Data Article

Title: Short-term forecasting of hierarchical time series in electricity consumption: An Application using South African data

Authors: [1]Mantombi Bashe, [2]Claris Shoko, [3]Thakhani Ravele and [3]Caston Sigauke
Affiliations:

[1]Eskom Holdings SOC Ltd, South Africa

[2]Department of Statistics, University of Botswana

[3]Department of Mathematical and Computational Sciences, University of Venda, Private Bag X5050, Thohoyandou, 0950, South Africa.

Contact email: caston.sigauke@univen.ac.za

Abstract
The article investigates the application of machine learning algorithms, Extreme Gradient Boosting (XGBoost) and Stochastic Gradient Boosting (SGB) for forecasting hierarchical time series data in the South African power generation sector. Additionally, the study explores reconciliation methods to ensure coherence in hierarchical forecasts. The data are stored in an Excel file.

Specifications Table

Subject area	*Hierarchical Time Series Forecasting*
More specific subject area	*Energy Modelling and Forecasting*
Type of data	*Excel file*
How data was acquired	*Provided and from the internet* https://www.eskom.co.za/dataportal/
Data format	*Filtered and analysed.*
Experimental factors	*N/A*
Experimental features	*N/A*
Data source location	*Eskom Data portal webpage*
Data accessibility	*The data is hosted on GitHub https://github.com/csigauke*
Related research article	*Short-term Forecasting of hierarchical time series in Power Generation: An Application using South African data*

Value of the Data

The data can be used for hierarchical time series forecasting in cross-sectional, temporal, or cross-temporal frameworks.

Data

The data used in this study is from Eskom, South Africa's power utility company, and it can be accessed from https://www.eskom.co.za/dataportal/. The data comprises renewable energy sources and non-renewable energies. From renewable energies, we have photovoltaic (PV), concentrated solar power (CSP), wind and other renewable energies. As for the non-renewable energies, we have thermal power, nuclear power, and gas (OCGT). All the data is measured hourly. The covariates used were hour of the day (hour), day of the week (day), nonlinear trend variable (noltrend) which we got by fitting the cubic regression spline model, and four differenced lagged data for each of the energy sources at lags 1,2,12 and 24, respectively. These covariates were used in the XGBoost and SGB models.

Experimental Design, Materials, and Methods

Data used in the study is from Eskom, South Africa's power utility company.

Acknowledgements

The authors acknowledge Eskom for providing the data.

Funding sources

No funding was provided for this project.

References

Eskom https://www.eskom.co.za/dataportal/ (Accessed on 20 June 2024).