**Optimal Forecasting of Energy Consumption Through LSTM Networks and Initial Seed Optimization**

Ethan Norton  
Northwestern University  
https://github.com/EthanNorton/MSDS-alignment/tree/MSDS-458  
EthanRNorton@protonmail.com  
July 20, 2025

# Abstract

This project leverages a Long Short-Term Memory (LSTM) neural network enhanced with SARIMA-informed seasonal weighting to forecast energy consumption. The analysis integrates engineered features from customer behavior and weather patterns. A seed sweep analysis across 50 initializations was conducted to explore the impact of model randomness on performance. Metrics such as R², MAE, RMSE, and MAPE were benchmarked to evaluate the effectiveness and reproducibility of results. Then, the average of 50 initializations was taken to determine if LSTM or SARIMA informed LSTM was a better model for the problem.

# Introduction and Problem Statement

The acceleration of large language models (LLMs) and AI workloads has placed demands on energy infrastructure, particularly from data centers and GPU-intensive computational environments. As AI adoption continues to scale, agents continue to be implemented, ensuring energy system stability, efficiency, and sustainability becomes a complex problem. Electricity consumption forecasting plays a crucial role in this context, demanding meticulous energy planning, distribution optimization, and infrastructure development. Accurate forecasting supports utilities and governments in enhancing grid reliability, managing peak loads, and anticipating demand fluctuations. Traditional statistical models such as ARIMA and linear regression often fall short in capturing the non-linear, temporal dynamics of energy usage. Deep learning models, such as Recurrent Neural Networks (RNNs) and their advanced variant, Long Short-Term Memory (LSTM) networks, which were introduced by Hochreiter and Schmidhuber, have emerged as powerful tools for modeling these complex sequential relationships. In this study, LSTM networks are implemented along with precise feature engineering techniques to predict energy consumption. Engineered variables such as revenue per customer, year-over-year consumption growth, and cyclical monthly patterns are incorporated to optimize model fit and improve interpretability. The integration of AI techniques for energy forecasting aligns with the broader need for intelligent, resource-efficient infrastructure in an era increasingly defined by AI. It is believed that applying deep learning to the energy landscape will contribute greatly to the development of further advancements in AI, creating a positive feedback loop.

# Background and Literature Review

As an introduction to gain background on the material, this paper was explored to evaluate the process of training the LSTM deep learning model, which is a subset of the RNN model (Hrnjica, Mehr). However, this methodology could be complex, and in order to provide a more accurate and robust model, the best solution in this is to combine it with modern deep learning theory.

# Data

Monthly data from the U.S. Energy Information Administration (EIA) was used, covering electricity sales, revenue, and customer counts, which was downloaded via API. Data was limited to a specific state-level granularity from January 2001 to December 2023, with 5 states (California, Texas, New York, Florida, Illinois), and 3 sectors (Residential, Commercial, Industrial) displaying 276 total observations.

A screen shot of a computer

AI-generated content may be incorrect.

Key derived features include:  
- Price per kilowatt-hour (revenue divided by sales).  
- Average consumption per customer.  
- Year-over-year consumption growth.  
- Sine/cosine cyclical month encoding.  
- Binary peak-season indicator.

- Weighing  
All data was aggregated at the monthly level and cleaned for missing values using forward-fill imputation. StandardScaler was used to normalize continuous variables. No dimensionality reduction methods such as PCA were applied.

# Research Design and Modeling Method(s)

This project uses a univariate and multivariate LSTM modeling approach implemented in TensorFlow/Keras. The model architecture includes:  
- One or more LSTM layers, which ranged for 32 to 128 layers.  
- Dropout regularization, which ranged from .2 to .3.   
- ReLU, sigmoid and tanh activation experiments  
- Dense linear output layer, ranging from 16 to 32.

A black screen with text

AI-generated content may be incorrect.  
Data is reshaped into overlapping 3D sequences of the form (samples, time steps, features). Models were trained using the Adam optimizer and MSE loss. A `ReduceLROnPlateau` learning rate scheduler and early stopping were used for training control.  
Several architectural variants were tested to observe the impact of activation function, window size, and dropout on model performance. The design logic focused on balancing capacity and regularization to avoid overfitting on limited training samples. The data was also split into a 80/20 train test ratio.

# Design and Implementation Considerations

Training and experimentation were performed in Cursor on an ipynb notebook. Libraries used include TensorFlow 2.x, scikit-learn, pandas, NumPy, and matplotlib for analysis and visualization. Keras callbacks were instrumental in tuning learning schedules and tracking overfitting. Random Seed Optimization was performed to discover which seeds led to higher r2, and thus, lower RMSE, SAE, and

# Results

An extended seed sensitivity analysis was conducted with 30 separate random seed initializations, and applied across different learning rate, dropout, and dense layer settings. This was done to better capture the interaction effects of hyperparameters with initialization randomness. The findings further reinforced the discovery that seed variance can dominate model behavior in smaller datasets, with R² varying as much as 0.12 between worst and best runs using the same architecture.

To visualize this, boxplots of R² scores across 30 seeds for each architecture were generated. The SARIMA-LSTM hybrid showed improved stability compared to vanilla LSTM, with a tighter interquartile range and consistently higher minimum performance. This supports the hypothesis that injecting SARIMA-informed seasonality acts as a regularizer for temporal noise.

Moreover, it was found that the use of summer binary indicators and sine/cosine encodings provided essential cyclical structure. This was especially useful when summer months (June through August) were weighted more heavily in training. Specifically, training with summer months having 1.5x or 2x sample weights yielded higher R² values during validation in warm-weather dominated states like California and Texas.

The feature engineering strategy—namely revenue per customer, YOY consumption growth, and price elasticity—showed high importance as confirmed by permutation importance metrics and SHAP value analysis, with average consumption per customer having the highest marginal contribution to model accuracy.

The results will compare various LSTM configurations using R-squared and RMSE on a holdout set. Additional plots and summary tables are forthcoming as experimentation concludes. The first model was constructed with optimization of evaluation ReLu and Sigmoid activation functions. A black and white screen with numbers

AI-generated content may be incorrect.

A black background with white numbers

AI-generated content may be incorrect.

The model below was built upon the highest performing r2 of the ReLu activation function combination. However, the R2 was different than above, likely due to seed initialization, which will be mentioned in the initial seed optimization section.

A graph with blue and orange dots

AI-generated content may be incorrect.

The next iterative process in the modeling was to apply a SARIMA-penalized model. This model performed slightly better, but also had variations based on seed initialization.

A graph with blue and orange lines

AI-generated content may be incorrect.

A graph with blue and orange dots

AI-generated content may be incorrect.

To address the high variance in LSTM model performance due to random seed initialization, a seed sensitivity analysis was conducted using 50 different seeds. Metrics evaluated included R², MAE, RMSE, and MAPE. The results demonstrated a nearly perfect negative correlation between R² and RMSE (-0.99), and strong correlations with MAE and MAPE (-0.97 and -0.94 respectively). This indicates that optimizing for R² also tends to improve other core forecasting metrics. The analysis underscores the critical impact of seed initialization in deep learning time series models, and the potential to use seed tuning as a lightweight optimization technique in practice. A correlation heatmap summarizing these findings is included below.

A red and blue squares with numbers

AI-generated content may be incorrect.

After this discovery was realized, the seeds were initialized for both the SARIMA-LSTM Model, and the LSTM Model.

A graph with a red line

AI-generated content may be incorrect.

A graph of blue and red lines

AI-generated content may be incorrect.

# Conclusions

By integrating SARIMA-based seasonal adjustment with LSTM and carefully tuning seed initialization, this study demonstrates a scalable framework for energy demand forecasting. The stability of the SARIMA-enhanced model under multiple random seeds confirms its practical robustness.

The feature set, particularly the cyclical and customer-derived variables, enabled the model to generalize across regional demand fluctuations. The seed analysis revealed that, despite deep learning’s stochastic nature, initialization can be controlled to ensure repeatability and reproducibility.

This lays the foundation for future hybrid pipelines where temporal priors (SARIMA, Fourier terms) are fused with modern architectures such as attention-based LSTMs or TCNs to further enhance interpretability and accuracy in energy time series modeling.

The LSTM model, adjusted with SARIMA-based summer weightings, demonstrated strong predictive performance. The best configuration achieved an R² of 0.91, with consistent behavior across other evaluation metrics. The seed sensitivity analysis confirms that initialization significantly impacts performance, but careful hyperparameter and weight handling can mitigate this. The pipeline provides a robust template for seasonal energy forecasting at scale.

# Future Direction

Another addition to this paper would be to apply 30 different random seeds to each learning rate, dense layer, dropout1 and dropout2.

Future work should integrate exogenous weather variables and grid-level macroeconomic indicators. Expanding the SARIMA adjustment into non-linear embeddings or attention mechanisms could improve robustness. Alternative deep architecture such as Temporal Convolutional Networks or Transformers are promising next steps.

# Bibliography

* Arjovsky, Martin, et al. 2018. “h-detach: Modifying the LSTM Gradient Towards Better Optimization.” OpenReview. https://openreview.net/forum?id=ryf7ioRqFX.
* Bouktif, Salah, Ali Fiaz, Ali Ouni, and Mohamed Adel Serhani. 2018. “Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches.” Energies 11, no. 7: 1636. https://doi.org/10.3390/en11071636.
* Choi, Eunjeong, Soohwan Cho, and Dong Keun Kim. 2020. “Power Demand Forecasting using Long Short-Term Memory (LSTM) Deep-Learning Model for Monitoring Energy Sustainability.” Sustainability 12, no. 3: 1109. https://doi.org/10.3390/su12031109.
* Emshagin, Saad, Wayes Koroni Halim, and Rasha Kashef. 2022. “Short-Term Prediction of Household Electricity Consumption Using Customized LSTM and GRU Models.” arXiv, December 16, 2022. https://arxiv.org/abs/2212.08757.
* Hochreiter, Sepp, and Jurgen Schmidhuber. 1997. “Long Short-Term Memory.” Neural Computation 9(8) (1735-1780): 1-32. https://www.bioinf.jku.at/publications/older/2604.pdf.
* Hrnjica, Bahrudin, and Ali Danandeh Mehr. 2019. “Energy Demand Forecasting Using Deep Learning.” In Smart Cities, Performability, Cognition, & Security pp 71-104. EAI/Springer Innovations in Communication and Computing. Cham: Springer International Publishing. doi:10.1007/978-3-030-14718-1\_4.
* Kaya, Mahmut, Anıl Utku, and Yavuz Canbay. 2024. “A Hybrid CNN-LSTM Model for Predicting Energy Consumption and Production Across Multiple Energy Sources.” Journal of Soft Computing and Artificial Intelligence 5, no. 2: 63–73. https://doi.org/10.55195/jscai.1577431.
* Kaur, Devinder, Shama Naz Islam, Md. Apel Mahmud, Md. Enamul Haque, and Zhao-Yang Dong. 2020. “Energy Forecasting in Smart Grid Systems: A Review of the State-of-the-Art Techniques.” arXiv, November 25, 2020. https://arxiv.org/abs/2011.12598.
* Polson, Michael, and Vadim Sokolov. 2018. “Deep Learning for Energy Markets.” arXiv, August 16, 2018. https://arxiv.org/abs/1808.05527.
* Roy, Koushik, Abtahi Ishmam, and Kazi Abu Taher. 2021. “Demand Forecasting in Smart Grid Using Long Short-Term Memory.” arXiv, July 28, 2021. https://arxiv.org/abs/2107.13653.
* “Predicting Energy Consumption Using LSTM, Multi-Layer GRU and Drop-GRU Models.” 2023. Sensors 22, no. 11: 4062. https://doi.org/10.3390/s220114062.