# Comparative Analysis of Statistical and Machine Learning Models for Electricity Demand Forecasting

Aleena Joseph (23264236)
Department of Computing,
Dublin City University
Dublin, Ireland
aleena.joseph6@mail.dcu.ie

Alfred Iyby (23268366)
Department of Computing,
Dublin City University,
Dublin, Ireland
alfred.iyby2@mail.dcu.ie

## Abstract

***This study investigates the performance of various machine learning algorithms for electricity demand forecasting, utilizing hourly consumption data from households in several Norwegian regions over two winter periods from early 2020 to spring 2021. The task involves predicting hourly electricity demand, crucial for efficient energy planning. Two forecasting methods are compared: (1) training individual models for each household, potentially achieving higher accuracy by capturing unique consumption patterns; and (2) training a single model across all households, enhancing scalability and generalization to unseen households. We compared traditional models such as ARIMA (AutoRegressive Integrated Moving Average) and SARIMA (Seasonal AutoRegressive Integrated Moving Average) with a hybrid deep learning approach combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. Performance is evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Results indicate that while ARIMA and SARIMA provide solid baselines with interpretable outcomes, the CNN-LSTM model significantly outperforms them in terms of accuracy, particularly in capturing intricate demand patterns. Additionally, survey data integration proves beneficial in specific scenarios, improving model performance. This comparative analysis highlights the strengths and limitations of each method, offering valuable insights into their applicability in various forecasting scenarios.***

## Introduction

Accurate forecasting of household electricity demand is the cornerstone of modern energy management systems, pivotal for ensuring operational efficiency, cost savings, and enhanced sustainability. With the growing complexity of consumption patterns and the increasing integration of renewable energy sources, traditional forecasting methods often fall short in capturing the dynamic and intricate nature of electricity usage. This challenge is particularly pronounced in regions, where seasonal variations significantly impact electricity demand, necessitating precise prediction during peak periods, especially in winter.

Forecasting also optimizes the use of energy storage systems, ensuring stored energy is available during peak demand periods. Personalized feedback on energy consumption patterns encourages households to adopt energy-saving behaviours. Demand Forecasting supports smart home systems that automatically adjust heating, cooling, and appliance usage based on predicted demand, enhancing energy efficiency. Policymakers can use forecasting data to plan and develop energy infrastructure more effectively, ensuring future demand growth is met with adequate supply. Utilities can better comply with regulatory requirements for reliability and efficiency by using accurate forecasts to meet mandated performance standards. It enables quicker and more effective response strategies during natural disasters or emergencies, ensuring critical services remain powered. Forecasting aids in developing contingency plans, ensuring alternative power sources can be quickly deployed if needed. Optimizing energy use and integrating more renewable sources through accurate forecasting contributes to lower greenhouse gas emissions, supporting environmental sustainability goals. Efficient energy management driven by accurate forecasts stimulates economic growth by reducing energy costs for businesses and consumers, leading to increased disposable income and investment in other sectors. Precise prediction enables smart home appliances to provide a smooth user experience by automatically modifying settings to maximize comfort and energy efficiency.

This research investigates the performance of various machine learning algorithms for household electricity demand forecasting, utilizing hourly consumption data from households in multiple Norwegian regions over two consecutive winter periods, from early 2020 to spring 2021. The study aims to identify the most effective method for forecasting hourly electricity demand at the household level by comparing traditional time series models, such as ARIMA and SARIMA, with a cutting-edge hybrid deep learning approach that combines Convolutional Neural Networks and Long Short-Term Memory networks.

Two distinct forecasting approaches are evaluated:

1. Individual Household Models: Training separate models for each household to capture unique consumption patterns, potentially achieving higher accuracy.

2. Unified Model for All Households: Training a single model across all households to enhance scalability and generalization, which is advantageous for large-scale implementations.

Performance is assessed using Mean Absolute Error, Mean Squared Error, and Root Mean Squared Error to provide a comprehensive evaluation of accuracy and robustness. Implementing accurate hourly electricity forecasting in households presents numerous practical benefits. Accurate forecasts help balance supply and demand, preventing blackouts and ensuring consistent power delivery. Improved load management minimizes the risk of grid overloading during peak demand periods. Anticipating demand allows for optimized utilization of power plants, reducing the need for expensive peaking power plants and lowering operational costs. Predictive insights facilitate better planning of maintenance activities, enhancing reliability and reducing downtime. Households benefit from dynamic pricing models that offer lower rates during off-peak hours, promoting efficient energy usage.

This comparative analysis of different forecasting methods highlights their respective strengths and limitations, providing valuable insights for stakeholders in the energy sector. The findings guide utility companies and policymakers in selecting the most appropriate forecasting techniques, ultimately contributing to more reliable, efficient, and sustainable energy management systems.

## Background

Electricity demand forecasting is a critical aspect of energy management and planning, with numerous studies exploring various methodologies to enhance prediction accuracy. This section reviews relevant literature on statistical and machine learning approaches for electricity consumption forecasting.

Traditional statistical models, such as ARIMA and SARIMA, have been widely utilized for time-series forecasting in energy consumption. Gellart et al. [8] compared ARIMA with TBATS (Trigonometric Seasonal, Box-Cox Transformation, ARMA Errors, Trend and Seasonal Components) for forecasting electricity consumption in smart homes. Their findings suggest that TBATS outperforms ARIMA due to its ability to handle complex seasonal patterns and provide automated model selection. Similarly, Bilgili and Pinar [13] employed SARIMA and LSTM models to forecast Türkiye's gross electricity consumption, demonstrating that while both models were effective, LSTM achieved slightly better accuracy.

Fumo et al. [16] explored regression analysis techniques for forecasting household energy consumption, highlighting the efficacy of linear, multiple linear, and quadratic regression models. Their research emphasized the importance of including external factors such as solar radiation and outdoor temperature to enhance model precision.

The limitations of statistical models in capturing non-linear and complex patterns have led to the adoption of machine learning techniques. Support Vector Regression (SVR) was employed by Zhong et al. [3] for building energy consumption prediction, showcasing improved robustness and generalization capacity. Ullah et al. [5] utilized a combination of Convolutional Neural Networks (CNN) and Bi-Directional Long Short-Term Memory (BDLSTM) networks to predict short-term residential power consumption, demonstrating superior performance over traditional statistical methods.

Ensemble learning techniques have also been explored to improve forecasting accuracy. Liang et al. [1] used a selective ensemble learning approach, integrating meteorological parameters to predict household power consumption. The Filtering Iterative Optimization Ensemble Strategy (FIOES) model demonstrated strong generalizability and higher accuracy compared to single base learner models.

Hybrid models combining different machine learning techniques have shown promising results. Yan et al. [17] proposed a hybrid model using Stationary Wavelet Transform (SWT) and LSTM networks for individual household energy consumption predictions. Their approach outperformed conventional methods like CNN-LSTM, LSTM, and SVR, particularly in terms of handling irregular patterns and volatility in data.

Deep learning models, particularly those combining CNNs and LSTMs, have gained attention for their ability to capture intricate patterns in time-series data. Kim and Cho [9] demonstrated the effectiveness of CNN-LSTM networks in predicting residential energy consumption. Jalali et al. [15] proposed a deep neuroevolution algorithm to optimize CNN architectures for load forecasting, achieving superior performance across various metrics compared to traditional algorithms like ARIMA, logistic regression, and random forest regression. Edge computing and federated learning have been proposed to address privacy concerns and computational overhead associated with deep learning models. Savi and Olivadese [6] introduced a federated learning approach for short-term energy consumption forecasting, showing significant improvements in performance and training efficiency compared to centralized models.

## Methodology

The goal of this study is to investigate the consumption of electricity in households and forecast demand per hour using temperature parameters. This study also analyses various household characteristics to determine if they have an impact on electricity consumption. The study uses a combination of statistical and neural network methodologies to compare their forecasting effectiveness. The experiment was conducted to analyse if the household change their power consumption based on the temperature and other contextual information from surveys.

### Dataset and Data Preparation

The population consist of households from Norwegian areas like Oslo, Bergen, Stavanger that consist of consumers of all ages. It includes hourly consumption data of all participating households along with other characteristics. The study was conducted on 4429 households and 1,35,15,600 rows of data. The dataset contains household characteristics such as Household ID, Date, Hour, Participation Phase, Demand kWh and temperature data varying from current hour to data from last 24, 48 and 72 hours. Each row contains information about the amount of electricity consumed by that household at that hour. It is then divided into two sets based on when the data was collected from the households. Demand_kWh ranged from as high as 9.8 to as low as 0.04 kW per hour. The consumption data for phase 1 was collected in winter 2019/20 and phase 2 was collected in winter 2020/21, with Phase 1 containing 75 days and Phase 2 containing 116 days of hourly demand consumption data.

Phase 1 consists of 742 households and Phase 2 consists of 4375 households from which 688 households (common ids) are filtered out to improve data purity and independence while also increasing analysis accuracy. The output is two sets with 54 and 3687 households, respectively. A third subset named common_records was also filtered out from the common elements to analyse the dependency on consumption. Apart from Phase 1 and Phase 2 subset common_records contain the data from both time periods and thus track the same households as they transition between phases or compare similar households under different experimental conditions. The datasets are processed and put through data cleaning operations to ensure errors and improve decision-making.

The resultant datasets are Phase 1 set which now consists of 54 households data and Phase 2 that has 3687 households. Two more additional datasets are created from Phase 1 and Phase 2 by using the survey questions. For both the phases relevant questions are filtered out that has potential impact on the consumption rate such as the *"household size, Number of residents in the household, If the house is electrically heated or not and whether an electric car is charged through the house".* Not all households have participated in the survey and therefore the number of households for each phase have decreased to 36 and 1697 respectively. The survey results are processed and converted into categorical format for better data handling and sampling.

In our analysis, we discovered significant multicollinearity among the predictor variables, particularly in the Phase 1 and Phase 2 dataset, as evidenced by high Variance Inflation Factor (VIF) values for several temperature-related features such as (*Temperature24, Temperature48, Temperature72)*, requiring further investigation and potential variable reduction to ensure model stability and accuracy. The chosen features for survey dataset — temperature, home_size, electric_car, number_of_people, and electrically_heater represent a comprehensive set of variables that capture temporal, environmental, household, and lifestyle factors, providing a solid foundation for analysing energy demand patterns and also having lower VIF values indicating less multicollinearity.

For this experiment, we are primarily focusing on the Phase 1 and Phase 2 sets. Three different methodologies are applied to these phases and results are analysed with and without survey responses. This will help us understand the impact of those characteristics on hourly demand. We have processed all the subsets and converted them into separate datasets, including:

- Phase 1
- Phase 2
- Phase 1 with survey columns
- Phase 2 with survey columns
- Common Records

The experiment has been done on 54 households from Phase 1 and 1000 households from Phase 2 using Method 1. Due to memory constraints, Method 2 will only consider 100 households.

The experiment aims to compare models that can handle time series data. As such, two statistical models and a machine learning model are used to analyse and assess the results of all three models to determine which model performs best. In this experiment, statistical models such as ARIMA and SARIMA have been used and for the machine learning algorithm, CNN-LSTM was used.

This section introduces the three models and how they are trained, all models are trained using two separate approaches. Method 1 focuses on training the models on each individual household and forecasting their demand; that is, the method treats each household as a separate model and trains on it, whereas Method 2 focuses on training the entire dataset as a single model, which is made up of multiple households. Both methods have been experimented on to find patterns and choose optimal solution, Method 1 will assist in adapting the model to a specific household providing personalized patterns, detailed analysis and high accuracy, whereas Method 2 will aid in training the model for much more generalisation, robustness scalability and statistical significance. For all the models, the dataset is split into training and test set with 80:20, 75:25 and 50:50 ratios for optimizing performance and avoid overfitting and underfitting.

**ARIMA**

ARIMA, is a popular statistical technique for forecasting time series data. This model's unique combination of autoregression, differencing, and moving average components enables it to effectively model a wide range of data patterns. By combining historical values and errors into its projections, ARIMA provides precise predictions necessary for decision-making in numerous sectors.

The experiment utilizes auto_arima from pmdarima library which automatically selects the best parameters (p,d,q) for the ARIMA model based on given data. This function significantly simplifies the process of identifying the most suitable model parameters. The model is trained on households individually and collectively for all ratios. To identify the most important aspects, we used a correlation analysis to find factors with significant relationships to consumption patterns. Following a preliminary analysis, which helped in refining our feature set Date, Hour and Temperature were selected. The selected features were integrated as exogeneous variables. After the model has been trained, it forecasts the consumption value for the remaining ratio i.e. if it is an 80:20 ratio, the model trains for 80 percent of the dataset and using the remaining 20 to predict and test its values.

After successfully predicting the model's accuracy is evaluated using several metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). These metrics provide information on the average magnitude of the model's prediction errors, with lower values indicating more performance.

**SARIMA**

SARIMA is an advanced statistical model designed to forecast time series data with seasonal patterns. This model extends the traditional ARIMA framework by integrating seasonal features, allowing it to handle fluctuations that occur at regular periods. SARIMA models, which incorporate both non-seasonal and seasonal components, can capture complicated behaviours in data, making them invaluable for properly predicting trends and cycles in domains like economics, meteorology, and energy consumption.

The model incorporates ARIMA model with seasonal properties and utilizes the same function from pmdarima library. The seasonal component can be enabled by setting the model to look for seasonal patterns during the model building process. While ARIMA utilizes p,d,q variables for finding patterns SARIMA uses p,d,q and P,D,Q which are seasonal parts of the model. A similar approach to that of ARIMA model is used for training and testing purposes. The model is trained for all the methods (1 & 2) and datasets (Phase 1 & 2 and Phase 1 & 2 with survey). After model training, the model summary is evaluated on the basis of statistical significance and goodness of fit. The same evaluation method is used for consistent benchmarking and model optimization.

**CNN – LSTM Network**

CNNs are proven to be highly effective for analysing time series data. For time series applications, CNNs leverage their convolutional network to identify and capture local patterns within data especially seasonality's which repeat over time. The best part of CNNs is that it can automatically detect and exploit time series patterns without the necessity for manual feature selection compared to ARIMA and SARIMA. The sequential data processing strength of LSTM along with CNN creates a hybrid model that can utilized efficiently for time series data.

In the CNN-LSTM architecture, CNN's is used to extract features from these multiple input series simultaneously reducing dimensionality and thus reducing the number or input features to LSTM layer. LSTMs help the model by maintaining internal states that help it to remember past data over long periods.

For the model implementation the feature set is first reshaped into a three-dimensional array that matches the requirements of CNN model and is split into required ratio. Both features and target variables are scaled using MinMaxScaler to normalize the data. The model architecture is implemented in such a way that a single LSTM layer with 64 units process the data along with two fully connected dense layers. The first layer has 50 neurons with ReLU activation function, and second layer has a single neuron to predict a single target variable. The model also uses and Adam optimizer with a custom learning rate of 0.001.

The model's performance is evaluated using the test set with test loss value. It also evaluates MSE, MAE, RMSE to assess the accuracy of predictions providing clear insights and making it easier for comparison with the other two models.

## Results:

This section presents the comprehensive results of our study on predictive models for household electricity usage. We evaluated the performance of three models— ARIMA, SARIMA, and CNN-LSTM—across different approaches and data splits. The following tables represent the results of each model's performance metrics for different datasets. Due to space constraints, we have not included common_records dataset results here.

**Method 1 - Phase 1**

| Performance Metrics | ARIMA | | | SARIMA | | | CNN-LSTM | | |
|---|---|---|---|---|---|---|---|---|---|
| | 80:20 | 75:25 | 50:50 | 80:20 | 75:25 | 50:50 | 80:20 | 75:25 | 50:50 |
| MSE | 0.92724 | 0.98628 | 0.91808 | 0.95663 | 1.01705 | 0.93542 | 0.78305 | 0.73994 | 0.89976 |
| MAE | 0.65347 | 0.67496 | 0.62449 | 0.66573 | 0.68782 | 0.63914 | 0.55696 | 0.53461 | 0.58079 |
| RMSE | 0.96293 | 0.99311 | 0.95816 | 0.97807 | 1.00849 | 0.96717 | 0.88490 | 0.86020 | 0.94855 |

Table 1

**Method 2 - Phase 1**

| Performance Metrics | ARIMA | | | SARIMA | | | CNN-LSTM | | |
|---|---|---|---|---|---|---|---|---|---|
| | 80:20 | 75:25 | 50:50 | 80:20 | 75:25 | 50:50 | 80:20 | 75:25 | 50:50 |
| MSE | 4.85628 | 6.42541 | 2.97439 | 4.85628 | 6.42541 | 2.97439 | 3.65729 | 3.43846 | 7.05351 |
| MAE | 1.77299 | 1.89162 | 1.27552 | 1.77299 | 1.89162 | 1.27552 | 1.90748 | 1.84878 | 2.63034 |
| RMSE | 2.20369 | 2.53483 | 1.72464 | 2.20369 | 2.53483 | 1.72464 | 1.91240 | 1.85430 | 2.65584 |

Table 2

**Method 1 - Phase 2**

| Performance Metrics | ARIMA | | | SARIMA | | | CNN-LSTM | | |
|---|---|---|---|---|---|---|---|---|---|
| | 80:20 | 75:25 | 50:50 | 80:20 | 75:25 | 50:50 | 80:20 | 75:25 | 50:50 |
| MSE | 1.27750 | 1.41282 | 2.32251 | 2.14462 | 2.56452 | 5.33939 | 0.89761 | 0.95337 | 1.00034 |
| MAE | 0.76932 | 0.79670 | 1.10692 | 0.88491 | 0.87179 | 1.33987 | 0.65127 | 0.66260 | 0.67843 |
| RMSE | 1.13026 | 1.18862 | 1.52398 | 1.46445 | 1.60141 | 2.31071 | 0.94742 | 0.97640 | 1.00017 |

Table 3

**Method 2 - Phase 2**

| Performance Metrics | ARIMA | | | SARIMA | | | CNN-LSTM | | |
|---|---|---|---|---|---|---|---|---|---|
| | 80:20 | 75:25 | 50:50 | 80:20 | 75:25 | 50:50 | 80:20 | 75:25 | 50:50 |
| MSE | 2.11496 | 2.12191 | 23.8280 | 2.11496 | 2.12191 | 23.8280 | 7.32452 | 7.03597 | 6.49043 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| MAE | 1.00869 | 0.96997 | 4.64164 | 1.00869 | 0.96997 | 4.64164 | 2.67869 | 2.62875 | 2.52699 |
| RMSE | 1.45429 | 1.45667 | 4.88140 | 1.45429 | 1.45667 | 4.88140 | 2.70638 | 2.65254 | 2.54763 |

Table 4

**Method 1 - Phase 1 with survey**

| Performance Metrics | ARIMA | | | SARIMA | | | CNN-LSTM | | |
|---|---|---|---|---|---|---|---|---|---|
| | 80:20 | 75:25 | 50:50 | 80:20 | 75:25 | 50:50 | 80:20 | 75:25 | 50:50 |
| MSE | 1.12411 | 1.19980 | 1.09771 | 1.13949 | 1.21468 | 1.10334 | 1.02552 | 0.91664 | 1.07407 |
| MAE | 0.74259 | 0.77036 | 0.70656 | 0.74763 | 0.77506 | 0.70954 | 0.64247 | 0.62557 | 0.67659 |
| RMSE | 1.06024 | 1.09535 | 1.04771 | 1.06747 | 1.10212 | 1.05040 | 1.01267 | 0.95741 | 1.03637 |

Table 5

**Method 2 - Phase 1 with survey**

| Performance Metrics | ARIMA | | | SARIMA | | | CNN-LSTM | | |
|---|---|---|---|---|---|---|---|---|---|
| | 80:20 | 75:25 | 50:50 | 80:20 | 75:25 | 50:50 | 80:20 | 75:25 | 50:50 |
| MSE | 7.66106 | 4.58318 | 4.83208 | 7.66106 | 4.58318 | 4.83208 | 12.5837 | 15.2010 | 7.60919 |
| MAE | 2.35453 | 1.67636 | 1.86041 | 2.35453 | 1.67636 | 1.86041 | 2.95979 | 3.12994 | 2.24768 |
| RMSE | 2.76786 | 2.14083 | 2.19820 | 2.76786 | 2.14083 | 2.19820 | 3.54736 | 3.89885 | 2.75847 |

Table 6

**Method 1 - Phase 2 with survey**

| Performance Metrics | ARIMA | | | SARIMA | | | CNN-LSTM | | |
|---|---|---|---|---|---|---|---|---|---|
| | 80:20 | 75:25 | 50:50 | 80:20 | 75:25 | 50:50 | 80:20 | 75:25 | 50:50 |
| MSE | 1.14906 | 1.25134 | 1.91033 | 1.93774 | 2.68937 | 4.40056 | 1.09121 | 1.12456 | 1.13307 |
| MAE | 0.69153 | 0.70602 | 0.95963 | 0.78711 | 0.80599 | 1.14858 | 0.66293 | 0.67521 | 0.68427 |
| RMSE | 1.07194 | 1.11863 | 1.38215 | 1.39202 | 1.63993 | 2.09775 | 1.04461 | 1.06045 | 1.06445 |

Table 7

**Method 2 - Phase 2 with survey**

| Performance Metrics | ARIMA | | | SARIMA | | | CNN-LSTM | | |
|---|---|---|---|---|---|---|---|---|---|
| | 80:20 | 75:25 | 50:50 | 80:20 | 75:25 | 50:50 | 80:20 | 75:25 | 50:50 |
| MSE | 8.20668 | 7.05287 | 7.13798 | 8.20668 | 7.05287 | 7.13798 | 9.74533 | 9.87396 | 9.30242 |
| MAE | 2.03446 | 1.86072 | 1.83243 | 2.03446 | 1.86072 | 1.83243 | 3.07775 | 3.09793 | 3.00654 |
| RMSE | 2.86473 | 2.65572 | 2.67170 | 2.86473 | 2.65572 | 2.67170 | 3.12175 | 3.14228 | 3.04998 |

Table 8

ARIMA and SARIMA generally performed well with specific data splits, indicating suitability for capturing linear trends and seasonal patterns in electricity usage data. CNN-LSTM excelled across most datasets and splits, particularly in Method 1. Its strength in handling complex, nonlinear relationships in large datasets was evident, making it a robust choice for detailed, predictive analytics. The inclusion of survey data had mixed effects on model performance. In some cases, it enhanced the models' ability to capture detailed behaviour patterns, while in others, it slightly reduced prediction accuracy, potentially due to overfitting or the introduction of noise. Method 1 provided higher accuracy per household but at the cost of increased complexity and computational resources due to the number of models managed. Method 2 offered scalability and ease of maintenance with a slight trade-off in individual accuracy, making it suitable for applications requiring generalized predictions across multiple households.

The CNN-LSTM model outperformed others in most setups, with notable efficiency in both methodological frameworks. Its ability to integrate and learn from large datasets with complex patterns was particularly advantageous. ARIMA and SARIMA showed strong performance in setups where the data characteristics matched their strengths, such as clear seasonal patterns or linear trends.

Data splits heavily influenced model performance, with larger portions of training data generally yielding better results. This highlights the importance of dataset size and composition in training predictive models. For complex predictive tasks involving large datasets or capturing intricate patterns, the CNN-LSTM model is recommended due to its superior performance. ARIMA and SARIMA should be utilized for applications with clear, consistent patterns, such as seasonal energy usage. The choice of methodology should be based on the end-use case: Method 1 is preferable for achieving high accuracy at the household level, while Method 2 is suitable for broader, less granular predictions. Additionally, it is essential to experiment with different training-test splits to find the optimal balance for each application, ensuring the models achieve neither underfitting nor overfitting.

This analysis emphasizes the importance of selecting the right model and method based on specific project requirements and data characteristics. By aligning the model and methodology with the intended application, organizations can significantly enhance the effectiveness and efficiency of their predictive analytics efforts in the energy sector.

## Challenges

This study on electricity demand forecasting using various machine learning algorithms encountered several challenges. Managing extensive datasets of hourly consumption from numerous households required significant computational resources and time. Training individual models for each household captured unique consumption patterns but was extremely time-consuming and required substantial computational power. Training a single model across all households aimed for scalability but was memory-intensive, posing potential issues with system resources when handling large datasets. Ensuring high-quality data preprocessing involved handling missing values, normalizing data, and integrating additional features like temperature and survey data, which added complexity and required meticulous attention. Balancing the need for detailed, household-specific models with the scalability of aggregated models was challenging. Achieving a model that could generalize well across different households and time periods required careful consideration of trade-offs. Incorporating survey data added contextual information but also introduced potential noise, requiring careful integration and validation to ensure its relevance and accuracy. These challenges highlight the complexity and resource demands of developing accurate and scalable electricity demand forecasting models.

## Conclusion

This comprehensive study evaluated the performance of various machine learning algorithms— ARIMA, SARIMA, and CNN-LSTM— across multiple datasets for electricity demand forecasting, using two different methods. The objective was to predict hourly electricity demand using data from households in Norwegian regions over two winter periods, enhancing energy management and planning through accurate forecasts. This study underscores the importance of selecting the appropriate algorithm and data split ratio tailored to specific datasets for optimal electricity demand forecasting. The CNN-LSTM model, particularly when using Method 1 and moderate to balanced data splits, proved highly effective for capturing intricate consumption patterns. ARIMA, however, demonstrated robustness and accuracy with larger training datasets, making it a reliable choice for scenarios with ample historical data. Including survey data generally enhanced the performance of models, especially CNN-LSTM, by providing additional contextual information. Training individual models for each household (Method 1) generally outperformed training a single model across all households (Method 2). This approach better captures unique consumption behaviors, leading to more accurate predictions.

By comparing both methods, it is clear that Method 1's individualized approach generally yields higher accuracy, emphasizing the necessity of capturing unique household consumption patterns for precise forecasts. These insights offer valuable guidance for practical applications in energy management and planning, advocating for customized approaches based on dataset characteristics and forecasting needs.

## References


[1] Kun Liang, Fei Liu , Yiyang Zhang, “Household Power Consumption Prediction Method Based on Selective Ensemble Learning”, IEEE Access, vol 8, pp. 95657 – 95666, 2020.

[2] Antonio Parejo, Stefano Bracco, Enrique Personal, Diego Francisco Larios , Federico Delfino, Carlos León, “Short-Term Power Forecasting Framework for Microgrids Using Combined Baseline and Regression Models”, *Applied Science 2021*, *11*(14), 6420; https://doi.org/10.3390/app11146420

[3] H. Zhong, J. Wang, H. Jia, Y. Mu, and S. Lv, ‘‘Vector field-based support vector regression for building energy consumption prediction,’’ Appl. Energy, vol. 242, pp. 403–414, May 2019.

[4] C. Tian, C. Li, G. Zhang, and Y. Lv, ‘‘Data driven parallel prediction of building energy consumption using generative adversarial nets,’’ Energy Buildings, vol. 186, pp. 230–243, Mar. 2019.

[5] F. U. M. Ullah, A. Ullah, I. U. Haq, S. Rho and S. W. Baik, "Short-Term Prediction of Residential Power Energy Consumption via CNN and Multi-Layer Bi-Directional LSTM Networks," in *IEEE Access*, vol. 8, pp. 123369-123380, 2020, doi: 10.1109/ACCESS.2019.2963045.

[6] M. Savi and F. Olivadese, "Short-Term Energy Consumption Forecasting at the Edge: A Federated Learning Approach," in IEEE Access, vol. 9, pp. 95949-95969, 2021, doi: 10.1109/ACCESS.2021.3094089

[7] D. A. Bashawyah and S. M. Qaisar, "Machine Learning Based Short-Term Load Forecasting for Smart Meter Energy Consumption Data in London Households," 2021 IEEE 12th International Conference on Electronics and Information Technologies (ELIT), Lviv, Ukraine, 2021, pp. 99-102, doi: 10.1109/ELIT53502.2021.9501104.

[8] Arpad Gellert, Ugo Fiore, Adrian Florea, Radu Chis, Francesco Palmieri, “Forecasting Electricity Consumption and Production in Smart Homes through Statistical Methods”, Sustainable Cities and Society, Volume 76, 2022, 103426, ISSN 2210-6707, https://doi.org/10.1016/j.scs.2021.103426

[9] T.-Y. Kim and S.-B. Cho, ‘‘Predicting residential energy consumption using CNN-LSTM neural networks,’’ Energy, vol. 182, pp. 72–81, Sep. 2019.

[10] C. Li, Z. Ding, D. Zhao, J. Yi, and G. Zhang, ‘‘Building energy consumption prediction: An extreme deep learning approach,’’ Energies, vol. 10, no. 10, p. 1525, Oct. 2017.

[11] L. Saad Saoud, H. Al-Marzouqi and R. Hussein, "Household Energy Consumption Prediction Using the Stationary Wavelet Transform and Transformers," in *IEEE Access*, vol. 10, pp. 5171-5183, 2022, doi: 10.1109/ACCESS.2022.3140818.

[12] Shamaila Iram, Hussian AL-Aqrabi, Hafiz Muhammad Shakeel, Hafiz Muhammed Athar Farid, Muhammad Riaz, Richard Hill, Prabanchan Vethathir, Tariq Alsboui, "An Innovative Machine Learning Technique for the Prediction of Weather Based Smart Home Energy Consumption," in *IEEE Access*, vol. 11, pp. 76300-76320, 2023, doi: 10.1109/ACCESS.2023.3287145.

[13]Mehmet Bilgili, Engin Pinar, “Gross electricity consumption forecasting using LSTM and SARIMA approaches: A case study of Türkiye”, Energy, Volume 284, 2023, 128575, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2023.128575.

[14]M. Fayaz, D. Kim, ‘‘A prediction methodology of energy consumption based on deep extreme learning machine and comparative analysis in residential buildings,’’ *Electronics*, vol. 7, no. 10, p. 222, Sep. 2018.

[15]S. M. J. Jalali, S. Ahmadian, A. Khosravi, M. Shafie-Khah, S. Nahavandi, and J. P. S. Catalao, ‘‘A novel evolutionary-based deep convolutional neural network model for intelligent load forecasting,’’ *IEEE Trans.Ind. Informat.*, vol. 17, no. 12, pp. 8243–8253, Dec. 2021, doi:10.1109/TII.2021.3065718.

[16]Fumo, Nelson & Biswas, Mantosh. (2015). Regression analysis for prediction of residential energy consumption. Renewable and Sustainable Energy Reviews. doi : 47. 332-343. 10.1016/j.rser.2015.03.035.

[17]K. Yan, W. Li, Z. Ji, M. Qi and Y. Du, "A Hybrid LSTM Neural Network for Energy Consumption Forecasting of Individual Households," in IEEE Access, vol. 7, pp. 157633-157642, 2019, doi: 10.1109/ACCESS.2019.2949065.