Time Series analysis and Forecasting: Household Power Consumption

Preethi Reddy Tera
Computer science
Utah State University
Logan, Utah
preethitera123@gmail.com

***Abstract*—Predicting individual household electric power consumption is paramount for fostering energy efficiency and sustainability. This project employs a diverse set of forecasting models, including ARIMA, Prophet, LSTM, and a hybrid LSTM-ARIMA approach, to provide accurate predictions for a single household's electricity usage. Feature selection techniques enhance the models, and the comprehensive analysis explores the implications of each methodology.**

With a dataset spanning nearly four years and featuring multivariate time-series characteristics, the project not only addresses the challenges of predicting energy consumption but also delves into the nuances of feature importance. Through meticulous evaluation, this study contributes valuable insights into household electricity usage patterns and the effectiveness of different prediction methodologies.

***Index Terms*—Electric Power Consumption Prediction, Household Energy Forecasting, ARIMA Modeling, Prophet Forecasting, LSTM Neural Networks, machine learning models, feature selection, time series analysis, hyper-parameter tuning,**

I. Introduction

In the contemporary landscape of energy consumption, understanding and predicting individual household electric power usage hold immense significance. This introduction sets the stage for the project's exploration of various predictive models and their applications in optimizing energy management.

The importance of predicting individual household electric power consumption is multi-faceted. For consumers, accurate predictions translate into reduced energy bills, improved comfort, increased energy awareness, and enhanced energy security. For energy providers, these predictions contribute to improved grid management, efficient resource allocation, enhanced demand response programs, and informed investment decisions.

This project employs a holistic approach, utilizing ARIMA, Prophet, LSTM, and a hybrid LSTM-ARIMA model, to capture the temporal dependencies and intricate relationships within the dataset. Feature selection techniques are incorporated to enhance the predictive accuracy of the models. The study also emphasizes the additional benefits of reduced greenhouse gas emissions, enhanced energy market efficiency, and the development of new technologies.

Through a detailed exploration of individual household electricity consumption prediction, this project aims to provide actionable insights for both consumers and energy providers. The findings contribute to the broader discourse on sustainable energy practices and efficient resource utilization in the evolving landscape of energy consumption.

II. Related Work

The realm of forecasting electricity consumption in households has been a focal point of considerable research activity, with a primary emphasis on leveraging time series data and advanced modeling techniques. Saab and colleagues (Saab et al. [2]) conducted an insightful study on the monthly electric energy consumption in Lebanon. Their methodology involved the application of ARIMA and AR(1) with a highpass filter as univariate modeling techniques. The outcomes of their research underscored the efficacy of the AR(1) highpass filter model, positioning it as the most effective for forecasting energy consumption in the specific context of Lebanon.

In a distinct geographical context, Zhu, Guo, and Feng ([1]) delved into the intricacies of household energy consumption in China spanning the years 1980 to 2009. Their investigation employed the construction VAR model, accompanied by the utilization of ARIMA and BVAR as forecasting methods. The study by Zhu et al. emphasized the critical importance of selecting appropriate techniques tailored to the specific characteristics of household energy consumption, underlining the need for nuanced modeling strategies.

Furthermore, the research by Chujai, N. Kerdprasop, and K. Kerdprasop ([4]) centered on a detailed time series analysis of household electric consumption. The study leveraged the Box and Jenkins method, renowned for its accuracy and applicability across various data movements. Employing ARIMA and ARMA models, the research aimed at identifying the most suitable forecasting period based on minimizing the values of AIC and RMSE, crucial metrics in time series analysis.

In summary, the reviewed literature collectively reflects a growing interest in employing advanced time series analysis and modeling techniques to forecast electricity consumption in households. These studies contribute significantly to the understanding of the field, emphasizing the need for model selection based on contextual characteristics for accurate predictions, thereby facilitating effective energy management and resource planning.

III. Methods

A. Data Collection

The data set comprises over 2 million measurements of electric power consumption in a household in Sceaux, France, collected at a one-minute sampling rate over nearly 4 years. With a focus on Physics and Chemistry, the multivariate, time-series data includes various electrical quantities and sub-metering values, making it suitable for regression and clustering tasks. The nine features encompass date, time, global active power, global reactive power, voltage, global intensity, and three sub-metering values representing different areas of the household. Notably, the data set contains about 1.25 percent missing values, and the provided expression calculates the active energy consumed by unmeasured electrical equipment. This valuable resource offers insights into household energy consumption patterns, though users should address missing data considerations for robust analysis.

B. Data Pre-processing

Data preprocessing is a crucial step in any analysis pipeline, and it's especially important for household power consumption data due to the potential for inconsistencies arising from various factors like data collection errors, entry mistakes, or storage issues. Identifying and addressing these inconsistencies is essential for ensuring accurate and reliable analysis results.Data preprocessing plays a vital role in ensuring the accuracy and reliability of any data analysis. This section details the steps taken to prepare the household power consumption data for further analysis.

After the raw data is first loaded to understand the data distribution, the frequency of entries per day is analyzed, revealing the minimum and maximum occurrences. Additionally, the number of entries per year is calculated for each year, providing insights into the data coverage. Lastly, the structure of the dataset is explored by examining the columns and the first few data entries, revealing date and time information.

1) Preparing for Time-Based Analysis: To facilitate efficient temporal analysis, the 'Date' column is converted to the datetime format, while the 'Time' column is converted to timedelta format. These two columns are then combined into a new 'Timestamp' column, enabling convenient time-based calculations and analysis.

2) Handling missing data: Missing values, initially denoted by '?', are replaced with the standard 'NaN' representation for easier handling. Additionally, numerical conversion is performed on relevant columns to ensure consistent data representation.

3) Imputing Missing Data: The presence of missing values is assessed by calculating the count for each column. For columns related to global power, voltage, intensity, and sub-metering, these missing values are imputed using the mean of the respective columns. This process minimizes bias and preserves valuable information present in the data.

C. Methodology

After completing the preprocessing phase, the analysis of the household power consumption data set involved two key methodologies: Seasonal Decomposition and Autocorrelation Function (ACF) Analysis.

1) Seasonal Decomposition: After completing the preprocessing, a comprehensive analysis of the household power consumption dataset was undertaken, primarily employing Seasonal Decomposition. This approach aimed to reveal temporal patterns and intrinsic dynamics within the dataset.The process initiated by setting the 'Timestamp' column as the index and converting it to datetime format. Following this, all numeric columns (excluding non-numeric ones) underwent seasonal decomposition using the additive model from the 'statsmodels.tsa.seasonal' library. The chosen decomposition period of 1440 minutes allowed for a detailed examination of minute-level patterns.

Due to the dataset's substantial size, visualizing patterns directly became challenging. To address this, a resampling technique on a weekly basis was employed, aggregating the data to provide a clearer overview of weekly trends. This approach enhanced clarity and facilitated the identification of broader patterns, with rounding applied to selected numeric columns for precision. The resampled data was then subjected to seasonal decomposition for key columns. Leveraging the 'statsmodels.api' library and the additive model, the decomposition revealed four distinct components: observed, trend, seasonal, and residual. These components collectively contributed to a nuanced understanding of the temporal dynamics within each variable.

In time-series analysis, we often decompose the data into different components to better understand the underlying patterns and characteristics of the data. The four components we analyzed for the time series are:

- Original: This is the actual observed data, which is the sum of the trend, seasonality, and residuals.
- Trend: This is the long-term pattern in the data, representing the general direction in which the series is moving.

It is the component that shows whether the series is increasing or decreasing over time.

- Seasonality: This is the pattern that repeats at fixed intervals over time, such as daily, weekly, monthly, or yearly. It is the component that shows the regular variations in the data due to external factors.
- Residuals: This is the component that captures the unexplained variation in the data that is not accounted for by the trend or seasonality. It represents the random fluctuations in the data that cannot be predicted.

The results of these 4 Time-series components can be seen in Figure. 9.

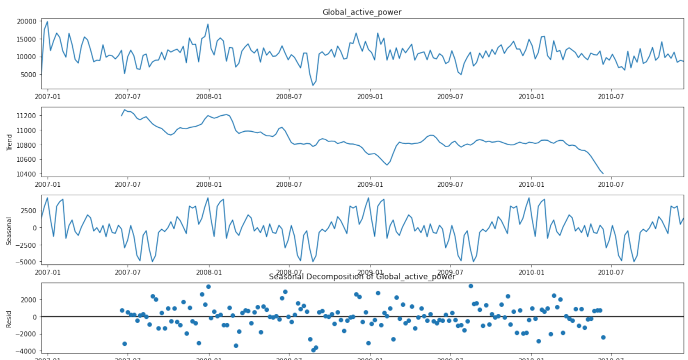

Fig. 1. Seasonal Decomposition of Global active power.

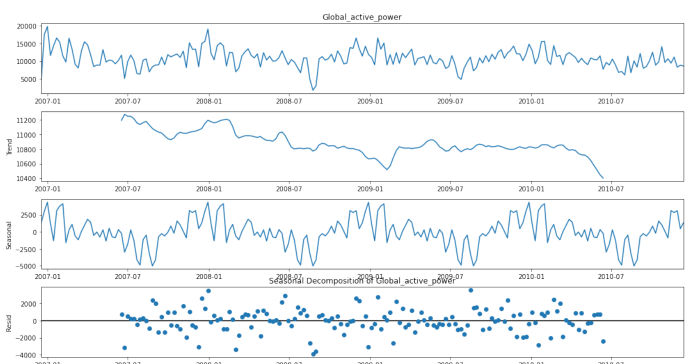

Fig. 2. Seasonal Decomposition of Global intensity.

Strong seasonal patterns were evident across various variables, offering crucial information for energy conservation strategies and system optimization. While the overall consumption of global active power and voltage remained relatively stable, some variables, such as global reactive power and certain sub-meterings, exhibited modest upward trends. These findings contributed to a nuanced understanding of household energy dynamics, paving the way for informed decision-making in energy management and efficiency. The small residual components further affirmed the effectiveness of the seasonal decomposition model in capturing the dataset's variation.

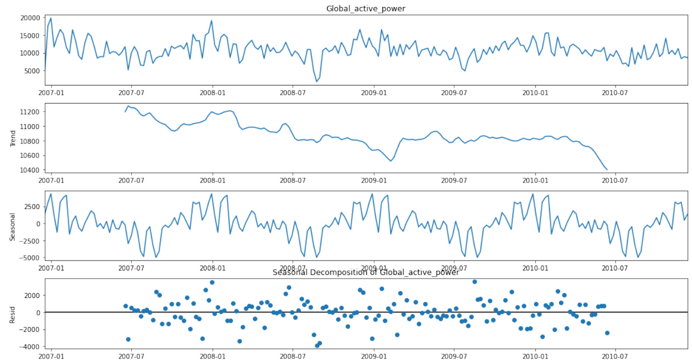

Fig. 3. Seasonal Decomposition of Sub metric - 3.

2) Autocorrelation Function (ACF): Following the initial preprocessing, a comprehensive analysis of the household power consumption dataset was conducted, employing Autocorrelation Function (ACF) analysis as an integral part of the methodology.The Autocorrelation Function (ACF) analysis played a crucial role in assessing the persistence of patterns over time. Key numeric columns were the focus of this analysis. Each column underwent ACF computation and visualization to reveal the correlation structure between the time series and its lagged values.

For every selected column, the ACF was computed utilizing the 'statsmodels.tsa.stattools.acf' function, incorporating the Fast Fourier Transform (FFT) for efficiency. The resulting ACF plots were then visualized using the 'statsmodels.graphics.tsa.plot acf' function, enabling a clear depiction of autocorrelation at different time lags.

The ACF analysis provided valuable insights into the temporal dependencies and autocorrelation patterns within the household power consumption data. The observed significant correlations at lag 1 across various variables suggested strong relationships between consecutive time points. Understanding these autocorrelations was crucial for predicting future values and identifying potential cyclical trends.

Moreover, the differences in correlation decay rates among variables highlighted variations in persistence. For instance, the slower decay in voltage and sub-metering variables implied a more sustained influence of past observations, possibly influenced by factors such as energy-efficient appliances or renewable energy sources.

These insights, together with seasonal decomposition results, provide a comprehensive understanding of the data, helping identify remaining seasonality and potentially uncovering additional patterns and trends for further investigation.

3) Feature Selection: The correlation matrix provides valuable insights into the relationships between different features in the dataset. In the context of our study on power consumption forecasting, the correlation matrix for the relevant features,

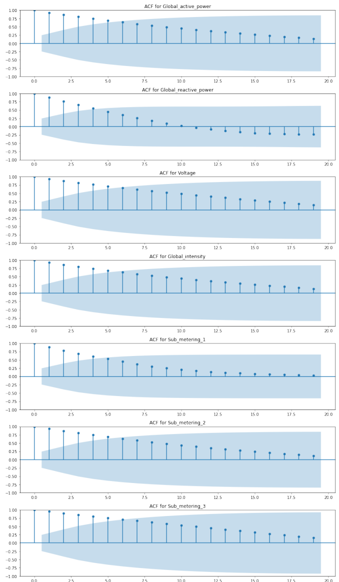

Fig. 4. Auto correlation

including ‘Global active power‘, ‘Global reactive power‘, ‘Voltage‘, ‘Global intensity‘, ‘Sub metering 1‘, ‘Sub metering 2‘, and ‘Sub metering 3‘, reveals strong interdependencies.

The correlation coefficients between ‘Global active power‘ and other features are notably high, indicating a robust linear relationship. Specifically, ‘Global intensity‘ exhibits an exceptionally high correlation of approximately 0.999, suggesting that it is almost perfectly correlated with ‘Global active power‘. This implies redundancy in information, as ‘Global intensity‘ essentially encapsulates the same information as ‘Global active power‘.

Given the high degree of multicollinearity among these features, a decision was made to omit certain features from the LSTM modeling process. Specifically, the exclusion of ‘Global intensity‘ was justified due to its extremely high correlation with the target variable (‘Global active power‘). This approach aimed at reducing computational complexity and potential overfitting caused by redundant features.

It is crucial to note that this feature selection process was driven by the specific characteristics of the dataset. In scenarios where features are highly correlated, careful consideration is essential to avoid introducing noise and ensuring the model's efficiency. The subsequent LSTM modeling focused on the selected features to streamline the forecasting process and enhance the model's interpretability.

D. Deep Learning Models

1) LSTM: In the application of LSTM for minute-level forecasting, the initial step involved preprocessing the household power consumption dataset. The ’Timestamp’ column was set as the index, and the ’Global active power’ column underwent Min-Max scaling for normalization. Subsequently, data preparation was carried out using a function to create input sequences and target values for the LSTM model, considering a time step of 10 minutes.

The LSTM model architecture consisted of an input layer with 50 LSTM units utilizing the ’relu’ activation function, followed by a dropout layer (20 persent) for regularization. The output layer comprised a Dense layer with a single unit. The model was compiled with the Adam optimizer and mean squared error loss. Training was conducted for 25 epochs with a batch size of 32, incorporating early stopping to monitor validation loss and restore the best weights.

Upon model convergence, predictions were made, and the results were evaluated using the Mean Squared Error (MSE). Additionally, the learning curve, visualizing training and validation losses across epochs, provided insights into the convergence and performance of the model over time. This detailed analysis highlighted the LSTM model's effectiveness in capturing minute-level patterns and its adaptability for short-term forecasting.

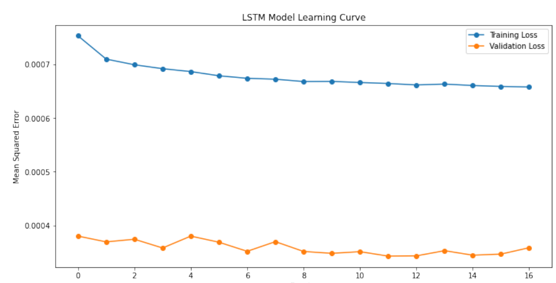

Fig. 5. Learning Curve for minutes data.

The learning curve shows that the LSTM model is able to learn the patterns in the training data and generalize to unseen data well. The training loss decreases rapidly, and the validation loss also decreases, but at a slower rate. This suggests that the model is able to learn the most important patterns in the data early on, but it continues to learn and improve over time. The learning curve also shows that the model is not overfitting the training data. Overfitting occurs when a model learns the training data too well and is unable to generalize to unseen data. In this case, the validation loss would continue to increase after the training loss starts to decrease. However, the validation loss continues to decrease in the learning curve, suggesting that the model is not overfitting. Overall, the learning curve shows that the LSTM model is able to learn the patterns in the household power consumption data and generalize to unseen data well. This suggests that the model could be used to accurately predict the global active power consumption of households in the future.

The LSTM modeling approach was extended to hourly forecasting by resampling the dataset to hourly intervals and normalizing the data. The LSTM architecture was adjusted accordingly, incorporating modifications such as a reduction in LSTM units to 25 and an increase in dropout to 30 percent for improved regularization. The hourly LSTM model demonstrated proficiency in capturing hourly patterns, as evidenced by the visualized predictions against actual values and the associated learning curve, which depicted the model's training and validation losses over epochs.

Finally, the LSTM model's forecasting capabilities were explored on a larger scale, focusing on daily and weekly predictions. The dataset was resampled accordingly, and separate LSTM models were created for daily and weekly forecasting. The models were trained and validated on their respective datasets, with predictions evaluated using metrics such as MSE, RMSE, and R-squared. Visualization of the results against actual values, coupled with learning curve analysis, provided a holistic understanding of the LSTM model's adaptability and forecasting performance across varying temporal resolutions.

2) ARIMA: The ARIMA modeling approach was applied to forecast household power consumption at different temporal resolutions, ranging from minute-level to monthly predictions. The initial step involved preparing the data by resampling the 'Global active power' column at various frequencies, such as hourly, daily, and monthly, to capture distinct temporal patterns.

For hourly forecasting, the ARIMA model was implemented using the SARIMAX class from the statsmodels library. Auto arima was employed to determine the optimal order for the ARIMA model, considering seasonal variations every 24 hours. The resulting model was fitted, and predictions were made for the entire dataset. Mean Squared Error (MSE) was used to evaluate the accuracy of the predictions, showcasing the model's performance at an hourly temporal resolution.

Similarly, daily and monthly forecasting were conducted following a similar procedure. The dataset was resampled accordingly, optimal ARIMA orders were determined, and the models were fitted and evaluated. The reported MSE values for daily and monthly predictions provided insights into the accuracy of the ARIMA models across these temporal resolutions.

To address the challenge of handling a large dataset, a focused analysis for the year 2007 was conducted at the minute-level. In this case, manual grid search was employed to identify the best ARIMA order for the 'Global active power' column. The data for the year 2007 was filtered, and the ARIMA model was iteratively tested with different orders, considering the trade-off between simplicity and accuracy. The selected model was then fitted to the 2007 data, and its performance was evaluated using various metrics, including MSE, RMSE, MAE, MAPE, and R-squared.

The ARIMA modeling results, across different temporal resolutions and for the year 2007, provided a comprehensive understanding of the model's forecasting capabilities. The reported metrics facilitated the comparison of performance at varying time intervals, offering valuable insights into the temporal dynamics of household power consumption and the adaptability of the ARIMA model to different forecasting horizons.

3) Prophet: The Prophet modeling technique was employed to forecast household power consumption at different temporal resolutions, including hourly, daily, and an overall prediction. Prophet, developed by Facebook, is known for its effectiveness in handling time series data with daily observations and multiple seasonality components.

For the hourly, daily, and weekly subsets, the 'Global active power' column was resampled accordingly, and the data was prepared for the Prophet model. The model was trained on 80 percent of the data, and the remaining 20 percent was used for out-of-sample evaluation. The model incorporated weekly and yearly seasonality components, allowing it to capture both short-term and long-term temporal patterns.

The evaluation of the Prophet model involved calculating the Root Mean Squared Error (RMSE) as a measure of prediction accuracy. Additionally, the model's components, including weekly and yearly seasonality, were visualized to provide insights into the temporal patterns captured by the model.

The Prophet model was also applied to the entire dataset, where the 'Global active power' column was prepared as 'ds' (timestamp) and 'y' (target variable). The model was initialized with MCMC samples set to 50, and predictions were made for an additional year beyond the existing data.

Analyzing the household power consumption data revealed a distinct seasonal pattern with higher consumption during winter months (January-April) due to increased heating needs. Weekend consumption also surpassed weekdays, potentially due to higher appliance usage at home. Additionally, evenings saw higher consumption compared to daytime hours, likely driven by lighting and increased appliance use in darker evenings. These findings suggest potential opportunities for targeted energy conservation strategies based on seasonal variations and daily routines.

The evaluation of the Prophet model on the entire dataset included assessing various metrics such as Mean Squared Error (MSE), RMSE, and R-squared. These metrics provided a comprehensive understanding of the model's performance in capturing the temporal dynamics of household power consumption.

The visualizations of the Prophet forecasts, coupled with the quantitative evaluation metrics, offered valuable insights into

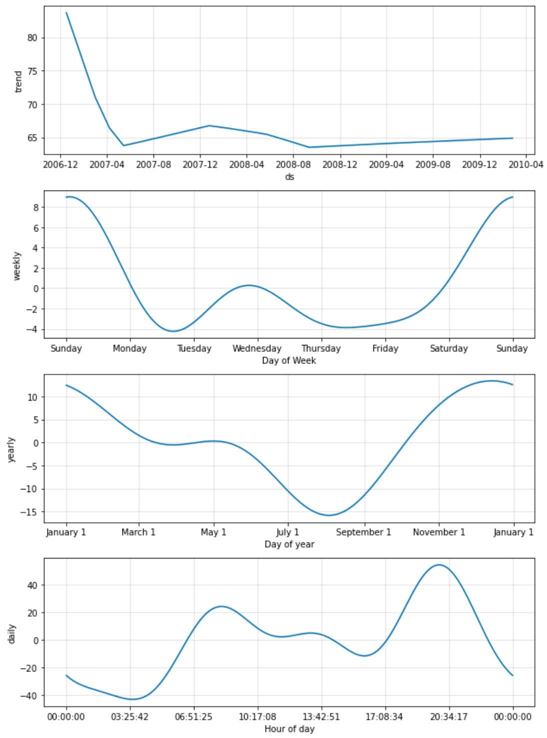

Fig. 6. Learning Curve.

the model's ability to capture both short-term fluctuations and long-term trends in the household power consumption dataset.

4) Hybrid Forecasting Approach using LSTM and Auto ARIMA: In the proposed Hybrid Forecasting Approach, a combination of Long Short-Term Memory (LSTM) neural networks and AutoRegressive Integrated Moving Average (ARIMA) models was employed to enhance the accuracy of household power consumption predictions. To refine the hourly, daily, and weekly forecasting, the dataset underwent resampling at different frequencies. This allowed for a nuanced understanding of the temporal patterns characteristic of each subset.

In the initial phase, the Auto ARIMA model was applied to the hourly dataset, with the best order determined through an automated selection process. Subsequently, a distinctive aspect of the methodology involved training an Auto ARIMA model on the residuals obtained from the disparity between the actual hourly data and LSTM predictions. This step aimed to capture any underlying patterns not fully discerned by the LSTM, creating a more comprehensive representation of the data.

For the LSTM modeling component, the hourly subset was normalized using Min-Max scaling, and LSTM sequences were formulated with a specified number of time steps. The LSTM model, characterized by two layers and incorporating dropout regularization, was trained to predict future values. Additionally, this process was replicated for the daily and weekly subsets, tailoring the LSTM models to capture both short-term and long-term patterns inherent in the respective data.

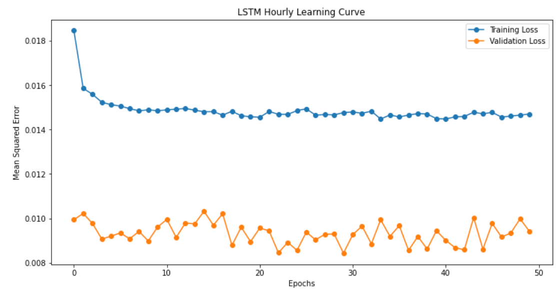

Fig. 7. Learning Curve.

The learning curve shows that the LSTM model can learn the patterns in the training data and generalize to unseen data well. The training loss decreases rapidly, and the validation loss also decreases, but at a slower rate. This suggests that the model can learn the most important patterns in the data early on. The learning curve also shows that the model is not overfitting the training data.

A pivotal aspect of this hybrid approach lies in the integration of the LSTM and ARIMA forecasts. Specifically, the Auto ARIMA model trained on residuals was employed to forecast the residuals for the hourly subset. Combining these residuals with the LSTM predictions resulted in a final hybrid forecast for hourly power consumption. For daily and weekly forecasting, the LSTM models for the respective subsets were utilized independently.

To assess the training progress and generalization capabilities of the LSTM models, learning curves were generated. These curves, visualized for hourly, daily, and weekly resolutions, provide valuable insights into the model's performance over epochs. Monitoring the training and validation losses over epochs aids in evaluating convergence, identifying overfitting, and ensuring the models effectively capture the intricacies of the data at different temporal scales. The inclusion of learning curves strengthens the transparency and interpretability of the proposed Hybrid Forecasting Approach, facilitating a comprehensive evaluation of its predictive capabilities.

IV Results

The forecasting performance of various models was rigorously assessed across different temporal resolutions, providing insights into their strengths and limitations.

The LSTM model exhibited exceptional accuracy when forecasting power consumption at the minute level. With a low Mean Squared Error (MSE) of 0.0607, a Root Mean Squared Error (RMSE) of 0.2463, and a robust R-squared (R2) score of 0.9450, the LSTM model demonstrated an ability to discern intricate patterns within high-frequency minute-level data.

Scaling the LSTM model to predict power consumption at different temporal resolutions revealed varying performance metrics. For hourly predictions, the model achieved an MSE of 1335.78 and an RMSE of 36.55. Daily predictions, while reasonably accurate with an MSE of 184985.27 and an RMSE of 430.10, showed a distinct drop in R-squared (R2) to 0.4630. Weekly predictions, while still valuable, exhibited a higher MSE of 6175406.32, an RMSE of 2485.04, and an R-squared of 0.3428.

The Prophet forecasting model, while capable, displayed comparatively lower accuracy than LSTM. The overall MSE was 0.8473, with an RMSE of 0.9205 and an R-squared of 0.2324. Subsets further emphasized the model's challenges, with hourly, daily, and weekly predictions yielding RMSE values of 41.68, 557.85, and 1717.52, respectively.

The Auto ARIMA models demonstrated effective forecasting capabilities across different temporal granularities. For hourly, daily, and monthly predictions, the best-fitted orders were (0, 1, 2), (2, 1, 2), and (1, 0, 0), respectively. Corresponding MSE values were 7.60, 191.23, and 516142.46, illustrating the model's adaptability to diverse temporal patterns.

For minute-level data in the year 2007, the ARIMA model presented a moderate performance with an MSE of 1.33, RMSE of 1.15, and a low R-squared of 0.0055, indicating challenges in capturing nuanced patterns for that specific year.

The Hybrid Forecasting Approach, integrating LSTM and Auto ARIMA, demonstrated promising results with an MSE of 1261.90, RMSE of 35.52, and an R-squared of 0.5611. The hybrid model showcased enhanced accuracy compared to individual models, highlighting the synergistic benefits of combining LSTM's ability to capture complex patterns and Auto ARIMA's adaptability to temporal variations. However, the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) for the hybrid approach are reported as NaN and infinity, respectively, signaling the need for further investigation and refinement in these metrics.

V Conclusion and Future Scope

In conclusion, our extensive exploration of forecasting models for power consumption has illuminated their distinctive strengths and weaknesses across various temporal resolutions. The LSTM model showcased exceptional accuracy in capturing minute-level patterns, underscoring its efficacy for high-frequency data. However, challenges emerged when extending the model to predict power consumption at hourly, daily, and weekly intervals, highlighting the need for tailored modeling strategies across different temporal granularities.

The Prophet model, while competent, encountered limitations in achieving precision comparable to LSTM, underscoring the importance of model selection based on the unique characteristics of the data. Auto ARIMA models proved to be versatile tools for forecasting, demonstrating effectiveness across different time scales. Their adaptability to hourly, daily, and monthly patterns enhances their value for diverse forecasting requirements.

The Hybrid Forecasting Approach, integrating LSTM and Auto ARIMA, presented promising results, surpassing individual models and revealing potential synergies when combining deep learning with traditional time-series analysis. Despite achieving enhanced accuracy, the hybrid model uncovered challenges in calculating Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), necessitating further refinement.

Future research should delve deeper into refining the Hybrid Forecasting Approach, particularly addressing challenges associated with MAE and MAPE calculations. Investigating and resolving these issues will contribute to a more comprehensive understanding of the model's performance.Moreover, exploring alternative deep learning architectures and ensemble techniques may further enhance the accuracy of forecasting models. Evaluating the models on diverse datasets and considering external factors, such as holidays or special events, can contribute to their robustness and adaptability.

As the field of forecasting evolves, integrating emerging technologies such as attention mechanisms and transformer architectures could offer novel insights and improvements. Additionally, exploring the impact of external factors like climate conditions and social events on power consumption patterns could lead to more nuanced and accurate predictions. In conclusion, this project lays the groundwork for future endeavors in refining forecasting models, fostering a deeper understanding of their capabilities and limitations, and propelling the field toward more accurate and adaptable solutions.

References

[1] Stefenon, Stefano Frizzo, et al. ”Aggregating prophet and seasonal trend decomposition for time series forecasting of Italian electricity spot prices.” Energies 16.3 (2023): 1371.

[2] J. Chujai, Pasapitch, Nittaya Kerdprasop, and Kittisak Kerdprasop. ”Time series analysis of household electric consumption with ARIMA and ARMA models.” Proceedings of the international multiconference of engineers and computer scientists. Vol. 1. Hong Kong: IAENG, 2013.

[3] F. Martinez Alvarez, A. Troncoso, J. C. Riquelme and J. S. Aguilar Ruiz, ”Energy Time Series Forecasting Based on Pattern Sequence Similarity,” in IEEE Transactions on Knowledge and Data Engineering, vol. 23, no. 8, pp. 1230-1243, Aug. 2011, doi: 10.1109/TKDE.2010.227.

[4] Gasparin, Alberto et al. “Deep Learning for Time Series Forecasting: The Electric Load Case.” CAAI Trans. Intell. Technol. 7 (2019): 1-25.

[5] Wang, Xinlin, Zhihao Yao, and Marios Papaefthymiou. ”A real-time electrical load forecasting and unsupervised anomaly detection framework.” Applied Energy 330 (2023): 120279.