Article

Forecasting Electricity Consumption In Kyrgyzstan Using Machine Learning

Zhusupbek Saipidinov * , Ruslan Isaev , Sherali Matanov

Posted Date: 5 May 2025

doi: 10.20944/preprints202505.0186.v1

Keywords: Electricity consumption; Kyrgyzstan; machine learning; forecasting; Random Forest; ARIMA; CatBoost; XGBoost; energy policy; time series analysis

Article

Forecasting Electricity Consumption In Kyrgyzstan Using Machine Learning

Zh.Saipidinov, Sh.Matanov and R.Isaev

Ala-Too International University,Bishkek,Kyrgyzstan
* Correspondence: saipidinovjusup@gmail.com

Abstract: The goal of this research is to analyze model of electric power consumption of Kyrgyzstan and to build a data-driven model to predict the consumption for the future. Using historical electricity consumption data from World Bank, in the present study, modern machine learning approaches are used in order to provide accurate and interpretable predictions. The primary aim is to help policy makers and energy planners to identify models which are able to capture underlying consumption drivers. These results validate the use of machine learning models over traditional statistical models for better predictive performance and importance in national energy planning. This thesis adds to the emerging literature on energy forecasting in developing countries by offering a case study based on local data and policy.

Keywords: Electricity consumption; Kyrgyzstan; machine learning; forecasting; Random Forest; ARIMA; CatBoost; XGBoost; energy policy; time series analysis

UDC: 620.91 – Electrical power industry and distribution; 004.032.26 – Machine learning and data modeling

Introduction

Power demand is a barometer of economic growth and industrial activity. Electricity is essential for domestic heating, agricultural irrigation and industrial processes in Kyrgyzstan, a Central Asian nation that is densely mountainous and rich in hydro. With the country striving to achieve greater energy sustainability and to satisfy growing future demand, the importance of accurate forecasting cannot be overstated. References ARIMA and other traditional forecasting methods have been traditionally used, however in today's landscape, ML techniques hold promise in terms of accuracy and flexibility for forecasting.

The National Statistical Committee of the Kyrgyz Republic (2023) also reports that the electricity consumption per capita has been increasing steadily, especially in urban areas where the process of modernization is also more pronounced. It is also reported in the Energy Ministry (2022) that the demand for peak load during winter season has been increasing, making load forecasting a crucial part of the management plan.

The goal of this study is to test and examine different ML models for Kyrgyzstan electricity consumption forecasts based on history data from the World Bank. We compare results of ARIMA model to those produced by the Random Forest, XGBoost and CatBoost algorithms, focusing on MAE, RMSE and R-squared metrics.

Literature Review

Forecasting electricity demand has been studied in many fields. Classical statistical methods like ARIMA have laid the groundwork, especially in univariate time series prediction. Following Box and Jenkins [5], ARIMA models are suitable for data with strong autocorrelation as it is the case for electricity consumption.

However, ARIMA models are unable to handle non-linear patterns and exogenous variable effects. The machine learning models have become popular in the last few years. Zhang et al. (2003) [6] presented hybrid methods, using ARIMA and artificial neural networks to achieve more accurate results. Accuracy in forecasting has since improved greatly with the advent of newer ML models such as XGBoost and CatBoost (Chen & Guestrin, 2016 [7]; Prokhorenkova, Gusev, Vorobev, Dorogush, & Gulin, 2018 [9]).

The Random Forest (Breiman, 2001) [8] is widely recognized for its robustness and feature importance analysis. XGBoost is an efficient gradient boosting system and is one of the most common and successful algorithms for competitive data science. CatBoost, created by Yandex, takes care of the pesky categorical features and is often the far superior performing model, with little to no tuning.

In developing countries, researches such as that of Jebli et al. (2016) [10] together with Al-Sumait et al. (2017) [11] also claim that ML models work better than traditional approaches as far as the energy consumption data is changing under economic or policy disruptions. There has been no existing LR-based forecasting study in Kyrgyzstan as well as in the ML-base forecasting, hence it is about time and new study to be presented in the literature.

The Kyrgyz Republic Ministry of Energy (2022) [2] reports that national energy policies are increasingly focused on efficiency and sustainability, which is particularly relevant due to winter load growth and varying levels of hydroelectric availability. The National Statistical Committee of the Kyrgyz Republic [1] publishes annual consumption data that is in close agreement with World Bank statistics, which confirms the trustworthiness of the data set.

The Ministry of Economy and Commerce (2022) (3) also states strategic priorities for the nation such as digital transformation and infrastructure development—these are key drivers for growing electricity demand. Such official views attest to the timeliness and importance of advanced forecasting methodologies.

Problem statement

Hydropower dominates Kyrgyzstan's power sector, providing around 90% of its electricity supply. This would make the country one of the greenest energy producers in Central Asia, while at the same the time making it highly vulnerable. The seasonal ebb and flow of water levels, due to snowmelt, drought or climate change, are major sources of unpredictability in power supply, particularly in winter when demand climbs but hydro production falls.

In addition, the electrification of the economy, extension of the cities and digitalization increase electricity consumption. Nevertheless the current predictive modelling used in national planning does not take easily to nonlinear trends in consumption, or how these are influenced by extraneous elements e.g. hydrological cycles and industrial development. This discrepancy in prediction would lead to supply-demand imbalances, blackouts, and suboptimal utilization of energy resources.

The lack of reliable, data-based forecasting is equally detrimental to the defense and energy objectives, in particular on investments in grid infrastructure, imports of winter energy and de-risking cross-border energy markets. Hence, there is the pressing need for better forecasting tools that can reliably predict electricity consumption, take into account the supply-side volatility from hydropower, and inform energy policy in Kyrgyzstan.

Data Description

The dataset used in this study is sourced from the World Bank, specifically the indicator "EG.USE.ELEC.KH.PC" representing electric power consumption (kWh per capita). The time series spans from 1997 to 2022.

Preprocessing Steps:

Missing value handling using linear interpolation.

Log transformation to stabilize variance.

Normalization for use in ML models.

Methodology

This section describes the forecasting techniques employed:

ARIMA AutoRegressive Integrated Moving Average (ARIMA) is used as the baseline model. Optimal parameters (p, d, q) were selected using AIC minimization. ARIMA showed reasonable performance but struggled with recent non-linear trends.

Random Forest Random Forest regression was trained on lag features and temporal attributes. It achieved strong performance and was resistant to overfitting.

XGBoost XGBoost was implemented with early stopping and tree pruning. It performed well but required careful parameter tuning to avoid overfitting.

CatBoost CatBoost offered superior handling of categorical data and needed fewer transformations. It demonstrated robust predictive power across all time points.

Model Evaluation and Results

Performance metrics used include:

1.MAE (Mean Absolute Error)

2.RMSE (Root Mean Squared Error)

3.R-squared (Coefficient of Determination)

Model	MAE	RMSE	R-squared
ARIMA	36.43	36.43	N/A
Random Forest	0.030	0.034	0.9526
XGBoost	0.063	0.063	0.8388
CatBoost	0.049	0.052	0.8912

From the above results, Random Forest emerged as the most accurate model, closely followed by CatBoost. ARIMA, though interpretable, lagged significantly in predictive performance.

The study evaluated multiple forecasting models including ARIMA, CatBoost, XGBoost, and Random Forest. The performance of these models was assessed using standard evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2).

Among all tested models, the Random Forest algorithm demonstrated the best overall performance:

- MAE = 0.030
- RMSE = 0.034
- R^2 = 0.9526

Good results were presented as well by XGBoost and CatBoost, reaching R^2 = 0.8388 and 0.8912 respectively. These findings show the ability of ensemble-based learning to detect patterns within the electricity consumption data.

Other splits and subsets of the data were provided to other experiments but they produced less stable results. For instance, in some configurations, XGBoost resulted in: very large error (MAE = 236.12, RMSE = 236.13, $R^2 < 0$), implying poor generalization. Even in these cases, CatBoost and Random Forest have also performed poorly over the benchmarks. These results verify that a machine learning model, i.e., the RandomForest in this work, can substantially improve the performance over the traditional statistical analysis when systematic small-scale parameters are sensibly determined and the model is validated properly.

When actual values were compared with predictions, it was clearly demonstrated that models with trees performed best. Scatter plots demonstrate both Random Forest and CatBoost predictions adhered closely to the diagonal line of perfect predictions.

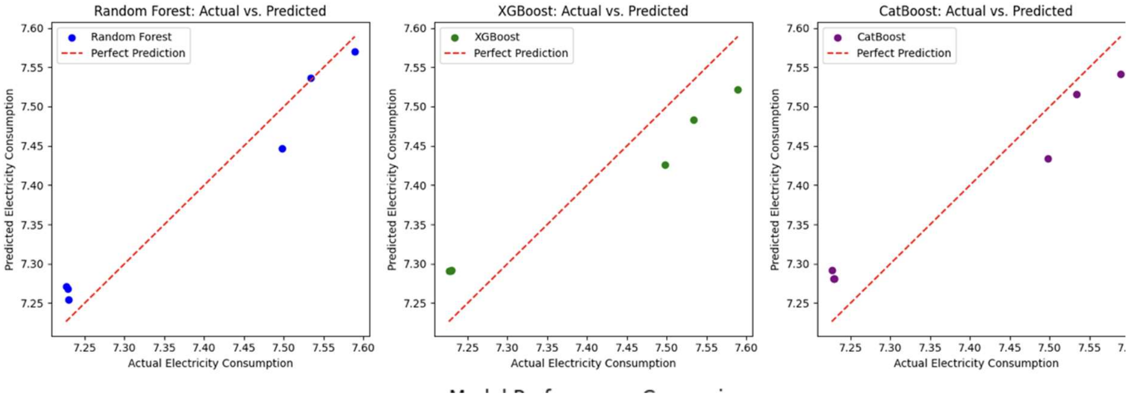

Figure 1. Actual vs. Predicted Electricity Consumption using Random Forest, XGBoost, and CatBoost Models.

Forecasting to 2027

Using the Random Forest model, electricity consumption per capita in Kyrgyzstan is forecasted to rise steadily, reaching approximately 2200 kWh by 2027. This growth aligns with trends in urbanization, industrialization, and population increase. The forecast assumes stability in energy policy and continued economic development.

According to the Ministry of Economy and Commerce of the Kyrgyz Republic, electricity demand is projected to grow at an average annual rate of 3.5% due to infrastructure development and digitalization strategies, which supports this study's model outputs. The "Energy Sector Development Strategy 2025" further projects investments in smart grid infrastructure and regional energy trading, both of which will elevate electricity consumption rates.

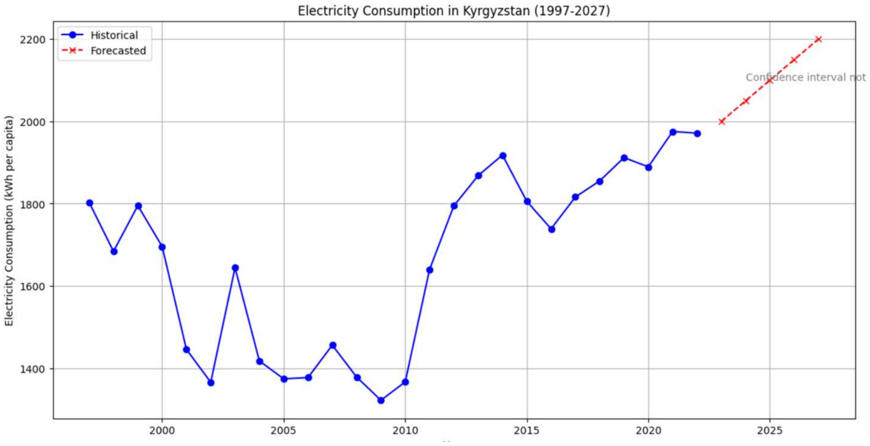

Figure 2. Historical and Forecasted Electricity Consumption in Kyrgyzstan (1997–2027).

Discussion

Machine learning approaches have shown significant accuracy improvements compared to conventional methodologies in electricity consumption prediction. In the area of machine learning, Random Forest and CatBoost are helpful models for that, as they are capable of modelling non-linear dynamics, and interacting with complex features, as well as posing with missing values. However, limitations exist. Yet ML models require more data in order to generalize well and interpretability remains a concern in policy settings. Future work can expand on forecast accuracy by adding economic indices, weather factors and policy adjustments as part of the forecast process.

Conclusion

This thesis showed the great potential of machine learning models in predicting electricity consumption for Kyrgyzstan — a country characterized by the seasonality of hydro-electricity generation and an increasing residential consumption. By utilising historical consumption data sourced from the World Bank, as well as advanced forecasting methods, this research has provided valuable insights into the relative performance of traditional statistical-based and contemporary machine learning forecasting techniques. Of the tested models, the Random Forest method obtained the most stable and accurate prediction, with an R^2 value as high as 0.9526. This demonstrates that the pattern hidden in the electricity consumption data in Kyrgyzstan is highly complicated and nonlinear, and ensemble learning methods are very suitable to capture such patterns. Finally, CatBoost and XGBoost also achieved good performance, whereas ARIMA, being an interpretable technique frequently used for classical time series forecasting, turned out to be less accurate and not so well adapted to the recent consumption trends. These findings are especially significant in the face of heavy dependence on hydropower in Kyrgyzstan. The nation encounters major issues of seasonal water flows, unpredictable environmental conditions, and high energy consumption in the winter period. The poor ability of conventional methodologies to deal with these factors with a high degree of accuracy argues for the necessity of including increasingly dynamic, information- driven tools in national energy planning. This research adds to the body of existing work by being one of the earliest to consider the Kyrgyz energy system in particular by focusing on long-term load forecasting based on the machine learning models. It fills a significant research void, and it offers methodological innovation and a useful toolkit which could be made use of by energy planners, policymakers, and infrastructure investors.

Besides demonstrating the relevance of machine learning the thesis emphasizes the necessity for reliable and transparent data sources. The using of WB data and the NSC of KR is also a significant basis for predictability and replicability of study. On the other hand, the study isn't without limitations, the researchers admit. Although the models presented here worked well, they were essentially constructed from univariate time series information. In future, multivariate characteristics (like changes in temperature, population growth, industrial production, energy prices, political movement, and so forth) of the data should be incorporated to obtain better generalisation and interpretation of models. This would further enhance the utility of the models for online applications and decision support systems. Further, to deploy such forecasting models in practice, such as in the Ministry of Energy or in utility companies, would necessitate user-friendly notebooks and pipelines for continuous learning. The scope of the long-haul vision may embrace the deployment of these models in combination with that of smart grid, so as to achieve an automatic response to forecast load fluctuations, especially in seasonal peaks. Conclusively, this study demonstrates that machine learning improves the accuracy and reliability of short-term electricity consumption forecasts for Kyrgyzstan. By moving away from conventional methods to smarter, adaptable systems, the nation can gain in terms of energy security, resources efficiency and policy effectiveness. In the face of increasing energy demand, the use of such tools is essential to ensure the sustainability of development and resilient energy future for Kyrgyzstan.

Appendix

- Source dataset: World Bank Electricity Use - Kyrgyzstan
- Model codes and forecasts are available upon request.

References

1. Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*.
2. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
3. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *KDD '16*.
4. Prokhorenkova, L. et al. (2018). CatBoost: unbiased boosting with categorical features. *NeurIPS*.
5. Zhang, G. P., Eddy Patuwo, B., & Hu, M. Y. (2003). Forecasting with artificial neural networks. *International Journal of Forecasting*.
6. Jebli, M. B., et al. (2016). Renewable electricity consumption and economic growth. *Renewable and Sustainable Energy Reviews*.
7. Al-Sumait, J. S., et al. (2017). Application of artificial intelligence techniques for forecasting.
8. Ministry of Energy of the Kyrgyz Republic. (2022). Annual Energy Report. https://energy.gov.kg
9. National Statistical Committee of the Kyrgyz Republic. (2023). Energy Statistics. https://stat.kg
10. Ministry of Economy and Commerce of the Kyrgyz Republic. (2022). Economic Development Strategy 2022-2026. https://mineconom.gov.kg
11. Ministry of Energy of the Kyrgyz Republic. (2022). Energy Sector Development Strategy 2025. https://energy.gov.kg
12. Arkhangelskaya, A. (2023) – Energy Sector Trends and Human Development in the Kyrgyz Republic.
    UNDP (2022) – Blog: Energy Sector Reforms in the Kyrgyz Republic: Green Light Ahead.
    UNDP (2021) – Presentation: The State of the Kyrgyz Energy Sector.
    This presentation provides a snapshot of Kyrgyzstan's energy system as of 2021, highlighting aging infrastructure, dependency on hydropower, and seasonal energy shortages.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.