The temperature is significant in demand because a temperature change creates a demand for cooling and heating. The information about usage timing is significant because usage is variable over a day based on peak usage and seasonal fluctuation. The type of structure (residential, commercial, and industrial) is significant in providing usage information we would anticipate, and occupation information is significant in approximating how many are in a location and how much power is going to get utilized.

To enhance the dataset even further, we incorporated features such as energy efficiency ratings, representing how efficiently power is utilized based on the type and size of the building. These ratings help identify buildings with potential advantages in power conservation. We also calculated demand peaks, representing the highest level of power usage over a specified timeframe, and provided insights on pressure points in the power system. These engineered features better inform what is driving power usage and set the stage for efficient modeling.

Model Selection and Training

The next phase was to select and train three distinct machine models to explore the energy consumption dataset, namely, Logistic Regression, Random Forest and XG-Boost algorithms. We commenced with a basic machine model, which is Logistic Regression, to discover linear connections in the dataset. It is a helpful model in detecting simple trends in the usage of power and is a reference with which other models with increased complexity are measured. Next, we applied a Random Forest Classifier, which is highly efficient in identifying non-linear relations and feature-feature relations. Not only is the method a robust prediction tool but also a feature importance tool, and we can observe what features are having the strongest impact on the consumption of energy. The ability of Random Forest to handle high-dimensional features without overfitting makes the method a valuable tool in such a type of analysis.

Lastly, we applied the XG-Boost Classifier, a high-precision and efficient gradient boost model. The XG-Boost is perfectly applicable to intricate relations and large-volume datasets and is therefore an appropriate instrument in our analysis of energy consumption.

The reason we applied ensemble learning through Random Forest and XG-Boost is that the models can enhance predictive accuracy through a collection of various models and reduce overfitting risks. The approach assists in enhancing accuracy in predictions of energy consumption patterns and peak demand scenarios.

Model Optimization and Performance Analysis

For model refinement, we carried out hyperparameter tuning using the Grid Search approach, a methodical procedure for attempting many different combinations of parameters to find the best combination for each model. The procedure optimizes model performance and efficiency with the tweaking of settings that have a notable effect on predictive power. Through tuning hyperparameters like the quantity of Random Forest trees and the learning rate of the XG-Boost model, we sought to enhance the predictive power of each model. Additionally, we made use of cross-validation techniques in order not to overfit and generalize models on unseen data. Cross-validation is a process in which we split the dataset in a train and validation set multiple times to evaluate model performance uniformly. By doing this, we obtained improved predictive power estimations and decreased the possibility that models would generalize only on the train information.

Evaluation Metrics

In evaluating models, we made use of differing metrics under the specific goals in prediction and forecasting peak demand. In prediction problems with a focus on the prediction of energy, we took on board Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE as primary metrics. RMSE is enlightening on the mean level of prediction error made, and MAE is a basic way in which we calculate mean errors and is important in assessing model performance.

For the classification issues of predicting peak demand, we employed measures such as precision, recall, the F1-score, and accuracy. Accuracy measures the overall correctness of the model, whereas precision and recall provide more insight into how well the model is doing at identifying peak demand events. The F1-score is the harmonic average of precision and recall and is a balanced performance metric, particularly in the case of imbalanced classes.

Finally, we undertook a comparative performance analysis to assess the trade-offs in interpretability and predictive ability among the models. While there is better interpretability in Logistic Regression, Random Forest and XG-Boost are better in predictive ability but with greater complexity. The evaluation helps guide decision-makers in determining the model that is most appropriate for fulfilling their requirement, balancing precision with the ability to understand the driving forces in the consumption trends.

V. Results and Analysis

Model Performance Evaluation

a) Logistic Regression Modelling

The code snippet demonstrated how a scikit-learn model is trained and evaluated in Python. It begins with modules such as Logistic Regression for the model and metrics such as classification report, confusion matrix, and accuracy score for evaluation. The model is initialized with a random state parameter and a max_iter parameter with a value of 1000. The model is trained on the training dataset, X-train, and y-train, with the fit method. Predictions are made on the test dataset, X-test, with logreg. Predict. The performance of the model is evaluated with the imported metrics. The confusion matrix, classification report with precision, recall, and F1-score, and the global accuracy are printed to get a complete picture of the predictive power of the model.

Output:

Table 1: Logistic Regression Classification Report

```
Classification Report:
               precision    recall  f1-score   support

         0.0       0.29      0.18      0.22        61
         1.0       0.20      0.08      0.12        61
         2.0       0.37      0.65      0.47        78

    accuracy                           0.34       200
   macro avg       0.29      0.31      0.27       200
weighted avg       0.29      0.34      0.29       200

Accuracy Score: 0.335
```

The table shows the outcome of a Logistic Regression model with a confusion matrix and a classification report on how the model performs on a multi-class classification with three categories of energy consumption. The confusion matrix shows a high rate of misclassification in correctly identifying instances with category 0 having been misclassified 41 times and category 1 having been misclassified 5 times, suggesting a high level of difficulty in distinguishing these classes. The classification report shows precision, recall, and the F1-score for every category with category 0 having a low precision level of 0.0 and a low F1-score level of 0.22, implying ineffective recognition. Category 2 is better with a level of recall at 0.47 and a level of F1-score at 0.63, suggesting a relatively high level of recognition by the model. The model's level of accuracy at 0.335 shows only a prediction rate of about a third of the instances correctly, and there is a need to refine the model with a better method of classification and a need to explore a better and more viable method of modeling to classify correctly in all categories.

b) Random Forest Modelling

The code snippet illustrated how a Random Forest Classifier is trained and tested with scikit-learn in Python. It begins with the importation of the Random Forest Classifier in sklearn. Ensemble. The classifier is initialized with a random state to keep results consistent and with 100 estimators (trees in the forest). The model is trained on the train set, X-train, and y-train, with the fit method. Predictions are made on the test set, X-test, with rf_clf.predict. The performance of the model is then calculated with such metrics as the confusion matrix, classification report (with precision, recall, and F1-score), and the accuracy score.

Output:

Table 2: Random Forest Classifier Results

```
Classification Report:
              precision    recall  f1-score   support

         0.0       0.38      0.21      0.27        61
         1.0       0.30      0.30      0.30        61
         2.0       0.36      0.49      0.41        78

    accuracy                           0.34       200
   macro avg       0.35      0.33      0.33       200
weighted avg       0.35      0.34      0.34       200

Accuracy Score: 0.345
```

The table displays the results of a Random Forest Classifier with a classification report and a confusion matrix to evaluate performance in classification prediction based on categories of energy usage. The confusion matrix reveals high misclassification with category 0 having been classified in error 32 times and category 1 having been classified in error 18 times, suggesting a challenge in differentiation among these categories. The classification report reveals performance in all categories with category 0 having a precision rate and F1-score rate of only 0.38 and 0.27, respectively, suggesting low predictive power. Category 1 reveals improved performance with a rate of recall and F1-score rate of 0.30 and 0.30, respectively, while category 2 reveals improved recognition with a rate of recall and F1-score rate of 0.49 and 0.36, respectively. The total score rate of accuracy at 0.345 reveals the model is correctly categorizing approximately 34.5% of the instances, suggesting a need for refinement. The feature ranking in order of importance reported in the Random Forest model could provide insights into what features are significant in prediction, guiding future refinement and adjustment in the model to improve classification performance in all categories.

c) XG-Boost Modelling

The code fragment trains and tests an XG-Boost Classifier with the Python library XG-Boost. It begins with the importation of the XGB-Classifier. The classifier is initialized with a random state to reproduce, with the parameter use_label_encoder set False in order not to risk having label encoding problems, and with the evaluation parameter set 'logloss'. The model is trained on the training dataset with the fit method with the training dataset, X-train, and y-train. Predictions are made on the test set, X-test, with xgb_clf.predict. The performance of the model is evaluated with the confusion matrix, classification report with precision, recall, and F1-score, and with the accuracy score, and printed to have a global evaluation of the predictive ability of the XG-Boost Classifier.

Output:

Table 3: XG-Boost Classifier Results

```
Classification Report:
              precision    recall  f1-score   support

         0.0       0.28      0.28      0.28        61
         1.0       0.30      0.28      0.29        61
         2.0       0.37      0.40      0.38        78

    accuracy                           0.33       200
   macro avg       0.32      0.32      0.32       200
weighted avg       0.32      0.33      0.32       200

Accuracy Score: 0.325
```

The table presents the results of an XG-Boost Classifier, which serves as the best-performing model for forecasting peak consumption periods. The confusion matrix illustrates the model's performance across three categories, revealing that category 0 was misclassified 27 times, while category 1 received 18 incorrect predictions, indicating difficulties in accurately distinguishing these classes. The classification report provides critical metrics, showing category 0 with a precision and recall of 0.28, resulting in an F1-score of 0.28, reflecting its inadequate predictive performance. Category 1 performs slightly better, with a precision of 0.30 and an F1-score of 0.29, while category 2 achieves the highest recall of 0.37 and an F1-score of 0.40, suggesting it is more reliably identified during peak periods. The overall accuracy score of 0.325 indicates that the model correctly classifies only 32.5% of instances, highlighting the need for further enhancements. Despite being the best among the tested models, this performance underscores the necessity for additional tuning or the exploration of alternative approaches to improve the model's effectiveness in predicting peak consumption accurately.

Comparison of All Models

The implemented code script in Python program compared the performance of three machine learning models - Random Forest, Logistic Regression, and XG-Boost - based on some important evaluation metrics. It begins with the required library imports - Matplotlib and Seaborn for plotting and Pandas for data manipulation. The outcomes - the accuracy, precision, recall, and F1-score - are held in a dictionary model results, extracted from the classification reports of the models. The dictionary is then converted into a Pandas Data Frame results_df for easier handling and printing. The program prints the data frame to obtain a table view of the performance of the models. It plots a grid of Seaborn bar plots in a 2x2 layout where each subplot is a metric plotted across the three models. It is a straightforward means of comparing how effective the models are. The subplots are labeled and the title is given with enhanced visualization using the viridis palette.

Output:

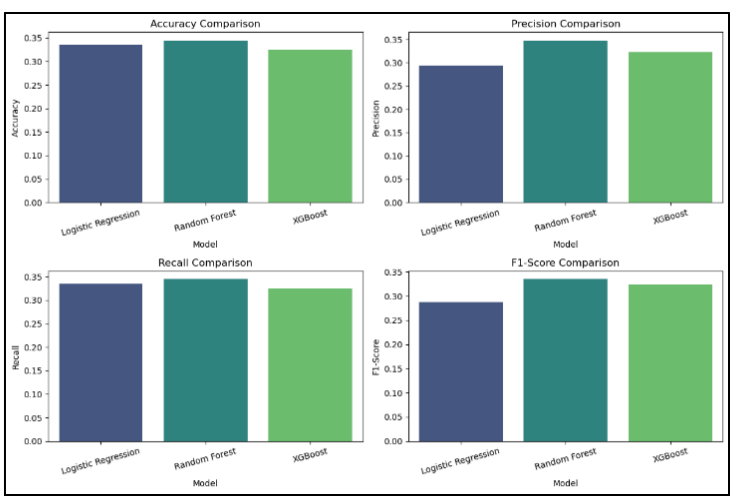

Figure 7:Depicts Model Comparison

The chart is a summary performance comparison among three machine models namely Logistic Regression, Random Forest, and XG-Boost with a special focus on forecasting times of highest consumption. The comparison on precision is such that there is low performance in all models with a score in Logistic Regression at 0.335, Random Forest at a score of 0.345, and a lower score in the case of XG-Boost at a score of 0.325. Despite these being relatively close in value, none of the models are predictive in a suitable way. The precision comparison is not unlike low precision in all models with none correctly labeled in the positive category. The comparison on recall is low in all models and is indicative of high false negative predictions. The balance on the F1-score is low in these models with none having a score greater than a score of 0.40. In general, the visualization is indicative of a need in the future for improved performance in models with improved feature engineering, improved tuning on the hyperparameter, or even experimenting with other models to get improved predictions on forecast times of highest consumption.

Policy and Sustainability Insight

Energy forecasting is a major in optimizing resource distribution in smart cities with a master plan for efficient power usage and sustainability. By relying on advanced predictive models, policymakers and municipal planners are in a position to forecast demand with greater precision. The fact that demand is forecast in advance makes provision and distribution of resources optimal with power generation and usage in sync with demand and wastage and grid instability minimized. An example is how precision forecasting is applied in informing generation timing and quantity from a variety of power generation resources to mitigate risks associated with power shortages in high usage times. Furthermore, forecasting facilitates better integration with power storage solutions because municipalities are in a position to identify the best times and durations in which power storage systems are charged and discharged based on forecast demand. By relying on data-driven approaches in power usage and distribution, smart cities not only maximize efficiency in their day-to-day operations but are also in a position to drive sustainability agendas such as greenhouse gas reduction and power resilience enhancement.

To maximize the benefits of energy forecasting, the incorporation of renewable resources in the power grid based on predictive models is important. Policymakers would need to prioritize the establishment of policies permitting the integration of renewables, such as wind and solar power, and could particularly suit cities seeking diversification away from fossil fuels. Predictive models are capable of identifying the best places and times where renewable generation is viable based on weather conditions, a record of current energy usage, and grid availability. If predictive models identify high chances of sunny weather, city planners are capable of incentivizing the placement of solar panels on buildings and commercial properties, harnessing the power of the sun while there is high sunlight. Furthermore, the incorporation of renewable resources involves having sound policies in support of grid modernization and smart technology implementation, such as smart grid infrastructure and demand response. These not only reinforce the potential of the city in harnessing fluctuating renewable generation but also involve the customer in the process of power management. By fostering collaboration among departments in the government, utility companies, and residents, cities are capable of designing a power system with high resilience and adaptability in meeting future issues.

VI. Practical Applications in the USA

Urban Energy Planning and Smart Grids

In major cities in the U.S. such as Los Angeles, Chicago, and New York, smart power forecasting based on AI is revolutionizing power distribution and power planning in cities. By utilizing advanced machine learning models, these cities can process vast amounts of information and predict power usage with high accuracy. It helps utility companies maximize power distribution in such a way that demand is fulfilled without overproduction and wastage is averted. In low-demand hours, excess power is either stored or redistributed, without the risk of power cuts or grid overload. Furthermore, smart grid construction is crucial to make power supply even more reliable. The smart grid is based on instantaneous information and automated systems to process power transmission, and in addition to reducing power loss, makes the power infrastructure even more secure. By incorporating forecasting based on AI in these systems, cities can better predict shifts in power demand, and generation and distribution strategies are formulated in seconds, resulting in a greener and more efficient power landscape in the city.

The incorporation of artificial intelligence (AI) in urban power planning has been a defining feature of modern-day power management in the USA. Major cities such as Los Angeles, Chicago, and New York are increasingly adopting AI-powered power forecasting technologies to rationalize power distribution. These technologies process vast amounts of information, varying from power usage trends in the past, weather forecasting, and grid performance in real-time, to predict power demand with precision. The utility companies are consequently in a position to allocate resources better, reduce operational costs, and eliminate the risk of power outages or power shortages. Moreover, the rollout of smart grids in the USA is getting a massive boost with the integration of AI technologies. Smart grids, reliant on high-end sensors, IoT technology, and machine learning models, enable two-way communication among power producers and power users. It ensures dynamic adjustment in power distribution, avoiding wastage and ensuring uniformity in power availability. For instance, in times of high demand, smart grids automatically reroute power where necessary, allowing a consistent and uninterrupted power supply to commercial and domestic users. The adaptation of such technologies is not only increasing power efficiency but also preparing the foundation for longer-lasting and power-efficient power infrastructure in cities.

Sustainable Infrastructure in the USA

The integration of machine learning insights in U.S. government-driven green construction is instrumental in driving sustainable construction in infrastructure. By utilizing data-driven approaches, policymakers are in a position to identify the optimal design methods and low-power technologies with high performance in buildings. Machine learning, for example, is capable of processing power usage and building information to guide insulation material selection, heating, and cooling systems, and integration with renewable power. The empirical research-driven strategy not only makes power-efficient construction possible but also helps in driving broader sustainability in cities.

Moreover, empowering local and state policymakers with the knowledge to design data-driven policies is instrumental in driving sustainability in societies. By utilizing predictive analytics, policymakers are in a position to better understand the power usage of their constituencies and design targeted policies aimed at power conservation, encouraging integration with renewable power, and enhancing awareness of sustainability efforts. The collaborative process in designing policies on power ensures optimal decision-making based on empirical information, and in turn, maximizes their efficiency and sustainability.

Impact on Integration with Renewable Energy

Leveraging machine learning in forecasting demand is crucial in optimizing the distribution of renewable resources, such as in California and Texas, where there are high concentrations of wind and solar resources. By forecasting accurately the trends in demand, utility firms are better positioned to accommodate the integration of renewable resources in the grid and balance generation and demand accordingly. An example is in sunny weather, where predictive models are capable of optimizing the usage of solar power and directing it where there is high demand. Supporting the clean transition envisioned under the Biden Administration with the power of AI-driven methods in managing power is a key area. By utilizing emerging technologies in grid upgrades, reducing reliance on fossil resources, and increasing access to clean resources, states are in a position to drive carbon-neutrality efforts substantially. These efforts not only drive country-wide sustainability efforts but also drive economic progress in the establishment of green sectors and jobs.

The transition to renewable power in the guise of wind and solar power is high on the agenda of the Biden Administration, and machine learning is playing a key role. California and Texas, among other major renewable power producers in the USA, are relying on ML to predict demand and balance wind and solar power distribution. For example, weather conditions are predicted with the assistance of ML models, and based on these predictions, the level of solar radiation and wind speed is calculated. Subsequently, grid managers are better positioned to balance demand and supply. It is a key consideration with renewable power, whose availability is weather-dependent. By relying on AI-driven energy management methods, the USA is taking a significant step in achieving clean power transition goals. These methods are in the guise of monitoring power grids in real-time, predictive power grid and other renewable power infrastructure maintenance, and dynamic pricing models rewarding users with lower costs to incentivize them to reschedule power usage away from peak times. In doing so, the USA is not only lowering reliance on fossil power but also optimizing power grid efficiency and stability.

Scalability in U.S.

Cities and States The implementation in high-power-consuming sectors such as commercial buildings and transportation brings scalability and efficiency advantages to U.S. states and cities. By tracking these sectors' information, decision-makers can identify trends and potential spots where energy could be saved, and can undertake targeted interventions to reduce consumption and costs. In commercial buildings, predictive models, for example, can inform building automation systems to modulate better on occupation, and in doing so, reduce usage substantially. In addition, the usage of AI in assisting power companies with demand-side management is crucial in decreasing peak loads and particular high-demand hours. By forecasting how and when usage is going to hit a high, utility companies can engage in demand response programs, encouraging users to curtail usage at such hours. Not only is pressure relieved on the grid but a culture is instated in users to instill longer-lasting efficient usage in the country. All these applications of predictive analytics and AI collectively are contributing to the building of a better-resilient, efficient, and sustainable future in power usage in cities and states in the U.S.

The scalability of AI-driven power solutions is a major consideration in their wide-scale implementation in U.S. states and cities. Predictive analytics with the power of AI is being utilized in high-power usage sectors such as commercial buildings and the transport segment to maximize power usage and efficiency and reduce costs. An example is in the usage of factory buildings and commercial buildings, where power usage is monitored and managed in real-time with the power of AI to maximize power usage. In a related way, the transport segment is benefiting from the power of AI-driven traffic management systems to reduce the usage of fuel and carbon emissions to maximize vehicle movement. On the demand side, power companies in the USA are implementing demand-side power usage with the power of AI, where power usage is shaped to reduce peak loads. It is achieved with methods such as the usage of time-of-use pricing, where power is utilized in low-demand hours, and demand response programs, where firms and residents voluntarily reduce power usage in high-demand hours. By scaling these solutions, the USA is not only optimizing efficiency in power usage but also creating a more adaptive and secure power system responsive to the evolving demands of the country's cities and states.

VII. Discussion and Future Directions in the USA

Challenges in AI-Driven Energy Forecasting in the USA

AI-driven energy forecasting in the U.S. is hindered by a variety of issues, with variability in consumption in different parts of the country owing to differences in weather. The Northeast would have a harsh winter and, accordingly, increased winter demand for warmth, while the Southwest would have increased cooling demand in summer. The variability in these consumption patterns makes developing a uniform forecasting model challenging, with what is applicable in a particular location not necessarily applicable in other locations. It is also necessary to eliminate the presence of biases in the consumption datasets, with these collected under varying state policies and regulations. Those with strict policies on energy efficiency would record lower usage, while the weaker policies in the states would not necessarily record the usage. These differences would result in inexact predictions and hinder the performance of the AI models, and strict consideration of the differences in the location and quality in forecasting is called for.

Limitations of the Study

This study is not without limitations, with most arising due to limitations in available data in smaller U.S. cities and rural areas. The fact that there is not much comprehensive information on power consumption in these locations tends to limit the generalizability of AI models because they would not necessarily portray the unique consumption patterns and power loads in these thinly settled places. Furthermore, integration with feeds in real-time from sensors in the Internet of Things (IoT) is a potential constraint. While models based on AI are quite good with existing data, feeds in real-time from sensors in the IoT would fine-tune the accuracy and reaction times in power forecasting. Without integration with such feeds, models would not necessarily handle sudden power demand and supply shifts, limiting their performance in a dynamic system. Overcoming these constraints would be crucial in creating efficient and responsive power forecasting models applicable to the diverse U.S. landscape in cities and locations.

Future Prospects

In the USA Future research potential in USA AI-powered energy forecasting is enormous and could significantly enhance the precision and resilience of the power system. One such area is research in deep learning techniques in obtaining better longer-run power demand forecasting in cities. Deep learning models with their ability to manage complex information could pick out subtle power usage patterns not perceivable with traditional methods. Another area is collaboration with the U.S. Department of Energy and major utility companies in finding a method to enhance power grid resilience with the potential of AI. These collaborations could witness the provision and access to resources and information with the design and implementation of improved models with a better response to instantaneous power demand. Moreover, research in policy-driven models based on AI could witness the design and implementation of a better and more efficient power distribution system based on the specific demand and requirement in a location. By focusing on these research fronts, concerned parties could move in the direction of a greener and more secure power future based on state-of-the-art technology and information-driven insights.

VIII. Conclusion

This research aims to develop advanced machine learning models with the potential to effectively predict energy consumption in cities. This involves not only identifying the key variables behind energy consumption but also selecting and fine-tuning machine learning algorithms that are most capable of understanding the dynamics of urban energy intricacies. This study focuses on the energy consumption patterns in the large cities of the United States, recognizing the diversity of challenges and opportunities presented by different geographic and demographic situations. The dataset used in this research project offered a comprehensive view of energy consumption across various fields of household, commercial, and industrial consumption, giving a holistic view of energy dynamics within cities. It integrated data collected from smart meters that offer granular electricity consumption patterns at the level of individual households and businesses with weather reports that detail ambient conditions governing energy demand, such as temperature and humidity fluctuations. Government energy records add historical context and policy information, further enhancing the dataset and enabling close analysis of trends and patterns in energy consumption. The next phase was to select and train three distinct machine models to explore the energy consumption dataset, namely, Logistic Regression, Random Forest, and XG-Boost algorithms. Random Forest outperformed Logistic Regression and XG-Boost slightly in terms of accuracy and other evaluation metrics. However, all models exhibit relatively low accuracy, suggesting the need for further tuning, feature engineering, or alternative models to improve predictions. In major cities in the U.S. such as Los Angeles, Chicago, and New York, smart power forecasting based on AI is revolutionizing power distribution and power planning in cities. By utilizing advanced machine learning models, these cities can process vast amounts of information and predict power usage with high accuracy. The incorporation of artificial intelligence (AI) in urban power planning has been a defining feature of modern-day power management in the USA. Major cities such as Los Angeles, Chicago, and New York increasingly adopt AI-powered power forecasting technologies to rationalize power distribution. The integration of machine learning insights in U.S. government-driven green construction is instrumental in driving sustainable construction in infrastructure. By utilizing data-driven approaches, policymakers are in a position to identify the optimal design methods and low-power technologies with high performance in buildings.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Abbass, M., Abbas, U., Jafri, R., Arif, S. M., & Akhai, S. (2025). AI and Machine Learning Applications in Sustainable Smart Cities. In Sustainable Smart Cities and the Future of Urban Development (pp. 1-32). IGI Global Scientific Publishing.

[2] Ahmed, A., Jakir, T., Mir, M. N. H., Zeeshan, M. A. F., Hossain, A., hoque Jui, A., & Hasan, M. S. (2025). Predicting Energy Consumption in Hospitals Using Machine Learning: A Data-Driven Approach to Energy Efficiency in the USA. *Journal of Computer Science and Technology Studies*, *7*(1), 199-219.

[3] Al Mukaddim, A., Mohaimin, M. R., Hider, M. A., Karmakar, M., Nasiruddin, M., Alam, S., & Anonna, F. R. (2024). Improving Rainfall Prediction Accuracy in the USA Using Advanced Machine Learning Techniques. *Journal of Environmental and Agricultural Studies*, *5*(3), 23-34.

[4] Anonna, F. R., Mohaimin, M. R., Ahmed, A., Nayeem, M. B., Akter, R., Alam, S., ... & Hossain, M. S. (2023). Machine Learning-Based Prediction of US CO2 Emissions: Developing Models for Forecasting and Sustainable Policy Formulation. *Journal of Environmental and Agricultural Studies*, *4*(3), 85-99.

[5] Barua, A., Karim, F., Islam, M. M., Das, N., Sumon, M. F. I., Rahman, A., ... & Khan, M. A. (2025). Optimizing Energy Consumption Patterns in Southern California: An AI-Driven Approach to Sustainable Resource Management. *Journal of Ecohumanism*, *4*(1), 2920-2935.

[6] Bansod, M. (2025). Advanced Predictive Analytics Driving Sustainable and Equitable Green Innovations Through Data-Driven Optimization and Strategic Insights. In Advancing Social Equity Through Accessible Green Innovation (pp. 1-16). IGI Global Scientific Publishing.

[7] Chowdhury, M. S. R., Islam, M. S., Al Montaser, M. A., Rasel, M. A. B., Barua, A., Chouksey, A., & Chowdhury, B. R. (2024). PREDICTIVE MODELING OF HOUSEHOLD ENERGY CONSUMPTION IN THE USA: THE ROLE OF MACHINE LEARNING AND SOCIOECONOMIC FACTORS. *The American Journal of Engineering and Technology*, *6*(12), 99-118.

[8] Cui, X., Lee, M., Uddin, M. N., Zhang, X., & Zakka, V. G. (2025). Analyzing different household energy use patterns using clustering and machine learning. Renewable and Sustainable Energy Reviews, 212, 115335.

[9] Ekeh, A. H., Apeh, C. E., Odionu, C. S., & Austin-Gabriel, B. (2025). Leveraging machine learning for environmental policy innovation: Advances in Data Analytics to address urban and ecological challenges. Gulf Journal of Advance Business Research, 3(2), 456-482.

[10] Ghallabi, F., Souissi, B., Du, A. M., & Ali, S. (2025). ESG stock markets and clean energy prices prediction: Insights from advanced machine learning. International Review of Financial Analysis, 97, 103889.

[11] Gazi, M. S., Barua, A., Karim, F., Siddiqui, M. I. H., Das, N., Islam, M. R., ... & Al Montaser, M. A. (2025). Machine Learning-Driven Analysis of Low-Carbon Technology Trade and Its Economic Impact in the USA. *Journal of Ecohumanism*, *4*(1), 4961-4984.

[12] Gera, R., Banerjee, S., Saratchandran, D. V., Arora, S., & Whig, A. (2025). Machine Learning for Environmental Sustainability in the Corporate World. In Driving Business Success Through Eco-Friendly Strategies (pp. 283-302). IGI Global Scientific Publishing.

[13] Hasan, M. R., Shawon, R. E. R., Rahman, A., Al Mukaddim, A., Khan, M. A., Hider, M. A., & Zeeshan, M. A. F. (2024). Optimizing Sustainable Supply Chains: Integrating Environmental Concerns and Carbon Footprint Reduction through AI-Enhanced Decision-Making in the USA. *Journal of Economics, Finance and Accounting Studies*, *6*(4), 57-71.

[14] Hasan, M. R. (2024). Revitalizing the electric grid: A machine learning paradigm for ensuring stability in the USA. *Journal of Computer Science and Technology Studies*, *6*(1), 141-154.

[15] Hossain, S., Hasanuzzaman, M., Hossain, M., Amjad, M. H. H., Shovon, M. S. S., Hossain, M. S., & Rahman, M. K. (2025). Forecasting Energy Consumption Trends with Machine Learning Models for Improved Accuracy and Resource Management in the USA. *Journal of Business and Management Studies*, *7*(1), 200-217.

[16] Hossain, M. S., Mohaimin, M. R., Alam, S., Rahman, M. A., Islam, M. R., Anonna, F. R., & Akter, R. (2025). AI-Powered Fault Prediction and Optimization in New Energy Vehicles (NEVs) for the US Market. *Journal of Computer Science and Technology Studies*, *7*(1), 01-16.

[17] Madabathula, C. T., Agrawal, K., Mehta, V., Kasarabada, S., Kommamuri, S. S., Liu, G., & Gao, J. (2025). Smart Green Energy Management for Campus: An Integrated Machine Learning and Reinforcement Learning Model. Smart Cities, 8(1), 30.

[18] Marella, B. C. C., & Palakurti, A. (2025). Harnessing Python for AI and Machine Learning: Techniques, Tools, and Green Solutions. In Advancing Social Equity Through Accessible Green Innovation (pp. 237-250). IGI Global Scientific Publishing.

[19] Mohammed, M., Mohammed, A., Ihmedee, F., Adam, T., Betar, B., & Gopinath, S. (2025). Emerging artificial intelligence methods in civil engineering: A Comprehensive Review. Al-Rafidain Journal of Engineering Sciences, 280-293.

[20] Ogunseye, O. O., Ajayi, O. T., Fabusoro, A., Abba, A. O., & Adepoju, B. (2025). Leveraging Artificial Intelligence for Advancing Key Sectors of National Growth and Development. Asian Journal of Current Research, 10(1), 45-55.

[21] Palaniappan, S., Logeswaran, R., Khanam, S., & Yujiao, Z. (2025). Machine learning model for predicting net environmental effects. Journal of Informatics and Web Engineering, 4(1), 243-253.

[22] Reza, S. A., Chowdhury, M. S. R., Hossain, S., Hasanuzzaman, M., Shawon, R. E. R., Chowdhury, B. R., & Rana, M. S. (2024). Global Plastic Waste Management: Analyzing Trends, Economic and Social Implications, and Predictive Modeling Using Artificial Intelligence. *Journal of Environmental and Agricultural Studies*, *5*(3), 42-58.

[23] Robi, R. K., & George, J. K. (2025). Application of Machine Learning Algorithms to Predict Urban Expansion. Journal of Urban Planning and Development, 151(2), 03125001.

[24] Sahar, S., Iqbal, Z., Mahtab, M., Zahra, S. A., & Azam, M. (2025). PREDICTIVE ANALYSIS OF ENERGY CONSUMPTION PATTERNS USING MACHINE LEARNING TECHNIQUES. Contemporary Journal of Social Science Review, 3(1), 517-529.

[25] Satpathy, I., Nayak, A., & Jain, V. (2025). The green city: Sustainable and smart urban living through artificial intelligence. In Utilizing Technology to Manage Territories (pp. 273-304). IGI Global.

[26] Shil, S. K., Chowdhury, M. S. R., Tannier, N. R., Tarafder, M. T. R., Akter, R., Gurung, N., & Sizan, M. M. H. (2024). Forecasting Electric Vehicle Adoption in the USA Using Machine Learning Models. *Journal of Computer Science and Technology Studies*, *6*(5), 61-74.

[27] Sumon, M. F. I., Osiujjaman, M., Khan, M. A., Rahman, A., Uddin, M. K., Pant, L., & Debnath, P. (2024). Environmental and Socio-Economic Impact Assessment of Renewable Energy Using Machine Learning Models. *Journal of Economics, Finance and Accounting Studies*, *6*(5), 112-122.

[28] Sumon, M. F. I., Rahman, A., Debnath, P., Mohaimin, M. R., Karmakar, M., Khan, M. A., & Dalim, H. M. (2024). Predictive Modeling of Water Quality and Sewage Systems: A Comparative Analysis and Economic Impact Assessment Using Machine Learning. *in Library*, *1*(3), 1-18.

[29] Sumsuzoha, M., Rana, M. S., Islam, M. S., Rahman, M. K., Karmakar, M., Hossain, M. S., & Shawon, R. E. R. (2024). LEVERAGING MACHINE LEARNING FOR RESOURCE OPTIMIZATION IN USA DATA CENTERS: A FOCUS ON INCOMPLETE DATA AND BUSINESS DEVELOPMENT. *The American Journal of Engineering and Technology*, *6*(12), 119-140.

[30] Vasenin, D. (2025). Machine Learning Techniques for Ensuring the Health of Citizens and the Environmental Sustainability of Buildings. Chicago: Sage Print