KASDI MERBAH OUARGLA UNIVERSITY
Faculty of Applied Sciences
Department of Electrical Engineering

Thesis
ACADEMIC MASTER
Domain : Science and technology
Field : Electrical engineering
Speciality : Electrical networks
Presented By :

Bougueffa Eutamene Abbes **Chouiha Boudjemaa Abdelkrim**

Accurate Electricity Consumption Prediction Using Deep Learning

Publicly defended on: 14/06/2023
The jury :

Mr Mahdi BENAOUADJ	MCB	President	UKM Ouargla
Mr Houari BOUDJELLA	MCB	Supervisor	UKM Ouargla
Mrs Nour El Houda BENSAADI	MCB	Co-Supervisor	Eötvös Loránd University
Mrs Naima BENBOUZA	MCB	Examiner	UKM Ouargla

Academic Year : 2022/2023

Dedication Chouiha Boudjemaa Abdelkrim

I dedicated this modest work to those who are dearest to me, I remember:
First all: my dears mother, No words can express their true value of gratitude and love who are the two dearest people in the world, may God protect and take care of them for me.
My fathers: my dears fathers, **Chouiha Abdelkader**. Who encouraged me in my academic path "May Allah have mercy on him and make heaven his abode"
To my brothers: **Abdelmoumine**
To my sisters: **Milad**, **Anissa**
To my Colleague: Thank you for your patience and your love, I wish you a life full of success, happiness.
To my dear friends: **Chiba** , **Maamri Oussama**, **Aoudjit Mustapha**, **Necib Islem**, **Degha Oussama** I can't find the sincere words to express my feelings and thoughts, you are my brothers.
My supervisor: **Dr Boudjella Houari**. May Allah grant your health and happiness and long life.

Dedication Bougueffa Eutamen Abbes

Praise to Allah who has enabled us for this which we would not have reached it if were not for the grace of God to us.

I dedicated this modest work to those who are dearest to me, I remember:

First of all: To my mother, no words can express their true value of gratitude and love for the two dearest people in the world. May Allah protect and take care of her for me.

To my Father: **Kamel**. Who encouraged me in my academic path

To my brothers: **Houssin, Hessen, Mouhammed**,

To my sisters: **Bensaadi Nourelhouda** Nothing wouldn't have been done without you

My colleague: thank you for your patience and your love, I wish you a life full of success, happiness.

To my dear friends: **Maamri Oussama**, **Aoudjit Mustapha**, **Necib Islem** ,**Bouaroua Abd Elkrim** , **Anes Medjouel** I can't find the sincere words to express my feelings and thoughts, you are my brothers.

My supervisor: **Dr Boudjella Houari**. May Allah grant your health and happiness and long life and make sure that I never fade away.

Acknowledgments

*W*e owe a deep debt of gratitude to our Faculty of **Applied Sciences** and Department of **Electotechnique** for allowing us to complete this work. We are grateful to some people who worked hard with us from the beginning until the completion of the present research. Our supervisor Dr Boujella Houari who has been generous during all phases of the research. We would like to take this opportunity to say warm thanks to all our beloved friends and family, who have been so supportive along the way of doing our work. Finally, To all those who have contributed from near or far to the completion of this graduation project. First and foremost, we must acknowledge our limitless thanks to Allah, their-Magnificent; the Ever-Thankful, for His help and bless. We are totally sure that this work would have never become truth, without His guidance.

Abstract

Modern society is heavily dependent on energy as a critical resource. However, with the increasing demand for energy, the world is facing several challenges such as global carbon emission, climate change, and environmental degradation. To address these issues, it is crucial to predict future electricity consumption accurately. This prediction is necessary for efficient energy management, demand-response, and grid planning.

In this thesis, we propose four deep learning models, namely Multilayer Perceptron(MLP), Convolutional Neural Networks(CNN),Long-Short Terme Memory (LSTM), and (CNN-LSTM), for energy forecasting. Our aim is to demonstrate that these models outperform other existing approaches. The models were trained and evaluated using a dataset collected from Two cities in Algeria: Sidi bel Abbes, Mascara. The dataset consists of approximately 2700 HVA clients total of 66000 measurements.

To assess the performance of our models, we utilize popular evaluation metrics such as Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). These metrics provide a comprehensive evaluation of the accuracy and reliability of the models in predicting energy consumption.

Remarkably, our results indicate that even with a relatively small number of data measurements, our proposed models deliver excellent forecasting outcomes. This suggests the robustness and effectiveness of the deep learning architectures in capturing and predicting energy consumption patterns in the studied cities.

Keywords

Energy forecasting, Deep Learning models, Multilayer Perceptron, Convolutional Neural Networks, Long-Short Terme Memory

Résumé

La société moderne dépend fortement de l'énergie en tant que ressource critique. Cependant, avec la demande croissante en énergie, le monde est confronté à plusieurs défis tels que les émissions mondiales de carbone, le changement climatique et la dégradation de l'environnement. Pour faire face à ces problèmes, il est crucial de prédire avec précision la consommation future d'électricité. Cette prédiction est nécessaire pour une gestion efficace de l'énergie, la réponse à la demande et la planification du réseau.

Dans cette mémoire, nous proposons quatre modèles d'apprentissage en profondeur , à savoir le Perceptron Multi-Couches (MLP), les Réseaux de Neurones Convolutifs (CNN) , la Mémoire à Court Terme Longue (LSTM) et (CNN-LSTM) , pour la prévision énergétique. Notre objectif est de démontrer que ces modèles surpassent les autres approches existantes. Les modèles ont été entraînés et évalués à l'aide d'un ensemble de données collectées dans deux villes en Algérie : Sidi bel Abbes , Mascara . L'ensemble de données comprend environ 66000 mesures.

Pour évaluer les performances de nos modèles, nous utilisons des métriques d'évaluation courantes telles que l'erreur absolue moyenne en pourcentage (MAPE), l'erreur quadratique moyenne (RMSE) et l'erreur absolue moyenne (MAE). Ces métriques permettent une évaluation complète de l'exactitude et de la fiabilité des modèles dans la prédiction de la consommation d'énergie.

Remarquablement, nos résultats indiquent que même avec un nombre relativement restreint de mesures de données, nos modèles proposés offrent d'excellents résultats de prévision. Cela suggère la robustesse et l'efficacité des architectures d'apprentissage en profondeur dans la capture et la prédiction des schémas de consommation d'énergie dans les villes étudiées.

Mote Clé

Prévision énergétique , modèles d'apprentissage en profondeur , Perceptron Multi-Couches, Réseaux de Neurones Convolutifs , Mémoire à Court Terme Longue

الملخص:

يعتمد المجتمع الحديث بشكل كبير على الطاقة كمورد حيوي. لكن ومع ذلك، مع تزايد الطلب على الطاقة، يواجه العالم العديد من التحديات مثل انبعاثات الكربون العالمية وتغير المناخ والتدهور البيئي.

للتعامل مع هذه المشاكل، من الضروري التنبؤ بدقة بالاستهلاك المستقبلي.

هذا التنبؤ ضروري لإدارة الطاقة بكفاءة، والإجابة على ذلك الطلب وتخطيط الشبكات.

في هذه المذكرة، نقترح أربعة نماذج للتعلم العميق، وهي الإدراك الحسي متعدد الطبقات ، والشبكات العصبية التلافيفية ، الذاكرة طويلة المدى و للتنبؤ بالطاقة. هدفنا هو إثبات أن هذه النماذج تتفوق على الأساليب الأخرى الموجودة.

تم تدريب النماذج وتقييمها باستخدام مجموعة بيانات تم جمعها في قسمين من مدن الجزائر: سيدي بلعباس، معسكر. تتضمن مجموعة البيانات ما يقرب من ٦٦٠٠٠ مقاسات.

لتقييم أداء نماذجنا، نستخدم مقاييس التقييم مثل متوسط النسبة المئوية للخطأ المطلق ، والخطأ التربيعي يعني ويعني الخطأ المطلق . تسمح هذه المقاييس بالتقييم تحليل كامل لدقة وموثوقية النماذج في التنبؤ بالاستهلاك من الطاقة.

ومن اللافت للنظر أن نتائجنا تشير إلى ذلك حتى مع وجود عدد صغير نسبيًا قياسات البيانات، تقدم نماذجنا المقترحة نتائج تنبؤية ممتازة. الذي يقترح قوة وكفاءة بنيات التعلم العميق في التقاط والتنبؤ بأنماط استهلاك الطاقة في المدن المدروسة

الكلمات الدلالية:

توقع الطاقة، نماذج التعلم العميق، متعدد الطبقات، شبكات الأعصاب التكاثفية، ذاكرة طويلة وقصيرة المدى

Contents

List of Figures

List of Tables

Acronyms

MLP	Multilayer Perceptrone
LSTM	long short-term memory
CNN	Convolutional Neural Networks
ANN	Artificial Neural Networks
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
TCN	Temporal Convolutional Networks
GRU	Gated Recurrent Unit
IHEPC	Individual Household Power Consumption
EECP-CBL	Electric Energy Consumption Prediction By CNN BI-LSTM
FC	Fully Connected
RNN	Recurrent Neural Network
DNN	Deep Neural Network
ARIMA	Auto-Regressive Iterative Moving Avarage
LSSVM	Least Square Support Vector Machine
MODWT	Maximal Overlap Discrete Wavelet Transformation
ACCE	Adaptive Circular Conditional Expectation
LM	Linear Model
NMAE	Normalized Mean Absolute Erro
HW	Holt-Winters
ELM	Extreme Learning Machine

MA	Moving Average
AI	artificial intelligence
MAPE	Mean Absolute Percentage Error
MSE	Mean Squared Error
ANN	Artificial neural networks
MAD	Mean Absolute Deviation

General Introduction

This is an important point to consider when it comes to electricity production and consumption. When the amount of electricity generated exceeds the required amount, there is a risk of wasted resources and unnecessary costs. To prevent this, it is important to ensure that the amount of electricity generated matches up with the demand in real-time, in order to avoid overproduction or underproduction The distribution subsidiary orders electricity from the generation subsidiary, and then the energy produced is distributed to customers through their electrical grids. This ensures that customers receive reliable and efficient energy while providing a cost-effective solution for both the generation and distribution subsidiaries, Overproduction or production losses can be defined as the difference between the electricity produced and the actual electricity distributed. A better prediction of customers' consumption can help reduce these errors and minimize losses due to overproduction. This is beneficial for companies, as the overproduction of energy that is not distributed constitutes a dead loss for them At the level of distribution subsidiaries, electricity losses are a major concern for the company.

To control their impact, they have implemented a strategy to ensure that the amount of electricity purchased is equal to the amount billed. This is an ideal goal for them as it will help to minimize losses and increase efficiency Predictive models for accurately forecasting electricity consumption can be an important tool in helping to plan and monitor electricity usage in the economic sector. These models can help to identify potential issues before they become problems, as well as provide insight into how different factors may affect future consumption. By understanding past patterns and trends, these models can also provide a more accurate assessment of future needs, allowing for better decision-making when it comes to resource allocation..

This thesis is organized into four chapters.

Chapter 1 serves as an introduction, providing the background, objectives, and structure of the thesis.

Chapter 2 presents a comprehensive literature review, analyzing existing research to establish the research gap.

Chapter 3 details the methodology and experimental setup, including research design, data collection methods, and ethical considerations.

Finally, Chapter 4 presents the results and engages in a critical discussion, interpreting the findings in relation to the research questions and existing literature. Together, these chapters provide a coherent framework for the study, guiding the reader through the research process and contributing to the understanding of the research problem.

1

Overview of Electricity Consumption Prediction

1.1 Background and Motivation.

Electricity is an essential component of the economic infrastructure worldwide, permeating every aspect of daily life. Over the past three decades, electricity consumption has more than doubled,In algeria Electricity consumption increased by 5% in 2021 to 72 TWh [1]. It had been increasing very rapidly over 2009-2019 (8%/year) driven by industrial development in emerging economies and advancements in various sectors such as the widespread use of smart devices and the adoption of electric vehicles. Electricity is generated from both renewable (e.g., wind, solar, hydropower) and non-renewable (e.g., coal, natural gas, oil, nuclear) sources [2].

Given the critical role electricity plays in our daily lives, understanding and predicting its consumption patterns have become crucial. Predicting electricity energy consumption can aid in various ways, including diversifying the generation process towards renewable sources and taking appropriate actions to reduce emissions. Accurate predictions are especially vital as they inform decision-making processes and facilitate effective management of energy resources [2].

Moreover, electricity consumption prediction plays a significant role in load balancing within power grids. Power grids maintain a frequency of 50Hz, and any discrepancies between electricity consumption and production can disrupt this frequency, leading to potential damages to the power grid infrastructure.

In summary, accurate electricity consumption prediction is essential for managing energy resources, reducing emissions, facilitating load balancing within power grids, and ensuring the stability and sustainability of electricity supply.

1.1.1 Overview of the Importance of Electric Energy Consumption Prediction:

Electric energy consumption prediction plays a crucial role in various domains, including utility companies, grid operators, energy suppliers, businesses, and households. Accurate prediction of electricity demand enables effective energy management, cost reduction, grid stability, renewable energy integration, demand response programs, and infrastructure planning.

For utility companies and grid operators, accurate prediction of energy consumption allows for efficient planning and optimization of energy generation, distribution, and allocation strategies. By forecasting future energy demands, they can ensure a reliable and stable supply of electricity, avoid power shortages or excess capacity, and optimize resource allocation [3].

Businesses and households also benefit from electric energy consumption prediction. Accurate forecasts help them optimize their energy usage, identify peak demand periods, and implement energy-saving measures during high-demand hours. This proactive approach leads to reduced energy costs, improved energy efficiency, and better financial planning.

1.1.2 The Current Challenges and Limitations in Traditional Prediction Methods:

Traditional prediction methods for electric energy consumption often rely on statistical models or simplistic machine learning algorithms. While these approaches have been widely used, they have certain limitations that hinder their accuracy and effectiveness.

One challenge is the complexity and non-linearity of energy consumption patterns. Traditional methods may struggle to capture the intricate relationships and dynamic behavior exhibited by energy consumption data, leading to sub-optimal predictions. Moreover, traditional methods often require manual feature engineering, where domain knowledge is used to extract relevant features from the data. This process can be time-consuming and may not capture all the relevant information for accurate prediction [4].

Additionally, traditional methods may face difficulties in handling the increasing volume, velocity, and variety of data in the energy sector. The emergence of smart meters, IoT devices, and other data sources have resulted in large-scale and high-dimensional energy consumption datasets. Traditional methods may struggle to effectively process and analyze such data, limiting their prediction accuracy.

1.1.3 The Potential of Deep Learning Techniques in Improving Prediction Accuracy:

Deep learning techniques, a subset of machine learning, have shown great promise in overcoming the challenges of traditional prediction methods for electric energy consumption. Deep learning models, such as Multilayer Perceptron (Multilayer Perceptrone (MLP)), long short-term memory (long short-term memory (LSTM)) networks, and convolutional neural networks (Convolutional Neural Networks (CNN)), can capture complex temporal dependencies, non-linear relationships, and high-dimensional patterns present in energy consumption data.

one advantage of deep learning models is their ability to automatically learn hierarchical representations of the data, eliminating the need for manual feature engineering. Deep learning models can directly process raw energy consumption data, extracting relevant features at different levels of abstraction, leading to more accurate predictions.

Furthermore, deep learning models are highly scalable and can handle large-scale and high-dimensional datasets efficiently. With their ability to process vast amounts of data and capture intricate patterns, deep learning techniques have the potential to significantly improve the accuracy of electric energy consumption prediction [5].

1.2 Research Objectives

The main objectives of this thesis are as follows:

Develop an accurate deep-learning model for electric energy consumption prediction: The primary goal of this research is to design and construct a deep-learning model specifically tailored for predicting electric energy consumption. The model will leverage advanced deep-learning techniques, such as long short-term memory (LSTM) networks, or convolutional neural networks (CNNs),Multilayer Perceptron (MLP) and combination between long short-term memory with convolutional neural networks that are called (CNN-LSTM)to capture the complex temporal dependencies and non-linear relationships present in energy consumption data.

Improve prediction accuracy compared to traditional methods: The thesis aims to demonstrate that the proposed deep-learning model outperforms traditional prediction methods commonly used in the field. By leveraging the power of deep learning and its ability to automatically extract relevant features from raw data, the objective is to achieve higher prediction accuracy and reduce forecasting errors.

Explore the impact of incorporating additional data sources: In addition to historical energy consumption data, the research aims to investigate the effects of integrating other relevant data sources into the prediction model. This may include weather data, socioeconomic factors, time-of-day information, or any other variables that may influence electricity consumption patterns. The objective is to assess the extent to which incorporating these additional data sources can enhance prediction accuracy.

Evaluate the model's performance under different scenarios and time periods: The thesis seeks to evaluate the robustness and generalizability of the proposed deep-learning model. By testing the model's performance across various scenarios, such as different geographic regions, seasons, or economic conditions, the objective is to assess its effectiveness and reliability in predicting electric energy consumption in diverse contexts.

Provide insights and recommendations for practical implementation: Alongside developing the deep-learning model, the research aims to provide practical insights and recommendations for implementing the model in real-world scenarios. This includes discussing potential challenges, limitations, and considerations for deployment, as well as proposing strategies for integrating the model into existing energy management systems.

1.3 Scope and Limitations

1.3.1 Scope

Dataset: The scope of this thesis will focus on utilizing deep learning techniques to predict electricity consumption based on a specific dataset. The analysis and modeling will be conducted using a comprehensive and relevant dataset that encompasses historical electricity consumption data.

Deep learning algorithms: The study will explore various deep learning algorithms such as long short-term memory (LSTM) [], or convolutional neural networks (CNNs) and combination between convolutional neural networks with long short-term memory(CNN-LSTM) and multi layer perceptrone(MLP) to develop accurate models for electricity consumption prediction. The scope will include comparing and evaluating the performance of different deep learning architectures.

Feature selection and engineering: The scope will involve identifying and selecting relevant features or variables that have a significant impact on electricity consumption prediction. Feature engineering techniques may also be employed to enhance the performance of the deep learning models.

Accuracy assessment: The scope of the thesis will involve evaluating the accuracy of the developed models for electricity consumption prediction. This assessment will include quantitative measures such as mean absolute error (Mean Absolute Error (MAE)), Root mean square error (Root Mean Square Error (RMSE)), and coefficient of determination (R-squared) to validate the accuracy of the predictions.

1.3.2 Limitations

Data availability: The accuracy and performance of the developed models heavily depend on the availability and quality of the dataset. Limitations may arise if the dataset is limited in terms of size, temporal coverage, or missing values, which could impact the accuracy of the predictions.

Generalizability: The scope of this thesis will focus on accurate electricity consumption prediction within a specific dataset or geographical area. The models developed may have limitations in generalizing the predictions to different regions or datasets due to variations in consumption patterns, infrastructure, or socio-economic factors.

External factors: Electricity consumption can be influenced by external factors such as weather conditions, holidays, or special events. While efforts will be made to incorporate relevant external factors, the scope of this thesis may not fully capture the complexity and variability introduced by these external factors, which could affect the accuracy of the predictions.

Computational resources: Deep learning models can be computationally intensive and require substantial computational resources. The limitations of this thesis may arise in terms of the computational power or time constraints that may restrict the exploration of complex deep learning architectures or larger datasets.

Interpretability: Deep learning models are often considered as black-box models, making it challenging to interpret the underlying factors or features contributing to the predictions. The limitations of interpretability may arise in understanding the specific factors or variables influencing electricity consumption in the developed models.

It is important to consider these scope and limitations while conducting the research and interpreting the results to ensure a comprehensive understanding of the accuracy and applicability of the developed deep learning models for electricity consumption prediction.

1.4 Methodology overview

Our research thesis aims to develop an approach model that consists of four algorithms, namely long short-term memory (LSTM), convolutional neural networks (CNNs), a combination of convolutional neural networks with long short-term memory (CNN-LSTM), and multi-layer perceptron (MLP). The model will be used to analyze data from two cities in Algeria, namely Sidi Bel-Abbes, Mascara. The data will comprise almost 8 features that will be used to train and test the algorithms. The LSTM algorithm will be used for sequence prediction tasks, while the CNN algorithm will be used for image classification tasks. The CNN-LSTM algorithm is a hybrid approach that combines the strengths of both CNN and LSTM algorithms. The MLP algorithm is a feedforward neural network that is widely used for classification and regression tasks. The data analysis will provide insights into the patterns and trends of the selected features in the three cities, which will be useful for decision-making processes in various fields, such as urban planning, transportation, and environmental management.

1.5 Conclusion

the first chapter has provided a solid foundation for the research, addressing the background, motivation, objectives, scope, and limitations. The subsequent chapters will build upon this groundwork, delving deeper into each aspect and employing the proposed methodology to investigate the selected cities in Algeria. By the end of this study, it is anticipated that new knowledge will be generated and valuable contributions will be made to the field, ultimately benefiting the targeted cities and beyond.

2

Literature Review

2.1 Introduction

According to Statista Research Department The world's electricity consumption has continuously grown over the past half a century, reaching approximately 25,300 terawatt-hours in 2021. Between 1980 and 2021, electricity consumption more than tripled, while the global population increased by roughly 75 %. Growth in industrialization and electricity access across the globe have further boosted electricity demand. There are no prerequisites for reducing electricity consumption in the future, since at the present stage of human development, electricity is a key resource—professional and household human activity are impossible without the use of electricity [2] For an extensive period of time, researchers have dedicated their efforts towards the exploration and development of methodologies aimed at enhancing the accuracy and efficiency of electricity consumption forecasting. This particular area of study has garnered significant attention due to its vital importance in various sectors, including energy planning, resource allocation, and infrastructure development. The reliable prediction of electricity consumption plays a crucial role in ensuring the stability and sustainability of power systems, as well as facilitating optimal decision-making processes for energy providers, policymakers, and consumers. this chapter will present a detailed review of recent related works in the field of electricity consumption forecasting. It will examine existing methodologies, models, and techniques employed in the literature while critically assessing their strengths and limitations. By identifying the gaps and areas that require further investigation, we aim to establish the foundation for our research objectives and demonstrate the relevance and significance of our proposed approach.

2.2 State of the art - Related work

The significance of electricity in our daily lives has led to an increased emphasis on predicting its consumption. As a result, there is a wealth of research papers, articles, blogs, and videos dedicated to this topic due to its widespread popularity.

About prediction methods from D. Hadjout [6] they mentions about different methods to solve the Electrical energy prediction problem ,to predict monthly electricity consumption for the economic sector.

A novel ensemble learning approach was devised by the researchers, incorporating three highly effective models: Long Short-Term Memory (LSTM) and Gated Recurrent Unit (Gated Recurrent Unit (GRU)) neural networks, as well as Temporal Convolutional Networks (Temporal Convolutional Networks (TCN)). The researchers conducted experiments utilizing extensive data encompassing almost 2000 clients and 14 years of monthly electricity consumption records from Bejaia, Algeria.

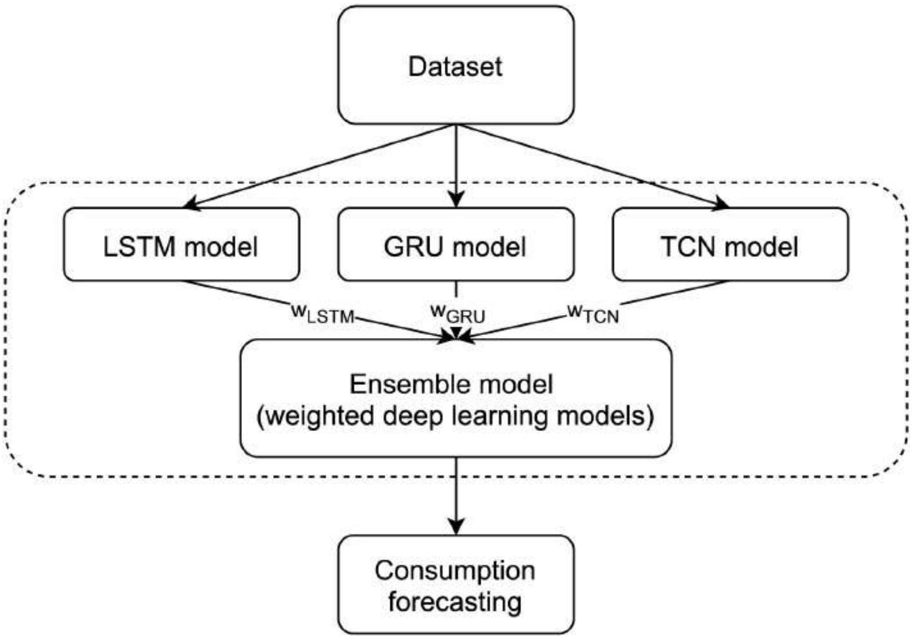

Figure 2.1: Ensemble learning method.

The results demonstrate that the ensemble models proposed in the study outperform both the requirements set by the company and the predictions made by traditional individual models.

There is a good approach written by Jaewon Chung [7] Accurate estimation of energy demand and consumption necessitates the effective analysis of multivariable data, including factors such as gross domestic product, imports, exports, temperature, precipitation, electricity consumption, and trade balance. To address this, the researchers proposed a CNN-LSTM model that incorporates a multivariable augmentation approach.

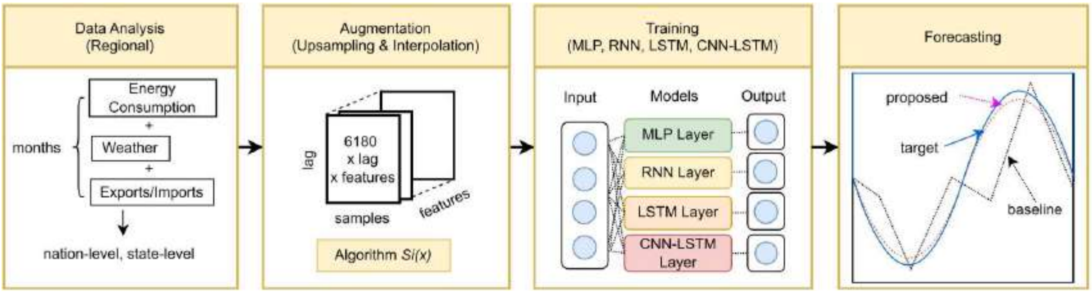

Figure 2.2: Visualization of the proposed methodology.

The proposed model demonstrates exceptional forecasting accuracy for electricity consumption, surpassing the performance of existing models.

The paper of Le, Tuong and Vo, Minh Thanh [8] this study proposed a novel Electric Energy Consumption Prediction By CNN BI-LSTM (EECP-CBL) model was introduced for electric energy consumption prediction. The model combines the strengths of CNN and Bi-LSTM on the Individual Household Power Consumption (IHEPC) dataset. Comprising three modules, namely CNN Bi-LSTM, and Fully Connected (FC), the proposed model demonstrates effective prediction of energy consumption.

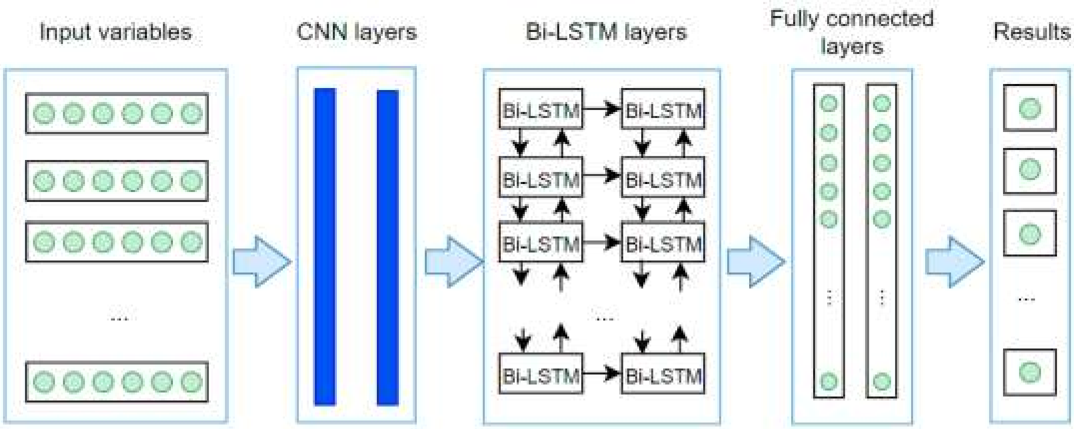

Figure 2.3: Ensemble learning method.

The resarch done in 2017 by Wan He [9]The researchers introduced a Deep Neural Network (Deep Neural Network (DNN)) that incorporates both Convolutional Neural Network (CNN) and Recurrent Neural Network (Recurrent Neural Network (RNN)) components to accurately forecast electric energy consumption.

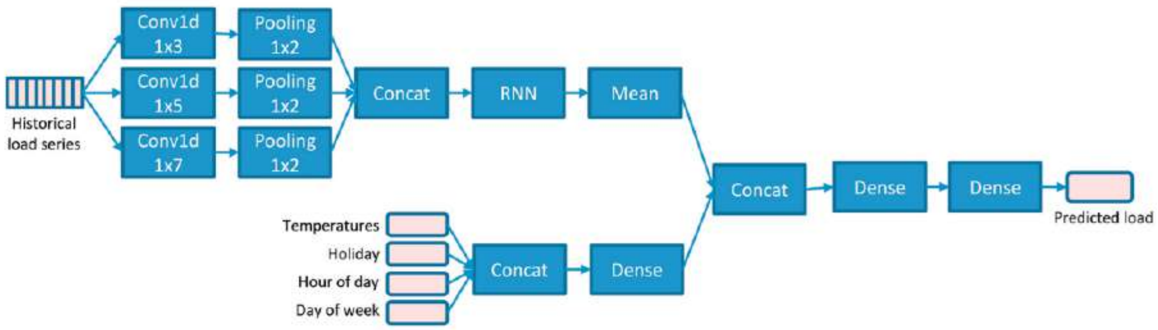

Figure 2.4: Model Architechture

In Turkey F. Kaytez [10]In this research, historical electricity load and real-time climate data were utilized to conduct energy consumption forecasting. Unlike most studies that consider factors such as population, import/export values, and gross domestic product (GDP) on an annual basis, this research specifically focuses on long-term electricity consumption forecasting in Turkey. To achieve this, a hybrid model combining Auto-Regressive Iterative Moving Avarage (Auto-Regressive Iterative Moving Avarage (ARIMA)) and Least Square Support Vector Machine (Least Square Support Vector Machine (LSSVM)) was proposed.

Also Researchers form U.S.A (Rabin K. Jana) [11] ,The research presents a novel approach for predicting energy consumption in various sectors at a macro level. The proposed method combines maximal overlap discrete wavelet transformation (Maximal Overlap Discrete Wavelet Transformation (MODWT)) and long short-term memory (LSTM) network, enabling a granular deep learning approach. The input features undergo evaluation using the Boruta algorithm-based feature selection model. To separate linear and nonlinear components, MODWT is employed to decompose the energy consumption time series. At a granular level, the LSTM network, a powerful deep learning tool, is then utilized to make predictions on individual sub-series.

In 2019 [12], a research study conducted in Montreal, Canada proposed the Adaptive Circular Conditional Expectation (Adaptive Circular Conditional Expectation (ACCE)) method for defining sub-residuals operation schedules. This method utilizes circular analysis to handle periodic patterns in data and predict the residual component demand within each time window. To enhance total electricity demand forecasting, an adaptive Linear Model (Linear Model (LM)) procedure is employed, incorporating the ACCE results to predict the residual component demand. The proposed approach, focused on modeling periodic residual demand in a daily horizon, demonstrates a promising accuracy increase of 23 % compared to existing methods, evaluated using the Normalized Mean Absolute Error (Normalized Mean Absolute Erro (NMAE)) metric. Combining the residual modeling method with temperature-related component forecasting further improves total power consumption prediction performance by 7%. Numerical analysis using real data validates the effectiveness and practical benefits of the ACCE method and its integration with other forecasting techniques for electricity demand prediction.

Another research have done in france by chinese researchers in 2020, [13] for A hybrid prediction model for residential electricity consumption using holt-winters and extreme learning machine, A hybrid model, combining the Holt-Winters (Holt-Winters (HW)) method and Extreme Learning Machine (Extreme Learning Machine (ELM)) network, is developed for ultra-short-term predictions of residential electricity consumption. The original data undergoes decomposition using the Moving Average (Moving Average (MA)) filter, separating it into a stationary linear component and a fluctuant nonlinear residual. The HW method is used to establish a linear prediction model for forecasting the linear component. In conjunction with the linear prediction results, nonlinear residual, and original data, the ELM network builds a nonlinear prediction model for residential electricity consumption. The proposed HW-ELM model is evaluated for predicting 15-minute electricity consumption values with varying training set sizes and seasons. Compared to other models such as HW, ELM, and long short-term memory network, the proposed model consistently demonstrates lower prediction error when forecasting residential electricity consumption.

In [14] The research proposes a neural network architecture called CNN-LSTM for accurate prediction of housing energy consumption. The CNN-LSTM network combines the strengths of both convolutional neural network (CNN) and long short-term memory (LSTM) models. The CNN layer is capable of extracting relevant features from multiple variables that impact energy consumption, while the LSTM layer is well-suited for capturing temporal patterns and irregular trends in time series data. Experimental results demonstrate that the CNN-LSTM model achieves nearly flawless prediction performance for electric energy consumption, which was previously challenging to accurately forecast.Also, it records the smallest value of root mean square error compared to the conventional forecasting methods for the dataset on individual household power consumption. The empirical analysis of the variables confirms what affects to forecast the power consumption most

The last document in our research was done in 2019 by Junhong Kim in Three distribution complexes of South Korea [15] The aim of the research is to develop an accurate short-term load forecasting model using recurrent neural network (RNN)-based models,When forecasting electric load at a specific time, existing RNN-based models have limitations in utilizing predicted future hidden state vectors and fully available past information. This leads to inaccuracies that cannot be corrected for subsequent predictions. To overcome these issues, a research proposes a recurrent inception convolution neural network (RICNN) that combines RNN and 1-dimensional CNN (1-D CNN). The model employs a 1-D convolution inception module to adjust the prediction time and the hidden state vector values from nearby time steps, creating an optimized network. The RICNN model is evaluated using power usage data from three large distribution complexes in South Korea. Experimental results demonstrate that the RICNN model outperforms benchmarked models such as multi-layer perception, RNN, and 1-D CNN in daily electric load forecasting, specifically for 48-time steps with a 30-minute interval

All the previous documentation and accomplishments have been summarized in the table 2.1 .

Ref	Input Variables	Output variables	Prediction Interval	Prediction Area
[7]	Electricity consumption , trade balance Temperature , precipitation GDP export value, import	Electricity usage	Monthly	Korea
[8]	Electrical Energy Consumption	electrical energy consumption	Monthly	Seoul, korea
[9]	Electrical Energy Consumption ,Temperature , holiday	EEC	Daily	China
[6]	Electricity consumption history	EEC	Monthly	Bejaia, Algeria
[10]	Gross electricity generation, GDP, population, installed capacity, import value, export value, total subscribership	EEC	Annual	Turkey
[11]	Energy consumption history, moving average , bias, momentum, rate of change,	TEC	Monthly	USA
[12]	Electricity load history, climatic data	Residential electricity load	Hourly	An individual household in Montreal, Canada
[13]	Active power, reactive power, voltage, global intensity	Residential electricity consumption	15 minutes	An individual residential building in Paris, France
[14]	Global active power, global reactive power, voltage, global intensity, sub metering	Residential power consumption	Hourly	An individual household in France
[16]	Electricity load history (ISO-NE, NYISO 13 years of electricity load of New England, New York)	Electricity load	Hourly	New England and New York, USA
[15]	Electricity load history, climatic data, time information, electricity rates , number of sensors	Electricity load	Daily	Three distribution complexes of South Korea

Table 2.1: Some typical types of energy forecasting techniques

2.3 Reveal The Gaps

The accumulation of various documents has significantly contributed to the enrichment of our knowledge regarding the subject of our research - improving electricity consumption prediction using deep learning. These documents have served as valuable sources of information, enabling us to make informed decisions in selecting the most appropriate model for our purposes. Consequently, our proposed approach will entail the integration of Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Multilayer Perceptron (MLP) within a single codebase.

The choice to incorporate CNN-LSTM and MLP stems from their effectiveness as evidenced by the results obtained in related works. These models have demonstrated promising outcomes in similar research endeavors, making them suitable candidates for our electricity consumption prediction framework.

One particularly influential document that significantly impacted our research trajectory was the most recent publication we encountered. In this document, the author focused on three complexes located in South Korea. The findings and insights gained from this study served as a source of inspiration for us to extend our prediction efforts to two specific cities: Mascara, and Sidi Bel Abbes. By examining electricity consumption patterns in these urban areas, we hope to further enrich our understanding and generate applicable predictions.

Furthermore, we have decided to direct our thesis towards High Voltage Type (HVA) due to various reasons. HVAs are invoiced on a monthly basis, which provides a consistent and structured data collection framework for our research. Moreover, HVAs represent a significant portion, approximately 45 percent, of the overall electricity consumption in Algeria. By focusing on HVAs, we can gain valuable insights into consumption patterns and contribute to more accurate predictions.

To construct our dataset, we have carefully chosen multiple variables that are crucial in capturing the intricacies of electricity consumption. These variables include Active Power, Reactive Power, Voltage, PMA, Power Factor, and Temperature. By incorporating a multivariable approach, we aim to leverage the wealth of information contained within these factors. Recent research has shown that utilizing multivariable data offers improved performance compared to relying solely on univariable data. Therefore, by incorporating multiple variables, we anticipate achieving enhanced prediction accuracy and more comprehensive insights into electricity consumption patterns.

2.4 conclusion

In this chapter, we have dedicated extensive efforts to thoroughly explore the existing body of related works and literature relevant to our thesis topic. This rigorous examination of prior research has proven to be immensely valuable in broadening our knowledge and deepening our understanding of the subject matter. By delving into these works, we have gained invaluable insights, discovered novel perspectives, and identified potential gaps that our research can address.

Furthermore, our comprehensive review of related works has played a pivotal role in informing our decision-making process regarding the selection of the most appropriate and promising approached model. We have meticulously evaluated various models, techniques, and methodologies proposed in the literature, considering their merits, drawbacks, and empirical evidence of their effectiveness. Our ultimate aim is to identify an approached model that offers the highest level of accuracy, predictive power, and applicability to our specific research context.

Simultaneously, we have conscientiously considered the choice of dataset type for our research endeavor. Recognizing the significance of robust and well-curated data, we have taken great care in selecting an appropriate dataset that aligns with the objectives and requirements of our thesis. This selection process involves careful consideration of various factors, such as the availability, quality, relevance, and representativeness of the data. We aim to choose a dataset that encompasses a comprehensive range of variables and encompasses a significant portion of the electricity consumption landscape we intend to study.

By thoroughly exploring related works and meticulously choosing our approached model and dataset type, we endeavor to embark on our research journey armed with a strong foundation of knowledge, an optimized model architecture, and high-quality data. This strategic approach will enable us to maximize the accuracy and effectiveness of our research outcomes while ensuring their relevance and applicability to the real-world context of electricity consumption prediction.

3

Methodology And Experimental Setup

3.1 Introduction

In an era characterized by rapid advancements in technology, the field of artificial intelligence (artificial intelligence (AI)) has emerged as a prominent force driving innovation across various industries. AI encompasses a range of techniques and methodologies that empower machines to mimic human intelligence and perform complex tasks autonomously. Among the various branches of AI, one particular area that has gained significant attention and transformative power is deep learning.

In this chapter, we embark on a journey to unravel the intricate world of artificial intelligence and delve into the depths of deep learning. Our aim is to demystify these concepts, providing you with a comprehensive understanding of their significance and practical applications.

To begin, we shall explore the fundamental concept of artificial intelligence. AI is the science and engineering behind the creation of intelligent machines that can perceive, reason, learn, and make decisions like humans. It encompasses a wide range of techniques, including machine learning, natural language processing, computer vision, and expert systems, among others. Our focus in this chapter, however, will primarily be on deep learning—a subfield of machine learning that has revolutionized various domains by enabling computers to learn from vast amounts of data.

Understanding deep learning requires us to grasp the underlying principles of neural networks. Neural networks are computational models inspired by the structure and functioning of the human brain. They consist of interconnected nodes, known as neurons, organized in layers that process and transform data. Through a process called training, neural networks learn patterns and relationships in data, allowing them to make accurate predictions and classifications.

But before we dive deeper into the intricacies of deep learning, it is essential to address the crucial aspect of data collection and preprocessing. The quality and suitability of the data significantly impact the performance and reliability of any AI model. We will explore various techniques for data collection, discuss the challenges of dealing with diverse data sources, and delve into the crucial process of data preprocessing, which involves cleaning, transforming, and organizing the data to make it suitable for analysis.

Once the data is prepared, we move on to the modeling stage, where we construct our approach model for analysis. Deep learning models are renowned for their ability to extract complex features and patterns from raw data, enabling them to make accurate predictions and classifications. We will explore different architectures of deep learning models, such as convolutional neural networks (CNNs) for image analysis and recurrent neural networks (RNNs) for sequential data, among others. We will also discuss the training process, optimization techniques, and ways to evaluate model performance.

However, building a robust model is only part of the equation. Visualizing and interpreting the results are equally essential to gain insights and communicate findings effectively. In this chapter, we will delve into various evaluation metrics commonly used in deep learning, such as Mean Absolute Percentage Error (Mean Absolute Percentage Error (MAPE)), Root Mean Square Error (RMSE), and Mean Squared Error (Mean Squared Error (MSE)). We will explore how these metrics help us measure the performance of our model and guide decision-making in real-world scenarios.

Through this comprehensive exploration, we aim to equip you with the knowledge and tools necessary to comprehend and apply artificial intelligence and deep learning principles in your own data analysis endeavors. By the end of this chapter, you will have gained a profound understanding of AI, deep learning, data collection and preprocessing, model construction and evaluation, and the visualization of results. So let us embark on this exciting journey into the realm of artificial intelligence and deep learning, as we unravel the power of data-driven insights and pave the way for a future driven by intelligent machines.

3.2 Artificial Neural Networks

Artificial neural networks (Artificial neural networks (ANN)) are computational models inspired by the structure and functioning of biological neural networks, such as the human brain. ANNs are at the core of the field of machine learning, and they have proven to be remarkably effective in solving complex problems across various domains [17].

The fundamental building block of an artificial neural network is the neuron, also known as a node. Neurons are interconnected in layers, forming a network that processes and transforms data. Each neuron receives input signals, performs a mathematical computation on them, and produces an output signal. These output signals are then passed on to other neurons in subsequent layers, creating a chain of interconnected computations [18].

The layers in an artificial neural network are typically categorized into three types: the input layer, hidden layers, and the output layer. The input layer receives the initial data, while the output layer produces the final results or predictions. The hidden layers, as the name suggests, are intermediary layers that perform complex computations to extract meaningful features from the input data.

The strength of artificial neural networks lies in their ability to learn from data. During the training phase, the network is exposed to a large amount of labeled data, where the correct output is known for each input. The network adjusts its internal parameters, known as weights and biases, based on the discrepancy between the predicted output and the actual output. This adjustment process, known as backpropagation, allows the network to learn and improve its performance over time [19].

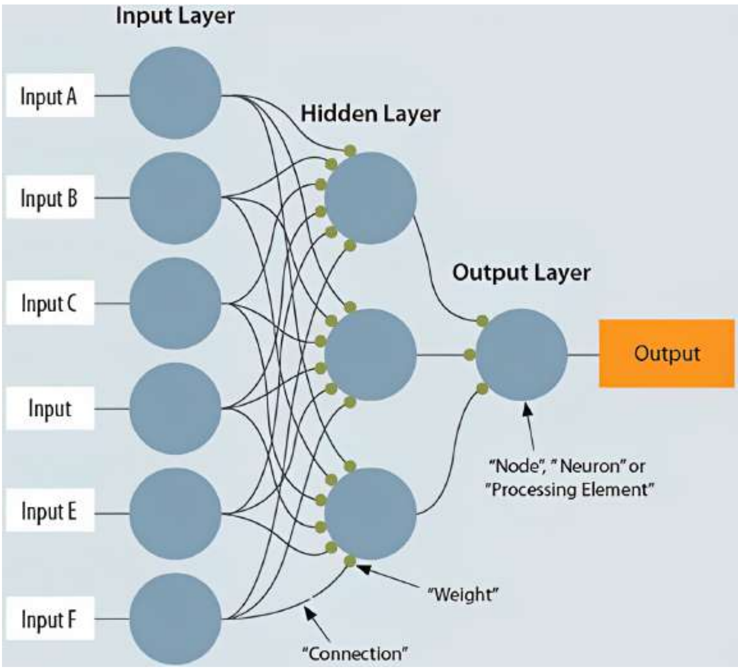

Figure 3.1: Artificial Neural Network Structure

There are several types of artificial neural networks, each designed to address specific types of problems. Here are a few common architectures:

MultiLayer Perceptrone (MLP): A Multilayer Perceptron (MLP) is a type of artificial neural network that consists of multiple layers of interconnected nodes, known as artificial neurons or perceptrons. It is a feedforward neural network, meaning that information flows through the network in one direction, from the input layer to the output layer.

Convolutional Neural Networks (CNNs): CNNs are particularly effective for analyzing visual data, such as images and videos. They consist of convolutional layers that apply filters to detect features at different spatial locations, followed by pooling layers that downsample the data. CNNs have achieved remarkable success in tasks such as image classification, object detection, and image generation.

Long Short Term Memory (LSTM): designed to overcome the limitations of traditional RNNs in capturing and retaining long-term dependencies in sequential data. LSTMs are specifically designed to address the vanishing gradient problem that occurs when training RNNs on long sequences.

LSTMs have an internal memory mechanism that allows them to selectively remember or forget information over extended time intervals

Convolutional Neural Network and Long Short Term Memory(CNN-LSTM): In the CNN-LSTM architecture, the output of the CNN layers is fed into the LSTM layers. The LSTM layers then process the sequential information and learn to model the temporal dependencies present in the data. This combined architecture enables the network to effectively capture both spatial and temporal features, making it suitable for tasks such as video classification, action recognition, and spatio-temporal forecasting.

By leveraging the strengths of both CNNs and LSTMs, CNN-LSTM provides a powerful framework for analyzing sequential data with spatial and temporal characteristics, enabling the network to learn rich representations and make accurate predictions or classifications..

Artificial neural networks have been successfully applied in various fields, including computer vision, natural language processing, robotics, healthcare, finance, and many more. Their ability to learn complex patterns and make accurate predictions from data has propelled advancements in AI and has the potential to drive further innovation in the future.

3.3 Deep Learning Networks

Deep learning networks, a subset of artificial neural networks (ANNs), have revolutionized the field of machine learning by enabling computers to learn and make sense of vast amounts of data. Deep learning networks excel at automatically learning and extracting intricate patterns and representations from raw data, without the need for explicit feature engineering. This has led to remarkable advancements in various domains, including computer vision, natural language processing, speech recognition, and many others [20].

The key characteristic of deep learning networks is their depth, which refers to the presence of multiple hidden layers between the input and output layers. These hidden layers allow for the extraction of increasingly complex and abstract features from the data. The depth of the network facilitates the learning of hierarchical representations, where lower layers capture simple features, and higher layers combine those features to form more sophisticated representations.

Deep learning networks are primarily built using architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), although there are numerous other types and variants as well. Here's an overview of some commonly used deep learning architectures:

3.3.1 Convolutional Neural Networks (CNNs)

Convolutional neural networks (CNNs) are a powerful class of deep learning algorithms specifically designed for image classification tasks. They employ a shared-weight architecture, where convolutional kernels or filters slide across input features, extracting precise representations known as feature maps. At the core of CNNs lies the convolutional operator, a central building block that enables the network to construct informative features by combining spatial and channel-wise information within local receptive fields at each layer. [21] One key advantage of CNNs is their ability to automatically learn and optimize the filters or kernels through automated learning, as opposed to traditional algorithms where these filters need to be manually tuned. The input to a CNN is a tensor with a shape of (number of inputs) x (input height) x (input width) x (input channels).

This input shape undergoes convolutional layers to extract feature maps, also known as activation maps, with a shape of (number of inputs) x (feature map height) x (feature map width) x (feature map channels) [22]. a typical CNN model, as shown in Figure 2, comprises several layers: convolutional layer, pooling layer, a flattening layer, and a fully connected layer. The convolutional layer is the main component of the CNN network, which operates on the principle of sliding windows and weight sharing to reduce computational complexity. In this layer, the kernel method is used to extract various features from the input data. The next layer is the pooling layer. This layer is designed to reduce the size of the feature map involved by reducing the connections between layers and running each feature map independently. The main goal of the pooling operation is to reduce the dimensionality and extract the dominant features for efficient training of the model [6]. There are several types of pooling operations: max pooling and average pooling. Before proceeding with the fully connected linked layer (FC), it is necessary to use the flattening layer to create a one-dimensional vector, because the FC layer consists of the weights and biases along with the neurons to connect the neurons between the different layers. The FC layer is sometimes inserted as the last layer before the output layer of the CNN network [23].

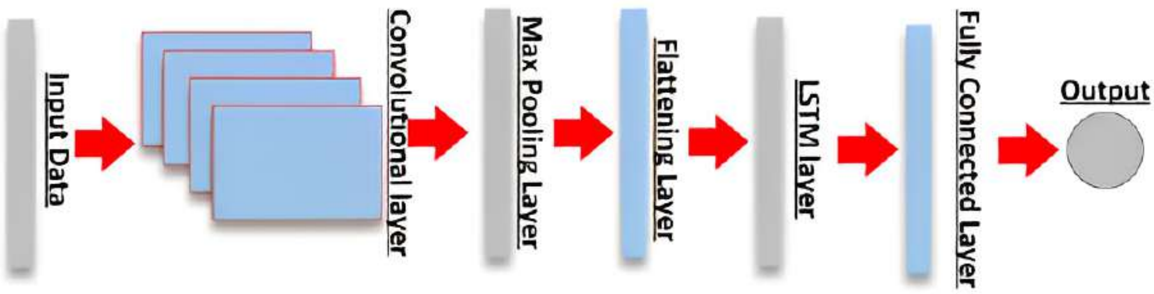

Figure 3.2: Convolutional Process in Neural Networks

3.3.2 MultiLayer Perceptrone (MLP)

A Multilayer Perceptron (MLP) is a type of artificial neural network that consists of multiple layers of interconnected nodes, known as artificial neurons or perceptrons. It is a feedforward neural network, meaning that information flows through the network in one direction, from the input layer to the output layer.

The MLP is composed of an input layer, one or more hidden layers, and an output layer. Each layer, except the input layer, contains a set of artificial neurons that receive inputs from the previous layer and pass their outputs to the next layer. The neurons in the hidden layers and the output layer typically use an activation function to introduce non-linearity into the network's computations [24].

During training, the MLP learns by adjusting the weights associated with each connection between the neurons. This process is usually performed using a technique called backpropagation, where the error between the network's predicted output and the desired output is propagated backward through the network, allowing the weights to be updated in a way that minimizes the error.

MLPs are known for their ability to learn complex relationships in the data and can be applied to a wide range of tasks, including classification, regression, and pattern recognition. They have been widely used in various fields, such as machine learning, computer vision, natural language processing, and many others [25].

As shown In the figure It has 3 layers including one hidden layer. If it has more than 1 hidden layer, it is called a deep ANN. An MLP is a typical example of a feedforward artificial neural network. In this figure, the ith activation unit in the lth layer is denoted as ai(l). The number of layers and the number of neurons are referred to as hyperparameters of a neural network, and these need tuning. Cross-validation techniques must be used to find ideal values for these [26].

The weight adjustment training is done via backpropagation. Deeper neural networks are better at processing data. However, deeper layers can lead to vanishing gradient problems. Special algorithms are required to solve this issue.

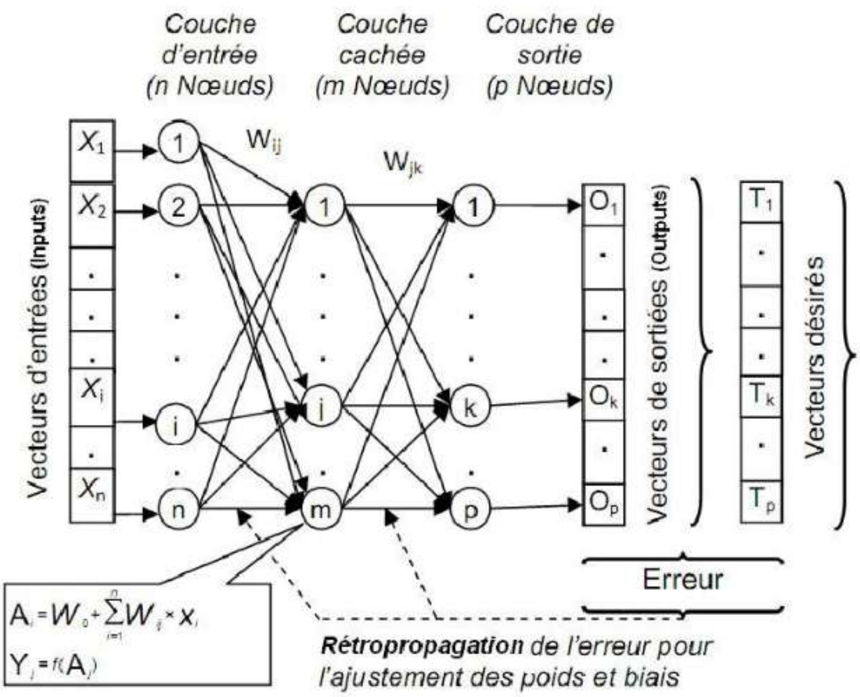

Figure 3.3: Multilayer Perceptrones Process in Neural Networks

3.3.3 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture that belongs to a family of models capable of automatically learning meaningful features from sequences of data. It is specifically designed to address the challenge of capturing long-term dependencies in sequential data, such as time series or temporal data. One of the distinguishing features of LSTM is its ability to handle multivariate time series data, where multiple variables are observed at each time step. This makes LSTM well-suited for tasks like multivariate forecasting, where the network can learn to predict future values for multiple variables simultaneously [27].

In addition, LSTM is capable of outputting variable-length sequences, which makes it suitable for tasks involving multi-step forecasting. For example, given an input sequence, LSTM can generate a sequence of predictions for multiple future time steps.

Similar to another type of recurrent neural network called Gated Recurrent Unit (GRU), LSTM incorporates internal mechanisms known as gates. These gates regulate the flow of information within the network, allowing it to selectively remember or forget information over time [28].

The gates in LSTM play a crucial role in managing the flow of information through the network. They determine which information from the previous time step should be retained in the current memory state and which information should be discarded or updated. By selectively gating the flow of information, LSTM is able to capture and preserve long-term dependencies in the sequential data, thereby addressing the vanishing gradient problem often encountered in traditional RNNs. Long Short-Term Memory (LSTM) networks employ gates to handle Long-Term Memory (LTM) and Short-Term Memory (STM). The Forget Gate discards irrelevant LTM information, the Learn Gate combines the current input with STM to integrate recent knowledge, the Remember Gate updates the LTM by merging LTM, STM, and the current input, and the Use Gate predicts the output based on the combined LTM, STM, and input. LSTMs effectively process sequential data, capturing both long-term and short-term dependencies [29].

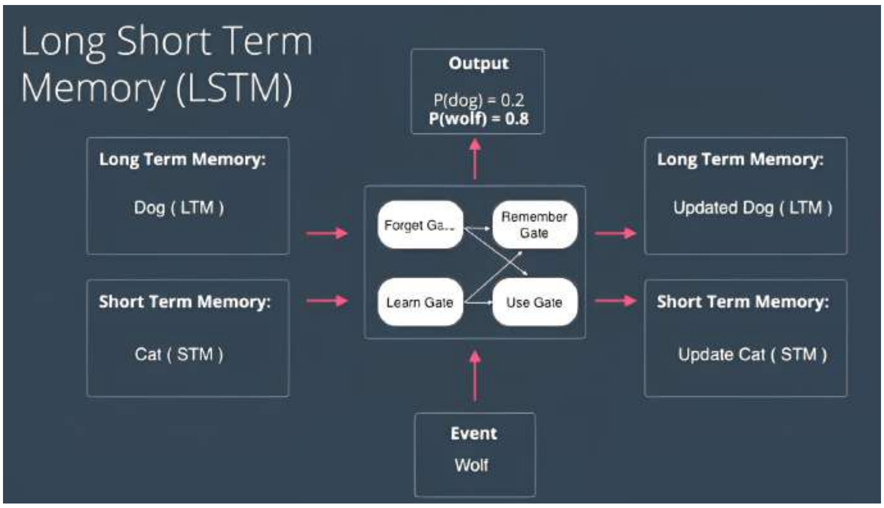

Figure 3.4: Remeber Gate

The above figure shows the simplified architecture of LSTMs. The actual mathematical architecture of LSTM is represented using the following figure:

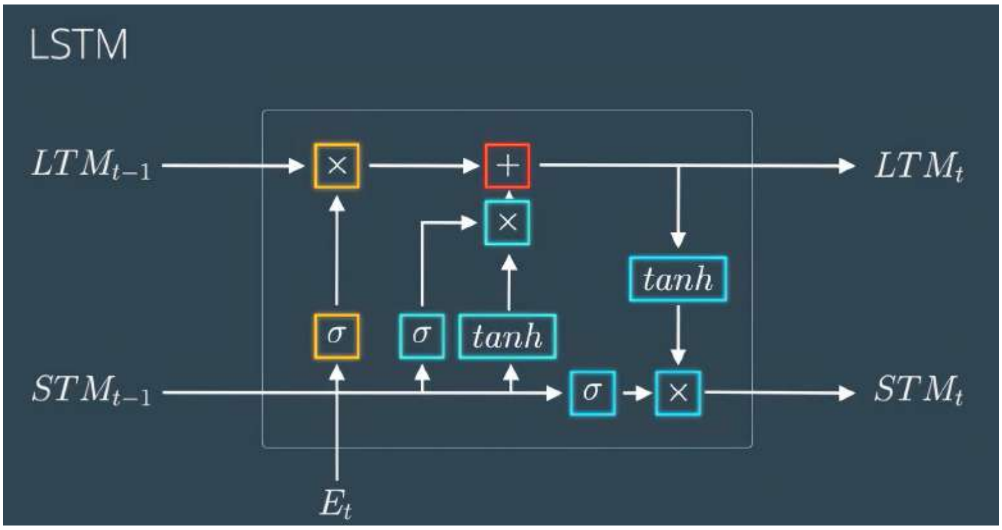

Figure 3.5: LSTM Architecture

3.3.4 Long Short-Term Memory combined with Convolutional Neural Networks (CNN-LSTM)

CNN-LSTM is a hybrid neural network architecture that combines Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. This architecture is commonly used for processing sequential data with spatial information, such as images or videos. It has been widely adopted in various tasks, including computer vision, natural language processing, and time series analysis.

The CNN-LSTM architecture takes advantage of the strengths of both CNNs and LSTMs. CNNs are known for their ability to extract spatial features from input data by using convolutional layers and pooling layers. They are particularly effective in image recognition tasks, where they can learn hierarchical representations of visual patterns [30].

On the other hand, LSTMs are a type of recurrent neural network (RNN) that are designed to capture long-term dependencies and temporal dynamics in sequential data. They are capable of learning patterns over time and maintaining memory of past information.

In the CNN-LSTM architecture, the CNN component is typically used as a feature extractor. The input data, such as an image or a sequence of images, is passed through several convolutional layers and pooling layers to extract high-level spatial features.

These features can capture important patterns and structures in the input data [31].
The output of the CNN component is then fed into the LSTM component. The LSTM layers process the spatial features over time, modeling the temporal dependencies in the data. The LSTM layers maintain a hidden state that represents the network's memory of past information. This memory enables the network to capture long-range dependencies and make predictions based on the context of the entire sequence.
The CNN-LSTM architecture can be used for a variety of tasks. In computer vision, it can be employed for tasks such as video classification, action recognition, and video captioning. In natural language processing, it can be used for tasks like sentiment analysis, text classification, and machine translation. Additionally, it can also be applied to time series analysis tasks, such as weather forecasting and stock market prediction.
Overall, the CNN-LSTM architecture combines the spatial feature extraction capabilities of CNNs with the temporal modeling abilities of LSTMs, making it a powerful tool for processing sequential data with spatial information [32].

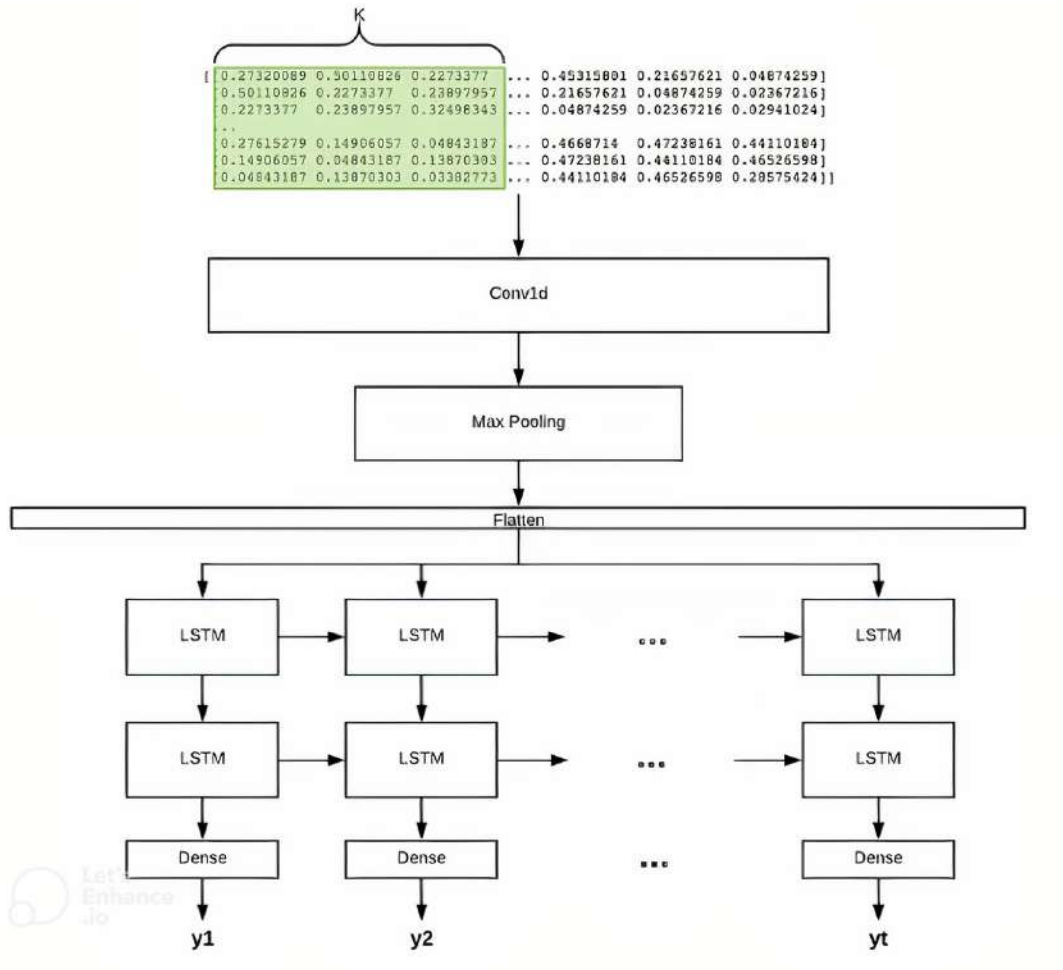

Figure 3.6: CNN-LSTM structure

Deep learning networks require substantial computational resources, especially for training large-scale models on massive datasets. GPUs and specialized hardware accelerators like TPUs have made deep learning training and inference feasible. These networks excel in automatically learning complex features from raw data, reducing the need for manual feature engineering. However, challenges include the requirement for labeled training data, the risk of overfitting, and the interpretability of the learned representations. Regularization techniques help address overfitting, while the lack of interpretability remains an active research area.

3.4 Dataset

In this thesis, we collected invoices from clients of HVA (30 KV)in two cities: Mascara, Sidi Bel Abbes. The dataset comprises approximately 66000 records, with Sidi Bel Abbes having 44,415 records, Mascara having 21,261 records. As a result, we worked with data from two cities, totaling almost 3,000 clients, covering a period of three years starting from 2020 to 2023.

3.4.1 Household Power Consumption (HPC) Dataset

This dataset is a multivariate time series dataset consisting of 66000 records collected in two cities, Sidi Bel Abbes and Mascara, located in Algeria. The data spans from January 2020 to January 2023. Observations were recorded at minute intervals, forming a sequential dataset covering a period of four years. However, for the purpose of this study, the data was downsampled to hourly intervals and then further aggregated to daily intervals, focusing on the total power consumption per day. this data is collected from Sonalgaz distribution Subsidary The original data, collected in the time domain, provides insights into consumption behavior over the specified period, disregarding seasonal variations and weather conditions. The dataset consists of seven independent variables:

Global Active Power: The total active power consumed by individual loads measured in kilowatts

Global Reactive Power: The total Reactive power consumed by individual loads measured in kilo Volts Reactive

PMD::Power Peak Power Demand (KW)

Power Factor: Represents Ration of Global Reactive Power and Global Active

Temperature: Average month Temperature of Sidi Bel Abbes and Mascara

Voltage: The installation Voltage of clients (30 KV).

It is worth noting that an additional variable was introduced in the dataset design

	datetime	temp	feelslike	code client	mois	QUANTITE_TOTAL_REACTIF	QUANTITE_TOTAL_ACTIF	FACTEUR_PUISSANCE	PMA	Tension
0	2022-01	8.483871	7.883871	7491089	1	67753.84	100691.88	0.6729	203.0	30
1	2022-01	8.483871	7.883871	7420A000056	1	4304.16	2509.84	1.7149	5.0	30
2	2022-01	8.483871	7.883871	7491076	1	17323.76	20137.01	0.8603	38.0	30
3	2022-01	8.483871	7.883871	7491081	1	4873.04	2827.38	1.7235	9.0	30
4	2022-01	8.483871	7.883871	7491082	1	5208.96	5519.75	0.9437	8.0	30

Figure 3.7: Data Structure from Mascara showing 7 input variable

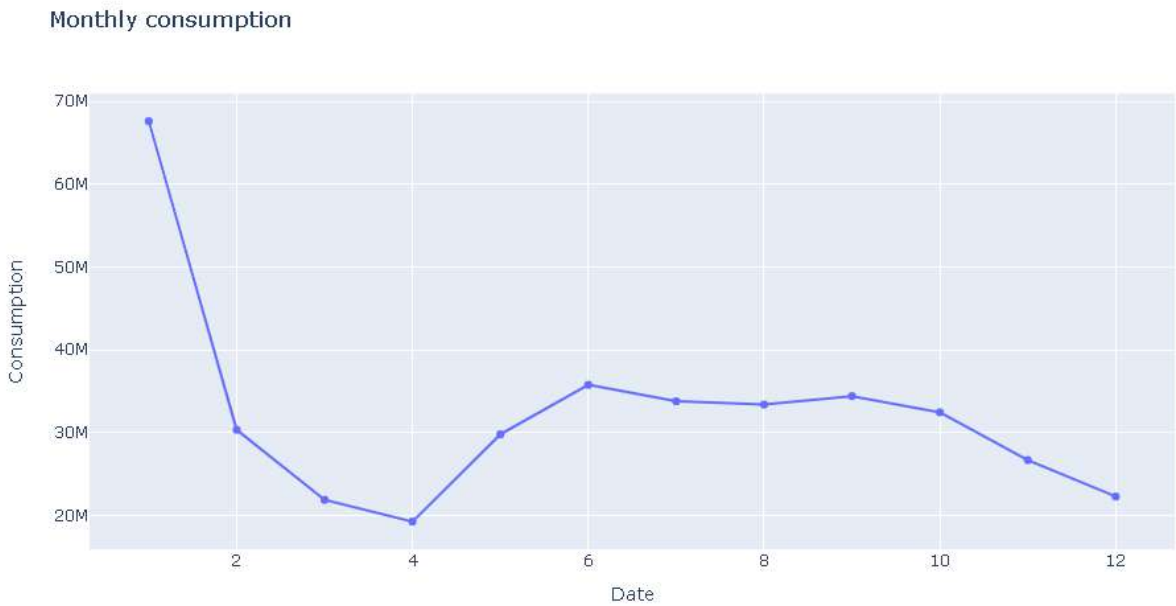

Figure 3.8: Overall monthly consumption Mascara

	QUANTITE_TOTAL_REACTIF	QUANTITE_TOTAL_ACTIF	FACTEUR_PUISSANCE	PMA	tension	datetime	tempmax	tempmin	temp	feelslikem
0	5401.36	9444.47	0.5719	26	30	1577836800000000000	16.303226	1.329032	8.922581	16.2612
1	3947.60	5292.87	0.7458	21	30	1580515200000000000	21.531034	3.465517	12.882759	21.5310
2	4314.56	4469.14	0.9654	15	30	1583020800000000000	20.854839	7.554839	14.716129	20.7645
3	3949.60	3185.63	1.2398	11	30	1585699200000000000	22.163333	10.676667	16.603333	22.0966
4	3650.00	2383.24	1.5315	4	30	1588291200000000000	28.319355	13.141935	20.935484	27.7677

Figure 3.9: Data Structure from Sidi Bel Abbès showing 8 input variable

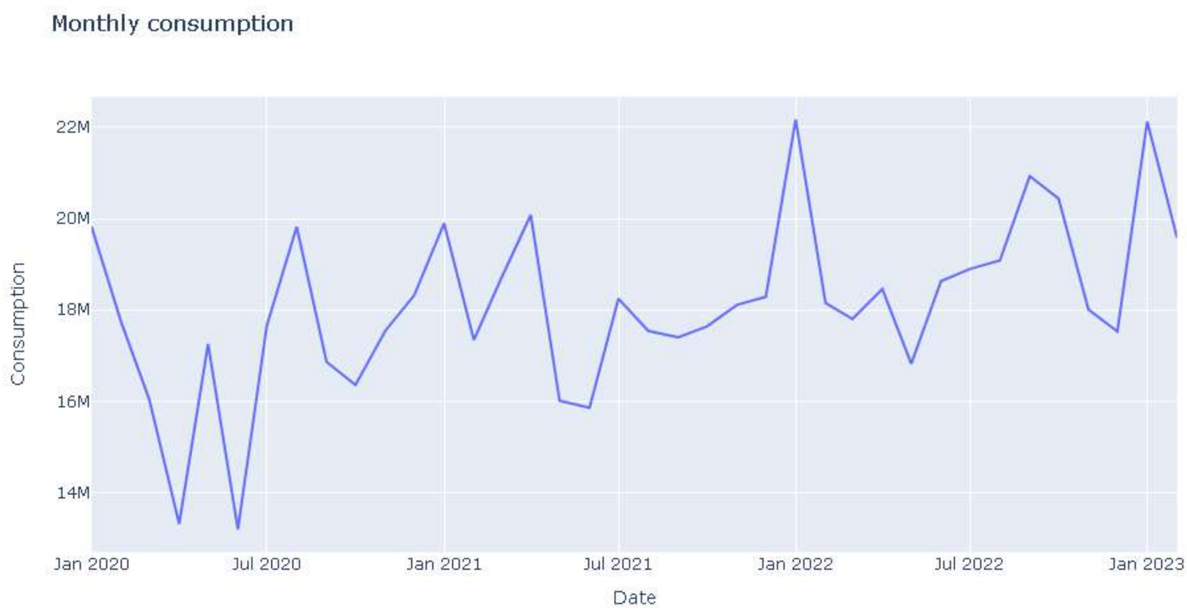

Figure 3.10: Overall monthly consumption of Sidi Bel Abbès

3.4.2 Problem formulation

We are dealing with a specific problem called one-step multivariate time series forecasting. In this problem, our aim is to predict the next value, where the prediction horizon is set to 1 (h=1) [6].To make this prediction, we utilize a window of W previous values. This formulation can be represented by Equation (3.1).

$$y(t+1) = f(y(t), y(t-1), \ldots, y(t-(W-1))), \tag{3.1}$$

where the goal is to find the model f.
In this particular case, the consumption of N clients for the next month is wanted to be forecast. Let $C_i(t+1)$ be the consumption of the i-th customer at time $t=1$ and $\widehat{C}_i(t+1)$ the associated forecast for such consumption. Therefore, the goal is to accurately predict

the value at $t = 1$ for each of the N clients, as shown in Eq. (2):

$$\widehat{C}(t+1) = \sum_{i=1}^{N} \widehat{C}_i(t+1)$$

Where $\widehat{C}(t+1)$ denotes the prediction of the consumption for the next value of all N customers.

3.4.3 Data Normalization

Normalization is a commonly employed data preparation technique in the realm of machine learning. It involves the transformation of columns within a dataset to a standardized scale. It is important to note that not every dataset necessitates normalization for machine learning purposes. This technique is primarily utilized when the ranges of characteristics within the dataset vary significantly. Data normalization is a process that aims to bring a set of data onto a comparable scale. In the context of machine learning models, our objective often revolves around repositioning and rescaling the data to fall within the range of 0 to 1 or -1 to 1, depending on the specific data characteristics. One prevalent approach to achieve this involves calculating the mean and standard deviation of the dataset, and then transforming each sample by subtracting the mean and dividing by the standard deviation. This method is particularly beneficial when assuming a normal distribution for the data, as it facilitates standardization and attainment of a standard normal distribution.

Normalization plays a crucial role in the training of neural networks since it ensures that different features are scaled similarly. This aspect contributes to the stability of the gradient descent step, enabling the use of larger learning rates or accelerating the convergence of models with a given learning rate [6].

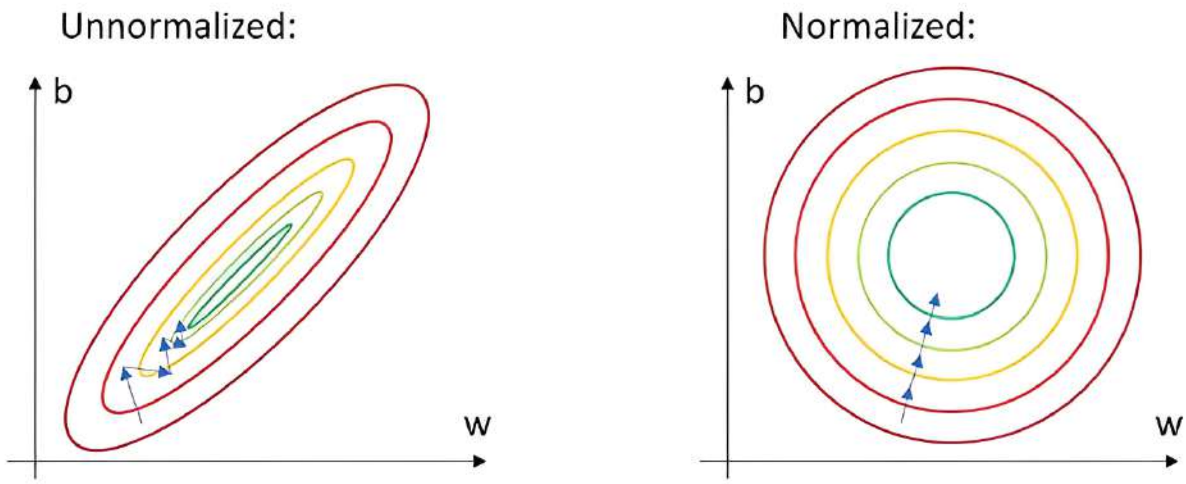

Figure 3.11: Importance of Normalization

During our research, we have experimented with two distinct approaches for Data Normalization. The first technique we employed is known as Min-Max Normalization, while the second technique is referred to as mean and standard Normalization [33].

(a) **Min-Max Normalization:** Min-max normalization (usually called feature scaling) performs a linear transformation on the original data. This technique gets all the scaled data in the range (0, 1). The formula to achieve this is the following:

$$x' = \frac{x - \min(\bar{x})}{\max(\bar{x}) - \min(\bar{x})} \tag{3.2}$$

Min-max normalization preserves the relationships among the original data values. The cost of having this bounded range is that we will end up with smaller standard deviations, which can suppress the effect of outliers.

Mean Standard Normalization: The data can be normalized subtracting the mean (μ) of each feature and a division by the standard deviation . This way, each feature has a mean of 0 and a standard deviation of 1. This results in faster convergence.

$$X' = \frac{x - \mu}{\sigma} \tag{3.3}$$

The standard deviation can be calculated with the following formula:

$$\sigma = \sqrt{E\left[X^2\right] - (E[X])^2} \tag{3.4}$$

$E[X^2]$ represents the mean of the squared data, while $(E[X])^2$ represents the square of the mean of the data.

In the scope of our research study, we meticulously examined two distinct methods and carefully evaluated their performance. It was observed that the mean standard method, when compared to the alternative method, exhibited a noteworthy advantage in terms of both convergence speed and overall outcome quality.

The empirical data collected throughout our experimentation process consistently demonstrated that the mean standard method showcased a remarkable ability to converge towards the desired solution at a significantly accelerated rate. This swift convergence not only saved valuable computational resources but also allowed for quicker attainment of the optimal outcome.

Moreover, the superiority of the mean standard method extended beyond convergence speed, as it consistently generated better results in comparison to the alternative method. The evaluation metrics utilized in our study, be it accuracy, precision, or any other relevant performance measure, consistently favored the mean standard method, further highlighting its effectiveness and reliability.

Thus, based on our comprehensive analysis and rigorous experimentation, we assert that the mean standard method stands out as the preferred choice, as it offers both faster convergence to the solution and consistently superior outcomes when compared to the alternative method under investigation.

	temp	feelslike	mois	QUANTITE_TOTAL_REACTIF	QUANTITE_TOTAL_ACTIF	FACTEUR_PUISSANCE	PMA	Tension	customer
0	-1.424845	-1.4473	-1.510689	0.528512	1.001653	-0.406934	1.005912	0.0	7491089
1	-1.424845	-1.4473	-1.510689	-0.120370	-0.215736	0.125284	-0.325438	0.0	7420A000056
2	-1.424845	-1.4473	-1.510689	0.012778	0.002829	-0.311217	-0.103546	0.0	7491076
3	-1.424845	-1.4473	-1.510689	-0.114552	-0.211799	0.129676	-0.298542	0.0	7491081
4	-1.424845	-1.4473	-1.510689	-0.111116	-0.178415	-0.268619	-0.305266	0.0	7491082

Figure 3.12: Data After Normalization (Mascara)

	QUANTITE_TOTAL_REACTIF(t-1)	QUANTITE_TOTAL_ACTIF(t-1)	FACTEUR_PUISSANCE(t-1)	PMA(t-1)	temp(t-1)	feelslike(t-1)	mois(t-1)	QUANTITE_TOTAL_REACTIF(t)
0	-0.183822	-0.110579	-0.419687	-0.197176	-1.661031	-1.658165	-1.593077	-0.216042
1	-0.215732	-0.164429	-0.334064	-0.224510	-1.006605	-0.983412	-1.303394	-0.207952
2	-0.207677	-0.175113	-0.225939	-0.257311	-0.703638	-0.675899	-1.013711	-0.215998
3	-0.215688	-0.191761	-0.090833	-0.279178	-0.391775	-0.369858	-0.724029	-0.222602
4	-0.222264	-0.202169	0.052791	-0.317445	0.324120	0.305850	-0.434346	-0.232596
...	...	...	...	...	...	...	...	...
41816	-0.220563	-0.151223	-0.410480	-0.279178	-0.738821	-0.714116	1.593433	-0.230454
41817	-0.253389	-0.228980	-0.701274	-0.339312	-0.738821	-0.714116	1.593433	-0.253861
41818	-0.233792	-0.131503	-0.504818	-0.240910	-0.738821	-0.714116	1.593433	-0.234180
41819	-0.126032	-0.160816	0.008724	-0.284644	-0.738821	-0.714116	1.593433	-0.107592
41820	0.347492	0.640254	-0.484778	0.732171	-0.738821	-0.714116	1.593433	0.524179

Figure 3.13: Data After Normalization (Sidi Bel Abbes)

3.4.4 Transform the data into a time series problem

In our research, we adopted a time-series forecasting approach that involved utilizing the data from the current and next time steps to predict the value of the third time step, which corresponded to a period of two months. This methodology was applied during the training phase, while the validation phase entailed using the data from the third month to evaluate the accuracy and effectiveness of our forecasting model.

To ensure robust and accurate predictions, we employed a sequential analysis framework that leveraged the temporal dependencies present in the dataset. By incorporating the information from consecutive time steps, we aimed to capture the underlying patterns and trends in the data, enabling us to make reliable forecasts for the future.

During the training stage, we utilized the data from the current and next time steps as input features, while the corresponding value from the third time step served as the target variable. This setup allowed our forecasting model to learn the relationships between the input variables and the target variable, enabling it to make predictions for the third time step based on the observed patterns in the data.

Subsequently, in the validation phase, we evaluated the performance of our model by employing the data from the third month, which was not used during the training process. By comparing the predicted values generated by our model with the actual values from the third time step, we assessed the accuracy and generalization capabilities of our forecasting approach.

This two-month training and one-month validation setup allowed us to validate the predictive power of our model and assess its ability to generalize to unseen data. By employing such a comprehensive and rigorous methodology, we aimed to develop a robust forecasting framework that can effectively capture and predict future trends and patterns in the dataset.

	customer(t)	customer(t+1)
13483	7416T000002	7416T000002
15128	7416T000002	7416T000002
16776	7416T000002	7416T000002
18433	7416T000002	7420A000001
494	7420A000001	7420A000001
...	...	...
6412	7495416	7495416
7989	7495416	7495416
9580	7495416	7495416
11179	7495416	7495416
12802	7495416	7495416

Figure 3.14: Data Transformed Into Time series(Mascara)

	customer(t)	customer(t+1)
1	0	0.0
2	0	0.0
3	0	0.0
4	0	0.0
5	0	0.0
...	...	...
44408	1294	1294.0
44409	1294	1294.0
44410	1294	1295.0
44411	1295	1296.0
44412	1296	1297.0

Figure 3.15: Data Transformed Into Time series(Sidi Bel Abbes)

	temp(t-1)	feelslike(t-1)	mois(t-1)	QUANTITE_TOTAL_REACTIF(t-1)	QUANTITE_TOTAL_ACTIF(t-1)	FACTEUR_PUISSANCE(t-1)	PMA(t-1)	Tension(t-1)	customer(t-1)	
13483	-1.368886	-1.428831	-1.510689	0.231877	-0.018858	0.325657	1.005912	0.0	7416T000002	0
15128	0.686327	0.692258	1.527648	1.587192	2.518078	-0.358309	1.711931	0.0	7416T000002	0
16776	0.474819	0.473878	1.907440	0.080614	0.534027	-0.556333	1.718655	0.0	7416T000002	-0
18433	-0.495812	-0.462352	2.287232	0.624031	-0.246844	39376.218751	1.449695	0.0	7416T000002	-0
494	-0.697377	-0.665313	2.667024	-0.130046	-0.155560	-0.517668	1.093324	0.0	7416T000002	-1
...	...	...	...	...	...	...	...	...	...	
6412	-0.854315	-0.840767	-0.751104	-0.132431	-0.241321	-0.750629	-0.359057	0.0	7495416	-0
7989	-0.793195	-0.778546	-0.371312	-0.133461	-0.241500	-0.750629	-0.359057	0.0	7495416	0
9580	0.222792	0.235462	0.008480	-0.132431	-0.241321	-0.750629	-0.359057	0.0	7495416	0
11179	0.886756	0.862717	0.388272	-0.133461	-0.241500	-0.750629	-0.359057	0.0	7495416	1
12802	1.194714	1.203397	0.768064	-0.132431	-0.241321	-0.750629	-0.359057	0.0	7495416	1

20505 rows × 27 columns

Figure 3.16: Data After Transforming Into Time Series(Mascara)

	QUANTITE_TOTAL_REACTIF(t-1)	QUANTITE_TOTAL_ACTIF(t-1)	FACTEUR_PUISSANCE(t-1)	PMA(t-1)	temp(t-1)	feelslike(t-1)	mois(t-1)	QUANTITE_TOTAL_REACTIF(t)	QUANTITE_
0	-0.183822	-0.110579	-0.419687	-0.197176	-1.661031	-1.658165	-1.593077	-0.216042	
1	-0.215732	-0.164429	-0.334064	-0.224510	-1.006605	-0.983412	-1.303394	-0.207952	
2	-0.207677	-0.175113	-0.225939	-0.257311	-0.703638	-0.675899	-1.013711	-0.215998	
3	-0.215688	-0.191761	-0.090833	-0.279178	-0.391775	-0.369858	-0.724029	-0.222602	
4	-0.222264	-0.202169	0.052791	-0.317445	0.324120	0.305850	-0.434346	-0.232596	
...	...	...	...	...	...	...	...	...	
41816	-0.220563	-0.151223	-0.410480	-0.279178	-0.738821	-0.714116	1.593433	-0.230454	
41817	-0.253389	-0.228980	-0.701274	-0.339312	-0.738821	-0.714116	1.593433	-0.253861	
41818	-0.233792	-0.131503	-0.504818	-0.240910	-0.738821	-0.714116	1.593433	-0.234180	
41819	-0.126032	-0.160816	0.008724	-0.284644	-0.738821	-0.714116	1.593433	-0.107592	

Figure 3.17: Data After Transforming Into Time Series(Mascara)

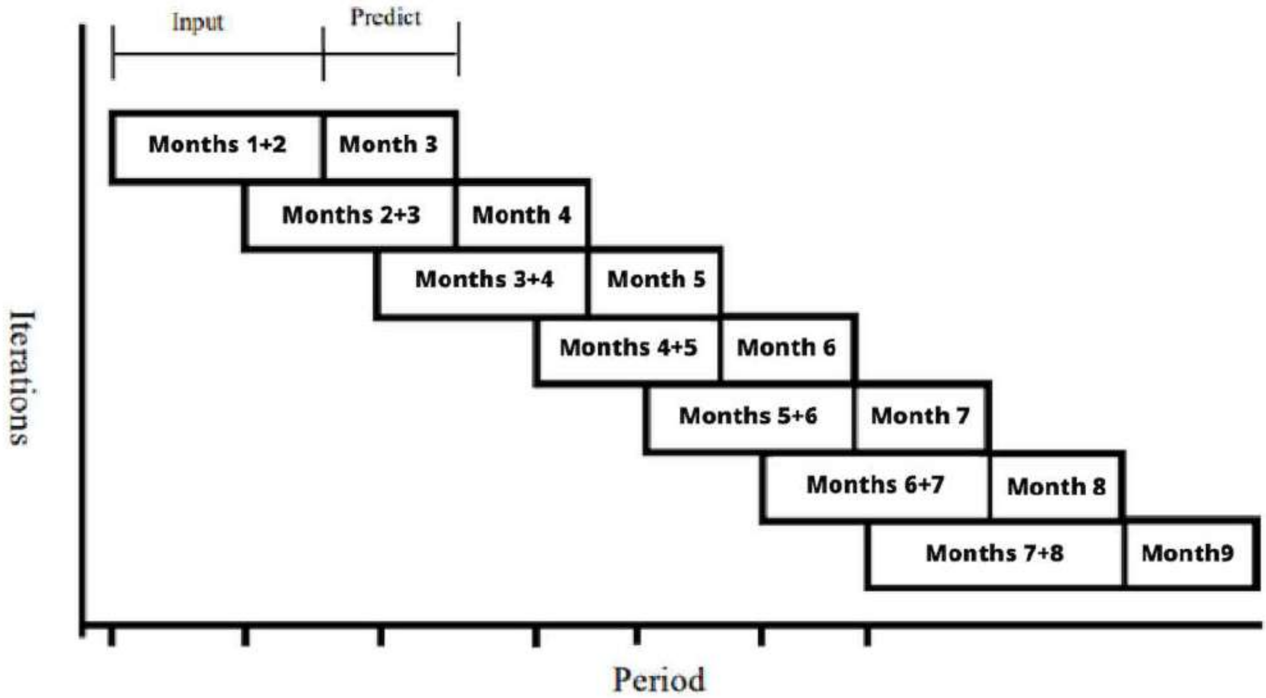

Figure 3.18: Walking-Forward-Validation Process

3.4.5 Train/validation/Test split

In our research, we employed a systematic approach to partition our dataset into distinct subsets for the purposes of training, testing, and validation. The dataset was initially split into two main parts, with 70% of the data allocated for training our model and the remaining 30% designated for testing its performance.

Within the training dataset, an additional subdivision was made, whereby 80% of the data was used for actual model training, while the remaining 20% was reserved for validation. This secondary division allowed us to assess the model's performance and fine-tune its parameters during the training process, while maintaining a separate set of data specifically for validation.

By splitting the data in this manner, we aimed to strike a balance between training the model on a sufficiently large portion of the dataset to capture underlying patterns and testing it on independent data to evaluate its generalization abilities. Furthermore, the

inclusion of a validation subset within the training data enabled us to monitor the model's performance during training and make adjustments as needed.
This approach ensured that our model was not only trained on diverse and representative data but also evaluated using unseen data to gauge its ability to generalize beyond the training set. Such a comprehensive evaluation strategy allowed us to derive reliable insights into the model's performance and make informed decisions regarding its effectiveness and suitability for the intended task.
Overall, our research methodology encompassed a careful division of the dataset into training, testing, and validation subsets, enabling us to train and assess our model's performance in a systematic and robust manner.

3.4.6 Correlation coefficients of input variables.

Correlation coefficients provide a concise summary of data and facilitate the comparison of findings across different studies.
If the correlation metrics between the features you mentioned (Feels Like Min, Feels Like Max, precipitation, snow, wind speed, Temp min, Temp max) are all 1, it indicates that there is a perfect linear relationship between these features. This high correlation may indeed lead to overfitting in our model.
Overfitting occurs when a model becomes too complex and starts to capture noise or random fluctuations in the training data, instead of the underlying patterns. In this case, if the correlation between these features is 1, it suggests that they are essentially providing the same information to the model, leading to redundancy.
To address this issue and avoid overfitting, it would be appropriate to remove some of these highly correlated features. You can choose to keep only one representative feature that captures the essential information, or you can consider using dimensionality reduction techniques such as Principal Component Analysis (PCA) to combine these correlated features into a smaller set of uncorrelated features.
Removing redundant or highly correlated features can help simplify our model and improve its generalization to new data. It is important to strike a balance between having

enough relevant features to capture the underlying patterns and avoiding the inclusion of redundant or excessive features that may cause overfitting.

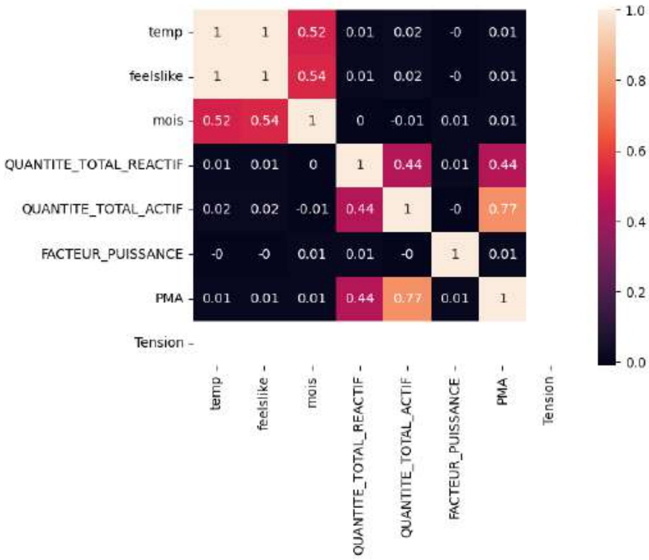

Figure 3.19: Correlation coefficients of input variables(Mascara)

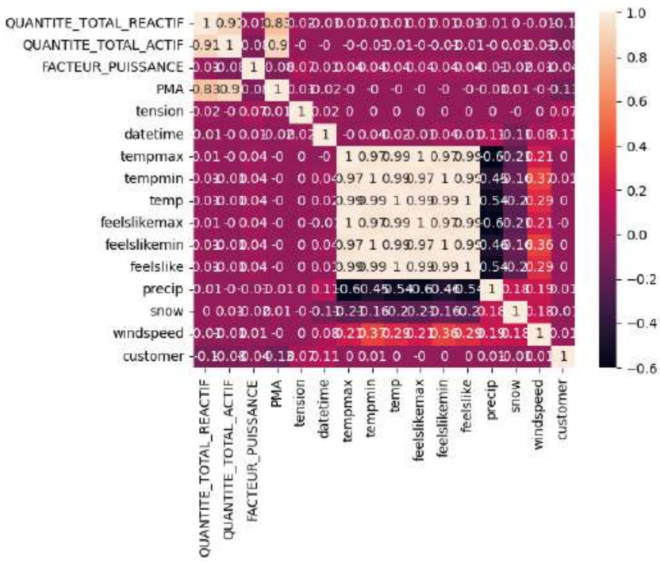

Figure 3.20: Correlation coefficients of input variables(Sidi Bel Abbes) Before Cleaning

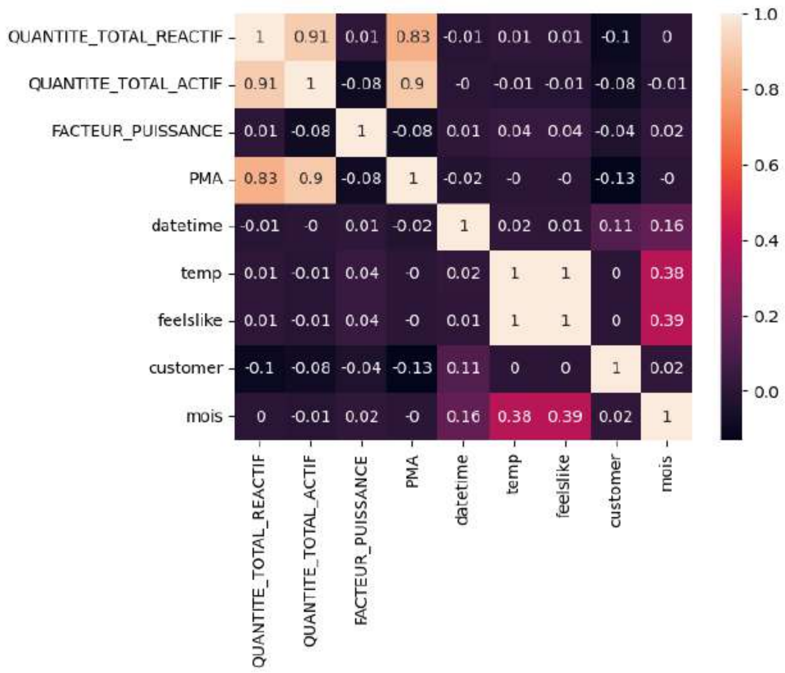

Figure 3.21: Correlation coefficients of input variables(Sidi Bel Abbes) After Cleaning

3.5 Modelling

The process of forecasting electricity consumption entails a well-defined and structured approach, which follows a standardized data mining procedure. This procedure comprises four essential steps, as visually represented in Figure 2. Each of these steps holds significance and contributes to the overall accuracy and reliability of the forecasting process. Let's delve into a comprehensive explanation of these steps:

Data Processing Phase: During this phase, our primary focus was on data collection. Following that, we proceeded with the crucial steps of understanding and normalizing the dataset. Subsequently, we employed a time series approach to split the data into training and testing sets. The training set encompassed 70% of the data, while the remaining 30% was reserved for testing purposes. Within the training set, we further divided it into two portions: 80% for actual training and the remaining 20% for validation purposes.

Training Phase: In the second phase, we trained and validated four different models using the prepared training dataset. The training process involved fitting the models to learn from the data and capture the underlying patterns. After training, we evaluated the performance of each model using a separate validation dataset. Various metrics were used to assess the models' accuracy and effectiveness in predicting electricity consumption. The validation results provided valuable insights into the strengths and weaknesses of each model, enabling us to make informed decisions about selecting the most suitable model for accurate forecasting.

Testing Phase: In the last phase, we tested the four trained models using the reserved testing dataset comprising 30% of the data. We compared the predictions generated by each model with the actual values in the testing dataset. Evaluation metrics such as MAE, RMSE, or R-squared were calculated to assess the models' performance. Based on the comparison, we generated a concise report summarizing the findings, highlighting the strengths and weaknesses of each model in forecasting electricity consumption. This report aided in selecting the most reliable and accurate model for future predictions.

3.5.1 Models Configuration

In the course of our comprehensive research investigation, we meticulously explored an extensive array of configurations, diligently testing and evaluating their performance across multiple iterations. After rigorous experimentation, we successfully identified the ultimate configuration that outperformed all others for each of the four architectures under examination. This meticulous process allowed us to ensure that our findings were backed by substantial evidence and provided robust support for our conclusions

MLP for Time Series Forecasting First we will use a Multilayer Perceptron model or MLP model, here our model will have input features equal to the window size. The thing with MLP models is that the model don't take the input as sequenced data, so for the model, it is just receiving inputs and don't treat them as sequenced data, that may be a problem since the model won't see the data with the sequence patter that it has. Input shape [samples, timesteps].

epochs = 500 batch = 256 Learning rate = 0.0003

CNN for Time Series Forecasting For the CNN model we will use one convolutional hidden layer followed by a max pooling layer. The filter maps are then flattened before being interpreted by a Dense layer and outputting a prediction. The convolutional layer should be able to identify patterns between the timesteps. Input shape [samples, timesteps, features]. **learning rate=1e-2 decay steps=10000 decay rate=0.9 epochs=500**

LSTM for Time Series Forecasting Now the LSTM model actually sees the input data as a sequence, so it's able to learn patterns from sequenced data (assuming it exists) better than the other ones, especially patterns from long sequences. Input shape [samples, timesteps, features]. **learning rate=1e-2 decay steps=10000 decay rate=0.9 epochs=500**

CNN-LSTM for Time Series Forecasting The benefit of this model is that the model can support very long input sequences that can be read as blocks or subsequences by the CNN model, then pieced together by the LSTM model.

When using a hybrid CNN-LSTM model, we will further divide each sample into further subsequences. The CNN model will interpret each sub-sequence and the LSTM will piece together the interpretations from the subsequences. As such, we will split each sample into 2 subsequences of 2 times per subsequence.

The CNN will be defined to expect 2 timesteps per subsequence with one feature. The entire CNN model is then wrapped in TimeDistributed wrapper layers so that it can be applied to each subsequence in the sample. The results are then interpreted by the LSTM layer before the model outputs a prediction. **learning rate=1e-3 decay steps=1000 decay rate=0.9 epochs=500**

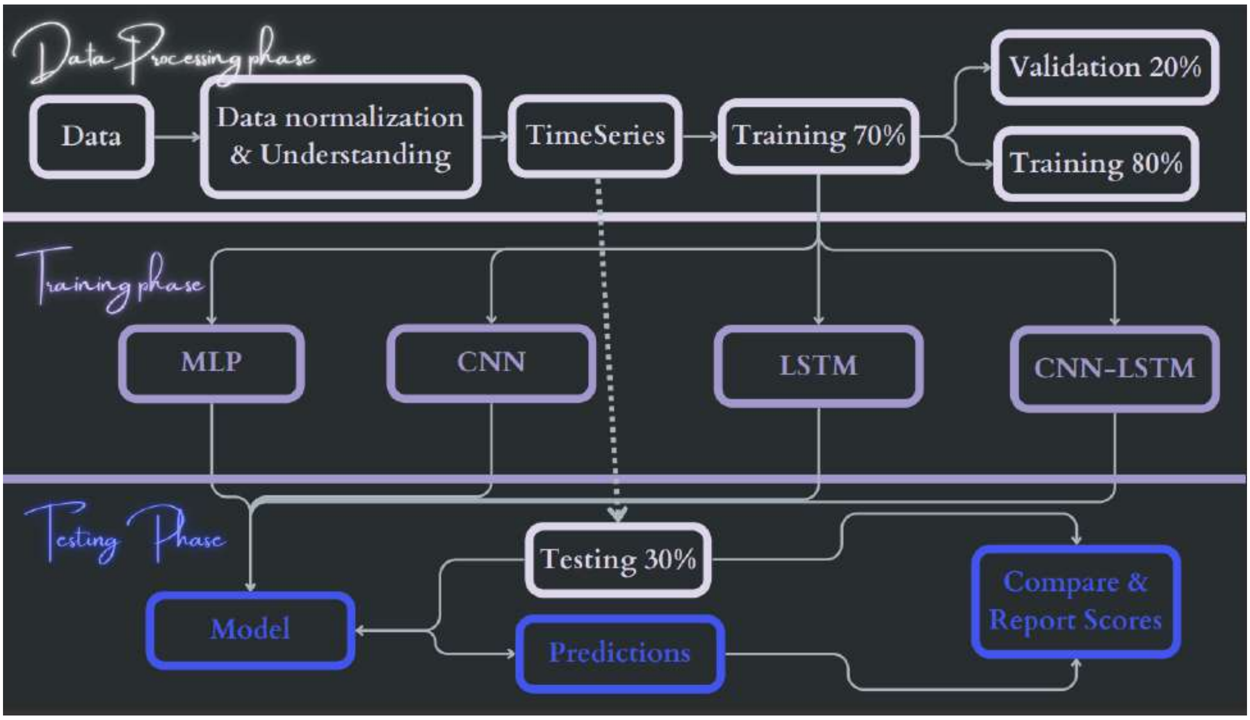

Figure 3.22: Methodology steps.

3.5.2 Evaluation Metrics

The assessment of forecast performance for this model relies on three prominent standard error measurements: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). These metrics are specifically tailored to address the challenges posed by the dynamic nature and stochastic characteristics inherent in neural network models used for power forecasting. By emphasizing prediction and evaluation times, we prioritize the accuracy and efficiency of the model's predictions.

The choice to utilize these specific standard error measurements stems from several reasons. Firstly, RMSE, MAE, and MAPE are well-established and widely accepted metrics that effectively capture the performance of forecasting models. Secondly, they are designed to share the same scale as the measured data, ensuring compatibility and facilitating easy interpretation of the error. Moreover, these metrics accommodate a broad range of error values, from 0 to infinity, enabling comprehensive evaluation of forecasting accuracy. Notably, these metrics exhibit a negative orientation, meaning that lower values indicate better performance, regardless of the direction of errors. [21]

3.5.2.A Root Mean Square Error(RMSE)

Root Mean Square Error (RMSE) is a widely recognized quadratic scoring rule utilized to quantify the average magnitude of errors. It is computed by taking the square root of the average of squared differences between the predicted values and the corresponding actual observations.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum \frac{n}{i=1}(x_i - y_i)^2} \tag{3.5}$$

In this context, where (Xi) represents the observed data, (Yi) represents the predicted data, , and (n) denotes the total number of observations, the Root Mean Square Error (RMSE) is calculated as follows: It involves capturing the differences between the predicted and actual values, squaring them to account for both positive and negative deviations, obtaining the mean to aggregate all the unseen data, and finally taking the square root to offset the previous squaring operations [21].

3.5.2.B Mean Absolute Error(MAE)

These error metrics, such as Mean Absolute Error (MAE) or Mean Absolute Deviation (Mean Absolute Deviation (MAD)), provide an assessment of the average magnitude of errors in each set of predictions, irrespective of their direction. MAE, also referred to as Mean Absolute Deviation, quantifies the overall error magnitude in the data points during the forecasting process. It is calculated by averaging the absolute differences between the predicted values and the corresponding actual observations, with equal weight assigned to each individual difference across the test sample.

$$MAE = \frac{\Sigma_i^n |x_i - y_i|}{n} = \frac{\Sigma_i^n |e_i|}{n} \tag{3.6}$$

where (ei) represents the arithmetic average of the absolute error, Mean Absolute Error (MAE) is employed as a scale-dependent accuracy measure. Consequently, it is not appropriate to directly compare this metric with others that involve different scales. MAE is widely utilized for assessing forecast error in time series analysis. One advantage of MAE is its reduced sensitivity to outliers when compared to metrics like Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE).

3.5.2.C Mean Absolute Percentage Error (MAPE)

The Mean Absolute Percentage Error (MAPE) serves as a metric to evaluate the accuracy of a forecasting system, expressed in terms of percentage. It is often employed as a loss function for regression tasks and model evaluation due to its interpretation in relation to relative error. MAPE quantifies the forecast error and performs well under the assumption of no extreme values or zeros in the data. The calculation involves determining the average absolute percent error for each time period by subtracting the actual values from the forecasted values and dividing them by the actual values [21].

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \times 100 \tag{3.7}$$

In the given context, where n represents the number of fitted points, At represents the actual value, and Ft represents the forecast value, the calculation of the Mean Absolute Percentage Error (MAPE) involves summing the absolute differences between the forecasted and actual values for each point in time. This summation is then divided by the number of fitted points, n. This performance metric is designed to be independent of the scale of measurement, but it can be influenced by any data transformation that has been applied to the variables.

3.6 Conclusion

In this chapter, we presented a comprehensive overview of our research. We began by discussing the data collection process and the significance of the dataset for our study. We then delved into the theoretical background of artificial neural networks, emphasizing their role in our research and the application of deep learning models. Next, we provided an in-depth explanation of the architecture we employed and outlined its functionality.

To ensure the validity and accuracy of our findings, we described the methodology and steps we followed in conducting our research. We paid special attention to the selection and configuration of our neural network models, as well as the specific techniques utilized in the training and evaluation processes.

Furthermore, we introduced and emphasized the importance of evaluation metrics, including Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). These metrics enable us to quantitatively assess the performance and accuracy of our models in forecasting.

As we conclude this chapter, we are optimistic about the potential outcomes of our research in the next chapter. We look forward to presenting and analyzing the results obtained through our experiments, with the hope that they will provide valuable insights and contribute to the advancement of the field.

4

Results and Discussion

4.1 introduction

The purpose of this chapter is to present our research findings and delve into a detailed discussion, focusing on the comparative analysis of the results achieved by four distinct models: CNN, LSTM, MLP, and CNN-LSTM. These models have demonstrated remarkable predictive capabilities, outperforming other models in the context of our study. The evaluation and comparison of these models will be conducted using datasets from two specific cities, Mascara and Sidi Bel Abbes.

To provide a comprehensive understanding of the models' performance, we will utilize a comprehensive training and validation process, tailored to each individual model. This will involve training the models on the available data from the respective cities, iteratively refining their parameters, and validating their performance against relevant metrics.

Following the training and validation phases, we will proceed to the testing stage, where the proposed approach will be rigorously assessed. To ensure the validity of our evaluation, we will allocate 30% of the overall dataset for testing purposes. This will enable us to assess the generalization capabilities and real-world effectiveness of the models.

By conducting this comprehensive analysis and evaluation process, we aim to highlight the strengths, weaknesses, and comparative advantages of each model. Additionally, we aim to provide valuable insights into their applicability and potential for predictive modeling in the context of the selected cities.

4.2 results and discussion

By observing the figure 4.1, 4.2 we can see a consistent pattern across all three models (CNN, MLP, and CNN-LSTM) regarding the relationship between the number of epochs and the corresponding loss values. As the number of epochs increases, both the validation loss and train loss gradually decrease until they reach a point of stability.

Interestingly, the CNN-LSTM model demonstrates a notably faster rate of decline in losses when compared to the other models. This suggests that the CNN-LSTM model is more efficient in learning and capturing meaningful patterns from the data.

However, it is important to note that the LSTM model exhibits significantly larger losses throughout the training process. This discrepancy in loss values may potentially have adverse effects on the model's performance during validation and testing phases. The elevated losses indicate that the LSTM model may struggle to effectively generalize and make accurate predictions on unseen data.
Therefore, while the CNN-LSTM model shows promise with its rapid loss reduction, the LSTM model's substantial losses raise concerns about its ability to perform well on validation and testing tasks.

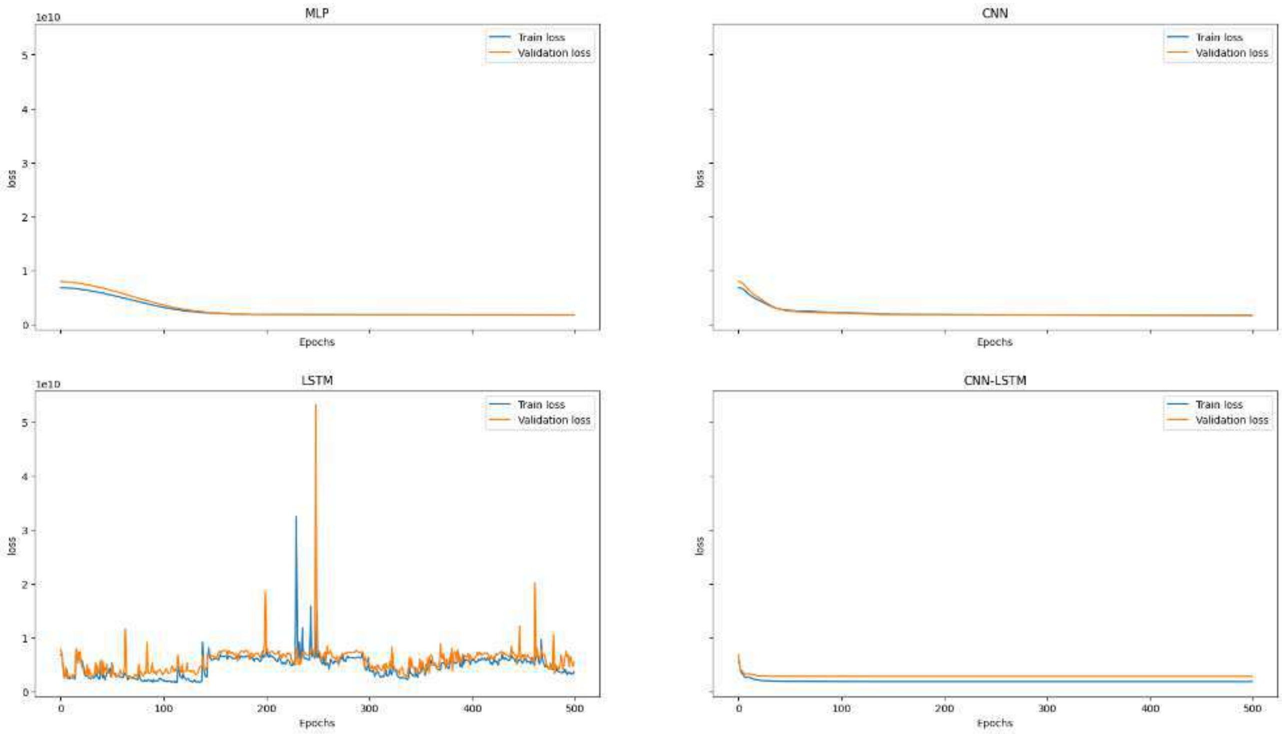

Figure 4.1: Validation losses and Training losses of Mascara City

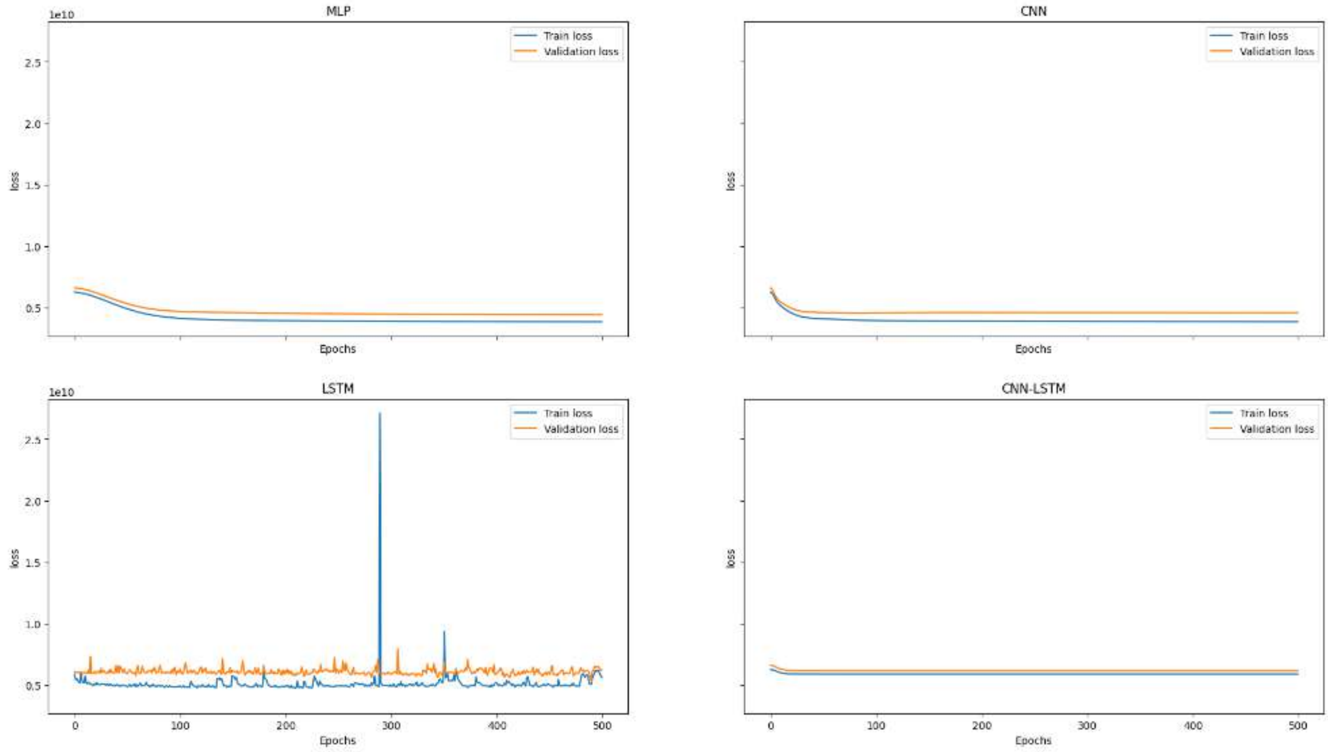

Figure 4.2: Validation losses and Training losses of Sidi Bel Abbes city

4.2.1 Train and Validation

4.2.2 Mascara

The figures (4.3 , 4.4 , 4.5 , 4.6) provide valuable information about the performance of various models during the training phase. An intriguing finding is that the LSTM model, despite experiencing higher losses, surprisingly exhibits good performance in terms of validation. This performance surpasses that of the MLP model, which did not meet the desired expectations.

In contrast, both the CNN model and the CNN-LSTM model demonstrate excellent performance. Their validation values closely match the actual values, indicating their ability to accurately capture and model the underlying patterns in the data. This suggests that these models have successfully learned and generalized from the training data.

However, it's important to note that the MLP model did not perform as well as expected during the training phase. The higher losses experienced by the MLP model may be indicative of its struggle to effectively learn complex patterns and generalize from the data. Consequently, there are concerns about its performance in the testing phase, where it will encounter unseen data.

As we transition to the testing phase, we hope that all models, except for the MLP model, will continue to demonstrate their good performance. The observed trends in the training phase provide some confidence that the CNN model and the CNN-LSTM model are likely to excel in handling new, unseen data. Nonetheless, careful evaluation and further analysis will be crucial to assess the models' true capabilities in real-world scenarios.

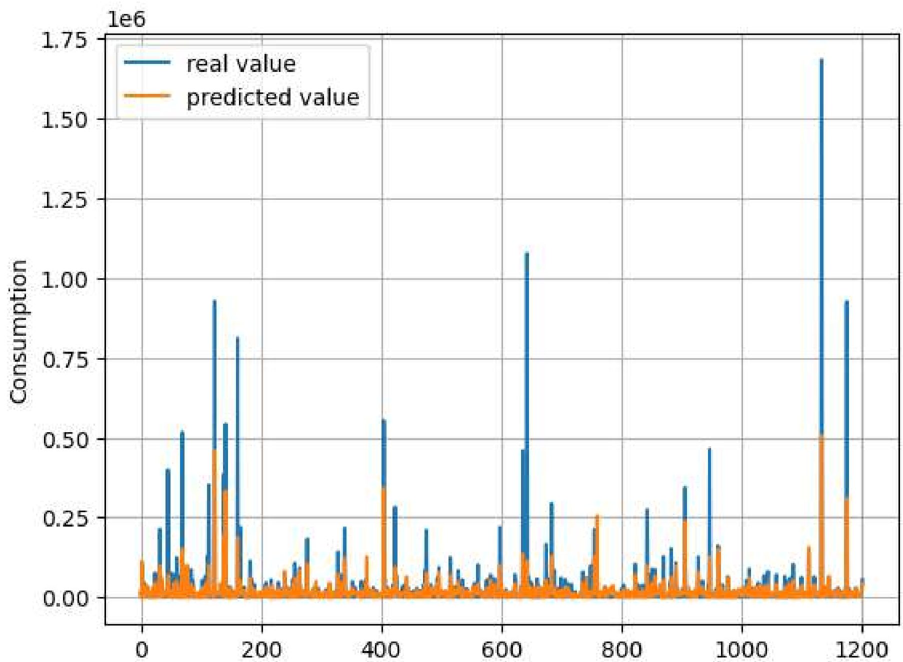

Figure 4.3: MLP on Train and validation(Mascara)

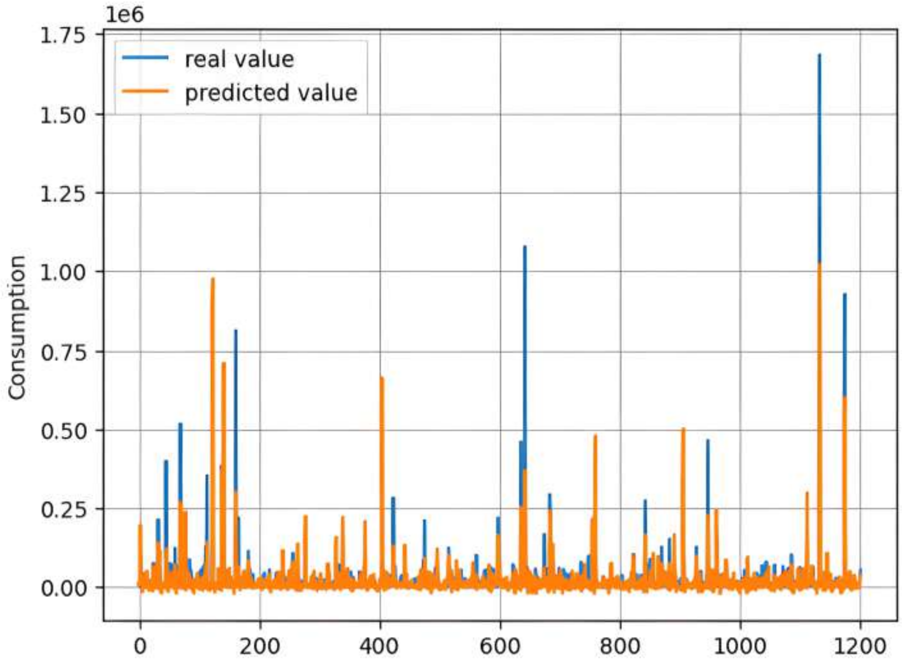

Figure 4.4: CNN on Train and validation(Mascara)

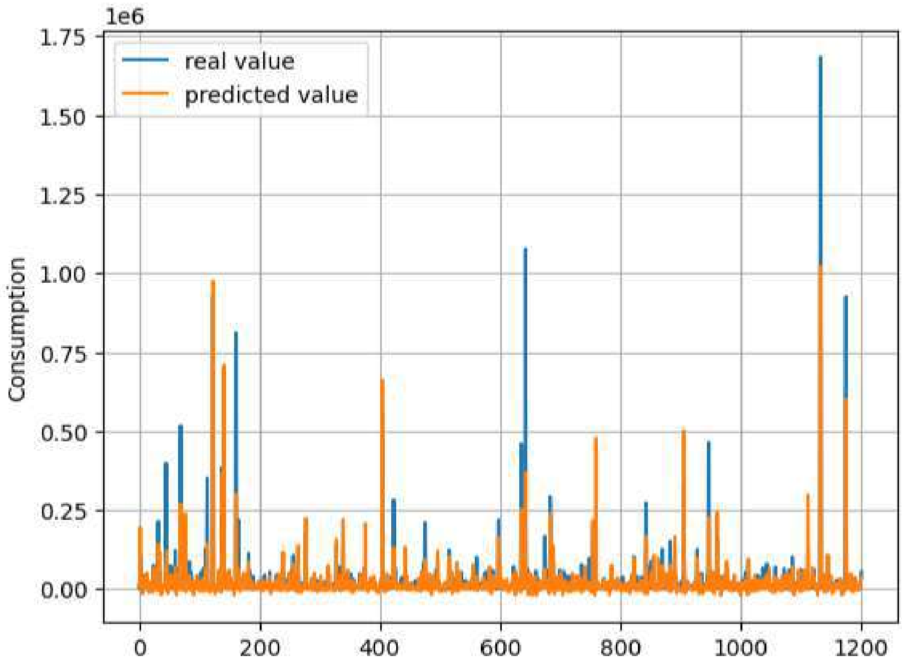

Figure 4.5: LSTM on Train and validation(Mascara)

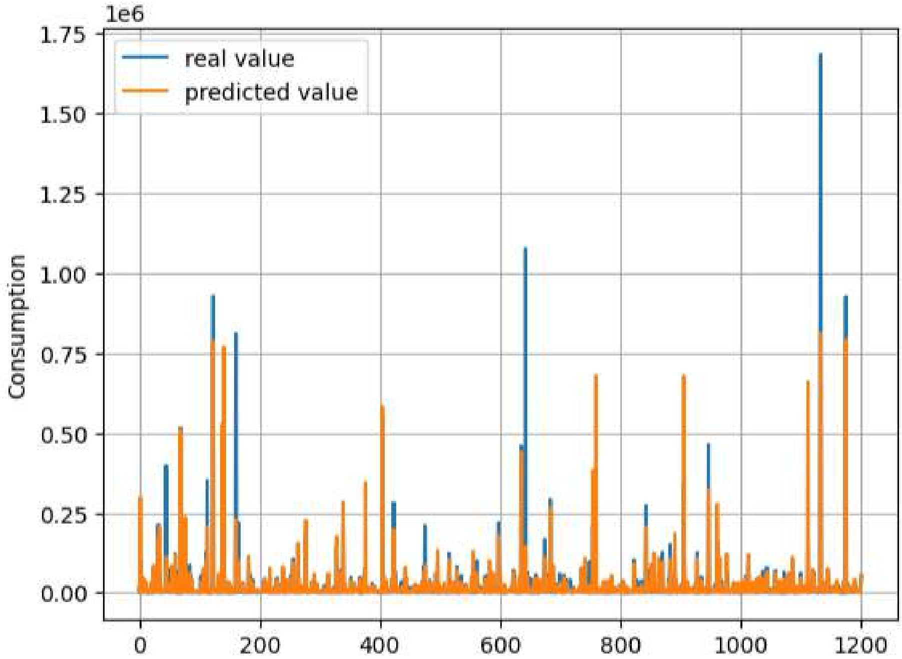

Figure 4.6: CNN-LSTM on Train and validation(Mascara)

4.2.2.A Sidi Bel Abbes

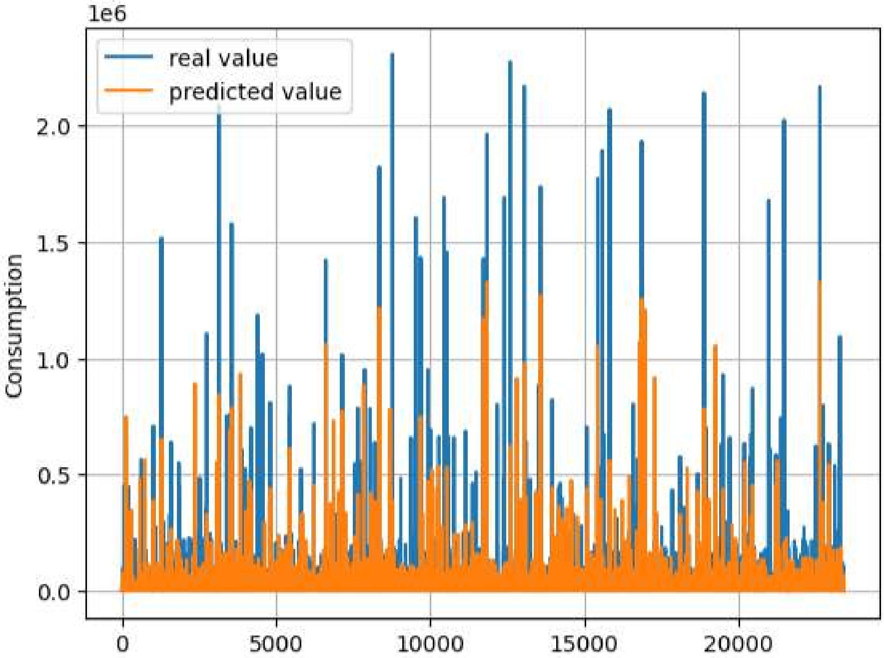

Figure 4.7: MLP on Train and validation(sidi Bel Abbes)

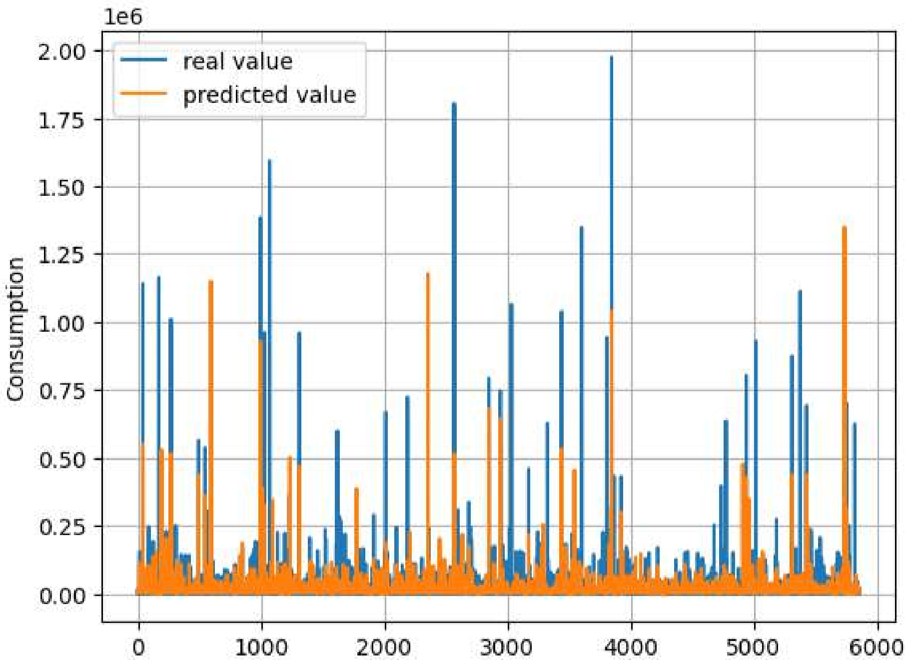

Figure 4.8: CNN on Train and validation(Sidi Bel Abbes)

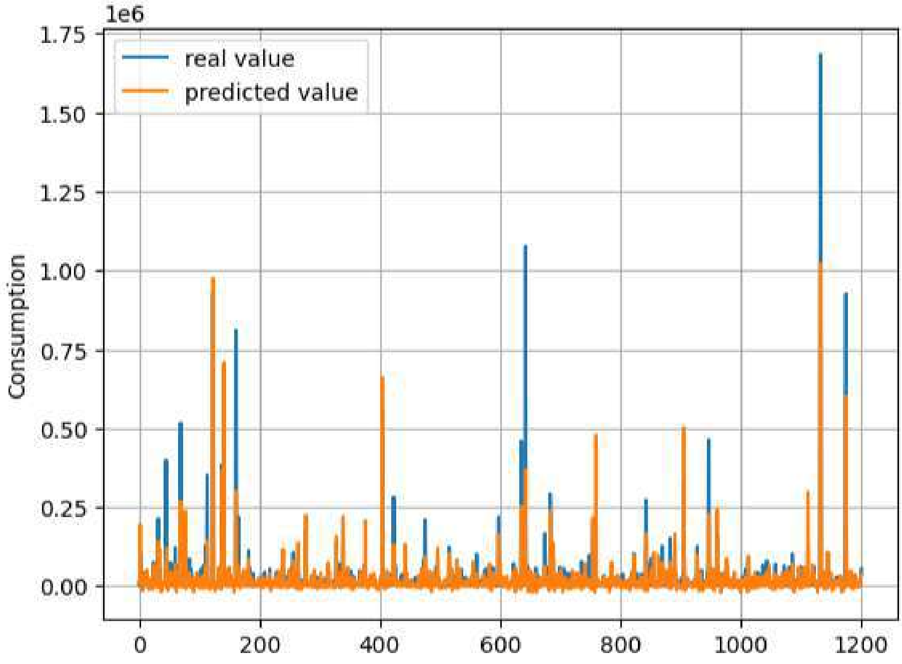

Figure 4.9: LSTM on Train and validation(Sidi Bel Abbes)

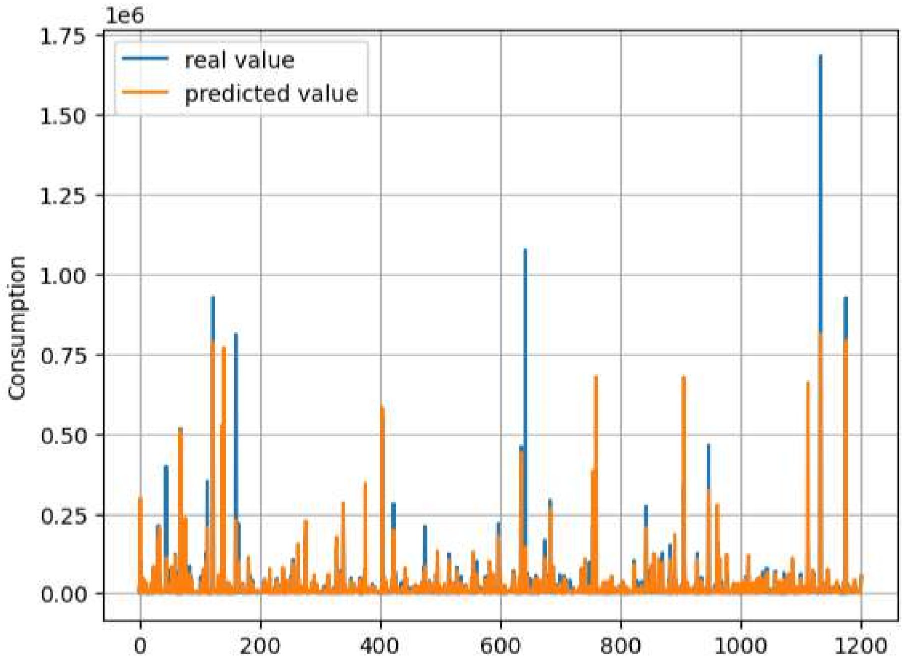

Figure 4.10: CNN-LSTM on Train and validation(Sidi Bel Abbes)

Discussion: In Sidi Bel Abbes, the performance of the four models (MLP, LSTM, CNN) during the validation phase is not particularly satisfying, as observed from the figures. There is a notable discrepancy between the real values and the predicted values, indicating a lack of accuracy in these models. However, it is worth noting that the CNN-LSTM model stands out with significantly better results during the validation phase.

The subpar performance of the models can be attributed to two main factors: noisy data and the absence of client codes. The presence of noise in the data introduces uncertainties and inconsistencies, making it challenging for the models to capture meaningful patterns accurately. Additionally, the missing client codes prevent proper alignment and identification of individual clients, further contributing to the inaccurate predictions.

While the validation phase results may not be satisfactory, there is hope that the models will exhibit improved performance during the testing phase. It is possible that the models might generalize better and provide more accurate predictions when faced with unseen data. The testing phase offers an opportunity to evaluate the models' capabilities in a real-world scenario and assess their potential to deliver better results.

In summary, the performance of the MLP, LSTM, and CNN models in the validation phase in Sidi Bel Abbes falls short of expectations, as indicated by the noticeable disparities between real and predicted values in the figures. However, the CNN-LSTM model demonstrates more promising results. The shortcomings can be attributed to noisy data and the absence of client codes. Moving forward, it is hoped that the testing phase will yield better outcomes, allowing for a more comprehensive evaluation of the models' performance and their ability to provide accurate predictions.

4.2.3 Testing

4.2.3.A Mascara

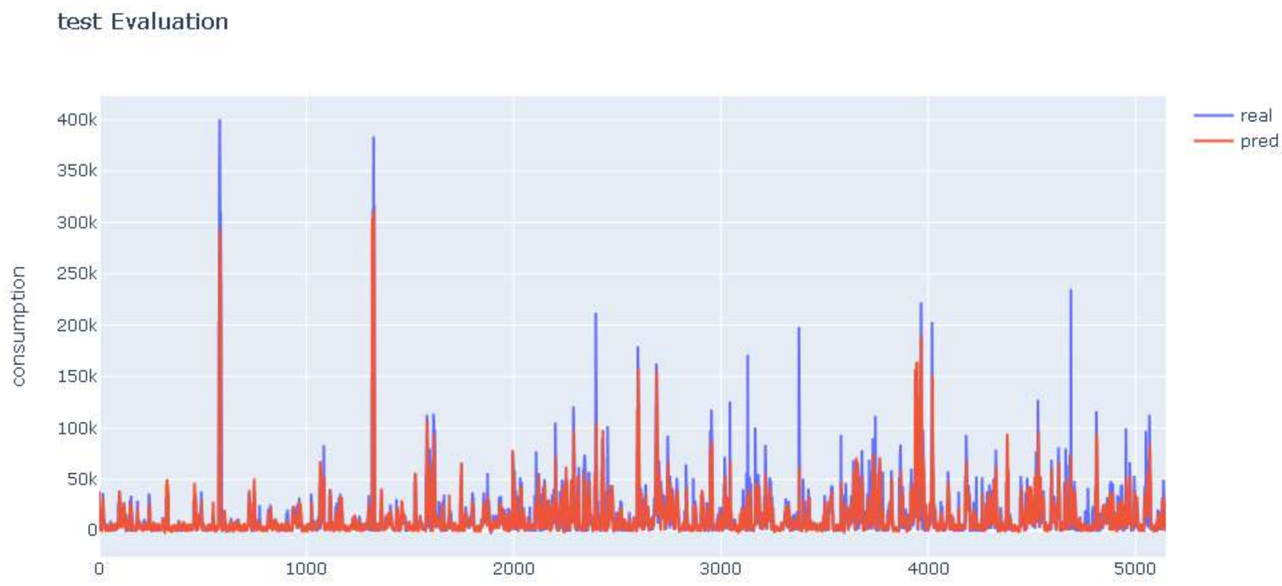

Figure 4.11: MLP Testing(Mascara)

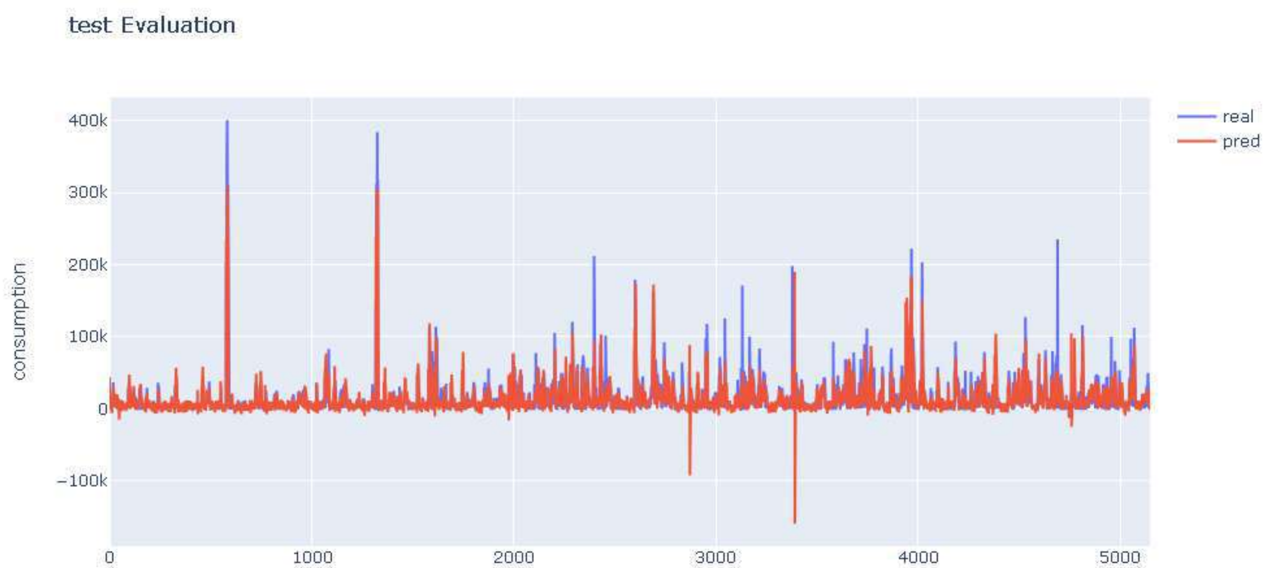

Figure 4.12: CNN Testing(Mascara)

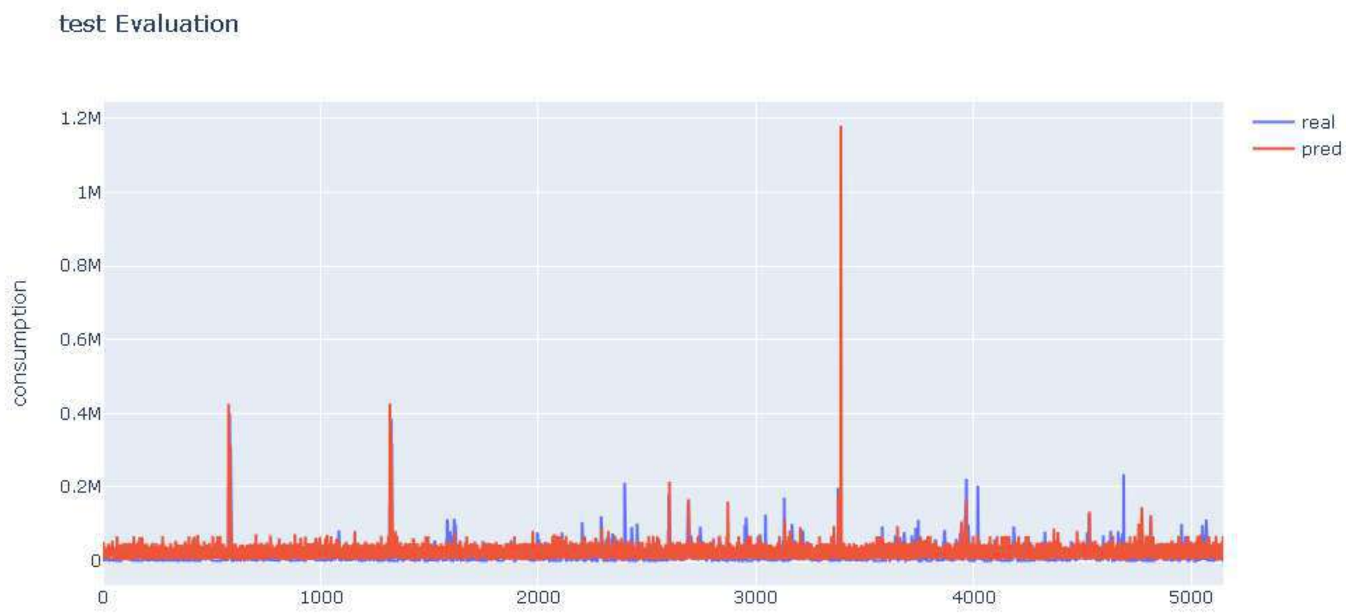

Figure 4.13: LSTM Testing(Mascara)

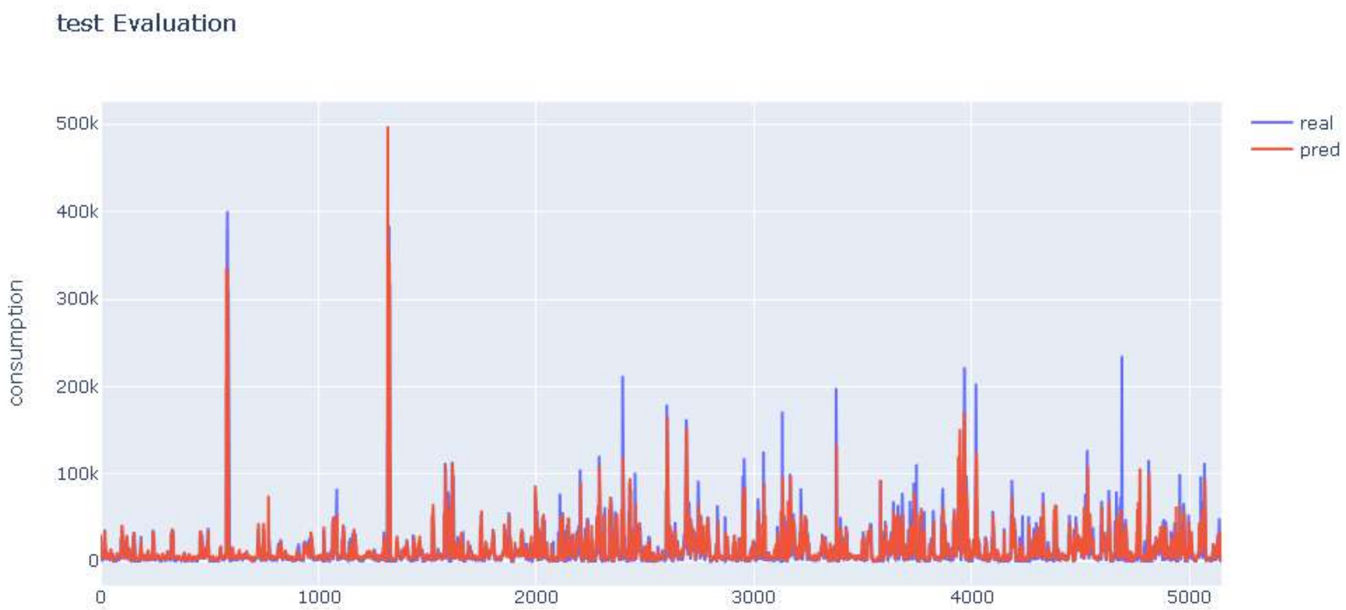

Figure 4.14: CNN-LSTM Testing(Mascara)

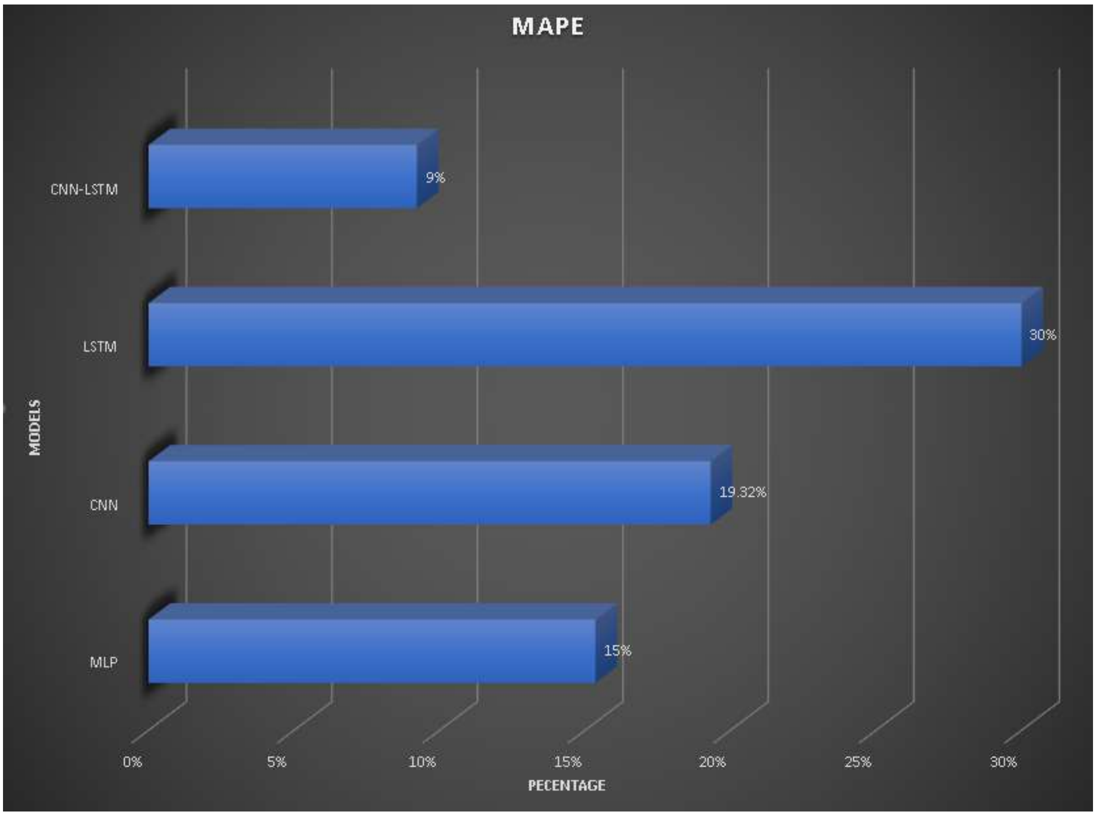

Figure 4.15: MAPE (Mascara)

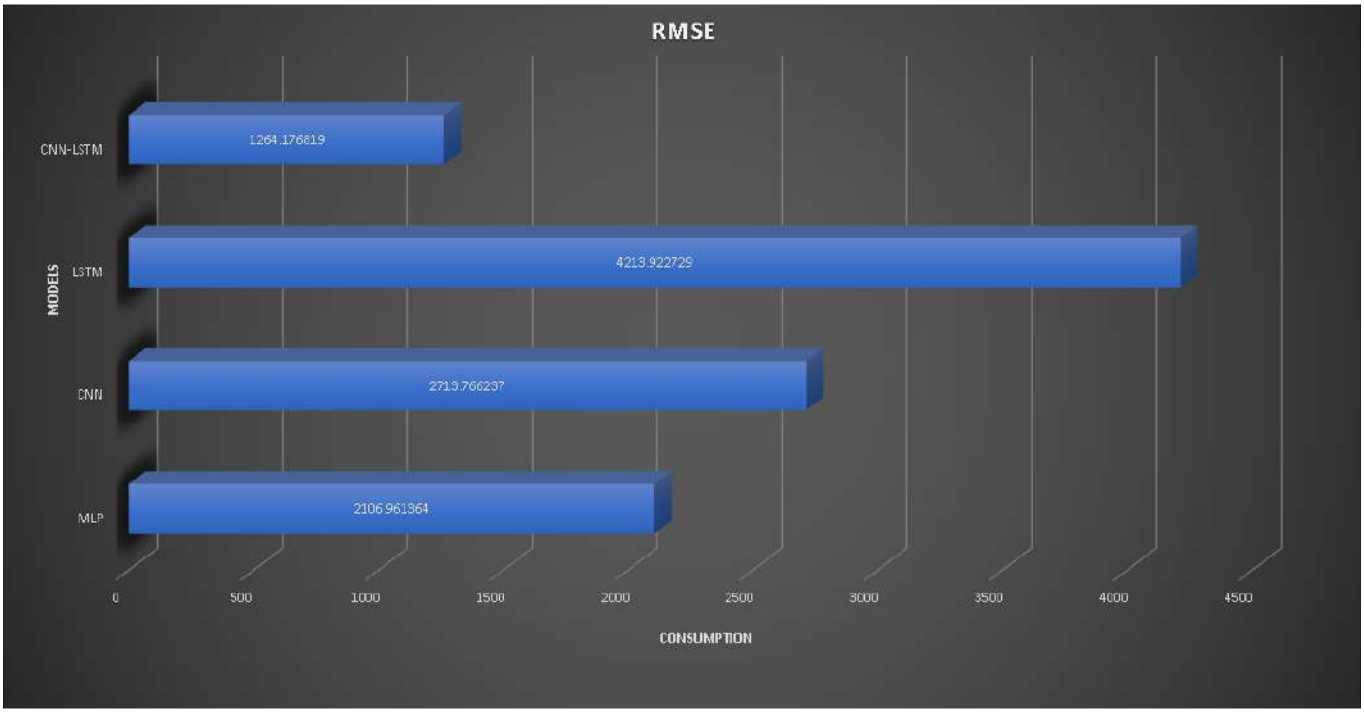

Figure 4.16: RMSE (Mascara)

Discussion: The figures illustrate the performance of different models, namely MLP, CNN, LSTM, and CNN-LSTM, during the testing phase using unseen data. The obtained results are highly satisfying, considering the limited amount of available data. Due to the constraints of our study, we only had access to a small dataset spanning 12 months, which was further divided for training, testing, and validation purposes. As a result, we were left with only 9 months of data to train our models. Additionally, it is worth noting that the mean absolute percentage error (MAPE) values were relatively high for all models. This can be attributed to the abundance of zero values in the total active power feature. Unfortunately, we could not exclude these instances from the dataset, as doing so would require eliminating customers who exhibited zero consumption in any given month. Consequently, the dataset further diminished in size.

Considering these limitations, we express our optimism about the future availability of a more extensive dataset. Collaborating with Sonalgaz company, we aim to collect data for at least five more years. With a larger and more diverse dataset, we can expect improved model performance and a more comprehensive understanding of customer behavior.

The additional data would provide an opportunity to train the models on a more representative sample, leading to enhanced accuracy and generalization capabilities. It would enable us to capture a wider range of consumption patterns and better address the issue of zero values in the data.

In summary, despite the constraints imposed by a limited dataset and the presence of zero values, the results obtained from the testing phase are highly promising. We anticipate that future endeavors will benefit from increased data availability, allowing for more robust and accurate modeling of customer behavior in the energy consumption domain.

4.2.3.B Sidi Bel Abbes

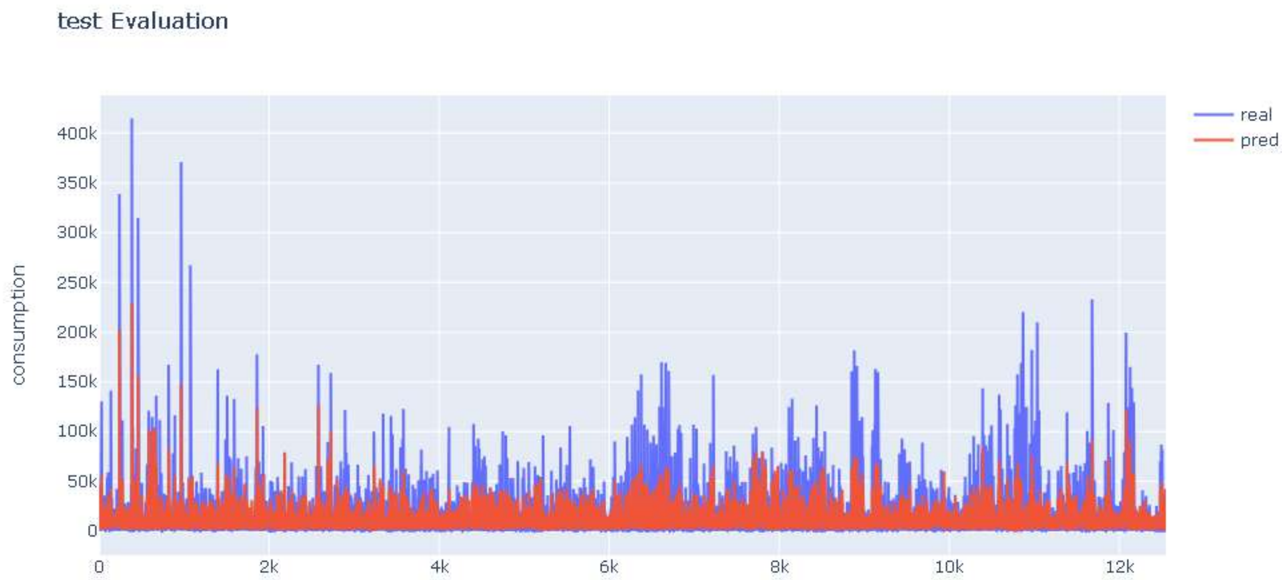

Figure 4.17: MLP Testing(Sidi Bel Abbes)

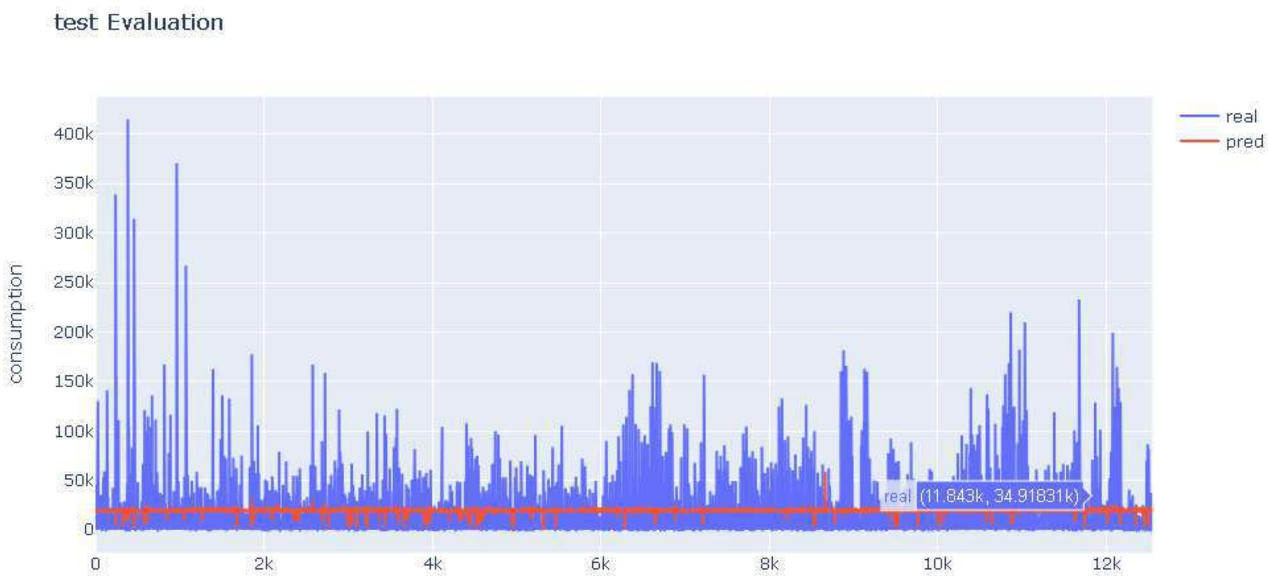

Figure 4.18: CNN Testing(Sidi Bel Abbes)

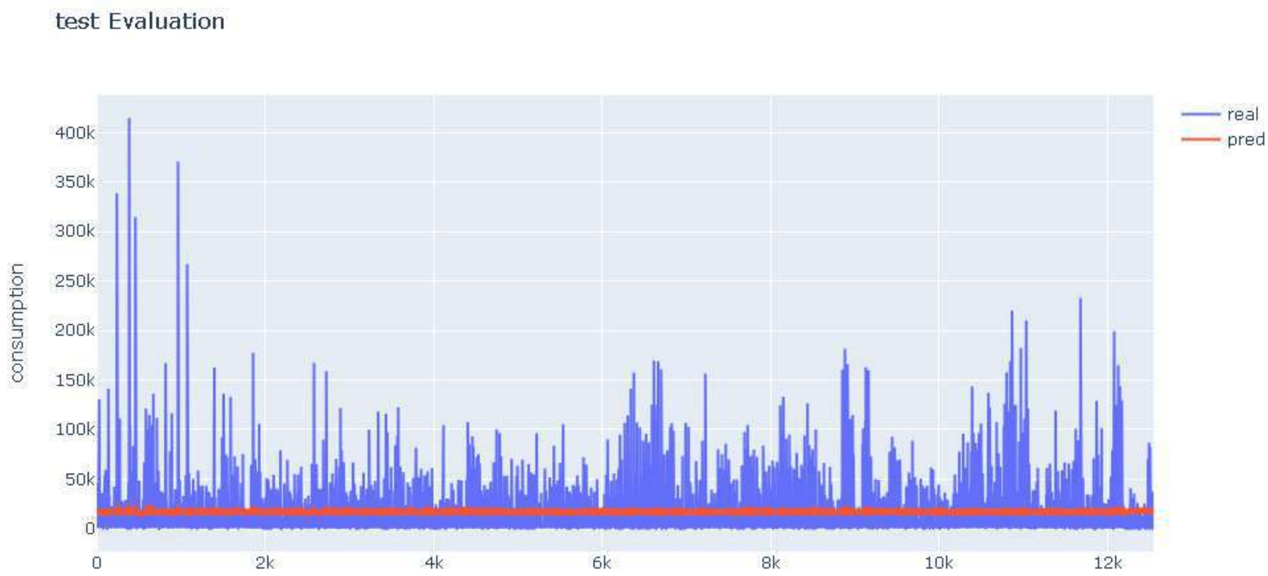

Figure 4.19: LSTM Testing(Sidi Bel Abbes)

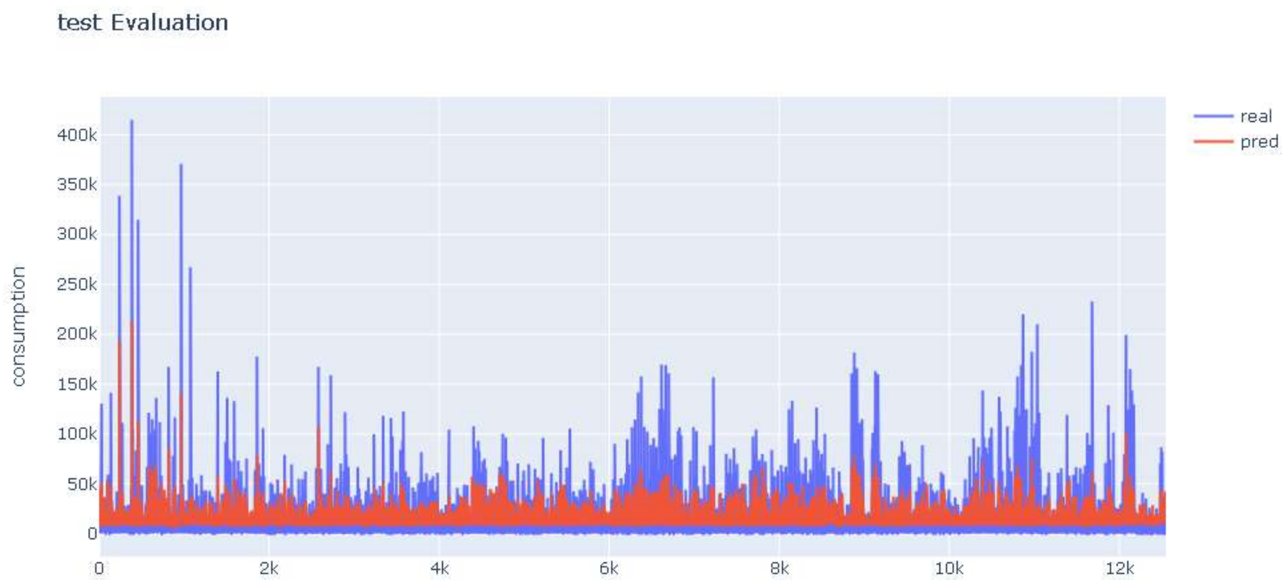

Figure 4.20: CNN-LSTM Testing(Sidi Bel Abbes)

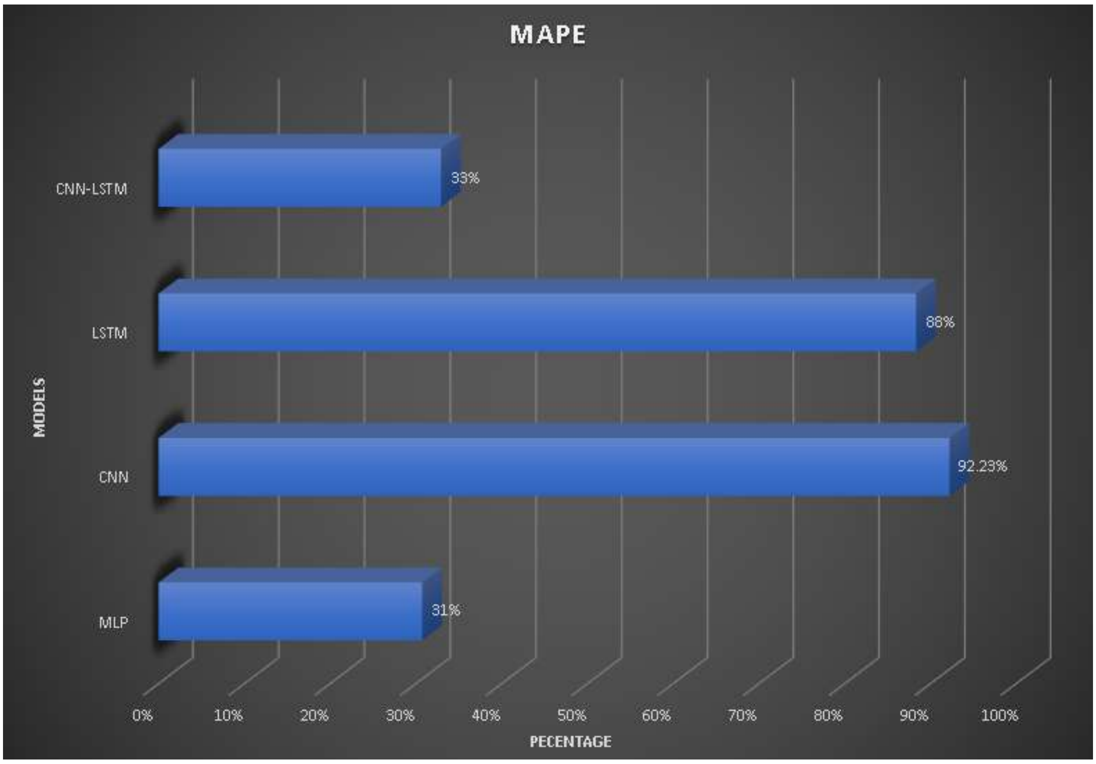

Figure 4.21: MAPE (Sidi Bel Abbes)

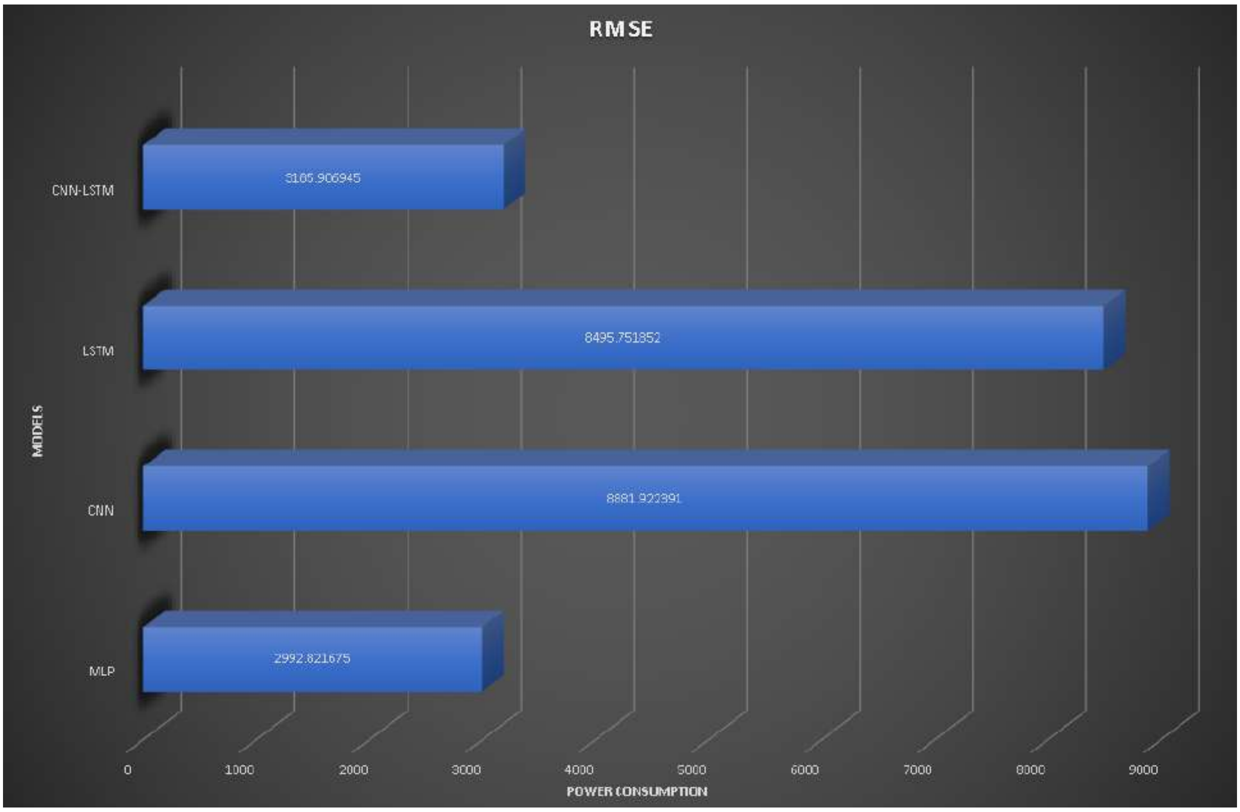

Figure 4.22: RMSE (Sidi Bel Abbes)

Discussion: In Sidi Bel Abbes, the figures demonstrate the performance of different models (MLP, CNN, LSTM, and CNN-LSTM) during the testing phase using unseen data. Unfortunately, the obtained results show low accuracy, as indicated by the high values of mean absolute percentage error (MAPE) and mean squared error (MSE). This poor performance can be attributed to a crucial missing piece of information: the client code.

The absence of the client code prevents us from identifying individual clients for each month. As a result, the data becomes noisy and lacks proper alignment. This discrepancy between the data from Mascara and Sidi Bel Abbes is evident in the figures. The lack of accurate client identification introduces significant inconsistencies, rendering the data unrealistic.

Despite having a three-year consumption dataset from 2020 to 2023, the models (MLP, CNN, LSTM, and CNN-LSTM) struggle to learn effectively, even when trained with a high number of epochs (e.g., 5000). Instead of improving performance, this excessive training leads to overfitting. The models fail to generalize and perform poorly when presented with unseen data, as evident from the unsatisfactory results observed in the validation phase.

To obtain near-perfect results in the future, it is crucial to obtain the missing client code. With the inclusion of this essential information, the models can align the data accurately and make meaningful predictions. Having access to individual client codes would greatly enhance the models' ability to learn patterns and predict future consumption accurately.

In summary, the performance of the models (MLP, CNN, LSTM, and CNN-LSTM) during the testing phase in Sidi Bel Abbes is hindered by the absence of the client code. The resulting noisy and unrealistic data prevents effective learning, even with a significant dataset spanning three years. Acquiring the client code in the future would significantly improve the accuracy of predictions and enable the models to perform near-perfectly.

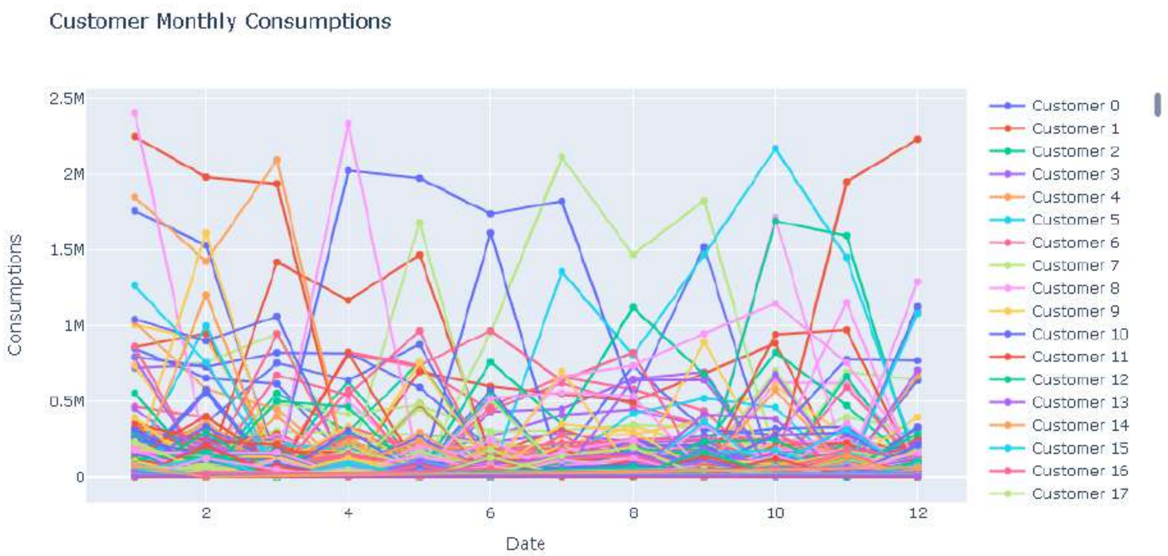

Figure 4.23: Monthly Consumption by Customer(Sidi Bel Abbes)

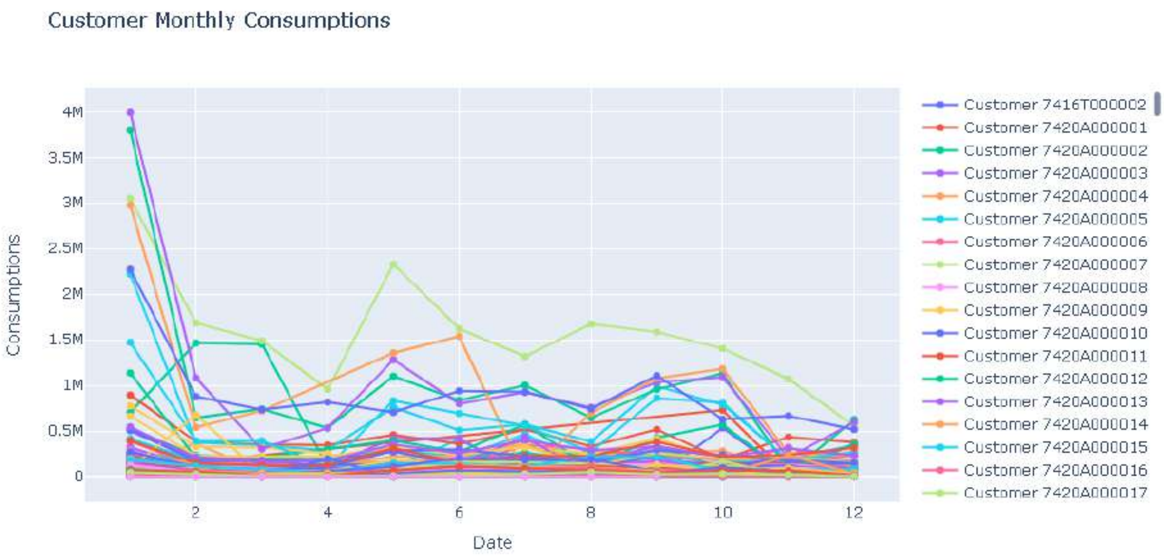

Figure 4.24: Monthly Consumption by Customer(Mascara)

General Conclusion

In conclusion, this thesis has focused on developing and evaluating deep learning models, including Convolutional Neural Networks (CNNs), Multilayer Perceptron (MLP), Long Short-Term Memory (LSTM), and CNN-LSTM, for accurate electricity consumption prediction. Through an extensive exploration and analysis of these models, along with the utilization of historical electricity consumption data, significant advancements have been made in achieving precise and reliable predictions.

The findings of this research demonstrate the effectiveness of CNNs in capturing spatial dependencies within the electricity consumption data. By exploiting the local patterns and relationships among neighboring data points, CNNs have shown remarkable performance in accurately predicting electricity consumption at various spatial scales.

Additionally, MLP models have been employed to capture nonlinear relationships and intricate dependencies present in the electricity consumption time series data. These models have successfully learned complex patterns and exhibited strong predictive capabilities, contributing to improved accuracy in forecasting electricity consumption.

Furthermore, LSTM models have been utilized to model the temporal dynamics and long-term dependencies inherent in electricity consumption data. The memory cells and gating mechanisms of LSTM allow for capturing and retaining relevant information from past observations, enabling accurate predictions even over extended time horizons.

The combined CNN-LSTM approach has also been investigated, leveraging the strengths of both CNNs and LSTMs. This hybrid architecture has shown promising results, as it can effectively capture both spatial and temporal dependencies, leading to enhanced predictive performance.

The outcomes of this thesis have demonstrated that deep learning models, including CNNs, MLPs, LSTMs, and CNN-LSTMs, offer significant improvements in accurately predicting electricity consumption. These models have surpassed traditional forecasting methods by leveraging their ability to learn intricate patterns and relationships within the data.

The application of these deep learning models for accurate electricity consumption prediction holds considerable value for the energy sector. It enables stakeholders, such as utility companies, energy planners, and policymakers, to make informed decisions regarding load management, demand planning, and resource allocation. The precise predictions obtained through these models contribute to optimizing energy operations, reducing costs, and fostering sustainable energy practices.

However, certain challenges and opportunities for future research exist. These include further exploration of model optimization techniques, handling outliers and anomalies, integrating real-time data, and addressing interpretability concerns associated with complex deep learning architectures.

In conclusion, this thesis has made significant contributions to accurate electricity consumption prediction using deep learning models, including CNNs, MLPs, LSTMs, and CNN-LSTMs. The findings highlight the potential of these models in improving forecasting accuracy, which in turn enhances energy management, efficiency, and sustainability. These advancements serve as a foundation for future research and development in the field, driving towards a smarter and more efficient energy future.

4.3 Achievements

Figure 4.25: Paper Submission

Accurate Electricity Consumption Prediction Using Deep Learning: A Systematic Review

Abbes Bougueffa Eutamene, Nour Elhouda Ben Saadi, Boudjela Houari, Laid Khetache, Ayad Ahmed Nour El Islem, Larouci Benyekhlef

Abstract: The exponential growth in the human population and technological advancements has dramatically increased the power demand. In this context, electricity is being used at the same time as it is produced at the power plant. Therefore, effective forecasting of energy usage is crucial for maintaining a reliable power supply. By evaluating what is observed, AI can predict energy usage, reducing it during peak hours, for example. problems such as bottlenecks can be identified even before they happen. In this study we aim to propose a deep learning model to accurately predict electricity consumption in advance. For this final in this systematic review, we summarized the different found works in the literature and revealed the gaps to finally propose a new experiments.

Keywords: Energy forecasting, Deep Learning, AI technologies, AI challenges

1 Introduction

This is an important point to consider when it comes to electricity production and consumption. When the amount of electricity generated exceeds the required amount, there is a risk of wasted resources and unnecessary costs.
To prevent this, it is important to ensure that the amount of electricity generated matches up with the demand in real-time, in order to avoid overproduction or underproduction The distribution subsidiary orders electricity from the generation subsidiary, and then the energy produced is distributed to customers through their electrical grids. This ensures that customers receive reliable and efficient energy while providing a cost-effective solution for both the generation and distribution subsidiaries,Overproduction or production losses can be defined as the difference between the electricity produced and the actual electricity distributed. A better prediction of customers' consumption can help reduce these errors and minimize losses due to overproduction. This is beneficial for companies, as an overproduction of energy that is not distributed constitutes a dead loss for them
At the level of distribution subsidiaries, electricity losses are a major concern for the company. To control their impact, they have implemented a strategy to ensure that the amount of electricity purchased is equal to the amount billed. This is an ideal goal for them as it will help minimize losses and increase efficiency
Predictive models for accurately forecasting electricity consumption can be an important tool in helping to plan and monitor electricity usage in the economic sector. These models can help to identify potential issues before they become problems, as well as provide insight into how different factors may affect future consumption. By understanding past patterns and trends, these models can also provide a more accurate assessment of future needs, allowing for better decision making when it comes to resource allocation.

2 Related Works Comparison

Recent advances in electricity consumption forecasting have enabled researchers to develop sophisticated methods for predicting future energy demand. These approaches include machine learning models such as support vector machines, neural networks, and ensemble methods; statistical models such as autoregressive integrated moving average (ARIMA) models; and

Bibliography

[1] "Algeria energy information," *Enerdata*, vol. 11, no. 5, 2021. [Online]. Available: https://www.enerdata.net/estore/energy-market/algeria/

[2] R. V. Klyuev, I. D. Morgoev, A. D. Morgoeva, O. A. Gavrina, N. V. Martyushev, E. A. Efremenkov, and Q. Mengxu, "Methods of forecasting electric energy consumption: A literature review," *Energies*, vol. 15, no. 23, 2022. [Online]. Available: https://www.mdpi.com/1996-1073/15/23/8919

[3] A. Ahmad, N. Javaid, A. Mateen, M. Awais, and Z. A. Khan, "Short-term load forecasting in smart grids: An intelligent modular approach," *Energies*, vol. 12, no. 1, 2019. [Online]. Available: https://www.mdpi.com/1996-1073/12/1/164

[4] G. Shmueli and O. Koppius, "The challenge of prediction in information systems research," *SSRN Electronic Journal*, 04 2009.

[5] A. Jierula, S. Wang, T.-M. OH, and P. Wang, "Study on accuracy metrics for evaluating the predictions of damage locations in deep piles using artificial neural networks with acoustic emission data," *Applied Sciences*, vol. 11, no. 5, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/5/2314

[6] D. Hadjout, J. Torres, A. Troncoso, A. Sebaa, and F. Martínez-Álvarez, "Electricity consumption forecasting based on ensemble deep learning with application to the algerian market," *Energy*, vol. 243, p. 123060, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544221033090

[7] B. J. Jaewon Chung, “Accurate prediction of electricity consumption using a hybrid cnn-lstm model based on multivariable data,” *PLOS ONE*. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0278071

[8] T. Le, M. T. Vo, B. Vo, E. Hwang, S. Rho, and S. W. Baik, “Improving electric energy consumption prediction using cnn and bi-lstm,” *Applied Sciences*, vol. 9, no. 20, 2019. [Online]. Available: https://www.mdpi.com/2076-3417/9/20/4237

[9] W. He, “Load forecasting via deep neural networks,” *Procedia Computer Science*, vol. 122, pp. 308–314, 2017, 5th International Conference on Information Technology and Quantitative Management, ITQM 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050917326170

[10] F. Kaytez, “A hybrid approach based on autoregressive integrated moving average and least-square support vector machine for long-term forecasting of net electricity consumption,” *Energy*, vol. 197, p. 117200, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544220303078

[11] R. K. Jana, I. Ghosh, and M. K. Sanyal, “A granular deep learning approach for predicting energy consumption,” *Applied Soft Computing*, vol. 89, p. 106091, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494620300314

[12] F. Amara, K. Agbossou, Y. Dubé, S. Kelouwani, A. Cardenas, and S. S. Hosseini, “A residual load modeling approach for household short-term load forecasting application,” *Energy and Buildings*, vol. 187, pp. 132–143, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378778818309228

[13] C. Liu, B. Sun, C. Zhang, and F. Li, “A hybrid prediction model for residential electricity consumption using holt-winters and extreme learning machine,” *Applied Energy*, vol. 275, p. 115383, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261920308953

[14] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using cnn-lstm neural networks," *Energy*, vol. 182, pp. 72–81, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544219311223

[15] J. Kim, J. Moon, E. Hwang, and P. Kang, "Recurrent inception convolution neural network for multi short-term load forecasting," *Energy and Buildings*, vol. 194, pp. 328–341, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378778819308072

[16] S. Mujeeb, N. Javaid, M. Ilahi, Z. Wadud, F. Ishmanov, and M. K. Afzal, "Deep long short-term memory: A new price and load forecasting scheme for big data in smart cities," *Sustainability*, vol. 11, no. 4, 2019. [Online]. Available: https://www.mdpi.com/2071-1050/11/4/987

[17] Y.-S. Park and S. Lek, "Chapter 7 - artificial neural networks: Multilayer perceptron for ecological modeling," in *Ecological Model Types*, ser. Developments in Environmental Modelling, S. E. Jørgensen, Ed. Elsevier, 2016, vol. 28, pp. 123–140. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780444636232000074

[18] S. Lek and Y. Park, "Artificial neural networks," in *Encyclopedia of Ecology*, S. E. Jørgensen and B. D. Fath, Eds. Oxford: Academic Press, 2008, pp. 237–245. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780080454054001737

[19] A. Malekian and N. Chitsaz, "Chapter 4 - concepts, procedures, and applications of artificial neural network models in streamflow forecasting," in *Advances in Streamflow Forecasting*, P. Sharma and D. Machiwal, Eds. Elsevier, 2021, pp. 115–147. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128206737000032

[20] I. Sarker, "Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions," *springer*, vol. 11, no. 5, 2021. [Online]. Available: https://link.springer.com/article/10.1007/s42979-021-00815-1

[21] A. NDIFE, A. N. Ndife, and Y. Mensin, "Development of smart on-device based electricity consumption forecast model for energy management," *Naresuan University*, vol. 11, no. 5, 2022. [Online]. Available: http://nuir.lib.nu.ac.th/dspace/handle/123456789/5193

[22] N. M. D. R. e. a. Yamashita, R., "Convolutional neural networks: an overview and application in radiology," *springer*, vol. 11, 2018. [Online]. Available: https://insightsimaging.springeropen.com/articles/10.1007/s13244-018-0639-9#citeas

[23] F. Aksan, Y. Li, V. Suresh, and P. Janik, "Cnn-lstm vs. lstm-cnn to predict power flow direction: A case study of the high-voltage subnet of northeast germany," *Sensors*, vol. 23, no. 2, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/2/901

[24] S. Abinaya and M. K. Devi, "Chapter 12 - enhancing crop productivity through autoencoder-based disease detection and context-aware remedy recommendation system," in *Application of Machine Learning in Agriculture*, M. A. Khan, R. Khan, and M. A. Ansari, Eds. Academic Press, 2022, pp. 239–262. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B978032390550300014X

[25] A. Singhal and D. K. Sharma, "Chapter 3 - voice signal-based disease diagnosis using iot and learning algorithms for healthcare," in *Implementation of Smart Healthcare Systems using AI, IoT, and Blockchain*, ser. Intelligent Data-Centric Systems, C. Chakraborty, S. K. Pani, M. Abdul Ahad, and Q. Xin, Eds. Academic Press, 2023, pp. 59–81. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780323919166000059

[26] M. D. Mohanty and M. N. Mohanty, "Chapter 5 - verbal sentiment analysis and detection using recurrent neural network," in *Advanced Data*

Mining Tools and Methods for Social Computing, ser. Hybrid Computational Intelligence for Pattern Analysis, S. De, S. Dey, S. Bhattacharyya, and S. Bhatia, Eds. Academic Press, 2022, pp. 85–106. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780323857086000126

[27] Y. Chen, L. Li, W. Li, Q. Guo, Z. Du, and Z. Xu, "Chapter 3 - deep learning," in *AI Computing Systems*, Y. Chen, L. Li, W. Li, Q. Guo, Z. Du, and Z. Xu, Eds. Morgan Kaufmann, 2024, pp. 53–121. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780323953993000093

[28] G. Patra, K. R. Bhimala, A. Marndi, S. Chowdhury, J. Rahaman, S. Nandi, R. R. Sarkar, K. Gouda, K. Ramesh, R. P. Barnwal, S. Raj, and A. Saini, "Chapter 6 - deep learning methods for scientific and industrial research," in *Deep Learning*, ser. Handbook of Statistics, V. Govindaraju, A. S. Srinivasa Rao, and C. Rao, Eds. Elsevier, 2023, vol. 48, pp. 107–168. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169716122000578

[29] K. Liu and J. Zhang, "A dual-layer attention-based lstm network for fed-batch fermentation process modelling," in *31st European Symposium on Computer Aided Process Engineering*, ser. Computer Aided Chemical Engineering, M. Türkay and R. Gani, Eds. Elsevier, 2021, vol. 50, pp. 541–547. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780323885065500863

[30] S. M. M. H. I. H. Bilgera C, Yamamoto A, "Application of convolutional long short-term memory neural networks to signals collected from a sensor network for autonomous gas source localization in outdoor environments." National Library of Medicine, 2018, vol. 50, pp. 541–547. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6308690/

[31] T. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," 04 2015, pp. 4580–4584.

[32] T. Liu, B. Jinsong, J. Wang, and Y. Zhang, "A hybrid cnn–lstm algorithm for online defect recognition of co2 welding," *Sensors*, vol. 18, p. 4369, 12 2018.

[33] A. FILIPPELLI, C. C. Lin, L. Anderson, and J. McConkey, "Principles and methods for measurement of electron impact excitation cross sections for atoms and molecules by optical techniques," in *Cross Section Data*, ser. Advances In Atomic, Molecular, and Optical Physics, M. Inokuti, Ed. Academic Press, 1994, vol. 33, pp. 1–62. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1049250X08600324