UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Electric Load Forecasting for Educational Buildings Using Machine Learning

Diploma Thesis

Koutsi Angeliki

Supervisor: Bargiotas Dimitrios

September 2025

UNIVERSITY OF THESSALY
SCHOOL OF ENGINEERING
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Electric Load Forecasting for Educational Buildings Using Machine Learning

Diploma Thesis

Koutsi Angeliki

Supervisor: Bargiotas Dimitrios

September 2025

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Πρόβλεψη Ηλεκτρικού Φορτίου για Κτίρια Εκπαίδευσης με Χρήση Μεθόδων Μηχανικής Μάθησης

Διπλωματική Εργασία

Κούτση Αγγελική

Επιβλέπων: Μπαργιώτας Δημήτριος

Σεπτέμβριος 2025

Approved by the Examination Committee:

Supervisor **Bargiotas Dimitrios**

Professor, Department of Electrical and Computer Engineering, University of Thessaly

Member **Daskalopulu Aspassia**

Associate Professor, Department of Electrical and Computer Engineering, University of Thessaly

Member **Fevgas Athanasios**

Laboratory Teaching Staff, Department of Electrical and Computer Engineering, University of Thessaly

Acknowledgements

I would like to sincerely thank my supervisor, Professor Bargiotas Dimitrios, for the guidance and support throughout the development of this thesis. His expertise and encouragement were essential in helping me overcome challenges and achieve the objectives of this work.

I am also grateful to the members of the Examination Committee, Professors Daskalopulu Aspassia and Fevgas Athanasios, for dedicating their time to evaluate my work and for providing valuable feedback and insights.

I would like to express my deepest gratitude to my family, whose love, patience, and constant encouragement have supported me throughout this journey. Their belief in me has been a source of strength during the most challenging moments.

I am equally thankful to my friends, who have stood by me with kindness, understanding, and motivation. Their support and companionship made the demanding times lighter and more enjoyable.

DISCLAIMER ON ACADEMIC ETHICS AND INTELLECTUAL PROPERTY RIGHTS

«Being fully aware of the implications of copyright laws, I expressly state that this diploma thesis, as well as the electronic files and source codes developed or modified in the course of this thesis, are solely the product of my personal work and do not infringe any rights of intellectual property, personality and personal data of third parties, do not contain work / contributions of third parties for which the permission of the authors / beneficiaries is required and are not a product of partial or complete plagiarism, while the sources used are limited to the bibliographic references only and meet the rules of scientific citing. The points where I have used ideas, text, files and / or sources of other authors are clearly mentioned in the text with the appropriate citation and the relevant complete reference is included in the bibliographic references section. I also declare that the results of the work have not been used to obtain another degree. I fully, individually and personally undertake all legal and administrative consequences that may arise in the event that it is proven, in the course of time, that this thesis or part of it does not belong to me because it is a product of plagiarism».

The declarant

Koutsi Angeliki

Diploma Thesis

Electric Load Forecasting for Educational Buildings Using Machine Learning

Koutsi Angeliki

Abstract

Electric load forecasting is an essential component of modern energy management, allowing utilities, regulators, and facility operators to estimate future electricity demand and plan appropriately. Precise projections facilitate the utilization of resources for generation and distribution, lower operating expenses, while improving system reliability, all of which contribute to more general sustainability objectives. In the building context, load forecasting can further guide energy saving initiatives, demand-side management programs, and peak load reduction strategies. Educational buildings, in particular, pose a unique forecasting issue because of their highly changeable usage patterns, which are affected by academic timetables, seasonal breaks, and shifting occupancy levels. These elements work together to produce a complicated and constantly shifting demand profile, as do the effects of external variables like temperature. Therefore, it is essential to comprehend and forecast these consumption trends in order to improve operational effectiveness, reduce energy waste, and facilitate the transition to more resilient and sustainable energy systems.

This thesis focuses on predicting the electricity consumption of educational buildings using historical load and weather data from the ASHRAE Great Energy Predictor III dataset. The study evaluates three different forecasting approaches—AutoRegressive Integrated Moving Average (ARIMA), Extreme Gradient Boosting (XGBoost), and Long Short-Term Memory (LSTM) neural networks. By comparing the performance of these models using standard accuracy metrics, the research aims to identify the most effective approach for short-term load forecasting.

The findings indicate that even though LSTM and ARIMA offer accurate predictions, XGBoost performs best, surpassing the other models by a significant margin with an RMSE of 2.25 kW and a MAPE of 2.39%. These results demonstrate the extent to which machine learning techniques –in particular, gradient boosting– work for predicting short-term electri-

cal loads in educational facilities.

Keywords:

Electricity, Energy, Load Forecasting, Buildings, ARIMA, XGBoost, LSTM

Διπλωματική Εργασία

Πρόβλεψη Ηλεκτρικού Φορτίου για Κτίρια Εκπαίδευσης με Χρήση Μεθόδων Μηχανικής Μάθησης

Κούτση Αγγελική

Περίληψη

Η πρόβλεψη ηλεκτρικού φορτίου αποτελεί βασικό στοιχείο της σύγχρονης ενεργειακής διαχείρισης, επιτρέποντας σε εταιρείες κοινής ωφέλειας, ρυθμιστικές αρχές και διαχειριστές εγκαταστάσεων να εκτιμούν τη μελλοντική ζήτηση ηλεκτρικής ενέργειας και να προγραμματίζουν ανάλογα. Οι ακριβείς προβλέψεις διευκολύνουν την ορθολογική αξιοποίηση των πόρων για την παραγωγή και διανομή, μειώνουν τα λειτουργικά έξοδα και βελτιώνουν την αξιοπιστία του συστήματος, συμβάλλοντας παράλληλα στην επίτευξη ευρύτερων στόχων βιωσιμότητας. Στο πλαίσιο των κτιρίων, η πρόβλεψη φορτίου μπορεί επιπλέον να καθοδηγήσει πρωτοβουλίες εξοικονόμησης ενέργειας, προγράμματα διαχείρισης ζήτησης και στρατηγικές μείωσης φορτίου αιχμής. Τα εκπαιδευτικά κτίρια, ειδικότερα, παρουσιάζουν μια ιδιαίτερη πρόκληση πρόβλεψης λόγω των ιδιαίτερα μεταβαλλόμενων προτύπων χρήσης τους, τα οποία επηρεάζονται από ακαδημαϊκά προγράμματα, εποχιακές διακοπές και μεταβαλλόμενα επίπεδα πληρότητας. Οι παράγοντες αυτοί, σε συνδυασμό με τις επιδράσεις εξωτερικών μεταβλητών όπως η θερμοκρασία, δημιουργούν ένα πολύπλοκο και συνεχώς μεταβαλλόμενο προφίλ ζήτησης. Συνεπώς, η κατανόηση και η πρόβλεψη αυτών των τάσεων κατανάλωσης είναι απαραίτητη για τη βελτίωση της λειτουργικής αποδοτικότητας, τη μείωση της ενεργειακής σπατάλης και τη διευκόλυνση της μετάβασης σε πιο ανθεκτικά και βιώσιμα ενεργειακά συστήματα.

Η παρούσα διπλωματική εργασία επικεντρώνεται στην πρόβλεψη της κατανάλωσης ηλεκτρικής ενέργειας σε εκπαιδευτικά κτίρια, χρησιμοποιώντας ιστορικά δεδομένα φορτίου και καιρικών συνθηκών από το σύνολο δεδομένων ASHRAE Great Energy Predictor III. Η μελέτη αξιολογεί τρεις διαφορετικές προσεγγίσεις πρόβλεψης — το στατιστικό μοντέλο AutoRegressive Integrated Moving Average (ARIMA), τον αλγόριθμο μηχανικής μάθησης Extreme Gradient Boosting (XGBoost) και τα νευρωνικά δίκτυα Long Short-Term Memory (LSTM). Μέσω της σύγκρισης της απόδοσης των μοντέλων με χρήση καθιερωμένων μετρικών ακρίβειας, η έρευνα στοχεύει στον εντοπισμό της πιο αποτελεσματικής μεθόδου για βραχυπρόθεσμη πρόβλεψη φορτίου.

Τα ευρήματα δείχνουν ότι, παρόλο που οι μέθοδοι LSTM και ARIMA παρέχουν ακριβείς προβλέψεις, η XGBoost αποδίδει καλύτερα, ξεπερνώντας σημαντικά τα υπόλοιπα μοντέλα με RMSE 2.25 kW και MAPE 2.39%. Αυτά τα αποτελέσματα καταδεικνύουν τον βαθμό στον οποίο οι τεχνικές μηχανικής μάθησης –στην συγκεκριμένη περίπτωση το gradient boosting– είναι αποτελεσματικές για την πρόβλεψη βραχυπρόθεσμων ηλεκτρικών φορτίων σε εκπαιδευτικές εγκαταστάσεις.

Λέξεις-κλειδιά:

Ηλεκτρισμός, Ενέργεια, Πρόβλεψη Φορτίου, Κτίρια, ARIMA, XGBoost, LSTM

Table of contents

List of figures

List of tables

Abbreviations

ADAM	Adaptive Moment Estimation
AI	Artificial Intelligence
AIC	Akaike Information Criterion
ARIMA	AutoRegressive Integrated Moving Average
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
CNN	Convolutional Neural Network
DL	Deep Learning
FNN	Feedforward Neural Network
kBTU	Thousand British Thermal Unit
kW	Kilowatt
LSTM	Long Short-Term Memory
LTLF	Long-Term Load Forecasting
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MTLF	Medium-Term Load Forecasting
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
STLF	Short-Term Load Forecasting
XGBoost	Extreme Gradient Boosting

Chapter 1

Introduction

Electrical power system management and optimization have advanced significantly as a result of the world's growing energy consumption. Buildings are among the industries that use the most power, making up a major portion of global energy demand. Therefore, in the larger framework of sustainability and climate change mitigation, the efficient operation of building energy systems is essential. Buildings, being complex and dynamic systems, display a wide range of energy consumption patterns that are impacted by a variety of factors such as occupancy behavior, use schedules, environmental conditions, and technological facilities. This complex nature makes it difficult to comprehend and successfully manage their energy demands. Unpredictability in demand, along with changes in external factors, such as weather, results in major uncertainty for energy providers and facility managers. Furthermore, new technologies like smart meters, renewable energy sources, and demand response programs are changing the way buildings use electricity. While these innovations provide several benefits, they also increase complexity and unpredictability. Therefore, it is essential to find trustworthy methods for comprehending and forecasting building energy usage in order to increase efficiency and promote a cleaner and smarter energy future. Overall, these features and complications emphasize the need for improved tools and approaches that can capture the multidimensional nature of building energy usage. Addressing these concerns is crucial for improving energy efficiency, lowering operational costs, and contributing to sustainable energy systems on both local and global scale.

1.1 Thesis Object

This thesis aims to tackle the difficulty of precisely predicting the power usage of educational facilities, which are distinguished by wildly fluctuating and inconsistent demand patterns. As previously stated, these patterns can be affected by academic calendars, seasonal breaks, varying occupancy, and weather conditions, all of which complicate the act of forecasting. The goal of this thesis is to provide an accurate estimate of short-term load in such scenarios, allowing for better energy planning, waste reduction, and operational efficiency.

To achieve this, the thesis makes use of historical energy consumption and weather data from the ASHRAE Great Energy Predictor III dataset. Three different forecasting approaches are applied and compared:

- ARIMA (AutoRegressive Integrated Moving Average), a classical statistical method that models time-series data by capturing trends and seasonality.
- XGBoost (Extreme Gradient Boosting), a machine learning method that builds an ensemble of decision trees and is known for its accuracy and robustness in handling complex data.
- LSTM (Long Short-Term Memory) neural networks, a type of deep learning model specifically designed to capture temporal dependencies in sequential data, making it suitable for load forecasting tasks.

By evaluating these approaches using standard accuracy metrics, the thesis seeks to establish which methodology delivers the most trustworthy forecasts in the context of educational buildings. In doing so, it also emphasizes the trade-offs between simplicity and interpretability (as in ARIMA), performance and flexibility (as in XGBoost), and modeling power at the cost of computational complexity (as in LSTM). The results are intended to give significant insights into each method's strengths and limits, as well as suggestions for future forecasting applications in building energy management.

1.1.1 Contribution

This thesis' contribution can be summarized as follows:

1. Study of the main problem. Specifically, the nature and patterns of energy consumption

in educational buildings, based on periods of buildings' employment during time and weather conditions.

2. Data filtering and preprocessing, which includes collecting, cleaning, and transforming the data of the ASHRAE Great Energy Predictor III dataset and turning them into the desired form for the forecasting.

3. Construction and implementation of the three models; ARIMA, XGBoost, and LSTM.

4. Performance evaluation using the metrics RMSE, MAE, and MAPE. In this way, the performance of each method in short-term load forecasting is assessed.

5. Comparing the evaluation metrics of the models to determine which gives the best result and addressing their benefits and limitations.

1.2 Structure of Thesis

The rest of this thesis is structured as follows: Chapter 2 presents the outline and importance of load forecasting using conventional statistical methods and modern ones, such as machine and deep learning. Chapter 3 provides a full overview of the ARIMA, XGBoost, and LSTM modeling methodologies. Chapter 4 describes the implementation including the dataset itself, data handling, and models used step-by-step. Moreover, it analyzes and examines the findings of the experiments, providing a comparative study of the three forecasting methods. Finally, Chapter 5 summarizes the main findings, considers the study's limitations, and suggests options for further research.

Chapter 2

Load Forecasting

2.1 Introduction

Electric load forecasting is the practice of estimating future energy demand using historical usage data, external factors like weather, and contextual indications such as occupancy patterns. For power systems to operate dependably and economically, accurate forecasting is essential because it facilitates tasks like managing demand response, scheduling generating units, and planning infrastructure investment [1]. With the increasing influence of renewable integration and decarbonization legislation, the significance of precise load forecasting in modern-day energy systems has increased.

Buildings account for a major portion of worldwide power consumption, making the building industry important in terms of load forecasting. Due to academic calendars, class schedules, exam periods, and holiday closures, educational buildings show special electricity use patterns [2]. They differ from other building categories due to the significant weekly and seasonal fluctuations created by these circumstances. Thus, forecast accuracy in educational settings has been demonstrated to increase when building-specific data, such as schedules and occupancy are included [2].

The emergence of improved metering technology and open access datasets has encouraged research on building-level solutions. In particular, the ASHRAE Great Energy Predictor III [3] dataset offers a thorough baseline to compare load forecasting techniques for various building types. This thesis focuses on educational buildings from the dataset and applies the forecasting approaches ARIMA, XGBoost and LSTM. Their results are evaluated by the metrics of root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The next sections of this chapter give an overview of load forecasting techniques, their importance, and the variety of methodologies available.

2.2 Load Forecasting Fundamentals

Load forecasting is the practice of predicting future power consumption over a set time period. Its main goal is to guarantee that supply and demand may be consistently balanced, enabling the safe, effective, and cost-effective operation of power systems [1]. For a variety of stakeholders, such as utility companies, grid operators, legislators, and building energy managers, reliable forecasts are highly desirable.

Forecasting is categorized in three major types based on the time intervals they cover. Short-term load forecasts (STLF), spanning from minutes to days, which are used to support real-life system operation control, demand response programs, and economic plans. Medium-term load forecasts (MTLF), from weeks up to months, for energy procurement strategies and maintenance planning in generation, transmission, and distribution systems. Long-term load forecasts (LTLF), ranging up to years, contributing in building an infrastructure development and capacity planning guide [1].

A number of factors affect load forecasts. The weather (temperature, humidity, solar intensity), calendar factors (day of the week, holidays, season of the year), and socioeconomic factors (population growth, business activity) are important drivers of electrical, heating, and cooling demand [1]. Operating schedules, occupancy trends, and climate conditions are all equally important at the building level.

Figure 2.1 shows a flow diagram of the method of load forecasting. Typical steps in the load forecasting pipeline include [1]:

1. Data acquisition: Assembling load data over time collectively with exogenous ones, like weather, occupancy, and calendar information.

2. Data preprocessing: Filtering data, dealing with missing values, and bringing the dataset into a desirable form.

3. Modeling: Building and training statistical, machine, or deep learning models to shape a relationship between inputs and outputs.

4. Evaluation: Evaluating forecast accuracy by comparing model behavior using parameters like RMSE, MAE, and MAPE.

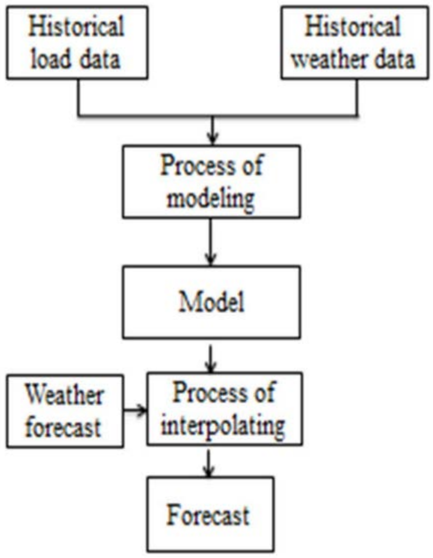

Figure 2.1: Flow chart of load forecasting method [1]

2.3 Statistical Methods

2.3.1 Overview

Among the first and most popular approaches for load forecasting are statistical methods. Using historical load patterns as a basis, these methods presume that statistical models with particular assumptions about linearity, stationarity, and noise can represent the connections between demand and explanatory variables [4]. Historically, statistical methods have been widely employed in utility forecasting practices, especially for short- and medium-term horizons, due to their ease of use, interpretability, and lower computational cost. Regression-based methods, autoregressive models, and seasonal models are some of the most frequently utilized statistical techniques. They are also helpful as benchmark models that are used to evaluate more complex deep learning and machine learning algorithms [4].

2.3.2 Autoregressive Integrated Moving Average (ARIMA)

The ARIMA model is one of the most largely used time-series forecasting methods. It combines three elements [5]:

- Autoregression (AR), which uses past data to predict future values.

- Integration (I), which makes use of data differencing to achieve stationarity.

- Moving Average (MA), which integrates linearly previous forecast errors to predict the error terms.

ARIMA models are prominently suitable for STLF when the examined series appear to be stationary –or can be converted into a stationary format– as they have strong autocorrelation [5]. However, they do not function favorably with nonlinear inputs, external influences, such as weather, and unexpected noise [5].

From a mathematical scope, each one of the ARIMA components uses a variable, p, d, and q, and it is often referred as ARIMA(p,d,q). Specifically:

- p: autoregressive lag observations (number of past values used for the prediction),

- d: degree of differencing (times the previous value was subtracted from the current value to remove trends or seasonality),

- q: moving average window (number of past forecast errors used in the model).

The general form of the ARIMA(p,d,q) is expressed as [5]:

$$\phi_p(B)(1-B)^d y_t = \theta_q(B)\epsilon_t$$

where:

- y_t is the load at time t,

- B is the backshift operator $(By_t = y_t - 1)$,

- $\phi(B) = 1 - \phi_1 B - ... - \phi_p B^p$ is the AR polynomial,

- $\theta(B) = 1 - \theta_1 B - ... - \theta_q B^q$ is the MA polynomial,

- ϵ_t is a white noise error term.

2.3.3 Advantages and Limitations

Statistical techniques, and ARIMA in particular, have been used for a long time in load forecasting because of their many benefits. Their interpretable and distinct parameters clarify the way the historical load patterns affect projections for the future. They are computationally efficient and relatively simple to build, which makes them appropriate for real-time or resource-constrained applications. Furthermore, when the load series is stationary or shows consistent seasonal patterns, these techniques function as trustworthy baseline models and frequently generate precise projections [4].

There are some significant restrictions, though. ARIMA is based on linear connections, which can be restrictive when modeling becomes complicated and nonlinear patterns are found in building electricity demands. They might need differencing or other changes in order to achieve their requirement that the underlying time series be stationary. Selecting parameters for seasonal and non-seasonal components can be delicate and time-consuming, demanding automated tuning processes or specific expertise. Lastly, these models are less effective than more recent machine learning or deep learning techniques due to their restricted scalability when working with high-dimensional datasets or when including several exogenous variables [4]. Despite these drawbacks, ARIMA is nevertheless important as an interpretable statistical standard that serves as a basis for evaluating the effectiveness of advanced forecasting techniques.

2.4 Machine Learning Methods

2.4.1 Overview

Machine learning (ML) is a type of computational approach that learns patterns from data without being explicitly programmed with predefined rules [6]. In contrast to conventional statistical models, ML algorithms are able to detect complicated relations, high-dimensional dependencies, and nonlinear correlations between input data and the target variable [1].

Classic statistical approaches, like ARIMA, frequently struggle with nonlinearities, many exogenous factors, and big datasets. Hence, ML techniques are becoming more and more prominent in the field of electric load forecasting [1]. To increase forecast accuracy, ML typically combines a variety of data sources, including historical load, weather, occupancy, and calendar data. Compared with classic techniques, these methods are able to automatically identify patterns in data, eliminating the need for manual feature engineering [6].

Notable machine learning approaches used in load forecasting are [6]:

- Decision Trees (DTs): Recursive partitioning of input features to predict load.
- Random Forests (RF): Assembly of decision trees that improve precision and reduce overfitting.
- Gradient Boosting Machines (GBM): Sequentially built trees that minimize predictive error using gradient descent.
- Support Vector Regression (SVR): Maps input data into a high-dimensional space to detect nonlinearities.

Gradient boosting algorithms are particularly useful in load forecasting because they offer strong predictive power along with robustness to diverse characteristics and missing data. This thesis focuses on XGBoost since it is one of the most popular and efficient gradient boosting solutions.

2.4.2 XGBoost

Extreme Gradient Boosting (XGBoost) is a gradient-boosted decision tree algorithm that iteratively builds an ensemble of trees, where each new tree attempts to correct the errors of the previous trees [7].

Mathematically, the model predicts a target y_i as [8]:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F}$$

where:

- x_i is the feature vector for sample i,
- f_i is a regression tree in the ensemble,
- $\mathcal{F}$ is the space of regression trees,
- K is the number of trees.

The objective function to minimize includes a loss term and a regularization term:

$$\mathcal{L}(\phi) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k)$$

where l s a differentiable loss function and $\Omega(f_k)$ penalizes tree complexity to avoid overfitting.

Because of its adaptability, XGBoost can effectively handle big datasets, missing values, and heterogeneous characteristics through parallel processing. Furthermore, it provides feature importance analysis, which is highly useful in building energy forecasting since it enables the evaluation of the input variables which have the most impact on predictions [8].

2.4.3 Advantages and Limitations

High predictive accuracy is provided by machine learning techniques, such as XGBoost, especially for complicated, nonlinear, and high-dimensional datasets. They are able to record complex interactions between load and variables that influence it including occupancy, temperature, and historic consumption. Furthermore, XGBoost's integrated regularization helps avoid overfitting, and it is resilient to multicollinearity and missing data [1].

However, these techniques are not without their limitations. To get the best results, they frequently need to have their hyperparameters (such as learning rate, tree depth, and number of trees) adjusted. The overall structure is still less interpretable than statistical models, even when XGBoost offers feature importance. Furthermore, ML models may not function effectively on small datasets and typically call for more data and computational power than conventional methods. Despite these difficulties, the precision and adaptability of machine learning techniques like XGBoost have made them a top choice for load forecasting [1].

2.5 Deep Learning Methods

2.5.1 Overview

Deep learning (DL) is a form of machine learning that uses several hidden layer artificial neural networks to automatically learn hierarchical data representations. Deep learning can immediately extract highly nonlinear, complicated temporal and spatial correlations from raw data, in contrast to standard statistical models that depend on explicit assumptions like linearity or stationarity [1].

DL approaches are very useful for electric load forecasting as load patterns are frequently impacted by a variety of external factors (such as weather, occupancy, and holidays) and show significant temporal dependencies (hourly, daily, weekly, and seasonal) [1]. Such long-term trends may be difficult for classic statistical and ML models to identify, but deep learning –particularly recurrent architectures– has shown better results in this area.

Popular DL approaches for load forecasting include feedforward neural networks (FNNs), convolutional neural networks (CNNs) for feature extraction, and recurrent neural networks (RNNs) for sequence modeling [1].

In sequence modeling applications like load forecasting, RNNs and particularly LSTM have proven to be highly effective algorithms. The present thesis utilizes LSTM networks, which are able to manage complicated temporal dependencies and preserve major historical information, to simulate the sequential patterns of energy consumption in educational buildings. In the following section is presented a framework for LSTMs.

2.5.2 Long Short-Term Memory (LSTM)

The LSTM network is a form of recurrent neural network (RNN) that was created to address the disappearing and expanding gradient issues that limit the ability of ordinary RNNs to capture long-term dependencies. In order to ”remember” essential past data over extensive sequences and ”forget” unimportant parts LSTMs rely on gates to control the information flow through the network [9].

A unit in the LSTM network is composed of three gates [9]:

- Forget gate: decides which data should be deleted from the prior cell state.
- Input gate: chooses which additional data should be included in the cell state.
- Output gate: determines what part of the cell state should be output as the hidden state.

The equations of a single LSTM cell are [9]:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right)$$

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right)$$

$$\tilde{C}_t = \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right)$$

$$h_t = o_t \odot \tanh(C_t)$$

where:

- f_t, i_t, o_t are the forget, input, and output gates,
- h_t is the hidden state,
- x_t is the input at time step t,
- C_t is the cell state,
- W_f, W_i, W_o are weight matrices for the corresponding inputs of each gate type,
- σ is the sigmoid activation and $\odot$ denotes element-wise multiplication.

Generally, when performing load forecasting, the LSTM takes as input a series of historical data and other attributes like weather or time features and predicts the following load values. It is particularly suitable for identifying recurring patterns and seasonal changes in electricity demand because it retains a memory of long-term dependencies [9].

2.5.3 Advantages and Limitations

There are several benefits to using DL methods for load forecasting, especially LSTM networks. They are quite effective at simulating long-term dependencies and nonlinear relations, which are typical in building electrical power usage. LSTMs are more adaptable than statistical techniques because they can capture seasonal cycles and short-term changes without the requirement for explicit feature engineering. Furthermore, they can make use of larger datasets thanks to their capacity to handle massive amounts of data, which improves predicting accuracy [1].

However, there are some disadvantages in DL models. Effective generalization necessitates vast quantities of data, which are not always accessible in practice. In comparison to conventional techniques, they also require a large amount of computing power for training and tuning. Another issue is interpretability: LSTMs function as ”black boxes”, making it challenging to comprehend how particular variables affect predictions, in contrast to statistical models or even other ML methods. Lastly, the selection of hyperparameters (number of layers, neurons, learning rate) has a significant impact on their performance, involving careful modeling and testing [1].

2.6 Evaluation Metrics

The accuracy of forecasts in relation to actual load values must be measured using suitable error metrics in order to assess the accuracy of forecasting models. Error metrics offer a mathematical foundation for evaluating the practical applicability of various models, guiding model selection, and comparing them in the context of electric load forecasting. It is crucial to use strong metrics since forecasting errors have a direct impact on choices about energy planning, scheduling, and operational efficiency.

This thesis refers to three commonly used evaluation metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Each of these emphasize on different aspects of model accuracy.

RMSE

The RMSE is the square root of the average of squared difference between predicted and actual values:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}$$

where y_t is the actual, $\hat{y}_t$ is the predicted load, and n is the number of observations [10]. RMSE is affected more by larger errors because of the squaring, so it is sensitive with abnormal values. RMSE gives a kW value. In the implementation it was also used the Normalized RMSE (%RMSE), which is the RMSE divided by the mean load value for a more understandable insight.

MAE

The MAE calculates the average of the absolute differences between predicted and actual values:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|$$

where y_t is the actual, $\hat{y}_t$ is the predicted load, and n is the number of observations. MAE is less susceptible to anomalies and treats every variation proportionately because it does not square the errors as RMSE does [10]. As a result, MAE offers a clearer understanding of the average error magnitude and is a more reliable indicator of overall prediction accuracy.

MAPE

The MAPE provides a straightforward way of interpreting model performance by expressing the forecast error as a percentage:

$$MAPE = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

where y_t is the actual, $\hat{y}_t$ is the predicted load, and n is the number of observations. Since MAPE shows the average % difference from real values, it is scale-independent and simple to understand. It is limited, though, when real values are near zero since this might result in excessively high percentage errors [10].

When combined, these three measures offer a thorough assessment of the model's performance. While MAE and MAPE offer comprehensible computations of average error in both absolute and relative terms, RMSE highlights big variances and permits comparison across multiple scales. This thesis provides an objective evaluation of the used models models for electric load forecasting by applying this combination of metrics.

Chapter 3

Methodology

3.1 Dataset Description

The dataset used for the implementation is obtained from the ASHRAE Great Energy Predictor III competition on Kaggle platform on 2019 [3]. It was developed for evaluating energy prediction methods and has data of a variety of energy types across a large group of commercial buildings.

The initial dataset includes over 20 million readings attained from meters in 1,448 buildings located in different areas. There are four meter types: Electricity (0), Chilled Water (1), Steam (2), and Hot Water (3).

Every register has a distinct timestamp, meter type, meter reading, and building ID. There are also additional files with building and weather data. The former include information about each building like site ID, primary use, square feet, year of construction, and floor counter. As for the weather data, it was collected from local stations and contains values such as air temperature, dew temperature, wind speed, cloud coverage, and precipitation depth.

For the purposes of this thesis, there were set some filtering criteria, resulting in the analysis to be limited to:

1. Educational buildings: Buildings with education as their primary use, to focus on a particular group with some specific consumption patterns.

2. Electricity meter readings: Data acquisition only from meter type 0, as the main target is electric load forecasting.

After applying these filters to the data, the dataset constricted to 549 educational buildings, out of which 537 had available electricity meter data. These serve as a foundation for the forecasting experiments presented in this thesis.

The principal fields used in the study are the timestamp (time of measurements in hour intervals), the meter reading (electricity load consumption value), the building ID (unique number identifier for each building), and the weather data.

Due to its ability to capture both the temporal dynamics of power usage and the impact of external factors like weather, this dataset offers a rich basis for forecasting model development and testing.

3.2 Data Preprocessing

Preprocessing was necessary to prepare the dataset for the forecasting. This stage made sure that the raw data –which included abnormalities and missing values– was converted into a consistent and helpful format that could be used with statistical, ML, and DL models. Three primary steps comprised the preparation workflow to get the data into a desired form: data cleaning, feature engineering, and train-test separation.

3.2.1 Data Cleaning

Certain information, including year of construction and floor counter, in the original dataset could be relevant building descriptors. Nonetheless, a significant percentage of missing data were present in these two areas. Thus, these variables were removed from the study because it was not possible to do credible imputation without adding bias.

In the original dataset, the meter readings is in kBTU (thousand British Thermal Units). It was used a unit transformation, due to the fact that the focus of this study is on power usage in kilowatts (kW), by multiplying every reading with the factor 0.2931, as:

$$1kBTU/h = 0.29307107kW$$

Meter readings and weather data were also examined for missing entries. Examination of the meter readings revealed no notable irregularities (such as severe spikes or negative values). Consequently, the values were applied as given. For weather data, missing values were handled with common imputation methods (forward/backward filling), keeping a continuous flow in the time series.

3.2.2 Feature Engineering

To improve the dataset's predicting potential, additional characteristics were extracted from timestamp and weather observations.

Time-based features

The following features were generated from the timestamp:

- hour: records trends of consumption throughout the day,
- day of week: discriminates weekdays from weekends,
- month: depicts variations in demand depending on the season,
- lag_1: power consumption of the previous hour, to trace temporal dependencies,
- rolling_3h: a consumption moving average of three hours in order to smooth out variations and record ongoing patterns.

Weather Features

Weather has a big impact on the amount of energy buildings use. In this study, the following meteorological variables was selected among the provided from the weather stations:

- air temperature
- dew temperature
- cloud coverage
- precipitation depth (1 hour)
- wind speed

These factors were chosen because they account for environmental and thermal aspects that can have a big influence on the total energy a facility uses, especially in educational institutions with different occupancy patterns.

3.2.3 Train-Test Split

The dataset was divided into training and testing subsets chronologically in order to assess predicting performance accurately. This ensures that models are trained on historical data and evaluated on future periods that are not yet observed. An 80–20 split was conducted in this study, with 80% of the data being used for model training and the other 20% set aside for testing. This ratio achieves an ideal balance between preserving a strong test set for assessing generalization performance and providing the models with enough historical data for learning.

3.3 Model Descriptions

This section outlines the forecasting models used in this research. Three methods, ARIMA (statistical), XGBoost (ML), and LSTM (DL), were chosen in order to consider methodologies from various categories. Each model was trained to capture the temporal dynamics of power consumption using the dataset mentioned above. The main ideas of each approach and the particular implementation details used in this thesis are covered in the next subsections.

3.3.1 ARIMA

Data Preparation

For the ARIMA model, only the timestamp and meter_reading fields were kept, because it uses only past consumption values to predict the future ones, not correlating the results with external factors, like time or weather features. Since modeling the total power consumption of educational buildings was the goal, the total mean of the electricity consumption of all buildings at each hour used to form a global series. An auxiliary daily series was created by adding up the hourly data into daily average totals in order to eliminate short-term noise and concentrate on more general demand trends. The daily series served as a framework for parameter selection and training of ARIMA model.

Parameter Selection

The selection of the ARIMA parameters (p, d, q) was conducted by using a grid search method. Each parameter could range from 0 to 4 and all potential combinations were examined. First, each of the possible models was fitted to the daily series and then the selection was accomplished using the Akaike Information Criterion (AIC) [5], according to which:

$$AIC = -2\ln(\hat{L}) + 2k$$

where $\hat{L}$ is the likelihood of the model and k is the number of estimated parameters. The model that was finally selected was the one with the lowest AIC value. The ARIMA model order elicited with the assistance of this procedure is (p, d, q) = (4, 1, 4)

Model Training and Forecasting

The selected ARIMA model was fitted on the training subset of the daily series. To guarantee stationarity applied first-order differencing (d=1), resulting in capturing both autoregressive and moving average dynamics. In autoregressive terms (p=4), the past four days' values are used to predict the next day's load and in moving average terms (q=4), the past four days' forecast errors are used to improve the next prediction. The predictions produced were executed on the test subset for the same time period and length.

3.3.2 XGBoost

Feature Selection

XGBoost model uses both time and meteorological features besides raw meter readings. The selected attributes are:

- Time features: hour of day, day of week, month.
- Weather features: air temperature, dew temperature, wind speed, precipitation depth (1-hour), and cloud coverage.
- Lag features: previous hour load (lag_1) and 3-hour rolling average (rolling_3h).

The variables features were picked in order to record periodic trends, weather-related impacts on power consumption, and short-term time dependencies.

Model Setup and Parameter Tuning

The basic model is a `XGBRegressor` with the squared error objective function. For a more accurate performance, a parameter grid was explored with a `RandomizedSearchCV` function. The hyperparameters tuned are:

- n_estimators: number of boosting trees,
- max_depth: maximum tree depth,
- learning_rate: step size shrinkage to prevent overfitting,
- subsample: fraction of training samples used per tree,
- colsample_bytree: fraction of features used per tree,
- min_child_weight: minimum weight of each child,
- reg_alpha: L1 regularization (lasso). Helps with feature selection,
- reg_lambda: L2 regularization (ridge). Helps reduce overfitting.

Min_child_weight, reg_alpha, and reg_lambda are used for regularization while the remaining are the core hyperparameters.

Furthermore, it was applied a cross-validation in the time series with a 5-fold `TimeSeriesSplit` using the RMSE as the optimization principle, i.e. the data was trained in five splits preserving chronological order and making sure that the future data was not used in training. This way temporal consistency is maintained while evaluating models.

Forecasting

The trained model was employed for developing forecasts on the test set. To illustrate and analyze the patterns in actual and expected power consumption, forecasted hourly numbers were then summed up to daily averages.

3.3.3 LSTM

Feature Selection and Scaling

The LSTM model used the same time and weather features as the XGBoost model. Thus, for time features the attributes employed are hour of the day, day of week, and month, for weather features are air temperature, dew temperature, wind speed, precipitation depth (1-hour), and cloud coverage, and for the lag features previous hour load (lag_1) and 3-hour rolling average (rolling_3h). The target element, which is to be predicted, is the meter_reading variable.

To prevent data leaking, all features were normalized using a `MinMaxScaler` before training and applied independently to the training and test sets. For a better convergence during neural network training, the desired variable and input parameters were both scaled into the [0,1] range.

Sequence Generation

LSTMs need sequential inputs and by using a time series generator supervised learning samples were created. Through the use of a 24-hour lookback window, the model was able to forecast the upcoming load value by utilizing the previous 24-hour feature readings. The training process also used a batch size of 128, at each training step, the network processes 128 training samples together.

Model Architecture and Training

The LSTM model architecture consists of:

- A primary LSTM layer with 64 units and return sequences enabled,
- A dropout layer (rate = 0.2) to mitigate overfitting,
- A secondary LSTM layer with 32 units,
- A dropout layer (rate = 0.2),
- A final dense layer with a single output neuron for regression.

For the network training, the Adam (Adaptive Moment Estimation) optimizer was employed with a MSE loss function and MAE served as an additional performance evaluation metric. Training was conducted for up to 10 epochs, each of which represented a full run through the training dataset, with an early termination callback based on validation loss (patience = 5), for the network to iteratively modify its weights and gradually optimize the loss function. When no progress was shown, the best weights were restored.

Forecasting

Predictions were produced on the test set using the trained model. Restoring the original kilowatt scale, required inverse transformation of the predicted values. Lastly, the anticipated hourly loads were combined into daily averages to make comparison and visualization easier.

3.4 Evaluation Metrics

To evaluate the forecasting models' performance, the evaluation metrics (RMSE, MAE, and MAPE) outlined in Section 2.6 were used. As mentioned, in order to give a more comprehensible view, Normalized RMSE (%RMSE) is also presented. Since they represent the absolute amount of errors (RMSE, MAE) as well as their relative size (%RMSE, MAPE), these measures were selected. MAPE and %RMSE offer a scale-independent evaluation of prediction errors in relation to the mean daily electrical load of 74.8 kW over the period of the testing. When combined, they offer a thorough assessment of forecast accuracy that maintains a balance between interpretability and sensitivity to significant deviations.

All implementations were carried out in Python using the Google Colaboratory (Colab) environment, which provides cloud-based computational resources suitable for machine and deep learning experiments. The main libraries used for data handling were `pandas` and `numpy`, for ARIMA modeling `statsmodels`, for the gradient boosting method `xgboost`, and for LSTM implementation `tensorflow/keras`.

Chapter 4

Results & Discussion

This chapter discusses the results of the implementation of the load forecasting models mentioned earlier. The assessment of the methods was done through the chosen error metrics and the results are presented and analyzed in relation with power consumption prediction especially for educational buildings.

Figure 4.1 shows the daily mean power usage of educational buildings over the period of time studied to establish the scene for the forecasting models. The figure illustrates general consumption trends, such as seasonal and sporadic changes, which are major obstacles towards an accurate forecasting.

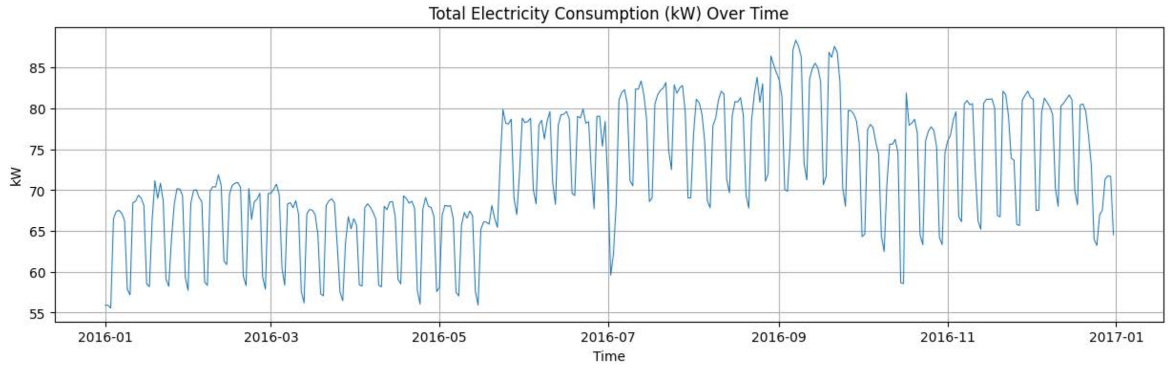

Figure 4.1: Actual Total Electricity Consumption (kW)

4.1 ARIMA

As discussed in Chapter 3 the parameters most suitable to the ARIMA model developed are (p, d, q) = (4, 1, 4).

Figure 4.2 presents a comparison between the actual observed values during the test period and the daily electricity usage predicted by ARIMA. Smaller changes and abrupt spikes are sometimes not entirely replicated by the model, but it does represent the general trend and significant fluctuations.

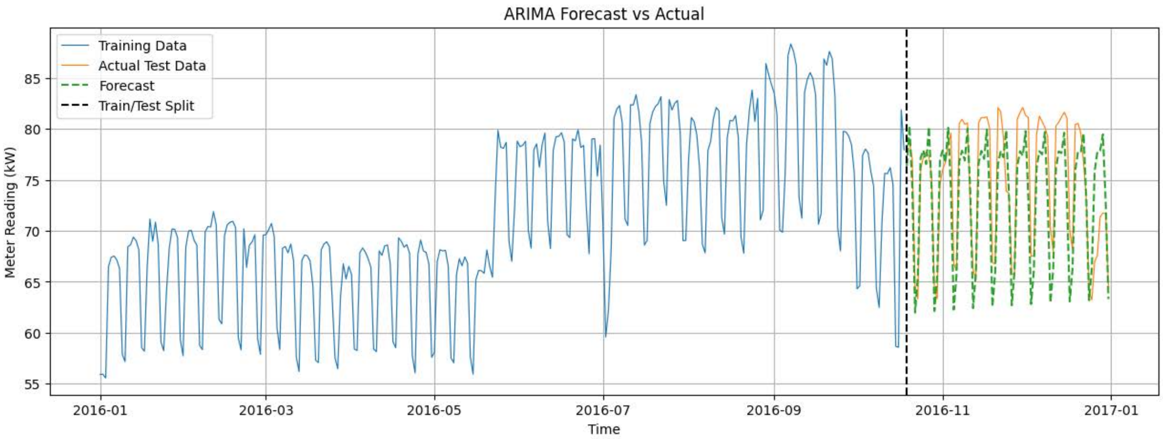

Figure 4.2: ARIMA Forecast vs Actual Electricity Consumption (kW)

Forecast accuracy was calculated with the evaluation metrics specified above. The results are shown in Table 4.1:

Table 4.1: ARIMA Evaluation Metrics

RMSE	15.33 kW
%RMSE	20.50 %
MAE	3.19 kW
MAPE	4.27 %
Mean Load	74.80 kW

The strong agreement between the predicted values and the actual test data demonstrated that the ARIMA model could accurately replicate the general trend and seasonal patterns of electricity usage in the educational facilities. The model produced projections that stayed within a moderate range of the observed load and successfully caught the recurring daily and weekly cycles. Some imbalances were noticeable though: the model had trouble adjusting to abrupt changes in usage, such as those brought on by inconsistent schedules or holidays, and peak demand levels were frequently underestimated. These drawbacks result from ARIMA's linear structure and dependence on historical data, which limit its capacity to take into consideration exogenous variables or non-linear impacts that have an important effect on power consumption.

The model shows good accuracy in capturing overall consumption behavior, with a MAPE of just 4.27%. Nonetheless, the comparatively elevated RMSE and %RMSE results demonstrate the model's weaknesses in managing abrupt load swings and nonlinear fluctuations. Apart from these drawbacks, ARIMA was a powerful baseline forecasting method that showed decent accuracy and emphasized on the significance of taking seasonality into account when predicting load. These results encourage the adoption of more adaptable strategies, such machine learning techniques, as discussed in the following section.

4.2 XGBoost

The XGBoost model was trained on meteorological variables (air temperature, dew temperature, wind speed, precipitation depth, cloud coverage), time-related features (hour, day of the week, month), and designed lag and rolling characteristics. Time-series cross-validation and randomized search were used to optimize the hyperparameters, and RMSE was used to determine the optimal configuration.

The optimal hyperparameters that produced the prediction described in this section are:

Table 4.2: XGBoost Hyperparameters

`n_estimators`	100
`max_depth`	7
`learning_rate`	0.05
`subsample`	1
`colsample_bytree`	1
`min_child_weight`	5
`reg_alpha`	0
`reg_lambda`	10

The comparison of the actual and predicted daily power usage throughout the test period is shown in Figure 4.3. Although there is still some underestimate during sudden fluctuations in demand, the XGBoost model matches more closely the observed values than ARIMA, with better peak and trough alignment.

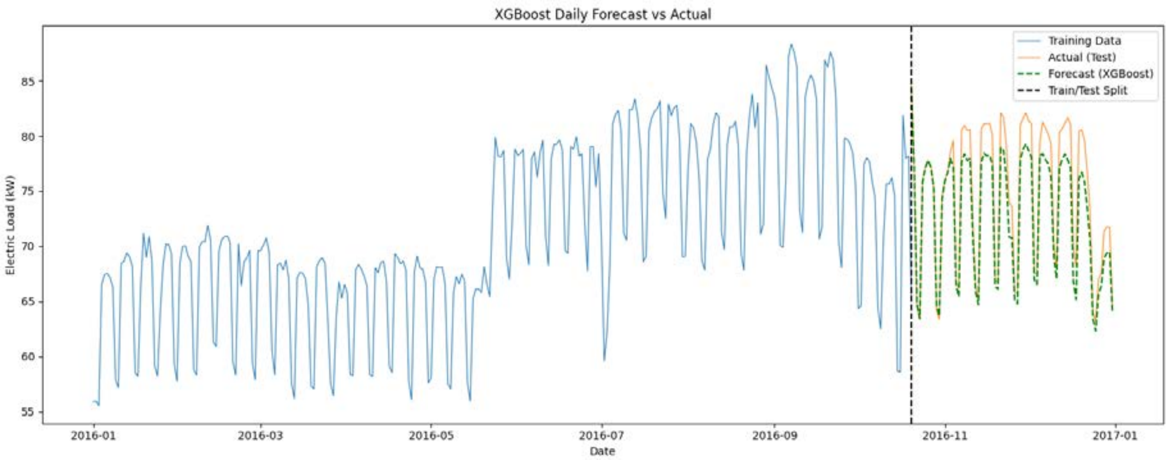

Figure 4.3: XGBoost Forecast vs Actual Electricity Consumption (kW)

The accuracy of the XGBoost modeled is showned in table 4.3. The percentage metrics %RMSE and MAPE are calculated in relation to the mean daily load 74.80 kW, as with ARIMA.

Table 4.3: XGBoost Evaluation Metrics

RMSE	2.25 kW
%RMSE	3.01 %
MAE	1.85 kW
MAPE	2.39 %
Mean Load	74.80 kW

From the figure, it is evident that the model successfully captures the general consumption patterns in the test period, particularly the cyclical weekday demand observed in educational buildings. The forecast aligns closely with the actual values, demonstrating the ability of XGBoost to generalize well from historical training data. However, some discrepancies are visible, most notably during periods of peak demand, where the model tends to underestimate the maximum load. In addition, the model exhibits a degree of smoothing, with sharp transitions in load not fully captured, suggesting that there is need for further incorporation of irregular usage behaviors.

The findings show that ML is superior at capturing nonlinear correlations between weather, temporal characteristics, and power use, with XGBoost achieving significantly lower errors than ARIMA. Despite that there are sporadic deviations at times of rapid fluctuation, the model offers a more accurate description of daily load variations, along with better peak prediction. These findings demonstrate how effectively tree-based models can predict power consumption and pave the way for the use of DL models, like LSTM networks, which are made to identify sequential trends in time series data.

4.3 LSTM

The LSTM model was trained with the same set of features as XGBoost, which included weather, time-related variables, and lagged load characteristics. MinMax normalization was used to scale the data, and a 24-hour lookback window was used to record daily sequential dependencies. The final model design included a dense output layer after two stacked LSTM layers (64 and 32 units) with dropout regularization (rate = 0.2). The Adam optimizer, mean squared error loss, and early stopping were used during the ten epochs of training.

The comparison of the actual and predicted daily power usage throughout the test period is displayed in Figure 4.4. The LSTM model can track the overall consumption trend, but it tends to smooth out abrupt swings and exhibits more deviations during times of high variability.

The evaluation metrics of the LSTM model are reported in Table 4.4. MAPE and %RMSE are both expressed in relation to the 74.80 kW mean daily load.

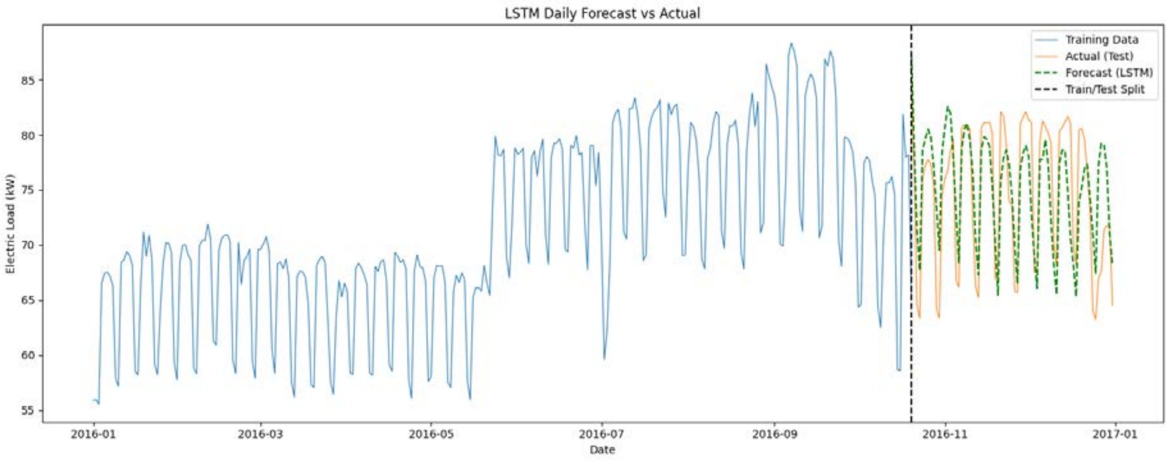

Figure 4.4: LSTM Forecast vs Actual Electricity Consumption (kW)

In its predictions, the LSTM model was able to reproduce the recurrent temporal patterns. The model's predictions correspond to the overall pattern and cyclical behavior of the real load levels. This illustrates how the LSTM can retain consistency with the actual data and learn short-term dependencies, especially during times of ordinary consumption. However,

Table 4.4: LSTM Evaluation Metrics

RMSE	19.39 kW
%RMSE	25.9 %
MAE	3.61 kW
MAPE	4.93 %
Mean Load	74.80 kW

there are occasionally disparities when the model undervalues the actual demand, particularly during periods of high load. This restriction implies that although the LSTM is good at simulating baseline consumption patterns, it is not as good at identifying sudden changes or non-linear spikes, which could be caused by outside variables that are not taken into account by the model (e.g. varied academic schedules, special occasions).

With significantly larger RMSE and error percentages, the LSTM model performs worse than ARIMA and XGBoost. The very brief training period, the limited number of epochs, or the absence of other sequential factors like occupancy patterns may have hindered the network's performance even if it was able to identify long-term relationships in the data. These findings demonstrate that deep learning did not perform better than machine learning in this study, indicating that tree-based models such as XGBoost may prove better for the specified dataset and forecasting horizon.

4.4 Comparative Evaluation

The following charts provide a summary of the assessment metrics of ARIMA, XGBoost, and LSTM in order to evaluate the forecasting models' overall performance. With the lowest errors across all assessment metrics (RMSE = 2.25 kW, MAE = 1.85 kW, MAPE = 2.39%, %RMSE = 3.01%), XGBoost performed noticeably better than the other two methods. This illustrates how effectively gradient boosting captures nonlinear relationships and makes use of both exogenous and time-related data.

The ARIMA model has a reasonably strong performance, with MAPE = 4.27% and RMSE = 15.33 kW. Its result shows the strength of statistical time series models in capturing broad temporal trends, but it also shows their weakness when the series have nonlinear interactions impacted by external variables like the weather.

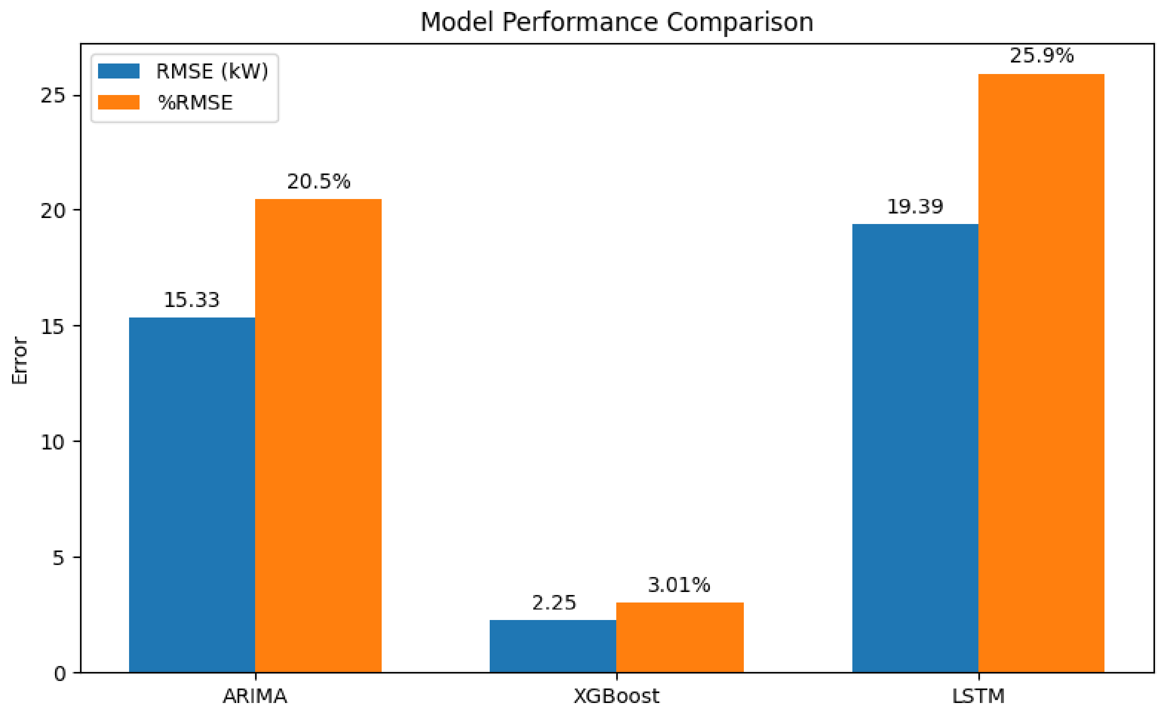

Figure 4.5: Model RMSE & %RMSE Comparison

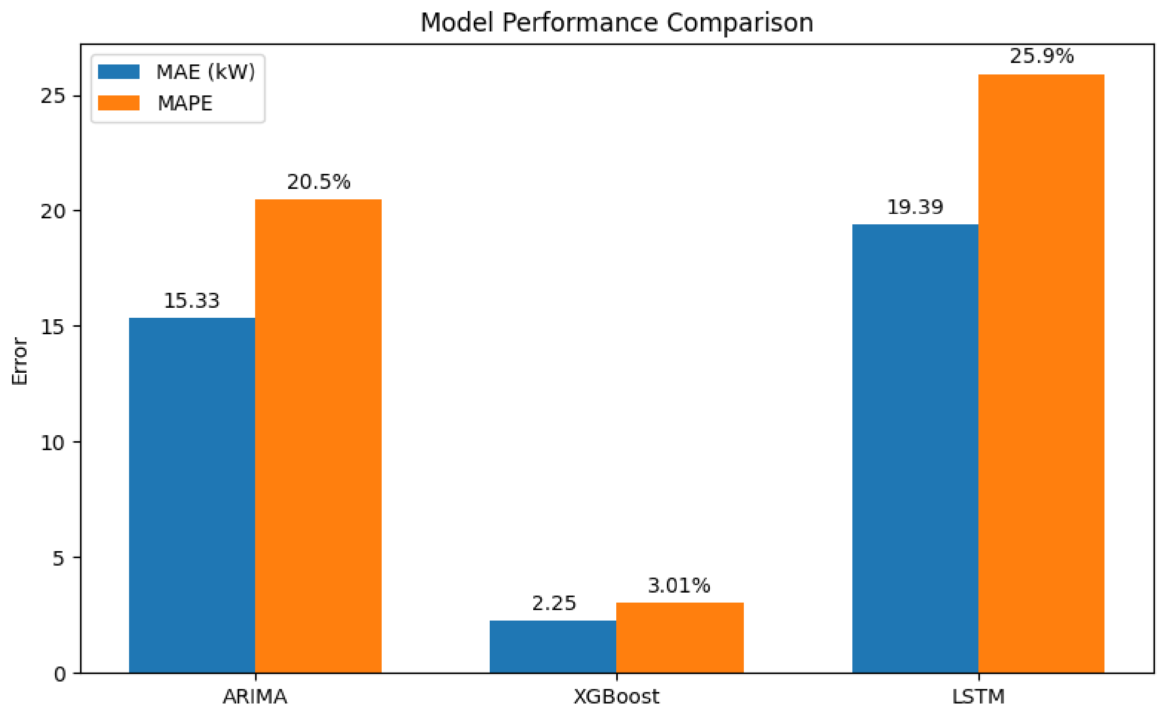

Figure 4.6: Model MAE & MAPE Comparison

In this configuration, the LSTM model performed worse than both ARIMA and XGBoost, with RMSE = 19.39 kW and MAPE = 4.93%. Although complicated temporal dynamics are an appropriate fit for DL techniques, the model setup, training time, and availability of extensive datasets all have a significant impact on the efficacy of such methods In this instance, the predicted accuracy of the LSTM may have been constrained by the few training epochs and rather simplistic design.

In general, according to the results, tree-based ensemble learning (XGBoost) is the best approach for short-term load forecasting in educational buildings, when considering the ASHRAE dataset and the chosen feature set.

Chapter 5

Conclusion

5.1 Summary of Findings

This thesis's main target was to apply statistical, machine learning, and deep learning methods to predict power consumption of educational buildings. The ASHRAE Great Energy Predictor III dataset was employed for supplying daily aggregated electrical power usage data, which was further utilized to test three models: ARIMA, XGBoost, and LSTM.

The findings show that the maximum predicting accuracy was attained using XGBoost, with evaluation metrics RMSE = 2.25 kW, %RMSE = 3.01%, MAE = 1.85 kW, and MAPE = 2.39%. Its capacity to record nonlinear correlations between load, temporal data, and meteorological conditions is the reason for its remarkable efficiency.

A strong baseline was supplied by the ARIMA model, which provided RMSE = 15.33 kW, MAE = 3.19 kW, MAPE = 4.27%, and %RMSE = 20.50%. Its linear form made it difficult to simulate abrupt fluctuations or nonlinear relationships in the load series, even if it was able to accurately depict the general temporal trend.

The LSTM model's metrics (RMSE = 19.39 kW, %RMSE = 25.9%, MAE = 3.61 kW, and MAPE = 4.93%) showed that it was less accurate than both ARIMA and XGBoost. Despite its stated purpose is for modeling sequential dependencies, LSTM may not have been able to properly capture the temporal nature of the data because to its relatively basic design and short training period.

All in all, the comparative analysis shows that while classical statistical models can still serve as a reasonable benchmark, machine learning techniques –especially tree-based models like XGBoost– are highly efficient for short-term load forecasting in educational buildings.

Deep learning, on the other hand, needs more data and careful tuning to outperform simpler methodologies.

In terms of practicality, the findings imply that XGBoost-based forecasting systems can positively impact energy management programs in educational facilities. Precise short-term load forecasts help facility managers support demand-side management activities, minimize wasteful energy usage during off-peak hours, and schedule HVAC systems optimally. Incorporating these predictions into smart campus energy systems can also help achieve sustainability and costs reduction.

5.2 Limitations

Besides the effective performance of ML for load forecasting in educational buildings, there are some drawbacks that should be discussed.

Firstly, the ASHRAE Great Energy Predictor III dataset, which was the source of data in the research, is considerably large and detailed but has insufficient or absent data. Excessive data gaps forced the exclusion of certain potentially useful characteristics, including number of floors of the building and year of construction. The lack of these factors made it more difficult for the models to accurately represent building-specific traits that might affect power usage.

Secondly, only weather and time-related predictors were used in the development of the models. The dataset does not include human variables like class schedules, holidays, and special events, which can have a significant impact on electricity consumption in educational buildings, even if these aspects offer useful information. The models may have been able to capture abrupt changes in consumption patterns if these characteristics had been included.

Lastly, only short-term forecasting was used for the evaluation. Even though this offers helpful insights for operational energy management, the models hadn't been tested for longer-term horizons, such weekly or monthly forecasting, which may have given additional significance for planning and policy purposes.

5.3 Future Work

Based on the conclusions of this thesis, numerous avenues for further research could be suggested.

Adding more operational and building-specific characteristics to the forecasting algorithms is one way to start. Variables that capture the human-driven features of power consumption in educational buildings, such as occupancy schedules, holiday times, and building usage periods, might remarkably increase forecast accuracy. Therefore, having access to more comprehensive datasets that incorporate these elements would improve the models' suitability.

Future research may potentially investigate higher-level designs. One of them is creating hybrid forecasting frameworks, that incorporate models to maximize their respective advantages. For instance, XGBoost or neural networks may be used to represent nonlinear residuals, whereas ARIMA could be employed to capture linear temporal trends.

Lastly, expanding the research to larger forecasting horizons, such weekly or monthly electricity use, might be beneficial. These predictions might offer useful information for strategic planning, budgeting, and sustainability projects at educational institutions in addition to daily energy management.

5.4 Final Remarks

This thesis used statistical, machine learning, and deep learning techniques to investigate the subject of short-term load forecasting in educational buildings. The study illustrated the relative efficacy of ARIMA, XGBoost, and LSTM models within a unified framework by applying them to the ASHRAE Great Energy Predictor III dataset.

According to the analysis, XGBoost performed better than ARIMA and LSTM, obtaining the lowest error levels for every evaluation metric. This result demonstrates how well tree-based ensemble approaches can incorporate temporal and weather-related data and capture intricate nonlinear interactions in energy forecasting tasks.

The results highlight the significance of model selection and feature engineering in load forecasting, despite constraints, such lacking building-specific data, limiting the work's total scope. They also highlight the expanding contribution of machine learning towards optimizing energy efficiency and promoting environmentally friendly building management in educational institutions.

In conclusion, the study lays the groundwork for further research into more sophisticated and integrated forecasting technologies while also supporting the continuous attempts to improve building energy consumption. Such methods can significantly save operating costs, improve sustainability, and support the shift to smarter energy systems by facilitating more precise forecasts of energy usage.

Bibliography

[1] Naqash Ahmad, Yazeed Ghadi, Muhammad Adnan, and Mansoor Ali. Load Forecasting Techniques for Power System: Research Challenges and Survey. *IEEE Access*, 10:71054–71090, May 2022.

[2] Joaquim Massana, Carles Pous, Llorenç Burgas, Joaquim Melendez, and Joan Colomer. Short-term load forecasting for non-residential buildings contrasting artificial occupancy attributes. *Energy and Buildings*, 130:519–531, Oct. 2016.

[3] ASHRAE Great Energy Predictor III. https://www.kaggle.com/competitions/ashrae-energy-prediction.

[4] S. Hadri, Y. Naitmalek, M. Najib, M. Bakhouya, Y. Fakhri, and M. Elaroussi. A Comparative Study of Predictive Approaches for Load Forecasting in Smart Buildings. *Procedia Computer Science*, 160:173–180, Nov. 2019.

[5] Ewa Chodakowska, Joanicjusz Nazarko, and Łukasz Nazarko. ARIMA Models in Electrical Load Forecasting and Their Robustness to Noise. *Energies*, 14(23):173–180, Nov. 2021.

[6] Darpan Pandey, Kamal Niwaria, and Bharti Chourasia. Machine learning algorithms: a review. *International Research Journal of Engineering and Technology*, 06(02), Feb. 2019.

[7] Athanasios Ioannis Arvanitidis, Dimitrios Bargiotas, Dimitrios Kontogiannis, Athanasios Fevgas, and Miltiadis Alamaniotis. Optimized Data-Driven Models for Short-Term Electricity Price Forecasting Based on Signal Decomposition and Clustering Techniques. *Energies*, 15(21), Oct. 2022.

[8] Yuanyuan Wang, Shanfeng Sun, Xiaoqiao Chen, Xiangjun Zeng, Yang Kong, Jun Chen, Yongsheng Guo, and Tingyuan Wang. Short-term load forecasting of industrial customers based on SVMD and XGBoost. *International Journal of Electrical Power & Energy Systems*, 129, Jul. 2021.

[9] Dimitrios Kontogiannis, Dimitrios Bargiotas, , and Aspassia Daskalopulu. Minutely Active Power Forecasting Models Using Neural Networks. *Sustainability*, 12(8), Apr. 2020.

[10] Dimitrios Kontogiannis, Dimitrios Bargiotas, , Aspassia Daskalopulu, Athanasios Ioannis Arvanitidis, and Lefteri H. Tsoukalas. Error Compensation Enhanced Day-Ahead Electricity Price Forecasting. *Energies*, 15(4), Feb. 2022.