Journal of Environmental and Agricultural Studies
ISSN: 710-1401
DOI: 10.32996/jeas
Journal Homepage: www.al-kindipublisher.com/index.php/jeas

JEAS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Harnessing Machine Learning to Analyze Energy Generation and Capacity Trends in the USA: A Comprehensive Study

Anchala Chouksey[1], Md Shihab Sadik Shovon[2], Md Raisul Islam[3], Bivash Ranjan Chowdhury[4], Mehedi Hasan Ridoy[5], Md Anisur Rahman[6], and Mohammad Hamid Hasan Amjad[7]

[1]Masters of Science in Financial Mathematics, University of North Texas, Denton, Texas
[2]Master of Science in Engineering Management, College of Graduate and Professional Studies, Trine University.
[3]Engineering Technology, Western Illinois University, Macomb, IL-61455
[4]MBA in Management Information System, International American University, Los Angeles, California, USA.
[5]MBA- Business Analytics, Gannon University, Erie, PA, USA.
[6]Engineering Technology, Western Illinois University, Macomb, IL-61455
[7]Master of Science in Engineering Management, College of Graduate and Professional Studies, Trine University.
Corresponding Author: Anchala Chouksey, **E-mail**: Anchalachouksey@my.unt.edu

| ABSTRACT
The constraints in conventional energy forecasting models in the USA are increasingly being appreciated as the energy landscape evolves. Such models are generally constructed on linear assumptions that do not capture energy generation and consumption patterns as dynamic and multi-dimensional. This research project aimed to develop effective machine-learning models to interpret historical trends in energy generation and capacity in the United States. With access to massive datasets for various energy sources—fuel-based to renewable and nuclear energy—this research endeavors to discover actionable information that can optimize energy output and improve grid stability. The dataset used is a comprehensive overview of energy production in America, integrating information from a variety of sources, ranging from fossil fuels to renewables and nuclear energy. This extensive dataset is sourced from reliable sources such as the U.S. Energy Information Administration (EIA) which supplies vast historical and current data on energy consumption and production patterns; the Federal Energy Regulatory Commission (FERC) which supplies information on regulatory frameworks and market forces that affect energy production; and smart grid management systems that deliver real-time data on energy flows and grid performance. Selecting appropriate machine learning models is crucial to process engineered features and derive actionable insights effectively. For this purpose, we use Logistic Regression, Random Forest Classifier, and the Support Vector Machines. The model performance table indicates how well three machine learning algorithms—Logistic Regression, Random Forest, and SVM—classify between renewable and non-renewable energy sources. All three models are highly accurate, with SVM having a slight edge, followed by Random Forest and lastly Logistic Regression. AI-powered energy forecasting is revolutionizing energy infrastructure planning in America through the provision of sophisticated information to guide investment in grid upgrades and expansion. As America transitions to dependence on renewable energy, AI-based forecasting is essential to optimize the integration of solar and wind energy production onto the grid. The variability of these renewable resources poses special challenges for grid managers, who need to balance demand and supply in real time. Moreover, AI can be used to identify gaps in renewable energy capacity and storage. The infusion of AI-based insights into the energy sector has far-reaching implications for U.S. policymakers, equipping them with fact-based tools to efficiently apply energy reforms. With evolving energy patterns, legislators need to adjust regulations to facilitate a seamless transition to a more sustainable energy system.

| KEYWORDS

Energy Generation, Machine Learning, Renewable Energy, Capacity Forecasting, Grid Optimization

| ARTICLE INFORMATION

ACCEPTED: 01 March 2025 **PUBLISHED:** 11 March 2025 **DOI:** 10.32996/jeas.2025.6.1.2

I. Introduction

Background and Background

According to Hasan (2024), the increase in electricity demand in America is a multi-dimensional challenge driven by a surge in population, economic growth, and a need for convenience in modern society. America is grappling with the twin challenge of meeting this increased demand and, at the same time, converting to cleaner sources of energy. Energy generation and capacity planning complexities are highlighted in America's challenge. The traditional energy sector, with a heavy reliance on fossil fuels, is transforming alternative energy sources such as solar, wind, and hydroelectric power to gain traction. This transformation has not been smooth, however; energy generation and capacity planning balance and grid reliability are crucial matters that have to be tackled with innovative strategies. Chowdhury et al. (2024), reported that merging alternative energy with the established grid system is a challenge in itself, with differences in generation patterns and a need for more advanced storage systems to balance demand and supply.

Increasing electricity consumption in the US is driven by a variety of factors ranging from technological advancements to population growth and electrification in sectors like transport. With the nation transitioning to a more sustainable energy system, with greater utilization of solar, wind, and hydroelectric power, energy generation and capacity planning are compounded (Barua et al. 2025). Energy companies and policymakers are tasked with reconciling multiple energy sources with reliability and resilience in integrating them into the grid system. This is important in light of most energy sources being intermittent and a stable power output is required to keep up with shifting demand patterns (Anonna et al., 2023).

Furthermore, Ahmed et al. (2025), indicated that traditional energy generation and capacity forecasting techniques have sometimes failed to capture nonlinearities and complexities in modern energy systems. Historical data often have intricate patterns due to multiple factors like weather patterns, regulation changes, and economic factors, which cannot be captured well using traditional models. Thus, there is a great demand for improved, AI-based forecasting methods to enhance decision-making in energy infrastructure planning and management. Gazi et al. (2025), posited that employing machine learning techniques is a promising method to enhance energy forecast accuracy and reliability and facilitate effective strategic planning for the future.

Problem Statement

As per Rahman et al. (2025), the constraints in conventional energy forecasting models are increasingly being appreciated as the energy landscape evolves. Such models are generally constructed on linear assumptions that do not capture energy generation and consumption patterns as dynamic and multi-dimensional. They can thus give inexact predictions, and demand and supply mismatches can occur, threatening grid stability. With policymakers and energy regulators under pressure to facilitate a transition to cleaner energy sources, constraints in conventional forecasting methods are a serious stumbling block to sustainable energy targets. There is therefore a critical need to explore and develop advanced analytical techniques that can leverage the strengths of machine learning to capture past energy generation patterns and predict future capacity demands (Sumon et al., 2024).

Hossain et al. (2025b), ascertained that traditional energy forecasting models have a hard time accurately projecting energy generation trends due to their complexity and nonlinear nature. These models typically rely on historical data and linear assumptions that do not capture dynamic interactions between energy sources, variations in demand, and external factors such as weather and economic conditions. This creates an imperative for more accurate, AI-based forecasting methods that can match changing energy trends and help enable the U.S. energy system. With its ability to handle large sets and find intricate patterns, machine learning is a viable solution for correcting such issues. By taking advantage of machine learning capabilities, stakeholders can develop more robust forecasting models that enhance decision-making and overall grid stability (Al Mukaddim et al., 2024).

Research Objective

This research project aims to develop effective machine-learning models to interpret historical trends in energy generation and capacity in the United States. With access to massive datasets for a variety of energy sources—fuel-based to renewable and nuclear energy—this research endeavors to discover actionable information that can optimize energy output and improve grid stability. Ultimately, the end goal is to develop a framework that optimizes not just energy forecast accuracy but also strategic decision-making for a variety of stakeholders ranging from policymakers to energy providers and regulators. With state-of-the-art machine learning algorithms, this research aims to contribute to energy transition and sustainability in America.

With advanced analytical techniques, this research intends to provide information that can optimize energy production and improve grid stability. More specifically, this research will focus on various energy sources, like fossil fuels, renewables, and nuclear energy, and how they contribute to America's energy mix. With this analysis, this research will search for significant trends and patterns that can inform policymakers and energy suppliers to make decisions to enhance America's energy grid reliability and sustainability.

Relevance and Scope

The scope of this research entails an in-depth exploration of various energy sources utilized across the United States, with a specific emphasis on fossil fuels, renewable energy, and nuclear energy. Employing machine learning techniques, this research strives to establish a sophisticated understanding of energy generation patterns and capacity trends that can inform future energy policies and infrastructural investments. Beyond theoretical significance, this research has real-world applicability in terms of energy production process maximization, grid performance enhancement, and enabling energy providers to meet changing consumers' needs. With a nation grappling with climate change issues and reducing greenhouse gas emissions, knowledge derived from this research will be instrumental in informing policies that promote a resilient and sustainable energy future. Ultimately, through bridging sophisticated analytical techniques and real energy issues, this research aims to enable stakeholders to manage effectively the complexities in the U.S. energy sector.

II. Literature Review

Trends in Energy Generation in the USA

Ahmad et al. (2022), postulated that the energy production scene in America has evolved remarkably in recent decades, with historical trends in fossil fuel, nuclear, and alternative energy production. In the mid-20th century, fossil fuels, namely coal and gas, dominated energy production, accounting for a majority percentage of overall electricity. Nuclear energy in the 1970s was introduced as a replacement for fossil fuels, with promises of cleaner energy production and lower greenhouse gas emissions. However, in later years, public opinion towards nuclear energy was wavering, particularly following big events such as the 1979 Three Mile Island accident and the 1986 Chornobyl disaster, leading to a slowdown in new nuclear plant construction (Goel et al., 2025). In contrast, in the early 21st century, there was a turning point towards alternative energy sources due to technological advancements, cost reduction, and greater awareness about climate change. Solar and wind energy have seen exponential growth with government policies from both state and federal governments to promote clean energy usage. Legislative measures such as the Energy Policy Act of 2005 and the American Recovery and Reinvestment Act of 2009 introduced tax credits and financing for alternative energy projects, further boosting a departure from traditional fossil fuels. Thus, the U.S. energy mix has diversified increasingly, with alternative energy accounting for an ever-increasing percentage of overall electricity output, a reflection of a collective commitment to sustainable energy consumption (Perera et al., 2024).

According to Karimi et al. (2024), policy-based changes in clean energy adoption have been crucial in revolutionizing the energy production sector. State-level Renewable Portfolio Standards and federal initiatives such as the Clean Power Plan have encouraged utility companies to invest in alternative energy and decommission coal-fired power plants. These policies have not only fueled solar and wind energy expansion but have facilitated the development of new technologies such as energy storage and smart grid systems. This interplay between government policies, market forces, and technological innovation has resulted in a more robust energy system better poised to serve a modern economy. Moreover, public sentiment has increasingly favored renewable energy, with public opinion polls indicating a preference for green energy systems over conventional fossil fuel systems. As per Rehan (2023), this cultural shift, coupled with economic benefits, has pushed America towards a cleaner energy future, although there are still challenges in fully integrating these alternative sources into the present power grid.

Capacity Planning and Grid Management Issues

As the energy generation industry evolves, capacity management and grid management present a challenging set of problems to be addressed to deliver a stable and efficient power supply. Balancing power generation demand and supply is at the heart of these problems with a greater proportion of intermittent alternative energy sources (Laturkar, 2024). Whereas traditional fossil fuel-based power stations can deliver electricity on demand, alternative energy sources such as solar and wind are weather and time-dependent. This unpredictability in power generation from alternative energy sources makes matching electricity demand from consumers with electricity supply a challenging process, which is highly variable throughout the day and season (Rizvi, 2023). This creates a daunting challenge for grid managers to balance grid stability with incorporating diverse energy sources. Advanced forecasting techniques and real-time monitoring systems have been utilized to predict demand patterns and realign energy production to match. These interventions, however, are costly to implement in terms of investment in infrastructure and technology and can pose a challenge to utilities and regulators (Shen et al., 2024).

Ukoba et al. (2024), found that regulatory frameworks have a significant role to play in energy capacity planning. Renewable energy adoption policies need to include considerations for grid infrastructure and capacity management. Regulatory bodies need to create guidelines that balance grid reliability with investments in renewable technologies. Complexity in such a framework can lead to inconsistencies and implementation challenges. For instance, interconnection procedures for renewable projects can vary substantively from state to state, making entry for new energy suppliers challenging. Additionally, a lack of standardized guidelines for managing grids nationally can hinder efforts to create an integrated and resilient energy system. Moreover, Yan et al. (2021), argued that a decentralized energy model with distributed generation and localized energy systems requires a rethink in traditional capacity planning. With the increased use of solar panels and energy storage systems, energy supply and demand patterns are being transformed, and new approaches to grid management and capacity forecasting are required.

Machine Learning in Energy

As per Ahmad et al. (2022), forecasting In energy production and capacity planning, machine learning has been a game-changer, opening new frontiers for predictive modeling approaches that can enhance forecast accuracy and optimize energy management. The application of machine learning algorithms in energy has grown in popularity due to their ability to handle large sets of data and find complex patterns that traditional statistical methods cannot. Predictive modeling approaches such as regression analysis, time series forecasting, and neural networks have been applied to improve demand forecasting, grid optimization, and load balancing. Such models can analyze historical energy usage data, weather patterns, and economic variables to give more precise predictions, and hence allow utilities to better decide on energy production and distribution (Arya, 2024).

According to Borunda et al. (2022), demand forecasting applications for artificial intelligence demonstrate how energy management processes can be revolutionized using machine learning. For instance, machine learning algorithms can be utilized to predict peak hours of demand with greater accuracy, and subsequently adjust generation strategies. Machine learning can also assist with grid optimization using real-time energy flow control and monitoring throughout the grid. Using data from smart meters and IoT sensors, machine learning algorithms can identify inefficiencies and give recommendations for making the overall grid more efficient. Chen (2024), reported that load balance is also optimized using machine learning since variations in energy demand can be predicted and output adjusted to match demand perfectly. These applications not only improve efficiency in operations but can also facilitate the integration of renewable energy sources into the grid, thereby improving sustainability and resilience.

Research Gaps

Duong et al. (2024), hold that despite promising advancements in machine learning applications in the energy sector, there exist vast research gaps, in particular concerning machine learning for capacity forecasting in the United States. While some studies have explored using machine learning for demand forecasting and grid optimization, there is a notable lack of research in capacity planning per se. This is indicative of a need for more research on how machine learning can be used to develop predictive models that solve specific problems in capacity forecasting. Additionally, extant research does not address the dynamic interactions between weather patterns, variations in demand, and policy interventions, which are significant drivers in energy generation and consumption (Fan et al., 2023).

To address these gaps, future research will have to develop AI-based models that integrate these multi-dimensional variables into capacity forecast models. By constructing integrated models that reflect the complexities of energy consumption and generation, researchers can provide actionable data that allows policymakers and energy companies to make informed decisions (Hanif & Mi, 2024). Furthermore, interdisciplinary collaboration between data scientists, energy economists, and policymakers will be required to ensure that machine learning deployments are best suited to meet specific U.S. energy sector demands. With the transformation to a more sustainable energy future accelerating, bridging these research gaps will be crucial in equipping stakeholders with tools to manage energy generation and capacity planning in a changing world (Husnain et al., 2024).

III. Data Collection and Exploration

The dataset used is a comprehensive overview of energy production in America, integrating information from a variety of sources, ranging from fossil fuels to renewables and nuclear energy. This extensive dataset is sourced from reliable sources such as the U.S. Energy Information Administration (EIA) which supplies vast historical and current data on energy consumption and production patterns; the Federal Energy Regulatory Commission (FERC) which supplies information on regulatory frameworks and market forces that affect energy production; and smart grid management systems that deliver real-time data on energy flows and grid performance. By combining all these different datasets, we aim to achieve a total understanding of energy production, facilitating in-depth analysis and informed decision-making for energy production optimization and capacity planning.

Feature Selection

S/No.	Feature/Attribute	Description
001.	**Energy Source**	A categorical feature indicating what energy source is being used for generation, e.g., fossil fuels (coal, gas, oil), renewables (solar, wind, hydro), or nuclear.
002.	**Actual Capacity (MW)**	A continuous measure representing a power plant's maximum output capacity in megawatts (MW).
003.	**General Capacity (MW)**	A continuous measure representing actual electricity generated in a specified period, in megawatt-hours (MWh).
004.	**Operating Hours**	A measure that shows how many hours a power plant is operating over a given period (e.g., a day, a month).
005.	**Fuel Cost**	A measure on a continuous scale expressing fuel cost for energy generation, typically in dollars per MMBtu (million British thermal units).
006.	**Levels of Emissions**	Continuous measures of levels of greenhouse gases (e.g., CO2) and nitrogen oxides (NOx) from power generation facilities, typically in metric tons.
007.	**Weather Conditions**	Categorical or continuous variables for weather factors that are pertinent to the generation peri
008.	**Load Demand (MWh)**	A continuous attribute indicating total electricity demand in a region or grid over a given period.
009.	**Location**	A categorical attribute is used to denote the geographical location of the power generation plant, e.g., state or region.

Data Preprocessing

The implemented code snippet demonstrated a typical data preprocessing workflow in Python using Pandas and Scikit-learn libraries. It began with importing modules and printing initial statistics regarding the dataset, i.e., data types and missing values. This procedure was followed by missing value imputation, imputing numerical columns with median and categorical columns with mode. The code then handled categorical variables with the target variable 'Energy_ encoded with Label Encoder and other categorical variables encoded with one-hot encoding. Scaling features is done with numerical columns standardized with Standard Scaler. Finally, the dataset is split into a training set and a test set using train_test_split with a specified test size, random state for reproducibility, and stratify to balance classes. The shape of the resulting datasets is then printed.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a crucial process in data analysis that entails describing and analyzing a dataset's salient features using visual and numerical methods to uncover patterns, outliers, and insights. EDA's principal goal is to gain a better understanding of data, and this can encompass assessing distributions, uncovering relationships between variables, detecting outliers, and identifying missing or erroneous data. By using statistical methods and graphical tools such as histograms, scatter plots, box plots, and correlation matrices, EDA provides a foundation for hypothesis generation and model selection and eventually guides further processes in data preprocessing, feature construction, and predictive modeling. This interactive and iterative process allows analysts to make informed decisions on data transformation and analysis strategies in a way that final models are sound and well-supported by underlying data structures.

a) Total Energy Generation/Capacity Trends

The code snippet is utilized to draw a line plot to show the trend in total energy generation/capacity from 2008 to 2023. It starts with establishing a consistent visual style with Seaborn and making settings like font size and figure size consistent. It then calculates total energy for every year by summing column values from 'F2000' to 'F2023' in a given data frame named data. Years are obtained from column names, and a line plot is drawn using matplotlib. Years are placed along the x-axis and total energy is along the y-axis with a marker for each point. Labels, titles, and grids are added for readability, and x-axis ticks are rotated for better readability. Finally, plt.show() is utilized to display the plot.

Output:

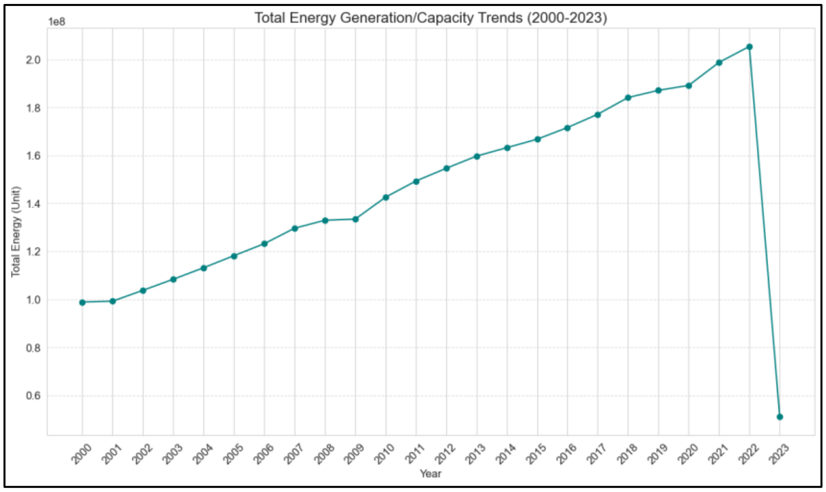


***Figure 1**: Total Energy Generation/Capacity Trends*

The graph representing total energy generation and capacity trends from 2000 to 2023 demonstrates striking variations in energy output over the years with a consistent increase in total energy generation from approximately 0.6 billion units in 2000 to a high level of approximately 2 billion units in 2022. This upward trend indicates a consistent boost in energy production capacity with technological advancements and greater utilization of renewable energy resources. However, the sharp drop in 2023 indicates a sharp decline in total energy generation, which may be due to a variety of factors such as changes in regulations, economic crises, or unforeseen disruptions in energy supply chains. This aberration indicates the vulnerabilities in energy production and capacity planning and highlights the imperative to further scrutinize factors responsible for such variations and their implications for future energy policies. Overall, the graph demonstrates the dynamic nature of the energy sector and calls for constant monitoring to ensure stability and sustainability in energy production.

b) Energy Generation by Technology

The implemented code plots a stacked area chart to show energy generation trends with different technologies from 2008 to 2023. It starts with aggregating data by 'Technology' and summing energy values for each year (ranging from 'F2000' to 'F2023'). It then retrieves years from column names. It then plots a stacked area chart using plt. Stack plot with years on the x-axis and the total energy values for each technology as the components of the stacked areas. Technology names are used as labels for the legend. The plot is titled "Energy Generation by Technology (2008-2023)" and has axes labeled. A legend is plotted outside the plot area for readability and the plot is displayed using plt.show(). Alpha is set to 0.8 to provide a slight transparency to the areas.

Output:

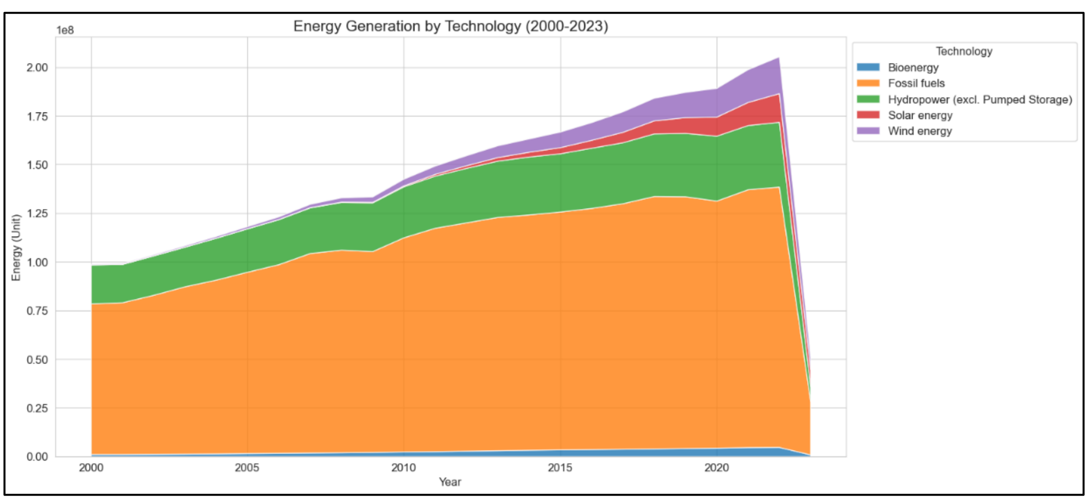


Figure 2: Energy Generation by Technology

The graph above illustrates energy production through technology from 2000 to 2023 and shows the transforming face of energy production in America, with varied energy sources contributing to the energy mix over time. Initially, fossil fuels led in the energy mix, but their share has gradually declined as renewable technologies have gained popularity, with solar and wind energy being good examples, which have seen a sharp increase since 2010. By 2023, solar energy has emerged as a significant contributor, indicating a successful shift to cleaner energy sources. Additionally, the line for hydropower is relatively stable, with bioenergy having slight variations. The sharp decline in overall energy production seen in 2023, particularly for fossil fuel contributions, is indicative of probable disruptions in the sector, possibly due to a shift in regulations or market forces in favor of renewable energy. This graph not only shows the increased share of renewable technologies in the energy mix but also indicates the need for ongoing innovation and investment in green energy technologies to meet future demands and environmental goals.

c) Missing Data Heatmap

The executed code produces a heatmap to present missing data in columns 'F2000' to 'F2023' in a Data Frame named data. It starts with a figure with a specified size. It then uses seaborn. Heatmap to plot the heatmap. A 16oolean Data Frame indicating whether or not each cell in specified columns is null is passed to the heatmap. The cbar is set to False to hide the color bar, and cmap is set to 'Viridis' to define color. The plot is titled "Missing Data Heatmap (2000-2023)", and the axes are titled 'Year' and 'Record Index'. Finally, plt.show() is called to display the plot. This is a helpful visualization for easily identifying patterns and locations of missing data across specified years in a datase'.

Output:

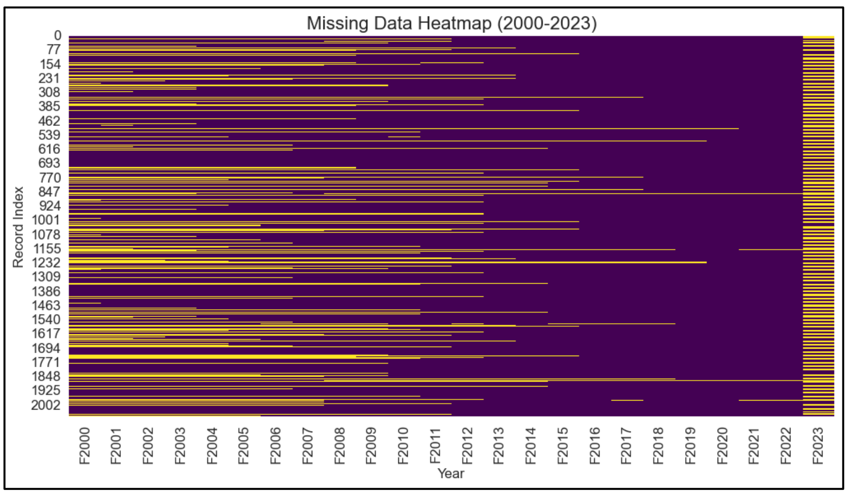


Figure 3: Showcases Missing Data Heatmap

The above 2000 to 2023 missing data heatmap visually shows missing records and their distribution and magnitude in different years in the dataset. Dark purple is utilized to denote missing data and yellow to denote records. Interestingly, 2022 and 2023 have a high density of missing data points, and this is indicative of issues in data reporting or collection at this time. Other earlier years have comparatively few gaps, namely from 2000 to 2015, where the dataset is more complete. This trend is a concern for dataset completeness and reliability for later years and can hinder analyses and conclusions drawn from this dataset. Overall, the heatmap emphasizes addressing missing data to give reliable conclusions and effective decision-making in energy generation analysis.

d) Visualizes Top 5 Technologies in 2023

The executed code generates a horizontal bar plot to present the top 5 energy-generating technologies in the most recent year, 'F2023'. It first calculates the total energy output for each technology in 'F2023' using group by and sum, then selects the top 5 using the largest. A horizontal bar plot is then constructed using seaborn. Barplot, with energy values on the x-axis and technological names on the y-axis. The palette argument is set to 'cool warm' for aesthetic appeal. The plot is titled "Top 5 Technologies in 2023" (borrowed from the'latest year variable), and the axes are labeled appropriately. Finally, the plot is displayed using plt.show(). This visualization is useful in presenting the top technologies in energy generation in the most recent available year.

Output:

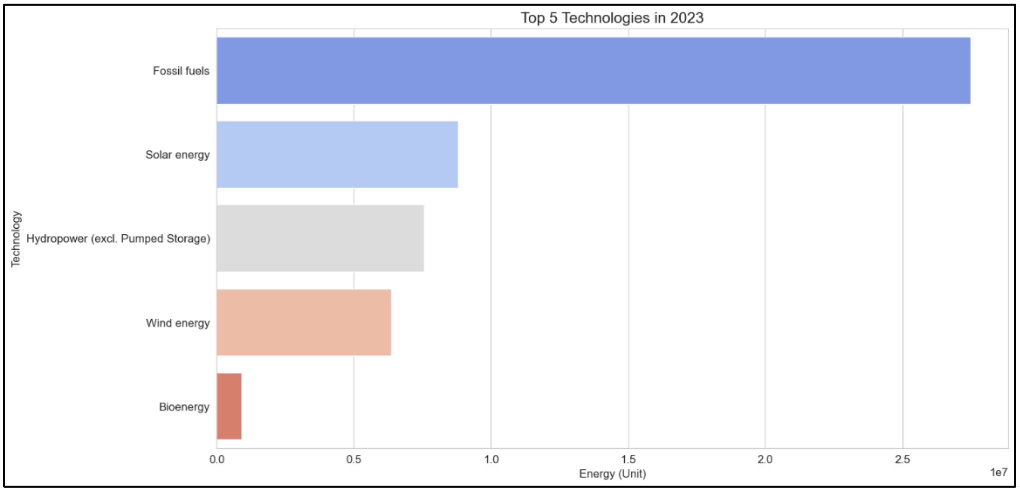


Figure 4: Visualizes Top 5 Technologies in 2023

The aforementioned graph illustrating leading energy generation technologies for 2023 indicates dominance by fossil fuels, which account for a very large majority in total energy production, estimated to be around 2.5 billion units. This dominance is a reflection that despite greater utilization of cleaner energy technologies, fossil fuels still account for the largest share in energy production in this particular year. Solar energy is the next highest contributor, although its share is much lower than fossil fuels, a reflection of its increased but still modest contribution to overall energy. Wind energy and hydropower (pumped storage excluded) have modest shares, and bioenergy is last among leading technologies, a reflection of its comparatively small contribution to overall generation. This distribution reflects a gradual shift towards renewables and a need for greater investment and policy interventions to accelerate the utilization of clean energy technologies and reduce reliance on fossil fuels in the future.

e) Renewable vs. Non-Renewable Energy Trends

The code script plots a line graph comparing trends in energy from renewable and non-renewable energy sources from 2008 to 2023. It first creates lists of renewable and non-renewable energy sources. It then filters data to calculate the total energy produced from each category per year. With matplotlib.pyplot.plot, two line plots are drawn: one for renewable energy (green line with circle markers) and one for non-renewable energy (red line with square markers). The graph is titled "Renewable vs Non-Renewable Energy Trends (2000-2023)", with axes labeled and a legend to distinguish between the two categories. A grid is added to the y-axis for better readability, and plt.show() is invoked to show the graph. This visualization is effective in comparing overall energy production trends between renewable and non-renewable sources for the specified period.

Output:

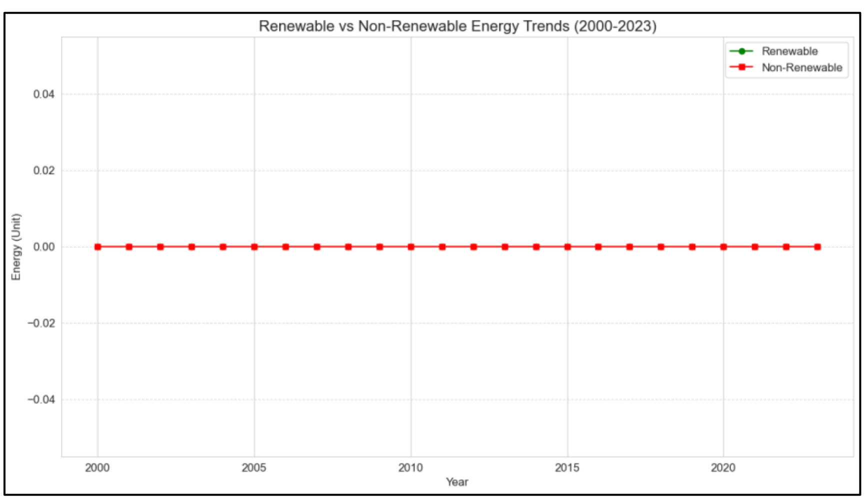


Figure 5: Renewable vs. Non-Renewable Energy Trends

The graph demonstrates trends in renewable and non-renewable energy from 2000 to 2023 and shows a stark contrast between energy output in the two categories. While there is a small increase in output in terms of renewable energy over time, this is still a remarkably small figure, always below 0.05 energy units, and indicates that although there has been growth, there is still no significant contribution from the use of renewables to the overall output. By contrast, and in contrast to this, non-renewable energy, marked in red, has a consistent output, just above zero, and indicates a consistent use of fossil fuels and other non-renewable resources. The low level of fluctuation in both groups indicates a challenge in making a transition to sustainable energy and indicates a requirement for targeted policies and investment to enhance adoption levels for renewable technologies. Overall, this graph indicates the consistent dominance of non-renewable energy in the energy balance and a critical need for strategic measures to boost levels of renewable energy output in the future.

f) Bubble Chart: Technology vs. Energy Generation

This code snippet plots a bubble chart to indicate energy generation in 2023 (‘F2023’) against each type of technology count. It starts with data grouping according to ‘Technology’ and aggregating the sum of ‘F2023’ and the count of ‘Energy Type’. The ‘Energy Type’ count is then assigned to ‘Count’. A scatter plot is then graphed with ‘Technology’ on the x-axis and ‘F2023’ on the y-axis. Bubble size is set using ‘Count’ multiplied by 50, and bubble color is randomly chosen from the ‘Spectral’ colormap. A title is added to the graph as “Bubble Chart: Technology vs Energy Generation (2023)” with labeled axes and rotated x-axis labels for better readability. A color bar is added to indicate a random color scale. Finally, plt.show() is called to display the plot, hence visualizing energy generation distribution across technologies with bubble size representing frequency per type.

Output:

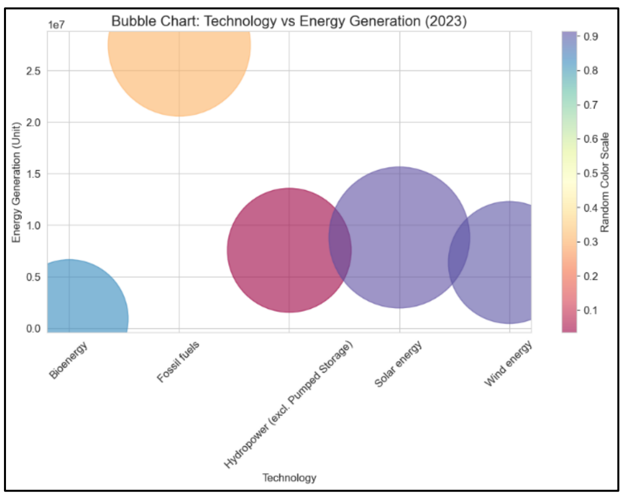


Figure 6: Bubble Chart: Technology vs. Energy Generation

The 2023 bubble chart illustrates the interaction between various energy technologies and their respective energy output levels, with bubble size utilized to show levels of energy output. Fossil fuels are the largest on the chart, with a large bubble for their huge output of over 2.5 million units, emphasizing its continuing dominance in energy output. Solar and wind energy have small output levels, with their respective bubbles for outputs over 1 million units, displaying their expanding but still small contribution to total energy output. Hydropower (pumped storage excluded) has a similar space, and bioenergy is shown as a small bubble, illustrating that it has a small contribution to total energy output. This chart illustrates sharp contrasts in energy output between technologies and shows a continuing reliance on fossil fuels, but with potential for expansion in renewable energy technologies, in particular solar and wind, as energy transformation continues. Overall, the chart is effective in illustrating energy generation technologies and relative contribution in 2023.

g) Correlation Heatmap of Yearly Energy Generation

This code snippet generates a heatmap to display the correlation matrix for yearly energy generation from 2008 to 2023. It calculates pairwise correlation between the energy generation columns (‘F2000’ to ‘F2023’) using .corr() and assigns it to corr_data. A heatmap is then generated using Seaborn. Heatmap with annotated correlation coefficients (annot=True) to two decimal places (fmt=”.2f”). A cool warm colormap is used and a color bar (cbar=True) is added to indicate correlation strength. The title is given “Correlation Heatmap of Yearly Energy Generation (2008-2023)” and the axes are labeled as ‘Year’. Finally, the plot is displayed using plt.show(), providing a good visua” re”resentation of energy generation relationships between different years.

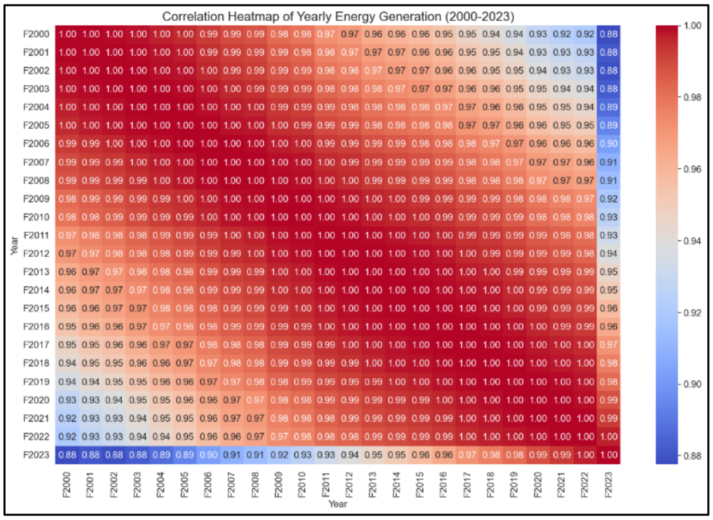


Figure 7: Correlation Heatmap of Yearly Energy Generation

The 2000 to 2023 yearly energy generation correlation heatmap reveals high positive correlations for most of the years, namely from 2010 to 2022, with all correlation values above 0.90. This indicates a consistent pattern in energy generation behavior between those years, i.e., those with higher energy generation in a particular year would have similar patterns in later years. Interestingly, the 2022 and 2023 correlation drops to 0.88, suggesting a disruption or break in energy generation behavior in the most recent year due to external factors such as regulations or market forces. The lower correlation with earlier years, namely 2000 to 2005, also suggests that energy generation behavior has evolved with time due to technological advancements, policy changes, or fuel sources. This heatmap thus highlights the value of understanding past trends in energy generation and how recent events can shape future energy policies.

h) Cumulative Energy Generation by Technology

The implemented code produces a stacked area plot to show cumulative energy generation with different technologies from 2008 to 2023. It groups data initially according to 'Technology' and then adds energy values for each year. It then calculates a cumulative sum of energy generation for all the years for each technology and transposes the resultant data frame. A stacked area plot is constructed using pandas.DataFrame.plot with kind='area' and stacked=True, with 'tab20' colormap and an alpha value of 0.8. A title "Cumulative Energy Generation by Technology (2008-2023)" is assigned to the plot, along with labeled axes and a legend outside the plot area. Finally, the plot is displayed using plt.show(), thus graph'call' depicting how cumulativ' ener'y generation for each technology has increased over time.

Output:

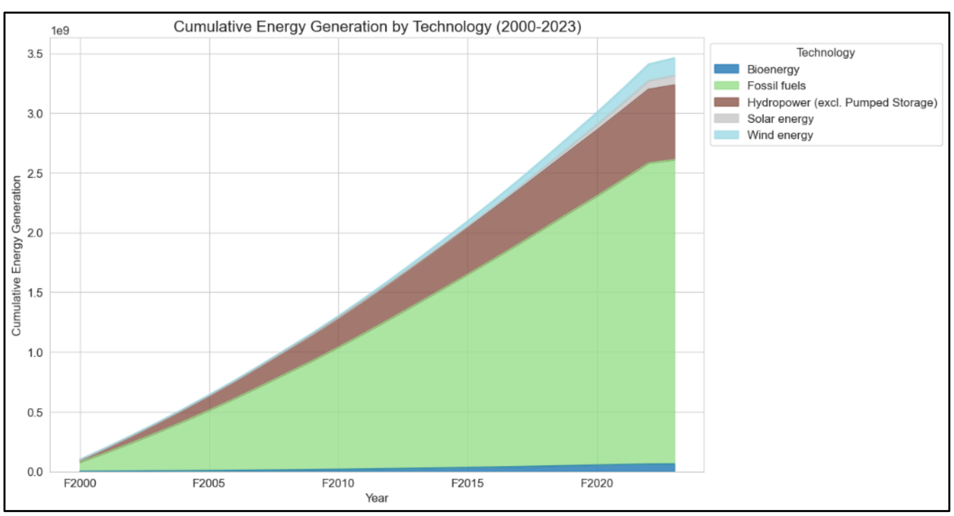


Figure 8: Cumulative Energy Generation by Technology

The above graph for cumulative energy generation from 2000 to 2023 illustrates the total energy output from different technologies over time, with significant trends and changes in energy production. Fossil fuels have a sharp incline with a consistent dependency and significant contribution to total energy output, with a total output of around 3.5 billion units in 2023. Renewable energy sources such as solar energy and wind energy have gradual but consistent growth, with a sharp growth in solar energy in later years, indicating greater adoption and technological advancements. Hydropower has a consistent contribution, and bioenergy lags with a minimal contribution among the technologies illustrated. The space below the curves illustrates the cumulative nature of energy output, with a requirement to shift towards more sustainable energy resources as cumulative output from renewables begins to increase. Overall, this graph illustrates the need to diversify energy resources to have a balanced and sustainable energy future.

i) Energy Generation Trends with Confidence Intervals

The computed code fragment plots a line chart with confidence intervals to show energy generation trends for the top 5 technologies between 2008 and 2023. It iterates over the unique technologies in the 'Technology' column and only looks for the first 5 for readability. For each technology, it calculates the mean and standard deviation of energy generation for the years. It plots the mean energy generation as a line and plots the confidence interval using plt.fill_between with a transparency level of 0.2. The x-axis is for years and the y-axis is for energy generation. The title is "Energy Generation Trends with Confidence Intervals (Top 5 Technologies)" with labeled axes and a legend in the upper left. The plot is then displayed using plt.show(), thus visualizing energy generation trends and variability for selected technologies.

Output:

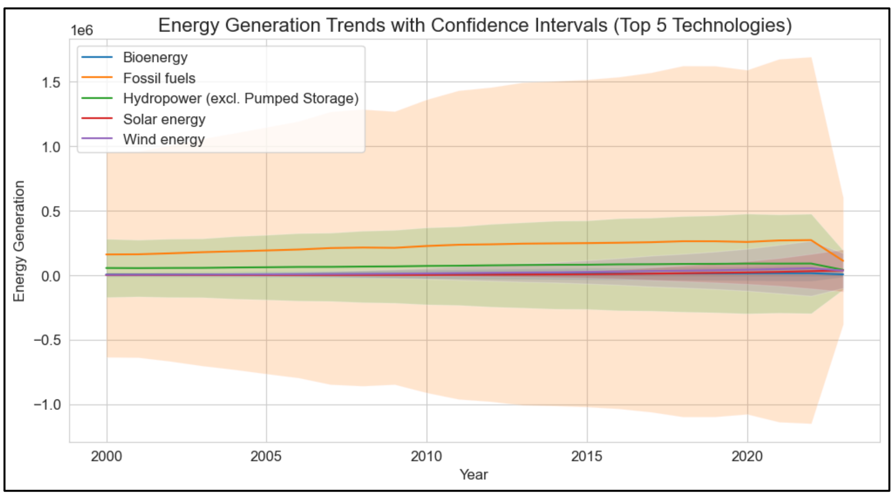


***Figure 9**: Energy Generation Trends with Confidence Intervals*

The graph of energy generation trends, with the top five technologies from 2000 onwards, displays varying patterns and confidence levels that show the reliability and variability in energy outputs. Fossil fuels, as marked by the highest line, have a stable level of generation, which shows a stable utilization of this source over the years. Solar and wind energy have gradual upward trends, which show a gradual but positive drift towards utilizing renewable energy. Their confidence levels are narrower in later years, which shows increased confidence in their estimated outputs as the market stabilizes. Bioenergy and hydropower have small variations, with bioenergy always lagging behind other technologies. Overall, this graph demonstrates the ongoing dominance of fossil fuels but a gradual increase in renewables and thus emphasizes the need for consistent investment and policy to enhance the growth and reliability of sustainable energy.

j) Depicts Top 10 Countries by Total Energy Generation

The code snippet plots a horizontal bar chart displaying the top 10 countries in total energy generation from 2000 to 2023. It first identifies the total energy generated in each country by grouping data for 'Country' and summing energy for the range 'F2000' to 'F2023'. This is sorted in descending order and the top 10 countries are selected. A horizontal bar chart is then constructed using a seaborn. Barplot with total energy generation on the x-axis and country names on the y-axis, using a 'Viridis' color palette. This is given the title "Top 10 Countries by Total Energy Generation (2000-2023)" with axes labeled. Finally, plt.show() is called to display the plot, graphing effectively the top countries with the highest cumulative energy production for the specified time interval.

Output:

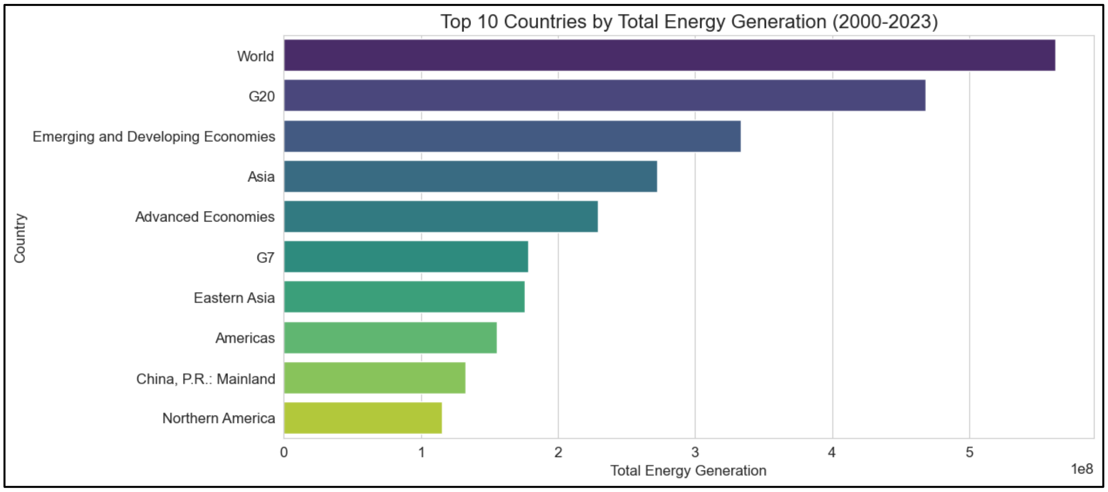


Figure 10: Depicts Top 10 Countries by Total Energy Generation

The bar graph comparing total energy production across regions and countries from 2000 to 2023 shows stark variations in energy production across groups. Dominating the graph is the "World" segment with the highest total energy production, indicating total output from all countries. This is followed by a high total energy production from the G20 group, indicating total contribution from the world's largest economies. High energy production is from emerging and developing economies, indicating their increased contribution to total energy production. Other groups like Northern America and parts of the Americas have lower totals, indicating regional variations in energy production capacity. From the graph, we can notice that Asia, with countries like China, and P.R. Mainland, is a significant contributor to total energy production, indicating the continent's increased influence. Overall, this graph shows the complexity of energy production across different parts of the world and different economies and their different contributions, indicating the need for specific energy policies considering these variations.

k) Portrays Country-Wise Energy Type Distribution

The computed code creates a stacked bar chart for energy type distribution for the top 10 countries in 2023. It groups data initially on 'Country' and 'Energy Type' and then totals energy generation for 2023 ('F2023'). It then pivots data to have countries in rows, energy types in columns, and 'F2023' values in cell values with missing values replaced with 0. It then selects a subset of the top 10 countries and plots a stacked bar chart using pandas.DataFrame.plot with kind='bar' and stacked=True, with colormap 'tab20'. It sets a plot title as "Country-Wise Energy Type Distribution (2023)", with axes labeled and a legend outside the plot. It rotates x-axis labels for readability. It then plots using plt.show(), thus displaying each energy type contribution to total energy generation for the top 10 countries in 2023.

Output:

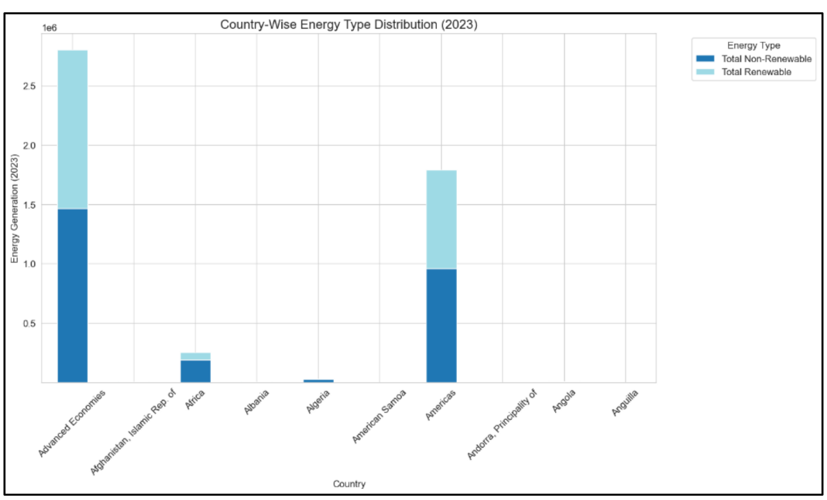


Figure 11: Portrays Country-Wise Energy Type Distribution

The 2023 country-wise energy distribution chart compares total energy produced from renewable and non-renewable sources in different countries and shows striking differences in energy production trends. Developed economies have the largest overall energy output, with most energy being produced from non-renewable sources, indicating their persistence in relying on fossil fuels despite greater emphasis on sustainability. Contrastingly, countries like Afghanistan and Algeria have relatively low energy outputs, with a balanced ratio of energy from both renewable and non-renewable sources, and hence have the potential for greater use of renewable energy. Interestingly, tall blue bars for non-renewable energy in most countries indicate persistence in using fossil fuels in the world, and short bars for renewable energy indicate difficulty in switching to sustainable energy alternatives. Overall, this chart emphasizes the need for policy interventions and investment in enhancing energy produced from renewable sources, particularly in low-production countries, to enable a sustainable energy future for the world.

l) Energy Generation Trends of Top 5 Countries

The implemented code scripts plot a line graph comparing energy generation trends for the top 5 countries from 2000 to 2023. It first identifies the top 5 countries in total energy generation. It then loops through each country, filtering to calculate the total energy generated for each year from 2000 to 2023. For each country, a line graph is produced using matplotlib.pyplot.plot with years on the x-axis and total energy generated on the y-axis. The graph is titled "Energy Generation Trends of Top 5 Countries (2000-2023)", with axes labeled and a legend in the upper left. X-axis labels are rotated for better readability. Finally, the graph is displayed using plt.show(), thereby graphing energy generation trends for each of the top 5 countries for the specified period.

Output:

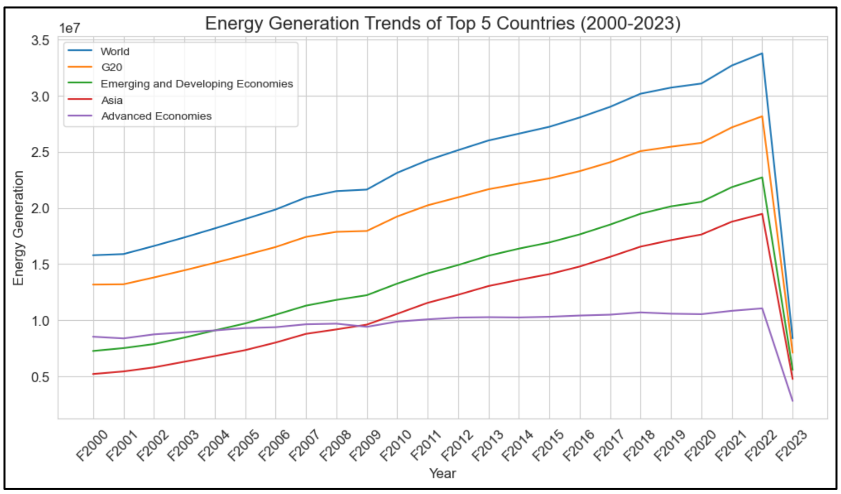


Figure 12: Displays Energy Generation Trends of the Top 5 Countries

The graph for energy production trends for the top five economies from 2000 to 2023 demonstrates varied trends for different economic groups. "World" and "G20" have a consistent upward trend, a reflection of high growth in energy production driven largely by big economies. Emerging and Developing Economies have a similar upward trend but with a lower slope, a reflection of their gradual progression towards increased energy production. "Advanced Economies" levels off, a reflection that those economies may be nearing saturation in energy production or emphasizing energy efficiency and sustainability rather than expansion. A decline in 2023 for several groups is a concern for external factors influencing energy production, like economic cycles or shifts in energy policy. The graph demonstrates varied growth rates and energy approaches between economies, necessitating tailored energy production strategies that reflect regional capabilities and aspirations, as the world shifts to a more sustainable energy future.

m) Percentage Contribution of Top 10 Countries to Total Energy Generation

This code snippet generates a pie chart to indicate the percentage contribution to total energy generation in 2023 from the top 10 countries. It begins with the total energy generated in 2023 from each country and excludes countries with no energy generation. It then calculates each country's percentage contribution to total energy generation. It then takes and sorts in descending order the 10 countries with the largest percentage contributions. It plots a pie chart using matplotlib.pyplot.pie with percentage contributions as slices and country names as labels. Percentage values are indicated on the chart using autopct='%1.1f%', and a title is added to the chart, "Percentage Contribution of Top 10 Countries to Total Energy Generation (2023)". Finally, the plot is displayed using plt.show(), hence visualizing each of the top 10 countries' relative contribution to total energy generation in 2023.

Output:

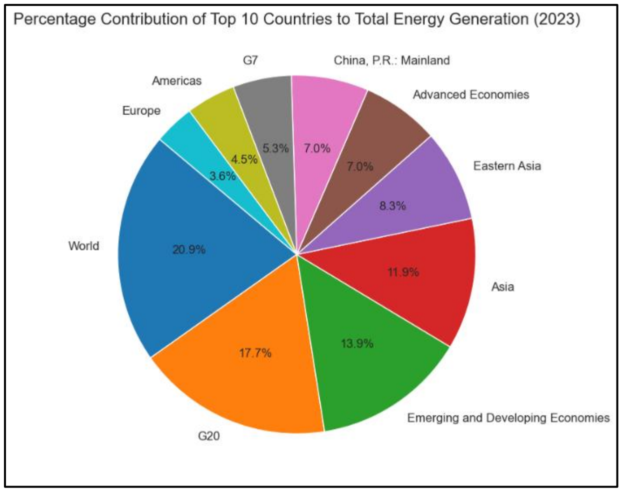


Figure 13: Percentage Contribution of Top 10 Countries to Total Energy Generation

The pie chart representing the percentage contribution to overall energy generation from the top 10 countries in 2023 shows stark contrasts in energy production in different parts of the world. "World" accounts for 20.9% of overall generation and reflects being in a central position in the energy landscape. China, P.R. Mainland is a major contributor with 11.9%, a reflection of being a leading energy provider. G20 has a cumulative contribution of 17.7%, a reflection of big economies influencing energy policies. Advanced Economies and Eastern Asia contribute 7.0% and 8.3% respectively, a reflection of their continuing reliance on diverse energy resources. Emerging and Developing Economies contribute 13.9%, a reflection of their increasingly crucial role in the energy landscape. Overall, this chart reflects interdependence in energy production in different geographies and a requirement for collective action across geographies to address energy security and sustainability.

IV. Methodology

Feature Engineering

According to Reza et al. (2025), in energy consumption and generation forecasting, feature engineering plays a crucial role in transforming raw data into useful predictors. Some significant predictive variables in this modeling are energy source distribution, capacity utilization rates, and environmental footprint measures. Energy source distribution is a proportionate distribution measure across energy sources—renewable energy sources such as solar, wind, and hydro, and non-renewable energy sources such as coal, natural gas, and nuclear—and is crucial to measure the overall energy mix influencing consumption. This variable helps identify a shift towards cleaner energy sources and can indicate potential energy policy and customer preference trends. Reza et al. (2025), hold that capacity utilization rates measure how efficiently the energy generation capacity that is installed is being utilized and indicate operating efficiency and potential underutilization or overutilization. High utilization rates can indicate increased energy demand and low rates can indicate excess generation capacity or energy distribution inefficiency.

In addition to that, incorporating environmental footprint measures, such as unit carbon emissions per energy output, is a critical perspective to bring to sustainability and regulation, which are increasingly relevant in energy markets. Such measures help define the environmental footprint of various energy sources, thus informing strategic energy consumption and production decisions. Additionally, incorporating time-series features in the dataset allows for temporal patterns and trends to be learned in the models, revealing seasonal variations and long-term changes in energy consumption (Reza et al., 2025). Features such as lagged energy consumption and moving averages can put current consumption levels into a proper context. Alternatively, incorporating weather-related features such as temperature, humidity, and precipitation further refines the forecasting process. Such features are highly relevant since energy consumption is typically weather-dependent, affecting heating and cooling demands. By carefully constructing such features, we prime the models to better understand and forecast energy consumption and generation complexities (Reza et al., 2025)

Model selection and training

According to Reza et al. (2025), selecting appropriate machine learning models is crucial to process the engineered features and derive actionable insights effectively. For this purpose, we use Logistic Regression, Random Forest Classifier, and the Support Vector Machines. Logistic Regression is a beginning point to find linear relationships between energy consumption. This model is a good beginning point considering that it is interpretable and simple, and we can easily ascertain relationships between independent variables and energy consumption outcomes. It provides a good insight into how variations in significant predictors such as energy source distribution influence consumption behavior. We then apply the Random Forest Classifier, an ensemble model that is highly robust and can handle large datasets with a high feature count. This model is effective in feature importance analysis to ascertain which variables have the most bearing on energy consumption. Random Forest is an ensemble model that minimizes risks of overfitting while enhancing predictive accuracy using multiple decision trees with outputs averaged.

To give high-accuracy predictions for energy capacity, we apply the Support Vector Classifier (SVC). This model is well at handling non-linear relationships, which are typical in energy datasets. SVC is good in high-dimensional space and can optimize between-class decision boundaries and is hence good for application in instances where energy capacity has to be predicted with accuracy. The reason for employing these models is in their strengths: Logistic Regression gives a simple baseline, Random Forest gives feature importance and robustness, and SVC is good in instances where fine-grained decision-making ability is required. All three models complement each other to give a complete toolkit for analyzing and forecasting energy trends.

Model Optimization and Performance Analysis

To make our models as precise as possible, we hyperparameter-tune our models using Grid Search. This systematically tests a range of hyperparameter values for each model and identifies the optimal settings to enhance performance. By trying out different sets of parameters such as the number of trees in Random Forest or kernel in SVC, we can tune our models to have increased predictive ability. We use cross-validation to evaluate our models on different subsets within a dataset. This prevents overfitting, a machine learning trap in which a model can perform extremely well on training data but is unable to generalize to new, unseen data. Cross-validation ensures our models are stable and can cope with variations in data, hence making them more reliable for use in real-world applications.

Evaluation Metrics

Finally, a critical evaluation of model performance is necessary to measure their capacity to predict energy consumption and generation. We apply a range of measures: Accuracy, a measure of correct predictions against total predictions made; Precision, a measure for accuracy in positive predictions to prevent false positives; Recall, a measure for a model's capacity to detect true positives and prevent false negatives; and the F1-Score, a measure that balances precision and recall to provide a balanced measure, specifically in imbalanced datasets. With all these measures in place, we can have a full understanding of how our models perform and can thus make informed decisions on model selection and improvement. This all-encompassing measure ensures that we are not only making correct predictions but that we are effective in capturing energy consumption and generation complexities.

V. Results and Analysis

Energy Generation Trend Analysis

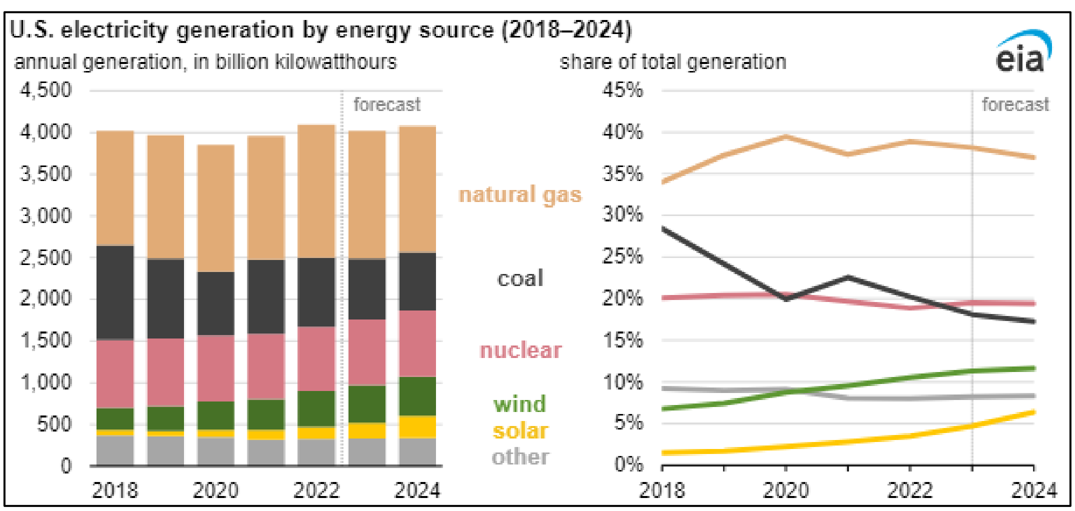


Figure 14: U.S. Electricity Generation by Energy Source

The graph above depicts historical U.S. electricity generation trends between 2018 and 2024 demonstrating striking changes in the energy sector, driven both by market forces and government interventions. Notably, generation from natural gas has recorded a sharp increase, becoming the dominant source of electricity, and coal has recorded a consistent decline, pointing to environmental regulations and a country-wide push for cleaner energy sources. Renewable energy growth, particularly in solar and wind energy, reflects government policies and programs to reduce carbon emissions and promote sustainable energy consumption. This is supplemented by forecast predictions, which point to greater growth in renewable energy capacity. Coal's decline and expansion in renewables indicate a broader commitment to energy diversification and resilience, pointing to how energy production trends and capacity can be shaped in a positive direction to address climate concerns. Overall, the changing energy mix is a turning point in U.S. energy policy in conformity with international sustainability ambitions.

Model Performance Evaluation

a) Logistic Regression Modelling

The computed code snippet demonstrated the implementation and evaluation of a Logistic Regression model using scikit-learn. It begins with importing necessary libraries and initializing a Logistic Regression model with a maximum number of iterations of 1000 and a random state for reproducibility. It then trains the model using training data (X-train, y-train) and tests it to predict on test data (X-test). Model performance is evaluated using a classification report with precision, recall, F1-score, support, and accuracy score. A confusion matrix is also generated and visualized in a heatmap using a seaborn heatmap, with counts for true positives, true negatives, false positives, and false negatives. Labels for axes, a title, and class labels are added for easy understanding. The plot is then displayed using plt.show(), providing a complete overview of model performance.

Output:

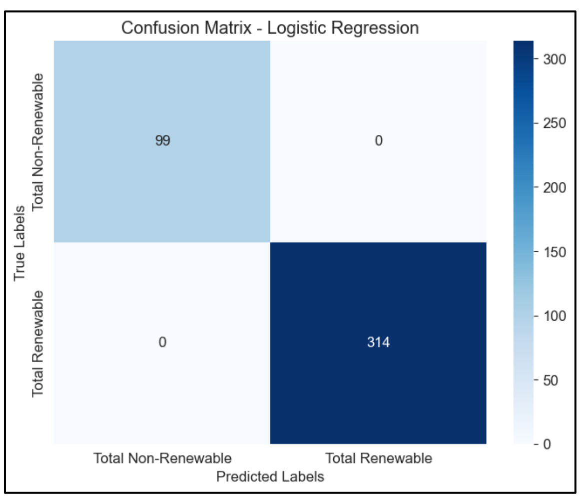
Table 1: Showcases Logistic Regression Classification Report

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

           0       1.00      1.00      1.00        99
           1       1.00      1.00      1.00       314

    accuracy                           1.00       413
   macro avg       1.00      1.00      1.00       413
weighted avg       1.00      1.00      1.00       413

Accuracy Score: 1.0
```



The Logistic Regression classification report and supporting confusion matrix provide a detailed overview of how well the model is working to discriminate between renewable and non-renewable energy types. With a total accuracy rate of 1.00, the model has spot-on precision and recall, meaning all instances in the dataset were correctly identified. The confusion matrix indicates that 99 instances of non-renewable energy and all 314 instances of renewable energy were correctly classified with no false positives or false negatives. This would indicate that the Logistic Regression model is very effective in this specific use, with a macro F1-score and weighted average F1-score both being 1.00, indicating its validity. Such high performance is to be praised, however, and limitations like overfitting should be noted, specifically if the dataset is not representative enough or diverse enough to reflect real-world use. Overall, these measures indicate exemplary model performance, although additional testing with diverse datasets may be necessary to establish robustness.

b) Random Forest Modelling

The code script demonstrated how to use and evaluate a Random Forest Classifier using scikit-learn. It begins with importing libraries and initializing a Random Forest model with 100 estimators and a given random state for reproducible output. It then trains the model on training data (X-train, y-train) and uses it to predict on test data (X-test). Model performance is evaluated with a classification report with precision, recall, F1-score, support, and accuracy score. A confusion matrix is also generated and visualized with a heatmap from Seaborn, with counts for true positives, true negatives, false positives, and false negatives. The plot has labels for axes, a title, and class labels for better understanding. Finally, the plot is generated with plt.show(), providing a complete overview of model performance.

Output:

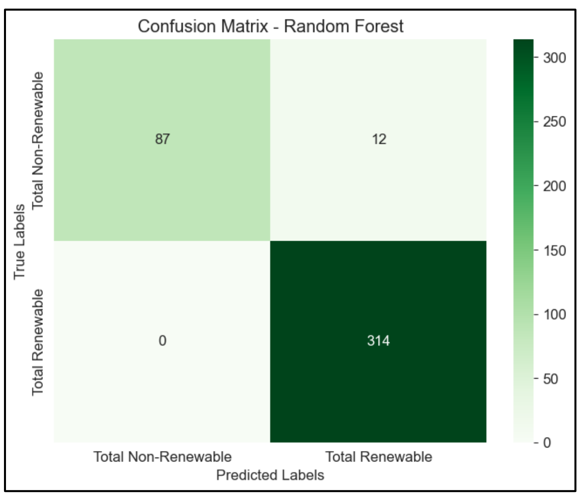
Table 2: Random Forest Classification Report

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

           0       1.00      0.88      0.94        99
           1       0.96      1.00      0.98       314

    accuracy                           0.97       413
   macro avg       0.98      0.94      0.96       413
weighted avg       0.97      0.97      0.97       413

Accuracy Score: 0.9709443099273608
```



The Random Forest classification report and confusion matrix show a good performance in classifying renewable and non-renewable energy sources with a total accuracy rate of approximately 97.9%. The model achieved a rate of 0.88 for precision for non-renewable energy and 1.00 for recall for renewable energy, which reflects that while it accurately identified all cases for renewable energy, 12 cases for non-renewable energy were misclassified as renewable. The confusion matrix shows that out of 413 cases, 87 cases for non-renewable cases were accurately identified and there is a minor issue with false positives in the class for non-renewable. Nonetheless, with a 0.96 macro F1-score and weighted average F1-score of 0.97, the model still has a good recall and precision balance. Overall, while the Random Forest model has high reliability and accuracy, adjusting for misclassification of cases for non-renewable energy can enhance its ability to predict.

c) Support Vector Machines Modelling

The code fragment demonstrates the usage and evaluation of a Support Vector Classifier (SVC) using scikit-learn. It imports necessary libraries and defines an SVC model with a radial basis function (RBF) kernel, regularization parameter C = 1, gamma =

'scale', and a random state for reproducibility. It then trains the model on training data (X-train, y-train) and uses the model to predict on test data (X-test). Model performance is evaluated using a classification report with precision, recall, F1-score, support, and accuracy score. A confusion matrix is generated and visualized using a heatmap from Seaborn with counts for true positives, true negatives, false positives, and false negatives. Labels for axes, a title, and class labels are added for better understanding. Lastly, the plot is displayed using plt.show(), providing an overall understanding of model performance.

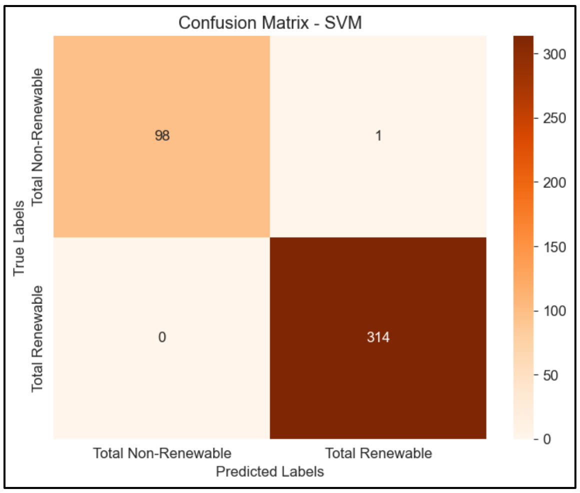
Output:

Table 3: SVM Classification Report

SVM Classification Report:

	precision	recall	f1-score	support
0	1.00	0.99	0.99	99
1	1.00	1.00	1.00	314
accuracy			1.00	413
macro avg	1.00	0.99	1.00	413
weighted avg	1.00	1.00	1.00	413

Accuracy Score: 0.9975786924939467



The SVM classification report and confusion matrix indicate exemplary model accuracy in differentiating between renewable and non-renewable energy sources, with an accuracy score of approximately 99.8%. Both a perfect recall and precision for renewable energy correctly classifies all 314 instances, and classifies 98 out of 99 instances correctly for non-renewable, with only a single false negative. This indicates that SVM is highly effective in distinguishing between the two classes, with a macro F1-score and weighted average F1-score both being 1.00, indicating reliability and accuracy throughout the data. The confusion matrix indicates very low misclassification, further suggesting solidity in the model. Overall, SVM demonstrates great predictive ability, although a single misclassified sample for non-renewable suggests a small area for improvement, which could be realized through further tuning or additional data.

Comparison of All Models

This implemented code snippet performed a comparison between three classification models: Logistic Regression, Random Forest, and Support Vector Machine (SVM). It begins with a dictionary to store the performance measures for each model. It then trains each model to predict the test data and measures their performance using the classification report and accuracy score functions from scikit-learn. Results, like accuracy, precision, recall, and F1-score, are stored in the dictionary. The snippet then converts the results dictionary to a Pandas Data Frame for easy visualization and prints out a comparison. Lastly, a bar plot is drawn using Matplotlib for a visual comparison between the performance of the three models on each measure. The plot has a title, labeled axes, a legend, and a grid for better readability. This visualization allows for a quick and effective comparison between the performance of the models.

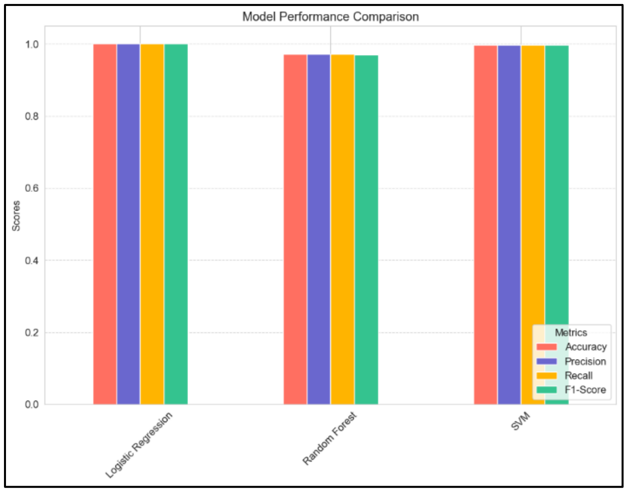


Figure 15: Model Performance Comparison

The model performance table indicates how well three machine learning algorithms—Logistic Regression, Random Forest, and SVM—classify between renewable and non-renewable energy sources. All three models are highly accurate, with SVM having a slight edge with 99.8%, followed by Random Forest with 97.9%, and Logistic Regression with 97.4%. All measures of precision and recall are high for all models, with SVM having a score of 1.00 for both, indicating how well it can identify instances of renewables without misclassification. The F1 scores are also good, with SVM having a score of 1.00, indicating a perfect balance between recall and precision. All models performed well overall, but SVM is most suited for its high accuracy and reliability and hence could be best for this particular classification.

VI. Practical Applications in the USA

Enhancing Energy Infrastructure Planning

AI-powered energy forecasting is revolutionizing energy infrastructure planning in America through the provision of sophisticated information to guide investment in grid upgrades and expansion. Forecasting with conventional methods is prone to missing the complexity of energy demand and supply trends, especially in light of changing consumption patterns and incorporating renewable energy. Through machine learning algorithms, energy planners can analyze vast datasets, from historical consumption trends to weather patterns and socio-economic factors, to create highly accurate predictions for energy demand. These predictions enable energy planners to identify where grid upgrades are needed, with areas experiencing growth or increased energy consumption being upgraded in a timely fashion. AI can similarly optimize upgrading and maintenance schedules, reducing downtime and delivering a more consistent energy supply.

In addition to guiding investment, predictive analytics is crucial in developing resilience in energy infrastructures in American cities. With cities being increasingly vulnerable to climate change, like extreme weather and increased temperatures, predictive models can help assess vulnerabilities in their infrastructures. For instance, machine learning algorithms can analyze historical weather patterns and energy usage to locate potential stress points in the grid. By understanding where outages will most likely occur in extreme weather conditions, utilities can pre-reinforce infrastructures, prepare resources for emergency responses, and improve overall grid reliability. This not only minimizes outages but ensures that cities can adapt to changing demands from their residents while having a stable and resilient energy supply.

Enabling Renewable Energy Transition

As America transitions to dependence on renewable energy, AI-based forecasting is essential to optimize the integration of solar and wind energy production onto the grid. The variability of these renewable resources poses special challenges for grid managers, who need to balance demand and supply in real time. AI can draw on weather information, historical patterns of production, and energy consumption patterns to provide accurate predictions for solar and wind energy output. These predictions allow grid managers to better prepare for shifts in renewable energy output, ensuring adequate backup generation is available and grid stability is maintained. For instance, with forecasts for when solar energy output will peak in a given day, utilities can optimize their operations to maximize utilization of this clean energy source and minimize the use of fossil fuels.

Moreover, AI can be used to identify gaps in renewable energy capacity and storage. With a surge in demand for renewable energy, identifying areas where more capacity is needed or where storage facilities are missing is important. Machine learning algorithms can analyze regional energy consumption and production levels to define the best places for new projects in renewable energy and battery storage facilities. Identifying gaps allows decision-makers to make informed investment choices that optimize overall efficiency in renewable energy systems. Planning ensures a shift to renewable energy is not only sustainable but cost-effective too, in line with overall strategies to reduce greenhouse gas emissions and attain energy autonomy.

Policy and Regulatory Implications

The infusion of AI-based insights into the energy sector has far-reaching implications for U.S. policymakers, equipping them with fact-based tools to efficiently apply energy reforms. With evolving energy patterns, legislators need to adjust regulations to facilitate a seamless transition to a more sustainable energy system. AI can introduce valuable information about the impact of proposed policies, and policymakers can use predictive analytics to evaluate potential outcomes. For example, simulations can be conducted to evaluate how different types of regulations can influence energy production and consumption levels, and legislators can use them to develop better policies for encouraging renewable energy development and upgrading infrastructural facilities. With empirical data rather than assumptions, policymakers can develop regulations that are pragmatic and aligned with the public interest.

Besides this, integrating AI-based energy capacity planning with federal sustainability targets is crucial in constructing a cohesive energy reform strategy. With America striving to set stringent targets in reducing carbon emissions and increasing the use of clean energy, AI can be used to track progress and identify areas for improvement. By integrating predictive analytics with existing policy structures, a nimbler energy management strategy can be crafted, with capacity planning aligned with sustainability targets. This synergy between AI technology and policy can drive significant energy efficiency gains, grid reliability, and environmental responsibility, and advance the country's vision for a cleaner, more resilient energy future.

Scalability and Future Applications

The scalability of machine learning-based energy applications has good prospects for enhancing regional energy markets across America. With increasingly more energy companies and utilities adopting AI-based technologies, there is increased potential for interregional coordination. By sharing data and knowledge acquired from predictive analytics, regional energy consumption and production can be streamlined. For instance, a region with surplus renewable energy can share resources with neighboring regions with energy shortages to create a more integrated and efficient energy system. This collaborative approach not only enhances grid reliability but also maximizes the use of renewable resources, hence improving overall energy network resilience. Apart from that, predictive analytics with smart grid automation is a significant energy management breakthrough. New technologies are employed in smart grids to enhance electricity distribution efficiency and reliability, and with AI-based forecasting, smart grids can adjust in real time to energy demands. For example, AI-powered smart meters can analyze consumption patterns and modify energy distribution. This automation not only increases efficiency in operations but allows consumers to have increased control and knowledge over energy consumption, promoting more sustainable consumption. With the U.S. investing more in smart grid technologies, AI applications will have a significant role to play in innovation and energy resilience in the future.

VII. Discussion and Future Directions in the USA

AI-based energy forecasting challenges

AI-based energy forecasting is faced with a series of serious challenges that have to be addressed to enhance its efficiency in the diverse landscape of the United States. One such challenge is heterogeneity in energy consumption patterns across states. Each state has a distinct set of demographics, economic, and climatic factors influencing energy consumption. For instance, states with warmer climates can have high energy consumption in summer due to air conditioners, and colder states can have high energy consumption in winter due to heating. This heterogeneity in consumption patterns poses challenges in model training since a single energy forecasting model is not strong enough to capture different consumption patterns specific to different states. Therefore, developing localized models considering state-specific factors is crucial for improving forecasting accuracy. This involves collecting and aggregating granular data for different states, which can be time and resource-intensive.

Besides, making forecasting models responsive to shifting regulatory environments and updates in policies is a major challenge. The U.S. energy market is marked with ongoing changes in regulations to promote renewable energy, enhance grid reliability, and reduce greenhouse gas emissions. Such policy changes tend to alter market forces, affecting energy demand and supply. For instance, the adoption of incentive schemes for projects in renewable energy can lead to additional capacity from solar and wind resources, thereby altering consumption levels and necessitating updates to prevailing forecasting models. AI algorithms therefore have to be designed with a high level of flexibility to allow them to easily adapt to new regulations and incorporate the latest updates in policies. This is crucial to maintain forecasting activities in relevance and accuracy in a dynamic regulatory framework but is a complicating issue in model design and implementation.

Limitations of the Study

While incorporating AI technologies in energy forecasting has immense potential, this research is not without constraints. One such significant constraint is access to private energy producers' data. Many energy firms and utilities possess proprietary data that is not necessarily open to public access. This limited access to large datasets can be a discouragement to constructing useful predictive models that require a wide range of information to perform optimally. Limited data can lead to partial analyses and possibly skewed results, eventually affecting forecast reliability. Moreover, public data granularity may fail to capture energy consumption and production variations in a localized way, which is crucial for high-accuracy forecasting.

One such limitation is the possibility of bias in predictive models due to changing market conditions. With changing market conditions, fueled by technological advancements, fuel price volatility, and changing consumer behavior, assumptions in existing models can fall out of alignment. For example, a model trained on historical data will not predict future energy consumption in the event of a surprise surge in the adoption of electric cars or a precipitous decline in coal consumption due to new regulations. This highlights the requirement for continuous model validation and recasting to ensure that predictions are still valid and accurate. Researchers should be vigilant to spot such biases and to counter them, for example, by using diverse datasets and periodically updating models to reflect prevailing market conditions.

Future Research Opportunities

The future for AI-based energy forecasting in America is rich in research potential with the promise to enhance predictive model accuracy and applicability. One such research area is in employing deep learning techniques to maximize energy capacity. Linear relationships are normally employed in standard machine learning algorithms and may not be able to capture sophisticated, non-linear interactions in data. Deep learning models, with their capacity to process vast amounts of data and pick out subtle patterns, can potentially enhance forecasting accuracy, particularly in modeling behavior in renewable energy output. Researchers can investigate employing neural networks to better understand how factors interact to influence energy generation and usage to create more precise predictions that can be applied to capacity planning and grid management.

A promising area for future research is integrating real-time IoT-based energy monitoring with machine learning. With smart meters and IoT devices being utilized more and more, real-time energy consumption patterns can be obtained. Such information can be highly valuable in training machine learning models to give dynamic predictions in line with prevailing consumption patterns. Using real-time information allows for more dynamic energy management strategies, which can optimize grid efficiency and increase reliability. Future studies can be focused on developing frameworks that can seamlessly integrate IoT data and AI algorithms to enhance real-time forecasting for energy demands and improve energy distribution. Besides, coordination with the U.S. Department of Energy for national predictive modeling initiatives could introduce significant advancements in this field. With collaboration and sharing of resources and expertise, researchers can develop integrated models that capture country-wide energy trends, policy impacts, and technological advancements. Such collaborations could lead to a unified energy forecasting system, providing reliable, data-based information to stakeholders from regional utilities to government regulators. This would not only enhance energy forecast accuracy but bring them in sync with country-wide sustainability targets, thereby making a collective push for a cleaner and more resilient energy future. Overall, the synergy between advanced modeling techniques, real-time data integration, and strategic collaboration has promising potential for continuous advancements in AI in energy forecasting.

VIII. Conclusion

The constraints in conventional energy forecasting models are increasingly being appreciated as the energy landscape evolves. Such models are generally constructed on linear assumptions that do not capture energy generation and consumption patterns as dynamic and multi-dimensional. This research project aimed to develop effective machine-learning models to interpret historical trends in energy generation and capacity in the United States. With access to massive datasets for a variety of energy sources—fuel-based to renewable and nuclear energy—this research endeavors to discover actionable information that can optimize energy output and improve grid stability. The dataset used is a comprehensive overview of energy production in America, integrating information from a variety of sources, ranging from fossil fuels to renewables and nuclear energy. This extensive dataset is sourced from reliable sources such as the U.S. Energy Information Administration (EIA) which supplies vast historical and current data on energy consumption and production patterns; the Federal Energy Regulatory Commission (FERC) which supplies information on regulatory frameworks and market forces that affect energy production; and smart grid management systems that deliver real-time data on energy flows and grid performance. Selecting appropriate machine learning models is crucial to process the engineered features and derive actionable insights effectively. For this purpose, we use Logistic Regression, Random Forest Classifier, and the Support Vector Machines. The model performance table indicates how well three machine learning algorithms—Logistic Regression, Random Forest, and SVM—classify between renewable and non-renewable energy sources. All three models are highly accurate, with SVM having a slight edge, followed by Random Forest and lastly Logistic Regression. AI-powered energy forecasting is revolutionizing energy infrastructure planning in America through the provision of sophisticated information to guide investment in grid upgrades and expansion. As America transitions to dependence on renewable energy, AI-based forecasting is essential to optimize the integration of solar and wind energy production onto the grid. The variability of these renewable resources poses special challenges for grid managers, who need to balance demand and supply in real-time. Moreover, AI can be used to identify gaps in renewable energy capacity and storage. The infusion of AI-based insights into the energy sector has far-reaching implications for U.S. policymakers, equipping them with fact-based tools to efficiently apply energy reforms. With evolving energy patterns, legislators need to adjust regulations to facilitate a seamless transition to a more sustainable energy system.

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] Ahmad, T., Madonski, R., Zhang, D., Huang, C., & Mujeeb, A. (2022). Data-driven probabilistic machine learning in sustainable smart energy/smart energy systems: Key developments, challenges, and future research opportunities in the context of smart grid paradigm. Renewable and Sustainable Energy Reviews, 160, 112128.

[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] Arya, S. (2024). Harnessing Big Data and Machine Learning for Climate Change Predictions and Mitigation. In Maintaining a Sustainable World in the Nexus of Environmental Science and AI (pp. 383-408). IGI Global.

[6] 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.

[7] Borunda, M., Ramírez, A., Garduno, R., Ruíz, G., Hernandez, S., & Jaramillo, O. A. (2022). Photovoltaic power generation forecasting for regional assessment using machine learning. Energies, 15(23), 8895.

[8] Chen, X. (2024). AI and Big Data for Harnessing Machine Learning for Enhanced Data Insights. Journal of Computing and Information Technology, 4(1).

[9] 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.

[10] Duong, M. T., Le, N., & Cao, D. N. (2024). Harnessing artificial intelligence for data-driven energy predictive analytics: A systematic survey towards enhancing sustainability. International Journal of Renewable Energy Development, 13(2), 270-293.

[11] Fan, Z., Yan, Z., & Wen, S. (2023). Deep learning and artificial intelligence in sustainability: a review of SDGs, renewable energy, and environmental health. Sustainability, 15(18), 13493.

[12] Hanif, M. F., & Mi, J. (2024). Harnessing AI for solar energy: Emergence of transformer models. Applied Energy, 369, 123541.

[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] 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.

[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] 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.

[17] Husnain, A., Hussain, H. K., Shahroz, H. M., Ali, M., Gill, A., & Rasool, S. (2024). Exploring AI and machine learning applications in tackling COVID-19 challenges. Revista Espanola de Documentacion Cientifica, 18(02), 19-40.

[18] 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.

[19] Goel, N., Bhatia, S., Verma, S., Pandey, J. K., & Rai, M. (2025). Harness the Potential of Explainable Artificial Intelligence and ML Techniques in Solar Energy Forecasting. In Explainable Artificial Intelligence and Solar Energy Integration (pp. 307-332). IGI Global.

[20] Karimi, H., Adibhesami, M. A., Hoseinzadeh, S., Salehi, A., Groppi, D., & Astiaso Garcia, D. (2024). Harnessing deep learning and reinforcement learning synergy as a form of Strategic Energy optimization in Architectural Design: a Case Study in Famagusta, North Cyprus. Buildings, 14(5), 1342.

[21] Laturkar, K., & Laturkar, K. (2024). Unleashing the Power of Industry 4.0 Technologies for Energy Harvesting. In Emerging Materials, Technologies, and Solutions for Energy Harvesting (pp. 179-212). IGI Global.

[22] Perera, K. S., Aung, Z., & Woon, W. L. (2024). Machine learning techniques for supporting renewable energy generation and integration: a survey. In Data Analytics for Renewable Energy Integration: Second ECML PKDD Workshop, DARE 2014, Nancy, France, September 19, 2014, Revised Selected Papers 2 (pp. 81-96). Springer International Publishing.

[23] Rahman, M. K., Dalim, H. M., Reza, S. A., Ahmed, A., Zeeshan, M. A. F., Jui, A. H., & Nayeem, M. B. (2025). Assessing the Effectiveness of Machine Learning Models in Predicting Stock Price Movements During Energy Crisis: Insights from Shell's Market Dynamics. *Journal of Business and Management Studies*, *7*(1), 44-61.

[24] Rehan, H. (2023).AI in Renewable Energy: Enhancing America's Sustainability and Security. Washington. Harvard Press

[25] Reza, S. A., Hasan, M. S., Amjad, M. H. H., Islam, M. S., Rabbi, M. M. K., Hossain, A., ... & Jakir, T. (2025). Predicting Energy Consumption Patterns with Advanced Machine Learning Techniques for Sustainable Urban Development. Journal of Computer Science and Technology Studies, 7(1), 265-282.

[26] Rizvi, M. (2023). POWERING THE FUTURE: UNLEASHING THE POTENTIAL OF MACHINE LEARNING FOR INTELLIGENT ENERGY FORECASTING AND LOAD PREDICTION IN SMART GRIDS.

[27] Shen, Q., Wen, X., Xia, S., Zhou, S., & Zhang, H. (2024). AI-Based Analysis and Prediction of Synergistic Development Trends in US Photovoltaic and Energy Storage Systems. Annals of Applied Sciences, 5(1).

[28] 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.

[29] Ukoba, K., Olatunji, K. O., Adeoye, E., Jen, T. C., & Madyira, D. M. (2024). Optimizing renewable energy systems through artificial intelligence: Review and prospects. Energy & Environment, 35(7), 3833-3879.

[30] Yan, Y., Borhani, T. N., Subraveti, S. G., Pai, K. N., Prasad, V., Rajendran, A., ... & Clough, P. T. (2021). Harnessing the power of machine learning for carbon capture, utilization, and storage (CCUS)–a state-of-the-art review. Energy & Environmental Science, 14(12), 6122-6157.