Time Series Challenge

Written Report

Jagruti Godambe

Matrikel Number: 23221403

Water and Heat Demand Forecasting

1 Abstract

This study investigates the use of deep learning methods to precisely forecast heat and water usage in a northern Denmark city or District Metered Area (DMA). Precise demand prediction is essential for maximizing the use of available resources, improving energy economy, and cutting expenses in a number of industries, including industry, agriculture, and urban planning. The study builds predictive models using pertinent meteorological data and an hourly dataset that includes heat and water usage.

A Long Short-Term Memory (LSTM) network , a kind of recurrent neural network especially skilled in time series forecasting, was put into place to estimate demand 24 hours in advance. Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) metrics were used to assess the LSTM model's performance. The findings show that the LSTM design produces precise and dependable forecasts by successfully capturing the temporal dynamics of heat and water demand.

The efficiency of LSTM networks for time series forecasting, especially in resource management applications, is demonstrated by this study. The efficacy of this methodology underscores its prospects for more refinements and wider implementations in intricate prognostic situations, thereby endorsing sustainable resource and infrastructure planning in both urban and rural contexts.

2 Introduction

The capacity to predict water and heat use with high accuracy is crucial in today's world of growing resource consciousness, especially in metropolitan areas where resource demand can vary greatly due to a variety of variables. Accurate demand forecasting is crucial for improving energy efficiency, cutting operating expenses, and allocating resources as best as possible. This is especially important in places like northern Denmark, where population density and climatic unpredictability make accurate forecasts essential for efficient resource management.

It is often difficult for traditional forecasting tools to account for the complex, nonlinear relationships found in time series data related to resource utilization. It is consequently becoming more and more important to have more sophisticated predictive models that can both capture these dynamics and anticipate future events with accuracy. This study forecasts a District Metered Area's (DMA) water and heat demand for a region in northern Denmark 24 hours in advance using deep learning techniques, specifically the Long Short-Term Memory (LSTM) network.

Time series forecasting benefits greatly from the recurrent neural network design known as LSTM, which can capture temporal patterns and long-term dependencies in data. Through the use of an extensive dataset comprising hourly water and heat usage measurements and pertinent meteorological data, this study seeks to create a model that greatly increases forecasting accuracy. The results of this study should help with more effective and sustainable resource management, which will help with a variety of applications ranging from agricultural management to urban planning.

In conclusion, this study not only tackles the technical difficulty of demand forecasting but also emphasizes the wider ramifications of using cutting-edge deep learning methods to solve resource management issues in practical settings. The technique employed, the conclusions achieved, and the possible implications of these discoveries for future forecasting endeavors are all covered in length in the sections that follow.

3 Methodology

3.1 Dataset Overview

Heat Demand Data for Urban Area : Training Data consist of hourly heat consumption for an urban area in kWh. This data is referred to as `Training_HeatDMA`. Meter Data has the quantity of heat meters in the city that are used to measure consumption is referred to file `HeatDMA_Number_of_Meters`. Testing Data consist of consumption with 24 ahead missing forecasting values, labeled as `Testing_HeatDMA`.

Water Demand Data for Urban and Rural Areas : Training Data consist of hourly heat consumption data for an urban and rural area in kWh. This data is referred to as `WaterDMA1_Training`and `WaterDMA2_Training`. Meter Data has the quantity of heat meters in the city that are used to measure consumption is indicated by the notation `WaterDMA1_Number_of_Meters` and `WaterDMA2_Number_of_Meters`. Data is similar to the heat demand data, missing values occur every eighth day, necessitating 24-hour ahead forecasts for each DMA, labeled as `WaterDMA1_Testing` and `WaterDMA2_Testing`.

3.2 Data Preprocessing

3.2.1 Data Loading and Cleaning

Separate three models were constructed for heat and water consumption, despite the fact that both forms of demand forecasting (water1, water2, and heat) used comparable preprocessing methods. Because the data columns were uniform, the initial preprocessing stages for the forecasting of both heat and water consumption were similar. The procedure that was followed was:

- Data Loading : Pandas DataFrames were used to import data from CSV files.
- Date Format Correction : Date format was standardized to ensure uniformity between datasets.
- Data Merging : Create a comprehensive dataset for each demand type by combining weather, meter, and consumption data on the time.

3.2.2 Linear Interpolation

Certain hourly heat consumption measurements have missing values during the preparation stage of the data. In order to mitigate this issue, we utilized linear interpolation, a technique that approximates missing data points by creating a straight line connecting current values falling within the same time frame. By successfully preserving the dataset's temporal continuity, this technique makes sure that the LSTM model was trained on accurate and consistent data. Notably, the statistics on water consumption had no missing values.

3.2.3 Feature Extraction

One important component that has been engineered in order to compute this feature and provide a normalized evaluation of meter efficiency, the total energy consumed is divided by the number of meters. Apart from that to enhance model performance in the process of forecasting urban water demand, a number of time-based variables were taken out of the timestamp.

- Consumption per Meter
- Daylight saving time
- The time of day
- The day of the week
- The month
- Whether or not the day occurs on a weekend

.

These temporal aspects represent the seasonal variations, human activities, and work schedules that are linked to cyclical patterns in water usage, which are typically more prominent in metropolitan contexts. These time-based characteristics were not included in rural models; they were solely utilized to forecast water consumption in urban ones.

3.2.4 Normalization: Min-Max Scaling

Min-Max The process of scaling involves converting numerical properties to a predetermined range, usually between 0 and 1, in order to normalize them. It guarantees that every feature adds the same amount to a model, particularly when features have various sizes or units. By modifying each feature's values in accordance with the dataset's minimum and maximum values, the scaling procedure improves the comparability of the features and boosts the efficiency of numerous machine learning methods.

3.2.5 Categorical One-Hot Encoding

This approach transforms categorical variables into binary vectors, creating a separate column for each category. For example, if a variable "Weather" can be "Sunny," or "Rainy," one-hot encoding generates three binary columns, each representing one weather condition. A value of '1' indicates

Data/Metrics	MAE	MAPE (%)	RMSE
Heat DMA	368.01	9.43	535.12
Water DMA 1	11.52	115.17	14.27
Water DMA 2	0.50	19.19	0.73

Fig. 1: Evaluation of Model Performance on Prediction Alongside Ground Truth

the presence of a category, while '0' indicates its absence. This method is essential for algorithms that require numerical inputs, as it prevents models from incorrectly interpreting categorical variables as ordinal.

3.3 Model Architecture

The forecasting models for Heat Demand, Water DMA 1, and Water DMA 2 are based on a similar Long Short-Term Memory (LSTM) neural network architecture. Each model processes a merged dataset containing relevant features such as consumption data, contributing meters, and weather data. While the core architecture remains consistent across all three models, certain parameters have been adjusted to better fit the specific characteristics of each dataset.

3.3.1 Heat Demand Forecasting (Urban)

The heat demand forecasting model also utilizes a 5-layer LSTM neural network architecture. The input layer takes in heat consumption data, contributing meters, and weather data. It includes two LSTM hidden layers, each with 256 neurons, followed by a linear transformation layer and an output layer that predicts heat consumption for the next 24 hours. The model is trained using the Adam optimizer with an MSE loss function, a batch size of 32, and a learning rate of 0.001.

3.3.2 Water Demand Forecasting (Rural)

The Water DMA 1 model is also based on an LSTM neural network, designed to predict water demand using similar features to the Heat Demand model. This model mirrors the architecture of the Heat Demand model, with two LSTM layers of 256 neurons each, followed by a linear transformation layer and an output layer that predicts the next 24-hour water demand. It shares the same training parameters, utilizing a batch size of 32 and a learning rate of 0.001, optimized using the Adam optimizer and MSE loss function.

3.3.3 Water Demand Forecasting (Urban)

The urban water demand forecasting model is structured around a 4-layer Long Short-Term Memory (LSTM) neural network. The input layer receives data from three sources: water consumption, contributing meters, and weather data. This is followed by a single LSTM hidden layer consisting of 256 neurons, designed to capture temporal patterns in the dataset. The LSTM output is passed through a linear layer and an output layer that predicts water consumption for the next 24 hours. The model is trained using the RMS optimizer with a Mean Squared Error (MSE) loss function, a batch size of 32, and a learning rate of 0.0006.

4 Model Evaluation

Across several District Metered Areas (DMAs), the model's performance was assessed using critical metrics like Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE). The assessment is summarized as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{1}$$

where $\hat{y}_i$ is the predicted value, y_i is the actual value, and n is the number of observations.

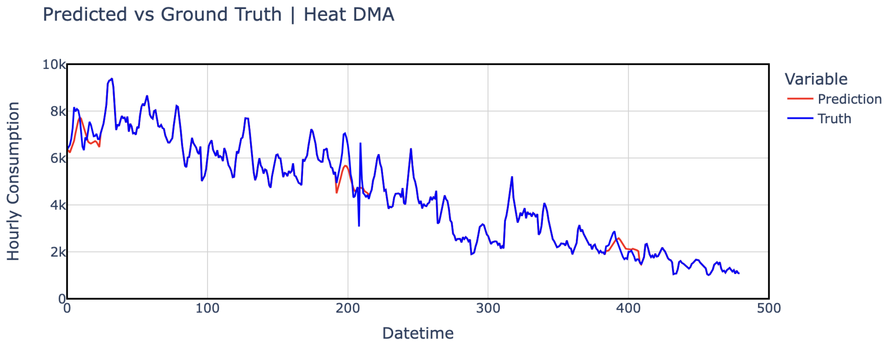
$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100\% \tag{2}$$

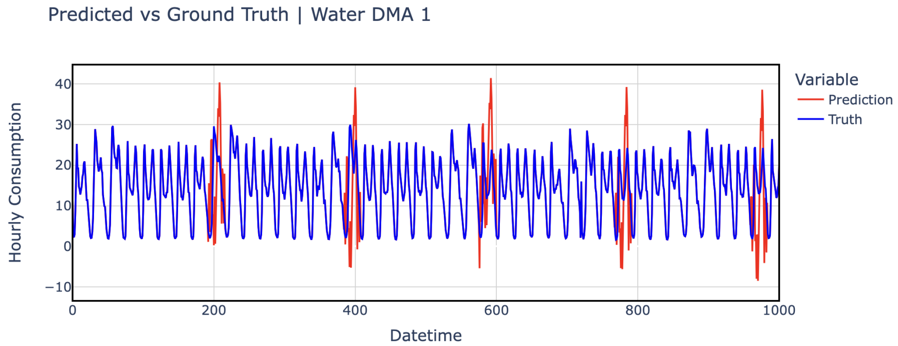
where $\hat{y}_i$ is the predicted value, y_i is the actual value, and n is the number of observations.

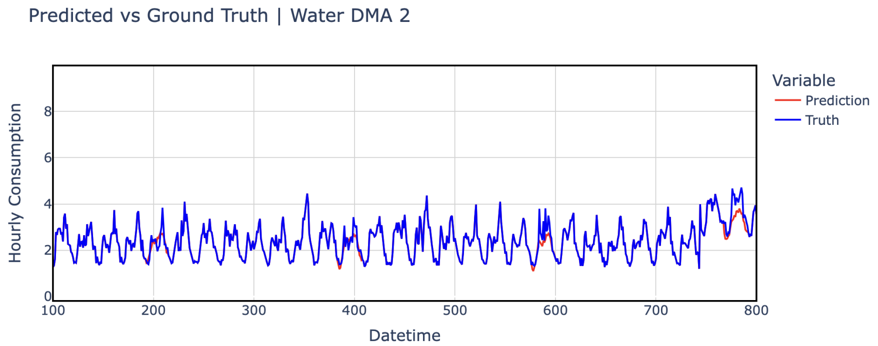
$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \tag{3}$$

where $\hat{y}_i$ is the predicted value, y_i is the actual value, and n is the number of observations.

Figure 1 provides the evaluation metrics for the model's performance on predictions, while Figures 2, 3, and 4 display the graphical results, where the







blue lines represent the ground truth values, and the red lines show the model's predictions. These visual comparisons highlight the model's predictive accuracy across different datasets.

5 Conclusion

LSTM models are highly effective for forecasting heat and water demand due to their ability to capture temporal and non-linear patterns in time series data. LSTM outperform traditional methods but require more data, computational power, and careful tuning.

6 References

- S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," in *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 15 Nov. 1997, doi:
- https://medium.com/@rebeen.jaff/what-is-lstm-introduction-to-long-short-term-memory-66bd3855b9ce
- https://medium.com/swlh/seasonality-analysis-and-forecast-in-time-series-b8fbba820327