Study on the Short-term Electricity load forecasting (Panama case) dataset using Machine Learning methods:

Based on the Data preprocessing experiment of the dataset, we suggest the following steps for a new scientific approach for this topic:

**Exploring Advanced Feature Engineering**

Given that electricity load data is temporal and exhibits seasonality, we can enhance feature engineering as follows:

*Lag Features:* We will create lag variables that reflect the load from the previous hour, day, or week. This captures autocorrelation in time-series data.

*Rolling Statistics:* We will compute rolling means, medians, or variances for specific time windows to smooth out short-term fluctuations.

*Holiday and Weather Data Integration:* Integrating external factors like holidays, weather conditions, or public events could help improve the prediction accuracy.

**ML Approach: Ensemble Learning**

An innovative approach to improve electricity load forecasting can involve ensemble learning using a combination of different models to capture distinct patterns in the data.

*Random Forest / Gradient Boosting:* These tree-based models can capture complex nonlinear relationships between time-series features and electricity load. Combining them with autoregressive models could enhance accuracy.

*LSTM (Long Short-Term Memory Networks):* As a time-series problem, using LSTM neural networks can be powerful for capturing both short-term and long-term dependencies.

*Hybrid Model (LSTM + XGBoost):* A hybrid model that first uses LSTM to capture time-series patterns, then applies XGBoost to residual errors, can help refine predictions.

**Proposed Methodology: Temporal Fusion Transformers (TFT)**

A more advanced approach could be using Temporal Fusion Transformers (TFT), a relatively new technique that has shown high performance in time-series forecasting tasks.

**Why TFT?**

*Interpretability:* TFT provides insight into the importance of input variables over time, which is crucial for load forecasting in understanding which factors impact load the most.

*Handling Complex Temporal Dynamics:* It combines multi-horizon forecasting with attention mechanisms, which can focus on relevant periods of time or significant trends.

*Handling Missing Data and Variable Input Lengths:* This model is well-suited for time-series data that may have irregular intervals or missing data points.

**The model architecture consists of three main parts:**

*Temporal Pattern Attention:* Identifies important time steps in the sequence and allows the model to focus on relevant historical patterns.

*Variable Selection Networks:* Assign importance to input variables dynamically.

*LSTM Encoder-Decoder Architecture:* Captures the sequential nature of time-series data while managing variable input lengths.

**Additional Enhancements**

*Hyperparameter Tuning with Bayesian Optimization:* Hyperparameter tuning using methods like Bayesian optimization (via libraries like Optuna) can help improve model performance.

*Ensemble Forecasting:* After applying different models (like LSTM, Random Forest, and TFT), an ensemble of the predictions can be generated to improve overall accuracy by averaging or using a weighted voting mechanism.

**Performance Metrics**

Once the new model is developed, we can compare its performance with existing models based on metrics such as:

*Mean Absolute Error (MAE)*

*Root Mean Squared Error (RMSE)*

*Mean Absolute Percentage Error (MAPE)*

**Conclusion**

For our expected approach, integrating a Temporal Fusion Transformer (TFT) for its advanced time-series modeling capabilities, combined with other ensemble techniques like LSTM and XGBoost, would provide an innovative direction for short-term electricity load forecasting. Adding external data sources (such as weather or holidays) and performing feature engineering would further enhance model performance.