Forecasting Future Renewable Energy Trends: A Comparative Analysis of Various Models

ROHIT[1], SAAHIL MISHRA[2] and SHRIYA UDUPA K[3]

1. *22BCS100, B.TECH, CSE, IIIT DHARWAD*
2. *22BCS105, B.TECH, CSE, IIIT DHARWAD*
3. *22BCS121, B.TECH, CSE, IIIT DHARWAD*

Abstract

In an increasingly renewable energy-oriented world, understanding current energy consumption patterns and projecting future trends is crucial. This article examines the use of solar and hydroelectric power in various countries, focusing on how economic indicators such as GDP and population influence consumption. Comparison of forecasting models, including the Random Forest Regressor (RFR), ARIMA, and Linear Regression, highlighting that RFR is the most accurate. The analysis uses synthetic data and K-fold cross-validation to enhance model robustness and prevent over fitting. XGBoost emerged as the leader in the prediction of future energy prices, with an advanced stacking ensemble approach that further improves prediction accuracy. Through visualizations, model comparisons, and forecasting techniques, this work aims to offer a comprehensive view of renewable energy trends, supporting sustainable energy policies and practices.

Introduction

In an increasingly renewable energy-oriented world, knowing the patterns of current energy consumption and how to project those patterns has never been more important. Two champions of this new direction are solar and hydroelectric power, replacing the old fossil fuel economy with much-needed alternatives. This article really delves cross-country on use, looking at some of the drivers of this consumption through GDP and population. Here is what we present at the table:

Key visual insights into solar and hydroelectric usage: We dissect the trends on solar and hydroelectric power consumption in various countries, paying major attention to key patterns that stand out.

This can be done using an economic impact analysis by comparing GDP and population with solar consumption, thereby getting a clearer view of how economic indicators tie into renewable energy use: basically, whether higher GDP or larger populations correlate with a boost in solar energy uptake.

Model comparison for solar forecasting: We are comparing the three big hitters in machine learning: Random Forest Regressor (RFR), ARIMA, and Linear Regression. Which one of these models will better predict solar energy consumption? RFR always wins the game in terms of accuracy.

Synthetic data and K-fold cross-validation: We generated synthetic data and applied K-fold cross-validation to make our models more robust so that there was no chance of overfitting.

The winning model is XGBoost since it predicts with the greatest accuracy the prices of future energy. Advanced Stacking ensemble: By fine-tuning our predictions through stacking ensemble techniques, we squeeze the highest possible accuracy in forecasting solar consumption, especially in synthetic data.

This is through a combination of visualizations, model comparisons, and forecasting techniques to unravel a more comprehensive picture of the trends in renewable energy sources around the world, hence providing insights that may push sustainable energy policies and practices.

Literature Review

Energy forecasting has gotten a lot better by mixing traditional methods with data about how people actually live. For example, in Ghana, a hybrid ARIMA-SVM model hit 96 percent accuracy by looking at household energy use and lifestyle data, though it struggled a bit when applied outside its region. Over in Europe, studies focus heavily on renewable energy and group countries by their green efforts. This shows how forecasting works well when tailored, but there's still a need for models that can tackle global trends effectively.

A study in the Journal of Electrical Systems and Information Technology uses a hybrid **ARIMA-SVM** model to predict household energy consumption with 96 percent accuracy. The model combines ARIMA's time-series capabilities with SVM's ability to detect nonlinear patterns, but it's geographically limited and doesn't include appliance-related data. The study also explores other methods like **KNN, GANs, and PCA** for anomaly detection. It highlights the importance of metrics like **RMSE and MAPE** and suggests that expanding the dataset and including appliance data could improve accuracy[1] . We also studied the use of **Self-Organizing Maps (SOM)** to analyze energy data from EU countries. The methodology includes data pre-processing to clean and prepare it for analysis, followed by visualization techniques to identify patterns. Correlation analysis is performed to explore relationships between energy indicators. The study then groups countries based on these indicators to observe trends and shifts. SOM helps to reveal hidden patterns in complex, high-dimensional data, offering insights into energy transitions[2] .A study used a **model-based approach** to assess the feasibility of a 100 percent renewable energy transition for Nepal and Bhutan by 2050. It employs scenario analysis, comparing energy costs with and without accounting for greenhouse gas costs. The model uses solar photovoltaics and hydropower as key components for energy generation, demonstrating potential cost reductions. Sensitivity analyses are suggested for the assumptions, and higher-resolution renewable energy data would be beneficial for improving the robustness of the findings[3] . One study investigates Saudi Arabia's rising electricity demand and strategies aligned with Vision 2030 to address greenhouse gas emissions. It evaluates**renewable energy options** like solar, wind, and carbon capture technologies, aiming to **reduce reliance on petroleum-based energy**. The methodology includes assessing renewable energy potential and combined-cycle power plants. However, the study lacks detailed cost analyses and doesn't address energy storage issues. Future research could benefit from including cost-benefit evaluations and decentralized energy solutions to strengthen the findings[4].There is also a study that introduces **Dynamic Modeling with Memory (DMWM)** to improve industrial energy predictions by addressing concept drift. This deep learning approach adapts better to changing datasets, supporting global energy efficiency goals. However, the paper lacks real-world applications and detailed scalability analysis. Adding case studies and computational complexity assessments would enhance the method's practicality for industrial use[5]. A study is such that it focuses on **HVAC** systems that make up 80 percent of energy use. It categorizes modeling approaches into physical, data-driven, and **physical-data fusion methods**, highlighting AI and smart sensors. Despite offering a comprehensive comparison, the study lacks real-world applications and scalability discussions. Including case studies and analyzing computational costs for large-scale use would make the findings more relevant[6].There is an article that evaluates greenhouse gas emissions in Saudi Arabia's cement industry using the **Vector Error Correction Model (VECM)** to analyze emissions and economic factors. It employs spatial and temporal modeling to assess trends and evaluate mitigation strategies like carbon capture. However, the study is mostly theoretical and lacks real-world validation. Incorporating pilot projects and cost-benefit analyses would provide practical insights and demonstrate the feasibility of the proposed strategies[7].

Usage of anonymized cellular network data to predict energy demand in Northern Italy involves analyzing human mobility patterns as a proxy for energy consumption. While the approach is effective, expanding the study geographically and temporally, and incorporating data from multiple telecom providers, would improve its generalizability[8]. One of the studies evaluates the savings potential of hybrid photovoltaic systems with energy storage across 32 European countries. The research uses models to assess various **PV capacities and storage configurations**, but lacks financial analyses and focuses only on capital cities. Future research should include cost-effectiveness assessments and broader regional data to enhance policy relevance[9]. Also there is an introduction to a model in one of the references which is used for predicting energy use in electric buses using **real-time data** from Malatya, Turkey. The study is limited by its geographic scope and lacks cost analysis. Expanding the study to other cities and including economic evaluations would increase its practical utility[10]. The study from India analyzes the impact of **air conditioning usage** on electricity demand, based on data from Hyderabad. While offering useful policy insights, the study is limited by its small sample size and focus on a single city. Expanding the dataset to other regions and incorporating additional variables, such as building insulation, would improve its applicability[11]. There is a study that uses a system dynamics model to **examine fuel consumption and emissions from road transportation** in Padang, Indonesia. It predicts a 34 percent reduction in emissions by 2050 through strategies such as improving public transportation. However, the study's focus on a single city and reliance on historical data limit its broader applicability. Expanding the model to include renewable energy scenarios and applying it to other regions would provide a more comprehensive understanding of transportation's role in emissions reduction[12].

To sum up, these studies show a lot of progress in predicting and managing energy use, but they still have some limitations. Many of the models, like the hybrid ARIMA-SVM, Self-Organizing Maps (SOM), and dynamic modeling methods, work well in specific situations but are limited by narrow geographic scopes, lack of real-world data, and missing key details. To improve, future research should expand these models to include more regions, update datasets with more detailed information, and consider the economic impacts. This would make the findings more relevant and useful for global energy policies and transitions

Regional Studies

Studies in Nepal and Bhutan have demonstrated the feasibility of a complete renewable energy transition by 2050, emphasizing solar photovoltaics and hydropower[3]. Saudi Arabia's Vision 2030 leverages renewable energy strategies to address rising electricity demand and greenhouse gas emissions[4].

Methodology

Data Preprocessing

To lay the groundwork for effective modeling, we executed a rigorous data preprocessing pipeline. This step ensured that the data was well-prepared, reliable, and suitable for model training. The preprocessing stage encompassed:

- **Data Cleaning:** We identified and addressed missing values by employing imputation techniques or discarding data points, standardized entries for consistency, and verified the quality of data to prevent inconsistencies during analysis.
- **Feature Scaling:** All numerical features were scaled to have uniform distributions, typically using standardization or min-max scaling. This step ensures that models sensitive to feature magnitudes, such as linear models and gradient boosting, perform optimally without any feature dominating due to its scale.
- **Encoding:** Categorical variables were transformed into numerical formats through techniques like one-hot encoding or label encoding, facilitating their use with machine learning algorithms that require numerical input.

To further enhance the model's ability to generalize, we introduced synthetic data. This synthetic augmentation improved model resilience, prevented overfitting, and strengthened the model's ability to adapt to new, unseen data. The incorporation of synthetic data played a pivotal role in preparing the model for real-world deployment.

Modeling Approach

The modeling phase involved constructing and training several machine learning models. The flowchart below illustrates the complete modeling pipeline, from data collection to prediction. Each step and component is described in detail:

Model Training: We trained the following models, chosen for their capabilities in handling different aspects of our data:

- **Random Forest (RF):** An ensemble method that leverages multiple decision trees to improve predictive accuracy and reduce over fitting. We also analyzed feature importance to identify which variables contributed most to the predictions.
- **XGBoost:** A highly efficient gradient boosting algorithm known for its robustness and performance in complex data sets. XGBoost was trained with hyper parameter tuning to optimize its learning rate, max depth, and regularization parameters.
- **Linear Regression (LR):** A simple model used as a baseline to assess the relationship between independent and dependent variables. We performed coefficient analysis to interpret feature contributions and validate linearity assumptions.

Model Evaluation: Post-training, we evaluated the models using a combination of performance metrics:

- **Root Mean Squared Error (RMSE):** A critical metric for measuring the average magnitude of prediction errors. It provides insight into how well the model predicts actual values, with lower RMSE values indicating better performance.
- **R-squared (R^2) Score:** Used to determine the proportion of variance in the dependent variable explained by the independent variables. This metric helps in assessing the overall fit of the regression model.
- **Precision and Recall:** For classification tasks, these metrics provided insights into the model's accuracy in identifying true positives and the ability to capture all relevant cases, respectively.
- **F1-Score:** The harmonic mean of precision and recall, providing a balance between the two for classification problems.

Prediction: The final step in the process was deploying the best-performing model to make predictions on new, unseen data. This step ensured the model's practical application for real-world scenarios and validated its effectiveness beyond the training set.

Dataset

For our project, we utilized a comprehensive dataset from Kaggle that offered global and country-specific energy consumption data, encompassing various energy sources with a strong emphasis on solar and hydroelectric energy. This dataset facilitated the analysis of energy consumption patterns and trends at both global and regional scales. Our primary focus was on solar energy consumption, and we conducted cross-country comparisons to understand the influence of factors such as GDP and population on the adoption and utilization of solar power.

In addition to the global dataset, we integrated data from reputable government sources, specifically the Ministry of New and Renewable Energy (MNRE) and the Indian Government's data portal. These sources provided detailed insights into energy consumption in India, with a particular focus on solar energy. The datasets included important metrics such as solar consumption, Levelized Cost of Electricity (LCOE), and pricing data. By filtering and curating the data to focus exclusively on solar energy, we were able to conduct a nuanced analysis of renewable energy trends within India and juxtapose them with global trends. This dual-source approach offered a comprehensive perspective on energy consumption, enhancing our ability to build a model capable of forecasting solar energy usage while accounting for economic and regional variations.

Data Preparation

Data cleaning was a pivotal step in preparing the dataset for accurate machine learning predictions. We began by addressing missing values using imputation techniques: the mean or median was employed for numerical columns, while the mode or placeholders were used for categorical data to maintain consistency and data integrity. Specifically, we focused on refining the solar consumption data and extracting relevant features from the Indian government datasets, such as LCOE and pricing information. This allowed us to construct a clean and well-targeted dataset that was suitable for in-depth analysis and machine learning model development.

The sources for our analysis were as follows:

Global Energy Data: The dataset from Kaggle, which provided comprehensive energy consumption data Kaggle Energy Dataset.

Indian Energy Data: Data from the Ministry of New and Renewable Energy (MNRE), which provided authoritative and region-specific energy statistics MNRE Renewable Energy Statistics, and data from the Indian Government's data portal Government Data Portal.

Afterward, we merged data from various sources, ensuring that all formats were consistent and removing any duplicates. We also handled outliers using visualization tools like box plots, deciding whether to remove or cap them based on their impact on model accuracy. Feature scaling was applied to ensure that all data was on the same scale, which was essential for machine learning models like Random Forest and Gradient Boosting. Finally, we checked the data for any formatting issues, ensuring that it was clean and ready for use in our predictive models, enabling more reliable insights into solar consumption trends.

- Solar Consumption Prediction Models
  - Linear Regression: This model was our go-to for a straightforward baseline, offering a simple prediction method to compare against more complex approaches. Linear regression finds the best-fitting line through the data by minimizing the sum of squared errors.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon \quad (1)$$

Where:

* y: Dependent variable (e.g., solar consumption)
* $x_1, x_2, \ldots, x_n$: Independent variables
* β_0: Intercept
* $\beta_1, \beta_2, \ldots, \beta_n$: Coefficients
* ϵ: Error term

The coefficients β_i are determined by minimizing the residual sum of squares (RSS):

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (2)$$

  - Random Forest Regressor (RFR): It quickly emerged as the top performer, effectively handling the complex and varied patterns in solar consumption. Random Forest's ability to work with non-linear data really made a difference here. It leverages an ensemble of decision trees, where each tree is trained on a random subset of the data and features, and predictions are averaged to improve robustness and reduce overfitting.
  - ARIMA: We tested ARIMA for its time-series capabilities, but it struggled with high error rates in this case, making it less reliable for our needs. The model's formulation:

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} \\ &+ \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t \end{aligned} \quad (3)$$

Where ϕ and θ represent the autoregressive and moving average coefficients, respectively, and ϵ_t is the white noise error term.

- Price Prediction Models
  - Linear Regression and Random Forest: While Random Forest brought versatility, Linear Regression ended up providing better accuracy in price predictions, with lower MSE and a stronger R^2 score. This was especially useful for understanding broader trends in pricing.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (5)$$

Where:

* MSE: Mean Squared Error
* R^2: Coefficient of Determination
* y_i: Actual values
* $\hat{y}_i$: Predicted values
* $\bar{y}$: Mean of actual values

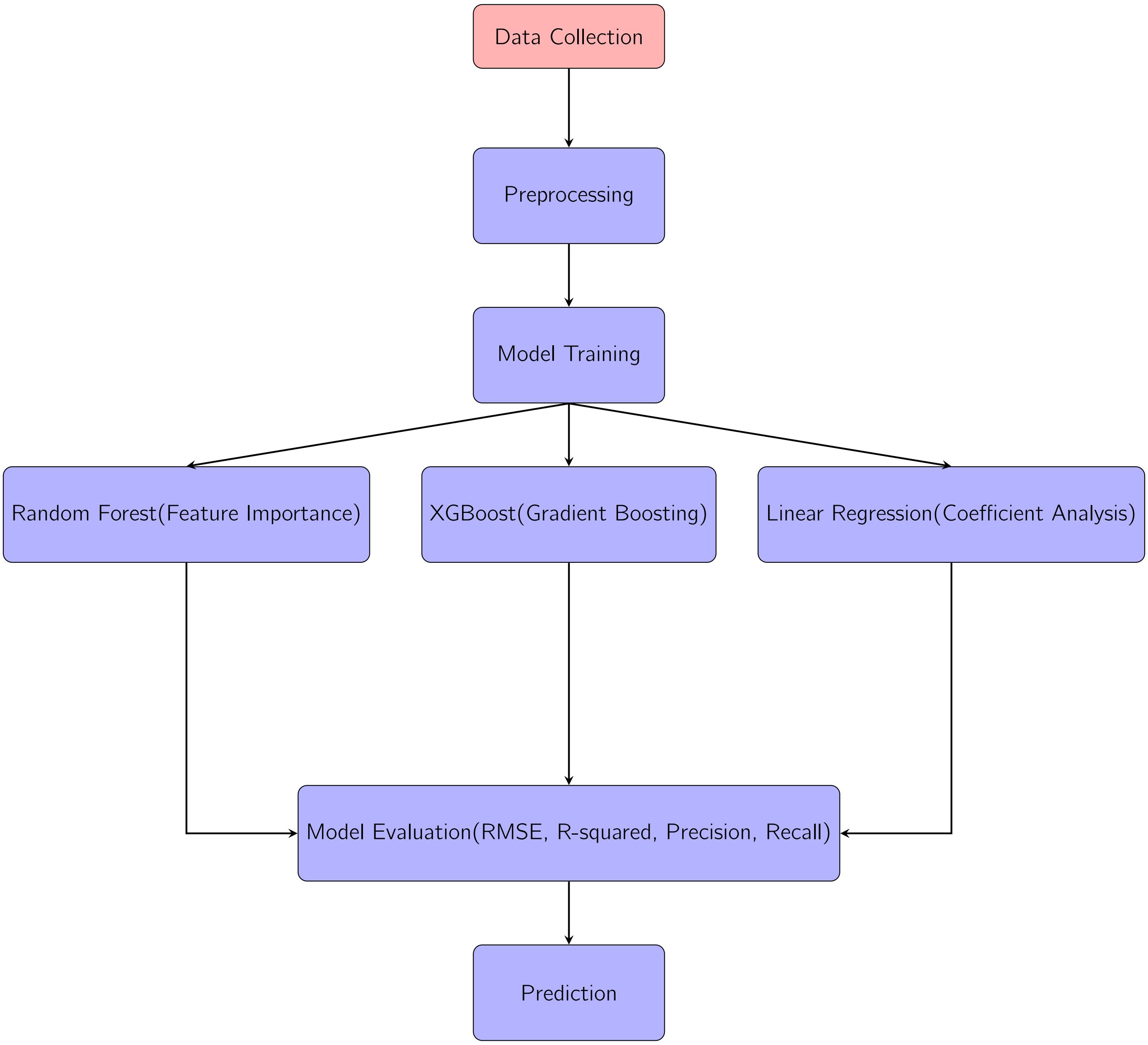


Figure 1: Complete Machine Learning Workflow

The R^2 score indicates how well the model explains the variability in the data. A higher R^2 value close to 1 signifies a better fit.

- Gradient Boosting: Another robust model for forecasting solar consumption, delivering strong results. Gradient Boosting constructs an additive model in a forward stage-wise fashion, where each new tree minimizes the residuals of the combined ensemble.

- XGBoost: Out shined the other models when it came to predicting prices. Its ability to fine-tune itself through iterative learning gave us a high level of accuracy, capturing the details in price fluctuations. XGBoost applies gradient boosting with regularization terms:

$$L = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{k} \Omega(f_j) \tag{6}$$

  Where $\Omega(f_j)$ is the complexity term for each tree f_j, and λ is a regularization parameter to control overfitting.

- Model Optimization

  - K-Fold Cross-Validation: We used K-Fold Cross-Validation to make our models more reliable by splitting and testing across various data folds. This way, the model learns to generalize well, reducing the risk of over-fitting on any single data subset. The dataset is divided into k folds, and each fold is used once as a validation set while the remaining $k-1$ folds are used for training.

– Stacking Ensemble: To get the best of all models, we stacked Linear Regression and RFR into an ensemble. This combined model allowed us to capture each model's strengths, resulting in more accurate and balanced predictions for solar consumption. The final prediction is computed as:

$$\hat{y}_{\text{final}} = \alpha_1 \hat{y}_1 + \alpha_2 \hat{y}_2 + \cdots + \alpha_n \hat{y}_n \quad (7)$$

Where α_i are the weights assigned to each base model $\hat{y}_i$, optimized to minimize the overall error.

Results and Analysis

Visualization of Energy Trends

We visualized the trends in solar and hydroelectric consumption in multiple countries, giving a clear picture of how energy sources compare globally. Scatter plots showed how GDP and population relate to solar consumption, helping us to see any direct correlations or trends between these factors. Figure 2 and Figure 3 provides us with the visualization of the global data.Figure 2 we have compared hydroelectricity and solar energy consumptions across different countries Figure 4 provides us with the socioeconomic comparison of India and its solar energy consumption.The metrics taken are GDP and population.

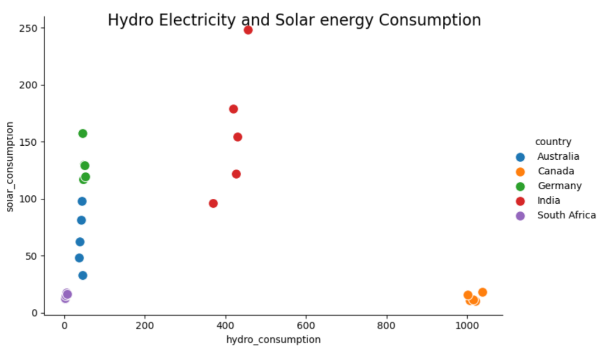


Figure 2: Hydro vs Solar

Figure 3 provides us with the visualization on the global data where we compare usage of solar energy by different countries over the years.

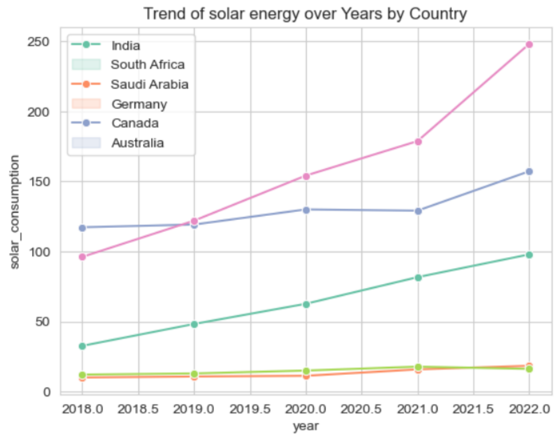
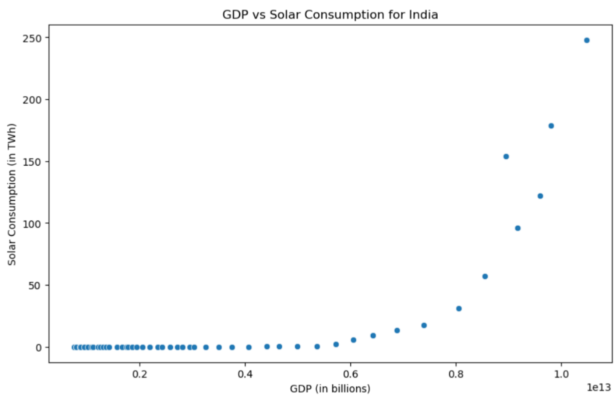
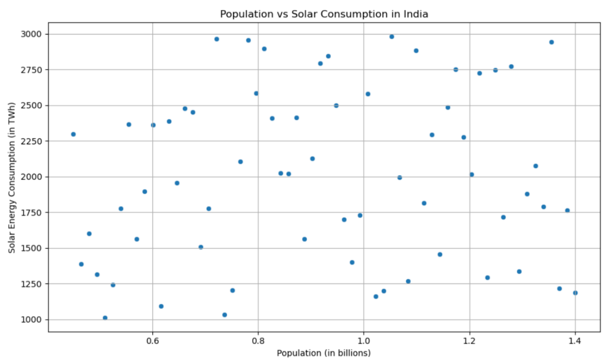


Figure 3: Solar consumption over different countries

Figure 4 provides us with the socio-economic comparison of India and its solar energy consumption.



(a) Solar consumption against GDP of India



(b) Solar consumption against Population of India

Figure 4: Comparison of Indian Energy Data

Solar Consumption Prediction

- Random Forest Regressor: Came through as the best model for solar consumption, scoring an R^2 of 0.65 and MSE of 2,076.75, proving its strength in prediction accuracy.

- ARIMA: Didn't perform well here, with a negative R^2 and high MSE, confirming it's not a good fit for this dataset.

- Linear Regression: Decent but limited, explaining only 14 percent of the variance and with higher error rates compared to RFR.

Model	Metric	Train Result	Test Result
Random Forest Regression	MSE	19968.09	45530.25
	R^2	0.9196	0.7337
Linear Regression	MSE	671.54	5756.48
	R^2	0.9973	0.9663

Table 1: Solar Consumption Model Performance Comparison

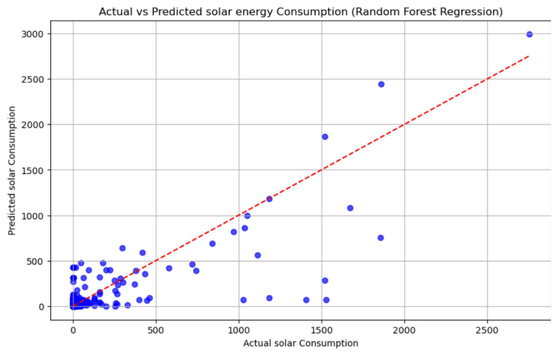


Figure 5: Random Forest Regressor

The Figure 5 describes the prediction given by the random forest regressor.This provided a better result compared to the other models used for predicting consumption on world data

Indian Dataset

Figure 6 provides us with comparison of generated power and the consumed power over the years in India Figure 7 provides us with the results of prediction of solar consumption with linear regression on Indian dataset.As analyzed we came to a conclusion with the historical data that over the years as the consumption increased the price of consumption decreased.

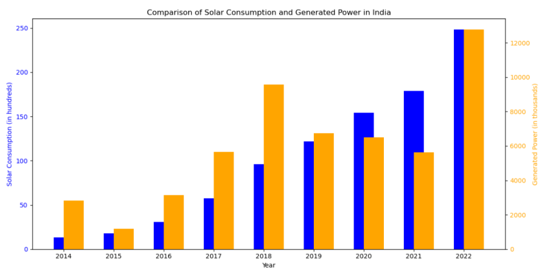
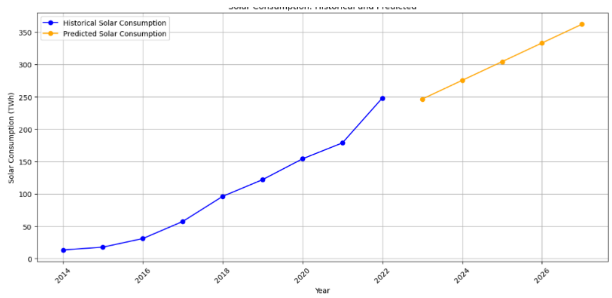


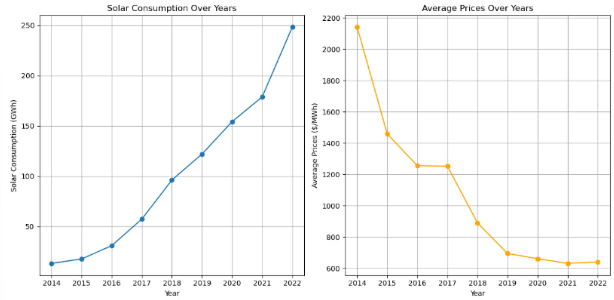
Figure 6: Solar power generated vs Solar electricity consumed over the years

Price Prediction

- Linear Regression: Linear Regression performed exceptionally well, achieving an impressive R^2 of 0.966 and an MSE of 5,756.48. Its high R^2 indicates that it explains the data trends effectively, making it a reliable option for understanding broader pricing patterns. However, while it's great for interpretation, it might struggle with more complex, non-linear relationships in the data.



(a) Solar consumption prediction with linear regression



(b) Solar consumption and prices over the years

Figure 7: Comparison of Indian Solar Energy Data solar consumption

- Random Forest Regressor: Random Forest did not perform as expected, with an MSE of 45,530.25 and a relatively low R^2 of 0.734. This suggests it couldn't capture the underlying patterns in price fluctuations effectively. Its tendency to overfit on training data might explain its weaker results in price prediction. Hyperparameter tuning or adding more relevant features could help improve its performance in future iterations.
- XGBoost: XGBoost emerged as the most accurate model, with a much lower MSE of 2,281.77 and MAE of 33.60, showing that it makes fewer errors than the other models. Although Linear Regression had a higher R^2, XGBoost's ability to handle complex relationships and minimize prediction errors makes it the better choice for precise forecasting. Its iterative learning process ensures better adaptability to diverse datasets, especially for trends with subtle non-linear variations. We considered XGBoost for minimizing prediction errors.XGBoost's robust handling of missing data, regularization to prevent overfitting, and ability to capture intricate patterns gave it an edge over Random Forest and Linear Regression. It strikes a balance between simplicity and complexity, making it ideal for both accuracy and

reliability.

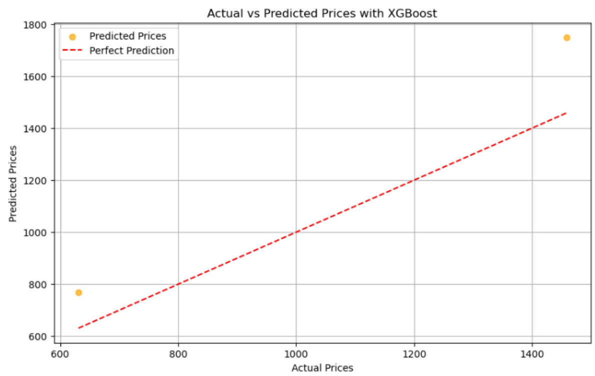


Figure 8: Solar prices prediction with XGBOOST

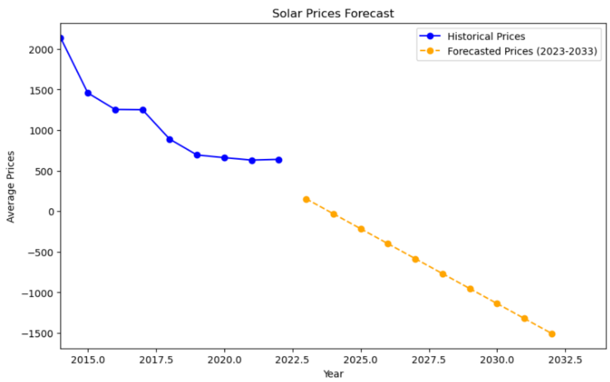


Figure 9: Solar prices over the years

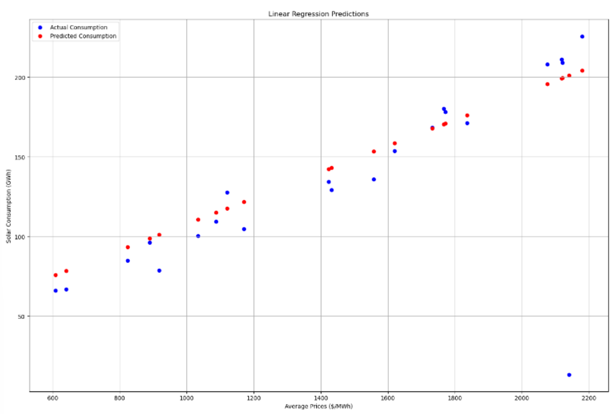


Figure 10: Prediction of Solar consumption and prices with linear regression

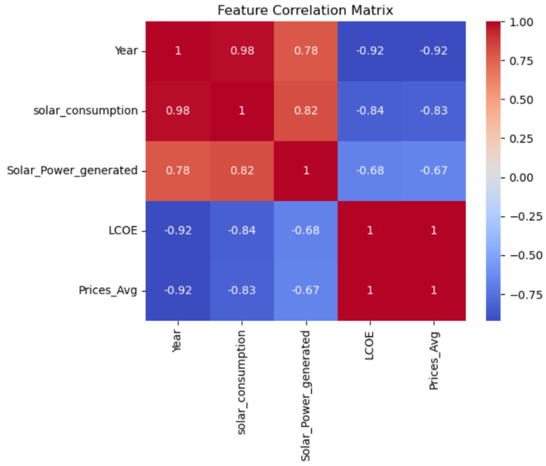


Figure 11: Correlation heatmap after cross-validation

Figure 8 gives us the highlights of the trend of solar prices over time, further validating the ability of XGBoost to capture dynamic variations accurately. While XGBoost is the most accurate here, combining it with Linear Regression in a stacking ensemble might leverage the strengths of both models—Linear Regression's interpretability and XGBoost's precision. Additionally, expanding the dataset with more economic and policy-related variables could further enhance accuracy.

Model	**Metric**	**Train Result**	**Test Result**
XGBoost	R^2	0.6510	0.6481
	MSE	1963.9307	2281.7702
	MAE	33.5979	47.7649
Linear Regression	R^2	0.8218	0.8800
	MSE	44249.0204	20523.3163
	MAE	179.5102	139.7857

Table 2: Model Performance Comparison

Comparison of solar consumption with prices

We also compared the solar consumption in India against the prices and also forecast future possible prices

- We used linear regression ,Random forest and Stacking Ensemble to predict future consumption and prices
- Out of the three models linear regression worked better with high accuracy and lower error rates compared to the others

Figure 10 shows us the prediction of solar consumption against prices under linear regression which provided with better accuracy than other models

Synthetic Data Analysis

Adding synthetic data and running cross-validation really boosted the model's overall performance. Gradient

Model	**Metric**	**Result**
Linear Regression	MSE	1735.88
	MAE	18.52
	R^2	0.42
Random Forest Regressor	MSE	2803.55
	MAE	25.63
	R^2	0.06
Stacking Ensemble Model	MSE	2333.05
	MAE	23.34
	R^2	0.22

Table 3: Performance Comparison of Linear Regression, Random Forest Regressor, and Stacking Ensemble Model

Boosting and the stacking ensemble models, in particular, benefited from this, delivering more consistent and robust predictions.

Model Performance

Table 4 highlights the performance of various models in solar energy prediction.

Table 4: Model Performance Comparison

Model	MSE	MAE	R^2
Random Forest	2076.75	25.63	0.65
Linear Regression	5756.48	18.52	0.42
XGBoost	2281.77	33.60	0.65

Conclusions

This project analyzed machine learning models for predicting solar consumption and prices in India. Random Forest Regressor proved to be the most effective for forecasting solar consumption, while Linear Regression excelled in predicting solar prices. XGBoost also performed well for price predictions, offering smaller errors, while the Stacking Ensemble model, though useful, did not outperform Linear Regression. Socio-economic factors like GDP and population were found to be significant drivers of solar consumption. The forecasting suggests an increase in solar consumption in the future, with prices following distinct trends. Incorporating synthetic data improved model accuracy, and extending the analysis to other renewable sources like wind and hydro could provide a more comprehensive view of India's energy landscape. This project demonstrates the value of Random Forest and Linear Regression in predicting energy trends and offers insights for future research and policy decisions in the renewable energy sector.The metrics were important for comparing the different models and figuring out which one made the most accurate predictions, so we could rely on them for forecasting solar consumption and prices.

This project not only contributes to the body of knowledge in energy analytics but also provides practical tools for optimizing energy planning and investment in India. In the long term, it could serve as a foundation for more sophisticated energy models that consider a wider array of variables, helping to shape a more sustainable energy future.

References

1. Seidu Agbor Abdul Rauf, Adebayo F. Adekoya. (2023). Exploring Household Energy Consumption and Forecasting Models: An Insight into Lifestyle Factors and Green Energy Transition. *Journal of Electrical Systems and Information Technology*, Springer.

2. Cristian Bucur, Bogdan George Tudorică, Simona-Vasilica Oprea, Dumitru Nancu, Dorel Mihail Duşmănescu. (2021). Insights Into Energy Indicators Analytics Towards European Green Energy Transition Using Statistics and Self-Organizing Maps. *IEEE Access*, IEEE.

3. Ashish Gulagi, Sanjeev Pathak, Dmitrii Bogdanov, Christian Breyer. (2021). Renewable Energy Transition for the Himalayan Countries Nepal and Bhutan: Pathways Towards Reliable, Affordable, and Sustainable Energy for All. *IEEE Access*, IEEE.

4. Syed Masiur Rahman, Fahad Saleh Mohammed Al-Ismail. (2021). Electricity Generation in Saudi Arabia: Tracing Opportunities and Challenges to Reducing Greenhouse Gas Emissions. *IEEE Access*, IEEE.

5. Abdulgani Kahraman, Muhammed Kotan, Mehmed Kantardzic. (2023). Dynamic Modeling with Integrated Concept Drift Detection for Predicting Real-Time Energy Consumption of Industrial Machines. *IEEE Transactions on Artificial Intelligence*, IEEE.

6. Xiao Han,Chaohai Zhang,Yujian Ye,Yi Tang. (2022). Physical-data Fusion Modeling Method for Energy Consumption Analysis of Smart Building. *IEEE Internet of Things Journal*, IEEE.

7. Zaid A. Khan, Babatunde A. Salami (2023). Dynamics of Greenhouse Gas Emissions from Cement Industries in Saudi Arabia—Challenges and Opportunities. *IEEE Access*, IEEE.

8. Andrey Bogomolov, Bruno Lepri, Roberto Larcher, Fabrizio Antonelli, Fabio Pianesi and Alex Pentland . (2016). Energy Consumption Prediction Using People Dynamics Derived from Cellular Network Data. *EPJ Data Science*, Springer.

9. Henrik Zsiborács, Nóra Hegedüsné Baranyai, Gábor Pintér, András Vincze. (2023). Exploring the Grid Electricity Saving Potentials of Using Hybrid

PV Systems in European Countries: A Comparative Study. *IEEE Access*, IEEE.

10. Yunus Emre Ekici, Ozan Akdağ, Teoman Karadağ, Ahmet Arif Aydin. (2023). A Novel Energy Consumption Prediction Model of Electric Buses Using Real-Time Big Data from Route, Environment, and Vehicle Parameters. *IEEE Access*, IEEE.

11. Pavan Ramapragada, Dharani Tejaswini, Vishal Garg, Jyotirmay Mathur, Rajat Gupta. (2022). Investigation on Air Conditioning Load Patterns and Electricity Consumption of Typical Residential Buildings in Tropical Wet and Dry Climate in India. *Energy Informatics*, Springer.

12. Iwan Sukarno, Hiroshi Matsumoto, Lusi Susanti. (2016). Transportation Energy Consumption and Emissions: A View from the City of Indonesia. *Future Cities and Environment*, Springer Nature.