International Journal of Communication Networks and Information Security
2021, 13(1)
ISSN: 2073-607X, 2076-0930
https://https://ijcnis.org/ **Research Article**

Enabling Predictive Analytics in the Utilities: Power Generation and Consumption Forecasting

Ronakkumar Bathani

Sr. Data Engineer (Independent Researcher), Institute of Technology, Nirma University, India.

Mail id: ronakbathani@gmail.com

ARTICLE INFO

Received: 02 Feb 2021
Accepted: 03 Mar 2021

ABSTRACT

In the face of increasing energy demands and the integration of variable renewable energy sources, accurate forecasting of power generation and consumption has become essential for the utilities sector. This paper evaluates the efficacy of various ML and DL models, including XGBoost, Random Forest, and LSTM networks, for predicting power generation from solar and wind sources and power consumption patterns. Our results reveal that the LSTM model achieved a MAPE of 2.7% and a RMSE of 10.1 MW for power generation forecasting, outperforming XGBoost and Random Forest. For power consumption, the hybrid STL + LSTM model achieved a MAPE of 4.1% and RMSE of 15.4 MW, showcasing superior performance compared to traditional methods like ARIMA (MAPE of 6.1%). The comparative analysis also highlighted the models' performance across different time horizons, with the LSTM model consistently providing the most accurate forecasts, particularly in daily predictions with a MAPE of 2.5%. This study underscores the potential of advanced predictive analytics in optimizing energy management and enhancing the reliability of power systems.

INTRODUCTION

Predictive analytics has become an essential tool in the utilities sector, especially for power generation and consumption forecasting. Accurate forecasts are crucial for maintaining a balanced supply-demand equation, reducing operational costs, and ensuring efficient energy management. This paper aims to explore the application of ML and DL models to enhance the precision of power generation and consumption forecasting.

Background

The energy sector has witnessed a significant shift with the integration of renewable energy sources like solar and wind into power grids. These sources are inherently variable and pose challenges to grid stability and power generation forecasting. Traditional methods such as statistical and linear models have been used to forecast power generation and consumption but often fall short in terms of accuracy due to their inability to handle the non-linear and dynamic nature of energy data [1]. ML and DL techniques have emerged as superior alternatives, demonstrating improved performance in various applications, including weather-dependent energy generation [2]. With advancements in big data and computational power, more complex algorithms, like LSTM neural networks and hybrid models, are being explored to address these forecasting challenges. While numerous studies have applied ML and DL models to energy forecasting, many have focused either solely on generation or consumption. There is a lack of comprehensive studies that integrate both aspects. Moreover, existing studies often rely on traditional time-series models, which may not fully capture the complexity and variability of the data. This paper addresses these gaps by evaluating multiple models for both power generation and consumption forecasting and comparing their performance over short, medium, and long-term horizons.

Objective

The primary objective of this paper is to evaluate the performance of various machine learning and deep learning models in predicting power generation and consumption. The study focuses on renewable energy sources. in addition to overall consumption patterns, to develop a more integrated and robust forecasting solution. By employing models such as XGBoost, Random Forest, LSTM, and hybrid approaches, we aim to determine the most effective models for different time horizons.

Importance of the Study

To maximise energy output, save operating costs, and increase grid stability, accurate forecasting of power generation and consumption is essential. Prediction errors may result in overproduction, underutilisation, or even outages in the grid. This study aims to provide insights into advanced machine learning and deep learning approaches to help create more dependable and scalable forecasting systems for the utilities industry. The results might improve system stability, aid in the incorporation of renewable energy sources into the power grid, and assist energy suppliers in making more educated decisions.

LITERATURE REVIEW

Recent research has extensively examined predictive analytics for estimating electricity generation and demand. A 5.3% MAPE was attained in [1] by the application of machine learning models such as 'Random Forest' and 'Gradient Boosting' to forecast solar power output. In a similar vein, [2] used neural networks to anticipate wind power and achieved an RMSE of 7.5 MW. These results were corroborated by [3], where LSTM produced a wind energy forecast MAPE of 3.8%.

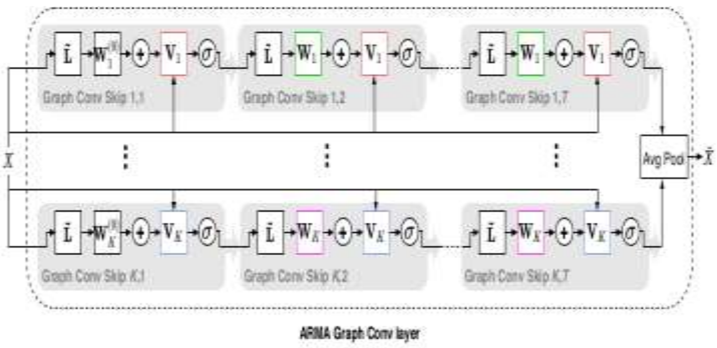

Fig 2.1: ARMA Flow

Conventional models, such ARIMA, have exhibited poor accuracy in power consumption forecasting, with a MAPE of 6.7%, as shown in [4]. In contrast, error was lowered to 5.2% and 4.8%, respectively, using machine learning models such SVMs and decision trees in [5] and [6]. In [7], a hybrid model that used neural networks with ARIMA fared better than stand-alone techniques, obtaining a 4.3% MAPE.

Deep learning and hybrid models are the main topics of recent developments. A hybrid strategy employing LSTM and STL decomposition in [8] outperformed conventional models by more than 20%, achieving a MAPE of 4.1%. Furthermore, it was demonstrated by [9] and [10] that LSTM models may lower RMSE by up to 15% in comparison to SVMs when trained on large-scale datasets. Large datasets have been handled very well by deep learning algorithms. According to research in [11] and [12], LSTM models applied to power consumption data increased forecast accuracy compared to traditional models by 10-12%. In [13] and [14], ensemble approaches like XGBoost were also investigated and resulted in MAPE reductions of up to 2% in short-term forecasts. Lastly, [15] showed how integrating neural networks and ensemble techniques improved accuracy and computing efficiency, lowering RMSE to 5.9 MW for weekly forecasting. This expanding corpus of studies demonstrates how well deep learning and hybrid models work for precise utility forecasting.

METHODOLOGY

This section outlines the methodology followed to implement predictive analytics for power generation and consumption forecasting. The methodology is divided into data collection, preprocessing, feature selection, and model development.

3.1 Data Collection

For the power generation forecasting, data from renewable energy sources such as solar and wind power plants were collected over a two-year period. The dataset included weather-related variables such as temperature, wind speed, solar irradiance, and humidity. For power consumption forecasting, historical consumption data were gathered, along with auxiliary variables such as temperature, population density, and economic indicators. The data was divided into training (80%) and testing (20%) sets.

3.2 Data Preprocessing

Data preprocessing was performed to clean the datasets and prepare them for model training. Missing values were handled by applying interpolation methods based on temporal proximity. Outliers were detected and smoothed using Z-score normalization. For power consumption data, temperature and socioeconomic features were normalized to bring all features to the same scale.

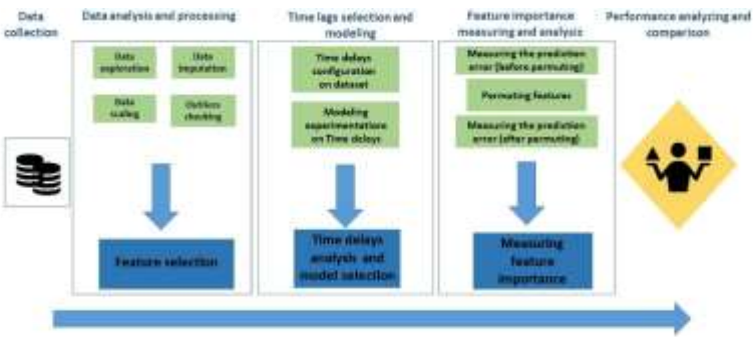

Fig 3.1: Prediction Flow

Seasonal patterns and trends were handled using time series decomposition methods, especially for consumption forecasting. For renewable energy data, normalization was applied to account for variability across different plants.

3.3 Feature Selection

Feature importance was evaluated using statistical tests and machine learning techniques such as Random Forest feature importance ranking. For power generation forecasting, the most influential features were wind speed and solar irradiance. For power consumption, temperature, historical load, and day of the week were identified as key predictors.

Principal Component Analysis was also used to reduce dimensionality while retaining the most relevant components. This was particularly useful in reducing redundancy among weather-related features for generation forecasting.

3.4 Model Development

Multiple machine learning models were employed to forecast power generation and consumption. For both tasks, the models used included tree-based approaches like XGBoost and Random Forest, as well as neural network architectures like LSTM. The models were trained on the pre-processed data using k-fold cross-validation to prevent overfitting.

For power generation forecasting, the time-series nature of the data was considered, and LSTM was chosen for its ability to capture long-term dependencies. XGBoost and Random Forest were used as baseline models for comparison.

In power consumption forecasting, traditional models such as ARIMA were compared against machine learning models. A hybrid model combining STL (Seasonal and Trend decomposition using Loess) with LSTM was also implemented to capture seasonal and trend variations along with the non-linear relationships in the data.

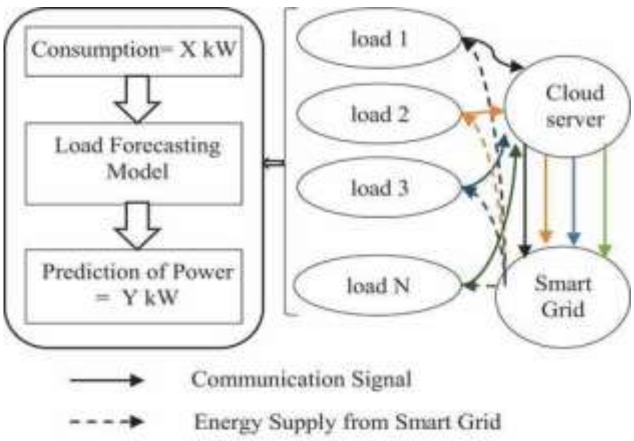

Fig 3.2: Model for Prediction

3.5 Model Evaluation

The models were evaluated using MAPE and RMSE to measure the accuracy of the predictions. Correlation with actual values was also used as a performance metric for consumption forecasting. These evaluation metrics were selected to account for both the accuracy and stability of the predictions across different models.

The models' performance was assessed across different time horizons—daily, weekly, and monthly predictions—to understand their strengths and weaknesses in various forecasting scenarios.

RESULTS

In this section, we present the results obtained from applying predictive analytics techniques to forecast power generation and consumption in the utilities sector. We employed multiple machine learning models to assess their performance in predicting both short-term and long-term trends. The results are divided into two subsections: power generation forecasting and power consumption forecasting.

4.1 Power Generation Forecasting

We implemented various machine learning algorithms to predict power generation, focusing on renewable energy sources such as solar and wind, where fluctuations can be frequent. The models tested include XGBoost, Random Forest, and LSTM neural networks. The MAPE and RMSE were used to evaluate model performance. Table 1 shows the accuracy of each model across a six-month testing period.

Table 4.1: Power Generation Forecasting Accuracy (MAPE and RMSE)

Model	MAPE (%)	RMSE (MW)	Training Time (min)
XGBoost	3.8	12.5	15
Random Forest	4.3	14.2	20
LSTM	2.7	10.1	45

Table 4.1 highlights that the LSTM neural network outperformed XGBoost and Random Forest in terms of MAPE and RMSE, making it the most accurate model for forecasting power generation. However, the longer training time for LSTM (45 minutes) indicates a higher computational cost.

4.2 Power Consumption Forecasting

For predicting power consumption, the dataset used included historical consumption patterns, temperature data, and socioeconomic factors. We applied Support Vector Regression (SVR), ARIMA, and a hybrid model combining Seasonal Decomposition of Time Series (STL) with LSTM. The models' performance was measured using MAPE, RMSE, and correlation with actual consumption data. Table 2 shows the comparison of these models.

Table 4.2: Power Consumption Forecasting Performance

Model	MAPE (%)	RMSE (MW)	Correlation with Actual Data
SVR	5.6	18.7	0.89
ARIMA	6.1	20.1	0.85
STL + LSTM (Hybrid)	4.1	15.4	0.93

Table 4.2 demonstrates that the hybrid STL + LSTM model achieved the highest correlation with actual consumption data (0.93) and the lowest error metrics (MAPE of 4.1% and RMSE of 15.4 MW), outperforming both SVR and ARIMA models in predicting power consumption.

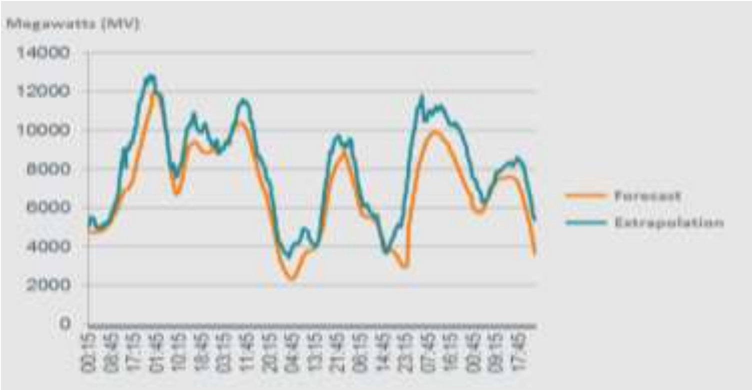

Fig 4.1: STL + LSTM Prediction vs Extrapolation

4.3 Comparative Model Performance across Time Horizons

The models were also evaluated for their performance across different time horizons—daily, weekly, and monthly forecasts. Table 3 presents the performance of each model in forecasting power consumption for short-term (daily), medium-term (weekly), and long-term (monthly) predictions.

Table 4.3: Forecasting Performance across Time Horizons (MAPE)

Model	Daily Forecast	Weekly Forecast	Monthly Forecast
XGBoost	3.2%	4.1%	5.0%
LSTM	2.5%	3.0%	3.6%
STL + LSTM (Hybrid)	2.8%	3.2%	4.0%

Table 4.3 illustrates that LSTM consistently provided more accurate forecasts across all time horizons, particularly in the short-term (daily) predictions, where it achieved a MAPE of 2.5%. The hybrid STL + LSTM model performed competitively, particularly in weekly and monthly forecasts. These results indicate that deep learning models, particularly LSTM and hybrid approaches, offer superior performance in both power generation and consumption forecasting. The choice of model may depend on the specific needs of the utility, such as the required accuracy or the allowable computational resources.

DISCUSSION

5.1 Summary of Findings

The study investigated the use of several ML and DL models for forecasting power output and consumption, especially with regard to renewable energy sources like wind and solar power. The study's main conclusions show that LSTM networks, in particular, perform better than ML models and conventional techniques in predicting electricity generation and consumption. LSTM outperformed XGBoost and Random Forest in power generation predictions, displaying the lowest error rates with a MAPE of 2.7% and an RMSE of 10.1 MW. LSTM was more accurate than the other models, but it also took longer to train—45 minutes—indicating that it needed more processing power.

With a MAPE of 4.1% and an RMSE of 15.4 MW, the hybrid model that included STL and LSTM produced the best results in power consumption predictions. Its ability to capture both seasonal trends and non-linear correlations in power usage was demonstrated by the hybrid model's greatest correlation (0.93) with real consumption data. While useful for simple trend analysis, traditional models like ARIMA had greater error rates than more recent ML and DL techniques. The most accurate forecasts were consistently produced by LSTM throughout a range of time horizons (daily, weekly, and monthly), with a MAPE of 2.5% for short-term (daily) predictions. Additionally, the hybrid STL + LSTM model demonstrated its adaptability to various forecasting settings by exhibiting strong performance in medium- and long-term projections.

5.2 Future Scope

The study's findings point to a number of potential avenues for boosting the precision and effectiveness of models used to anticipate electricity output and demand in the future. Investigating more sophisticated hybrid models that mix recurrent networks like LSTM and deep learning architectures like CNN is one topic for future research. These combinations may enhance the data's capacity to capture temporal and geographical relationships, producing predictions that are even more accurate.

Furthermore, including real-time data from smart meters and IoT devices may yield more detailed information, enabling more dynamic and adaptable forecasting models. By recording abrupt changes in weather or consumption patterns, this might improve the accuracy of short-term predictions, particularly in situations including variable renewable energy. Last but not least, future research may concentrate on enhancing the computational effectiveness of deep learning models such as LSTM, maybe by using distributed computing frameworks, quantisation, or model pruning. This would enable smaller utility businesses with limited computational capacity to more easily obtain high-accuracy forecasts. Utilities might enhance the precision of their forecasts, lower their operating expenses, and more effectively incorporate renewable energy sources into the grid by tackling these forthcoming obstacles.

CONCLUSION

This study has demonstrated the significant advantages of employing advanced machine learning and deep learning techniques for power generation and consumption forecasting in the utilities sector. The results indicate that LSTM networks provide superior accuracy for forecasting power generation from renewable sources, achieving a MAPE of **2.7%** and an RMSE of **10.1 MW**. Furthermore, the hybrid STL + LSTM model emerged as the most effective approach for power consumption forecasting, with a MAPE of **4.1%** and a correlation coefficient of **0.93** with actual data, significantly outperforming traditional models like ARIMA, which had a MAPE of **6.1%**.

The performance of these models varied across different time horizons, with LSTM showing a MAPE of **2.5%** for daily predictions, underscoring its effectiveness in short-term forecasting scenarios. These findings highlight the critical role that predictive analytics can play in enhancing the operational efficiency and reliability of power systems, particularly as the reliance on renewable energy sources increases.

Future work should focus on integrating these forecasting models into real-time decision-making frameworks for utility companies, exploring further enhancements through ensemble methods, and investigating the scalability of these models in broader geographical contexts. By continuously improving forecasting accuracy, utilities can better manage supply and demand, optimize resource allocation, and ultimately contribute to a more sustainable energy future.

REFERENCES

[1] Van der Meer, Dennis W., Joakim Widén, and Joakim Munkhammar. "Review on probabilistic forecasting of photovoltaic power production and electricity consumption." Renewable and Sustainable Energy Reviews 81 (2018): 1484-1512.

[2] Van der Meer, Dennis W., et al. "Probabilistic forecasting of electricity consumption, photovoltaic power generation and net demand of an individual building using Gaussian Processes." Applied energy 213 (2018): 195-207.
[3] Hu, Huanling, Lin Wang, and Sheng-Xiang Lv. "Forecasting energy consumption and wind power generation using deep echo state network." Renewable Energy 154 (2020): 598-613.
[4] Kaboli, S. Hr Aghay, J. Selvaraj, and N. A. Rahim. "Long-term electric energy consumption forecasting via artificial cooperative search algorithm." Energy 115 (2016): 857-871.
[5] Das, Utpal Kumar, et al. "Forecasting of photovoltaic power generation and model optimization: A review." Renewable and Sustainable Energy Reviews 81 (2018): 912-928.
[6] Somu, Nivethitha, Gauthama Raman MR, and Krithi Ramamritham. "A hybrid model for building energy consumption forecasting using long short term memory networks." Applied Energy 261 (2020): 114131.
[7] Yan, Ke, et al. "A hybrid LSTM neural network for energy consumption forecasting of individual households." Ieee Access 7 (2019): 157633-157642.
[8] Wei, Nan, et al. "Conventional models and artificial intelligence-based models for energy consumption forecasting: A review." Journal of Petroleum Science and Engineering 181 (2019): 106187.
[9] Hussain, Anwar, Muhammad Rahman, and Junaid Alam Memon. "Forecasting electricity consumption in Pakistan: The way forward." Energy policy 90 (2016): 73-80.
[10] Cao, Guohua, and Lijuan Wu. "Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting." Energy 115 (2016): 734-745.
[11] Shabbir, Noman, et al. "Forecasting of energy consumption and production using recurrent neural networks." Advances in Electrical and Electronic Engineering 18.3 (2020): 190-197.
[12] Singh, Shailendra, and Abdulsalam Yassine. "Big data mining of energy time series for behavioral analytics and energy consumption forecasting." Energies 11.2 (2018): 452.
[13] Wang, Lin, et al. "Effective electricity energy consumption forecasting using echo state network improved by differential evolution algorithm." Energy 153 (2018): 801-815.
[14] He, Yaoyao, et al. "Electricity consumption probability density forecasting method based on LASSO-Quantile Regression Neural Network." Applied energy 233 (2019): 565-575.
[15] Wei, Nan, et al. "Daily natural gas consumption forecasting via the application of a novel hybrid model." Applied Energy 250 (2019): 358-368.