People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
University of 8 Mai 1945Guelma

Faculty of Mathematics, Computer Science and Material Sciences.
Department Computer Science
Domiciliation laboratory Information and Communication Sciences and Technologies Laboratory (LabSTIC)

Thesis

Submitted in Candidacy for the Degree of *Doctorate in Third Cycle*

Field: Computer Science Stream: Mathematics and Computer Science
Speciality: Artificial Intelligence

Presented by:
Farah Rania

Title

Prediction of electric charge with deep learning models

Defended on : …………………......   Before the jury composed of:

Full name	**Rank**	**University**	
Mr NEMISSI Mohamed	Professor	Univ. of Guelma	President
Mr SERIDI Hamid	Professor	Univ. of Guelma	Supervisor
Mr FAROU Brahim	Professor	Univ. of Guelma	Co- supervisor
Mr KHADIR Tarek	Professor	Univ. of Annaba	Examiner
Mme ZEDADRA Ouarda	MCA	Univ. of Guelma	Examiner
Mr KOUAHLA Zineddine	Professor	Univ. of Guelma	Invited

Academic year: 2023/2024

Acknowledgement

I wish to extend my heartfelt gratitude to the individuals who played pivotal roles in bringing this work to fruition.

Foremost, I extend my sincere appreciation to Professor Hamid Seridi. His expertise, insightful guidance, and unwavering support have been invaluable throughout this project. His ability to inspire excellence has served as a constant beacon of motivation.

I also wish to express my deep gratitude to Professor Brahim Farou for his attentive mentorship and enlightening guidance, which significantly enriched this research. His dedication to the advancement of scientific inquiry has been a profound source of encouragement.

A special acknowledgment is due to Professor Zineddine Kouahla for his calm and reassuring presence, as well as his consistent encouragement. His supportive demeanor has fostered a positive and conducive working environment.

I am profoundly grateful to my father, my guiding light and role model, Professor Nadir Farah, for his unwavering support, wisdom, and inspirational guidance. His counsel and encouragement have been invaluable throughout my journey.

To my mother, I express deep appreciation for her unwavering support and boundless patience.

I extend heartfelt thanks to my friends, particularly Zahia, Lina, and Rania, and their families, for their unwavering friendship, steadfast support, and constant encouragement. Their presence has been a source of immeasurable strength throughout this endeavor.

Finally, I express my gratitude to all who have contributed, directly or indirectly, to the development of this work through their guidance, expertise, and support.

I am profoundly grateful for their invaluable collaboration and unwavering support throughout this academic odyssey.

Table of contents

List of figures

Abstract

Electricity forecasting is a strategic asset for optimizing the management of power grids, promoting the integration of renewable energies, reducing operating costs, and enhancing the efficient and sustainable use of electrical energy.

In order to satisfy demand in real time, energy network operators need to plan the production and distribution of electricity. This efficient operational planning is made possible by electricity load forecasting, which predicts the evolution of demand on different time scales (daily, weekly, and seasonal).

The thesis aims to improve intelligent energy management by providing an in-depth study and notable advances in the field of electrical load modeling. New approaches to the study form its foundation, and the application of deep learning techniques to reliably predict changes in electrical load is given particular attention.

The research work revolves around the evaluation and improvement of existing models, implementing deep neural network architectures such as convolutional neural networks (CNNs), and models with embedded attention mechanisms (Transformers). The use of complex datasets with rich temporal information content has allowed to capture the temporal dependencies inherent in electrical charge profiles.

Firstly, one of the contributions of this research consists of a rigorous methodical approach aimed at acquiring a thorough understanding of the database provided to represent by the Algerian electricity production company. An in-depth statistical study of over 10 years of electricity consumption was undertaken, analyzing in detail the characteristics and trends inherent in the dataset. This provided crucial insights into the distribution of core variables, potential correlations, and temporal patterns, laying the foundations for a comprehensive understanding of the underlying context.

Subsequently, the research focused on exploring different machine learning and deep learning architectures. Several approaches were tested and evaluated, with the emphasis on selecting the most appropriate models for a specific challenge. This step involved the exploration of various architectures, such as classical machine learning models like Support Vector Machine and Linear Regression as well as more sophisticated deep learning architectures, including convolutional neural networks (CNNs), Transformers, and other models.

The final contribution of this research is the design and implementation of an innovative system capable of dynamically switching to renewable energy sources, thus helping to reduce CO_2 emissions during power generation. This system intelligently integrates the use of gas, offering a cleaner, more ecological alternative while protecting human health. This approach is based on the implementation of an automatic switching mechanism between different energy sources, exploiting the advantages of renewable energies when CO_2 emissions have reached a critical threshold during electricity production. This strategy aims to minimize the overall carbon emissions footprint of power generation, responding to environmental imperatives and growing concerns about climate change.

These contributions converge to create a more efficient energy system. On the socio-economic plan, rigorous planning helps reduce costs, while minimizing energy losses. Enhanced planning gives power producers' greater control over electricity distribution and promotes more efficient management. In addition, in environmental terms, these advances contribute to a greener future by reducing pollutant emissions, thus working towards an atmosphere less impacted by pollution.

Résumé

La prévision de l'électricité est un atout stratégique pour optimiser la gestion des réseaux électriques, promouvoir l'intégration des énergies renouvelables, réduire les coûts d'exploitation et améliorer l'utilisation efficace et durable de l'énergie électrique. Afin de satisfaire la demande en temps réel, les opérateurs de réseaux énergétiques doivent planifier la production et la distribution d'électricité. Cette planification opérationnelle efficace est rendue possible par la prévision de la charge électrique, qui prédit l'évolution de la demande sur différentes échelles de temps (quotidienne, hebdomadaire et saisonnière).

La thèse vise à améliorer la gestion intelligente de l'énergie en fournissant une étude approfondie et des avancées notables dans le domaine de la modélisation de la charge électrique. De nouvelles approches d'étude forment sa base, et l'application de techniques d'apprentissage profond pour prédire de manière fiable les changements de charge électrique est particulièrement soulignée.

Les travaux de recherche portent sur l'évaluation et l'amélioration des modèles existants, la mise en œuvre d'architectures de réseaux neuronaux profonds telles que les réseaux neuronaux convolutifs (CNN) et les modèles avec des mécanismes d'attention intégrés (Transformers). L'utilisation d'ensembles de données complexes avec un contenu d'information temporelle riche a permis de capturer les dépendances temporelles inhérentes aux profils de charge électrique.

Tout d'abord, l'une des contributions de cette recherche consiste en une approche méthodique rigoureuse visant à acquérir une compréhension approfondie de la base de données fournie par la société algérienne de production d'électricité. Une étude statistique approfondie de plus de 10 ans de consommation d'électricité a été entreprise, analysant en détail les caractéristiques et tendances inhérentes à l'ensemble de données. Cela a fourni des aperçus cruciaux sur la distribution des variables principales, les corrélations potentielles et les tendances temporelles, posant les bases d'une compréhension globale du contexte sous-jacent.

Par la suite, la recherche s'est concentrée sur l'exploration de différentes architectures d'apprentissage automatique et d'apprentissage profond. Plusieurs approches ont été testées et évaluées, en mettant l'accent sur la sélection des modèles les plus appropriés pour un défi spécifique. Cette étape impliquait l'exploration de diverses architectures, telles que les modèles d'apprentissage automatique classiques comme la machine à vecteurs de support et la régression linéaire ainsi que des architectures d'apprentissage profond plus sophistiquées, y compris les réseaux neuronaux convolutifs (CNN), les Transformers et d'autres modèles.

La dernière contribution de cette recherche est la conception et la mise en œuvre d'un système innovant capable de passer dynamiquement à des sources d'énergie renouvelables, aidant ainsi à réduire les émissions de CO_2 lors de la production d'électricité. Ce système intègre intelligemment l'utilisation du gaz, offrant une alternative plus propre et plus écologique tout en protégeant la santé humaine. Cette approche est basée sur la mise en œuvre d'un mécanisme de commutation automatique entre différentes sources d'énergie,

exploitant les avantages des énergies renouvelables lorsque les émissions de CO2 ont atteint un seuil critique lors de la production d'électricité. Cette stratégie vise à minimiser l'empreinte carbone globale de la production d'électricité, répondant aux impératifs environnementaux et aux préoccupations croissantes concernant le changement climatique.

Ces contributions convergent pour créer un système énergétique plus efficace. Sur le plan socio-économique, une planification rigoureuse aide à réduire les coûts, tout en minimisant les pertes d'énergie. Une planification améliorée donne aux producteurs d'électricité un plus grand contrôle sur la distribution d'électricité et favorise une gestion plus efficace. De plus, sur le plan environnemental, ces avancées contribuent à un avenir plus vert en réduisant les émissions de polluants, travaillant ainsi vers une atmosphère moins impactée par la pollution.

ملخص

تعد التنبؤ بالكهرباء أصل استراتيجي لتحسين إدارة شبكات الطاقة، وتعزيز التكامل للطاقات المتجددة، وتقليل التكاليف التشغيلية، وتعزيز الاستخدام الكفء والمستدام للطاقة الكهربائية.

من أجل تلبية الطلب في الوقت الحقيقي، يحتاج مشغلو شبكات الطاقة إلى تخطيط إنتاج وتوزيع الكهرباء. يتمكن هذا التخطيط التشغيلي الفعّال من خلال توقع الحمولة الكهربائية، الذي يتنبأ بتطور الطلب على مختلف المدى الزمني (اليومي، الأسبوعي، والموسمي).

تهدف الرسالة إلى تحسين إدارة الطاقة الذكية من خلال دراسة عميقة وتقدم ملحوظ في مجال تصميم الحمولة الكهربائية. تشكل النهجات الجديدة للدراسة أساسها، ويولى اهتمام خاص لتطبيق تقنيات التعلم العميق للتنبؤ بتغيرات الحمولة الكهربائية بشكل موثوق.

تدور أعمال البحث حول تقييم وتحسين النماذج الحالية، وتنفيذ تصاميم لشبكات عصبية عميقة مثل الشبكات العصبية التكرارية ، والنماذج ذات آليات انتباه مضمنة .وقد سمح استخدام مجموعات بيانات معقدة ذات محتوى معلومات زمنية غني بالتقاط الاعتمادات الزمنية الكامنة في ملامح الشحن الكهربائي.

أولاً، تتضمن إحدى المساهمات في هذا البحث نهجًا منهجيًا دقيقًا يهدف إلى فهم شامل لقاعدة البيانات المقدمة لتمثيلها من قبل شركة إنتاج الكهرباء الجزائرية. تم إجراء دراسة إحصائية عميقة لأكثر من 10 سنوات من استهلاك الكهرباء، محللة بالتفصيل السمات والاتجاهات الكامنة في مجموعة البيانات. وقدمت هذه الدراسة رؤى حاسمة حول توزيع المتغيرات الأساسية، والارتباطات المحتملة، والأنماط الزمنية، مما أرست الأسس لفهم شامل للسياق الكامن.

ثم، ركز البحث على استكشاف مختلف تصاميم التعلم الآلي والتعلم العميق. تم اختبار وتقييم عدة نهج، مع التركيز على اختيار النماذج الأكثر ملاءمة للتحدي المحدد. تضمنت هذه الخطوة استكشاف مجموعة متنوعة من التصاميم، مثل النماذج التقليدية للتعلم الآلي مثل آلة الدعم الناقل والانحدار الخطي، بالإضافة إلى تصاميم التعلم العميق الأكثر تطورًا، بما في ذلك الشبكات العصبية التكرارية(CNN) ، والمحولات، وغيرها من النماذج.

تتمثل المساهمة النهائية لهذا البحث في تصميم وتنفيذ نظام مبتكر قادر على التحول ديناميكيًا إلى مصادر الطاقة المتجددة، مما يساعد في تقليل انبعاثات ثاني أكسيد الكربون أثناء توليد الطاقة. يدمج هذا النظام بذكاء استخدام الغاز، مقدمًا بديلاً نظيفًا وأكثر بيئية بينما يحمي صحة الإنسان. تعتمد هذه الطريقة على تنفيذ آلية تحويل تلقائي بين مصادر الطاقة المختلفة، استغلال فوائد الطاقات المتجددة عندما تصل انبعاثات ثاني أكسيد الكربون إلى عتبة

حرجة أثناء إنتاج الكهرباء. تهدف هذه الاستراتيجية إلى تقليل البصمة الكربونية العامة لتوليد الطاقة، ردًا على الضروريات البيئية والمخاوف المتزايدة بشأن التغيرات المناخية.

تتقاطع هذه المساهمات لإنشاء نظام طاقة أكثر كفاءة. من الناحية الاجتماعية والاقتصادية، يساعد التخطيط الدقيق على تقليل التكاليف، مع تقليل الخسائر الطاقية. يمنح التخطيط المحسن المنتجين الكهرباء مزيدًا من السيطرة على توزيع الكهرباء ويعزز الإدارة الأكثر كفاءة. بالإضافة إلى ذلك، من الناحية البيئية، تساهم هذه التقدمات في مستقبل أخضر من خلال تقليل انبعاثات الملوثات، مع العمل نحو جو أقل تأثرًا بالتلوث ..

Introduction

General Context

Modern life is powered by electrical energy, which powers many technologies as well as residences and industries. For this reason, it is imperative to rethink the strategy of production and management of this precious resource, which must be reconsidered in the light of the growing demand for electricity without causing loss or damage.

With the ability to improve operational efficiency, limit environmental impact, and enable more intelligent planning, electrical load prediction is becoming increasingly important in this context. Energy prediction is emerging as an innovative area, essential to satisfying the growing needs of an expanding society.

The rise of the digital revolution has had a significant impact on consumer habits, contributing to a growing demand for electricity. The explosion of digital devices, electronic appliances, and connected technologies has not only enriched our daily lives but has also resulted in a growing dependence on electricity.

The interaction of people with their environment has changed dramatically with the development of cell phones, computers, connected devices, and internet services. Electricity is now more in demand due to the energy requirements of these technologies. New factors have been added to daily energy consumption by habits such as video streaming, remote working, home robots, and constant connectivity.

Consequently, electricity consumption has increased significantly, boosted by the constant evolution of the digital revolution, so it becomes imperative to understand and anticipate these changes, also to develop advanced artificial intelligence methods. These technological advances offer a unique opportunity for energy producers, enabling them to maintain precise control over distribution while optimizing production to best respond to the diverse needs of the population and guarantee an electricity supply adapted to the new demands of the digital society.

Artificial intelligence is thus positioning itself as an essential catalyst for tackling the complex challenges of continually growing energy demand in an efficient and sustainable way.

Cost reduction is the socioeconomic advantage of these intelligent methods since accurate planning helps to minimize the costs associated with electricity generation, distribution and storage. It also helps avoid the costs associated with peak demand and the penalties associated with inefficient resource management.

Electricity prediction is a strategic tool for optimizing the management of electricity networks. The knowledge of the future peak consumption periods will enable a better distribution management and more efficient adjustment of energy production. This promotes more balanced demand management on the grid.

Moreover, anticipating variations in demand helps to maintain the stability of the power supply network. This helps avoid overproduction, which will result in energy loss, and underproduction, which will lead to potential blackouts, thus ensuring a reliable power supply.

Motivation

The motivation for this work comes from the recognition of the crucial challenges facing the energy sector, especially in light of the expanding demand and the need to create sustainable solutions. The constant increase in energy requirements, fueled by the growing use of electronic devices and modern technologies, calls for a proactive approach to anticipating and responding effectively to these demands.

The fundamental motivation is derived from the conviction that energy prediction, in conjunction with cutting-edge artificial intelligence techniques, offers an innovative and practical approach to addressing these challenges.

The application of these methods will be of considerable help to companies in the energy sector, who must not only satisfy the traditional needs of their customers, but also anticipate and adapt to new expectations, climate variations and population growth.

The objective of the work is to contribute to the development of intelligent and adaptive solutions to enable predictive management of electrical load by understanding the various issues related to energy production and distribution.

This will enable a more efficient electricity planning, optimize resource availability and reduce environmental impact, contributing to a greener, more sustainable energy future. The energy landscape that will result from all these advances will be able to both satisfy the demands of modern civilization and protect resources for generations to come.

Problematic

The challenge of this study relates to the management of a voluminous quantity of data from the company SONELGAZ. In light of this challenge, it is essential to develop rigorous methods for understanding, evaluating, and extracting valuable information from these data. To do this, thorough investigation and statistical analysis are required to unravel the complexity of the data provided.

Research Objectives and Contributions

In pursuit of this endeavor, this study aims to address the following inquiries:

How can we approach this mass of data efficiently to extract relevant information? How can we perform an in-depth study that discerns trends, correlations and seasonal peculiarities within consumer profiles?

The next challenge is to compare different models for electrical load prediction. **How can artificial intelligence methods be effectively integrated into electrical load prediction to optimize energy production and distribution in a sustainable way?**

And finally, **how can we design a system capable of switching between different energy sources, including renewables, to meet demand while minimizing CO2 emissions and promoting a cleaner environment?**

Therefore, it is essential to find creative answers to these problems, paying particular attention to the development of reliable predictive models, efficient energy management, and the transition to more ecologically responsible energy systems. The above questions form the unifying theme of this study and guide the development of practical, adaptable solutions.

The Algerian electricity and gas company (SONELGAZ) provided ten years of authentic data for this thesis, which were thoroughly tested, examined and studied to determine the factors influencing the electricity consumption of the country. To target and describe different electricity consumption profiles, the aim is to build models and provide essential information.

Particular attention will be focused on the development of short-term forecasting models. Several approaches will be developed, presented and compared with existing advanced electrical load forecasting techniques.

Thesis Roadmap

This thesis is structured into four chapters, each contributing to the in-depth understanding and improvement of electricity load forecasting techniques.

The first chapter introduces the background and main objectives of the thesis, highlighting the growing importance of electricity demand forecasting in the contemporary energy landscape.

The second chapter offers an in-depth introduction to artificial intelligence (AI) and machine learning, outlining the different types of learning. A literature review on energy demand forecasting is presented, highlighting the importance of such prediction. Emphasis is then placed on the use of machine learning and deep learning techniques, with a distinction between statistical methods, machine learning and those based on deep learning.

The third chapter is dedicated to a detailed analysis of data from the Algerian Electricity and Gas Company (SONELGAZ). This essential phase involves understanding the characteristics of the data, detecting trends and highlighting the factors influencing national electricity demand.

The fourth chapter forms the core of the thesis contributions, presenting different methods for forecasting electricity load, both machine learning and based on deep learning. The best-performing models are identified after thorough evaluation, and other approaches are introduced and compared with existing methods, with a focus on the development of short-term forecasting models.

Finally, the fifth chapter concludes the thesis by synthesizing all the results and contributions, while suggesting future prospects for research in the field of electricity load forecasting and the integration of digital technologies in the energy sector.

Chapter 1: State of the art

1.1 Introduction

Energy, the vital force that has shaped the evolution of humanity, has followed a remarkable trajectory through the ages. The history of energy is one of continuous expansion and transformation, from the use of fire by our ancestors to the current age, when industrial advances and digital technologies have driven energy demand to historically unprecedented levels.

Throughout the decades, this exponential growth has given rise to complex challenges, particularly in terms of efficient resource management and energy demand forecasting. It is in this context that time series have become an essential tool for projecting future energy needs due to their ability to identify trends and patterns in time series data.

The increasing use of electricity in homes, businesses and transport has transformed the way we relate to energy. Nowadays, seasonal variations, economic developments and technological advances all have a complex impact on consumption patterns. This is the situation in which time series analysis is beneficial.

Exploiting time series, we can examine past patterns of energy consumption, identifying seasonal trends, hourly variations, and recurring cycles. Time series enable consumption patterns to be modeled and forecast with greater accuracy, making it easier to make informed decisions to satisfy the growing needs of an ever-changing world.

Consequently, considerable research has been invested in the study of energy prediction using time series, with several objectives explored. The frameworks for many approaches to predicting trends, seasonal fluctuations and future progress in energy consumption have been established by this research, which is based on the field of time series applied to energy.

Significant advances in artificial intelligence (AI) models have led to a considerable evolution in the development of prediction in many fields. Among all the sectors benefiting from these developments, electricity load forecasting is emerging as a crucial area for efficient power grid management. The implementation of AI models in this field is receiving increased attention due to the growing need to accurately forecast energy consumption in order to ensure smooth distribution, anticipate fluctuations and manage production resources more efficiently for companies (Makridakis & Hyndman, 1974).

The growing demand for energy, resulting from the continuous expansion of technologies and the needs of modern society, makes it imperative to develop robust and intelligent methods for anticipating electricity consumption patterns. Artificial intelligence (AI) models, including neural networks, transformers and clustering approaches, are becoming highly effective in tackling this challenging subject. This will enable better management of energy resources and more efficient planning. In this chapter, the aim is to present several forecasting models widely used in the electricity industry.

Recent years have seen remarkable advances that have ushered in a new era in which Artificial Intelligence (AI) plays a key role in creating content and solving complex problems, touching on diverse fields such as pattern recognition, early disease detection, medical image analysis, and personalized treatment.

Significant progress has been made in the field of AI, illustrating its major impact in various sectors. Large Language Models (LLMs), such as GPT-3 (Brownet, 2020) and GPT-4 (OpenAI, 2023) developed by Open AI, stand out for their exceptional text generation and language comprehension capabilities, revolutionizing content creation and giving rise to renowned chat bots.

At the same time, models such as DALL-E (Aditya Ramesh & al., 2021) have transcended the boundaries between text and image, enabling the creation of high-resolution images and illustrations from natural language descriptions. These advances have considerably enriched the field of artistic and visual creation.

In the field of computer vision, the Efficient Net and ViT (Vision Transformer) (Testagrose & al., 2023) architectures bear witness to ongoing progress in this area. These models stand out for their exceptional performance in crucial tasks such as image classification and object detection, opening up new prospects for the application of AI in a variety of contexts.

Artificial intelligence (AI) has seen remarkable advances, particularly in the field of video games, underlining its ability to solve complex problems. A prominent example is Alpha Go (Silver & al, 2017), an AI capable of mastering the game of Go and outperforming champions, illustrating AI's depth in game strategy. Similarly, Dota 2 (Berner & al, 2019*),* highlighted exploits in video gaming, where AI systems managed to triumph over world e-sport champions, showcasing the adaptability of AI to dynamic, competitive environments.

Furthermore, Alpha Fold (Jumper, & Hassabis, 2022) marked a major achievement in revolutionizing the prediction of 3D protein structure models. This breakthrough has significant implications in the field of molecular biology, enabling a deeper understanding of protein structure and opening the way to new research possibilities and medical applications. These achievements reflect the constant evolution of AI and its significant impact in diverse fields, transcending the traditional boundaries of technology.

The development of AI models for medical image analysis is a surprising example in the medical field. Healthcare professionals can benefit from the effective use of deep neural networks to detect disease from radiological images (Sim & al., 2020).

The detection of certain forms of cancer from medical imaging scans is a concrete example of this advance. AI models, such as those based on convolutional neural networks (CNNs), have demonstrated an exceptional ability to detect subtle anomalies in images, enabling more effective early detection and intervention (Abiyev, & Ma'aitaH, 2018).

Surgical procedures have been revolutionized by AI-assisted surgical robots (Hussein & al, 2023), which offer exceptional dexterity and precision. These robots can be remotely supervised by surgeons, enabling a more precise and less invasive procedure.

The advantages of intelligent surgical robots include improved visualization, the ability to perform complex movements with great precision, and the potential to reduce scarring and recovery time for patients (Thai & al., 2020). In fields such as cardiac surgery, oncology surgery and other specialized interventions, these technologies have been used successfully.

1.2 Artificial Intelligence

1.2.1 Regression

A statistical model is a crucial mathematical tool for conceptualizing and explaining the intrinsic relationships between variables in a data set. They are used to make predictions, draw conclusions and make informed decisions, mainly in the fields of statistics, data science and machine learning. Complex issues, such as the dynamics of economic growth, market trends and medical results, are particularly appropriate for these models.

Various types of statistical models are available, covering specific categories such as regression models, time series models, Bayesian models, and many others. These structures are deployed to solve a variety of problems, including the prediction of future values, the identification of determining factors, and the estimation of causal relationships, among other applications.

ARIMA

ARIMA, which means "Auto Regressive Integrated Moving Average", is a class of statistical models widely used to model and forecast time series. ARIMA models combine the components of auto regression (AR), integration (I), and moving average (MA). Here is a detailed explanation of each component

Linear Regression

Linear regression is a statistical technique designed to model the linear relationship between a dependent variable (also known as the response variable) Y and one or more independent variables (also known as the explanatory variables or predictors) X (Nasteski, 2017). The main objective is to find the best straight line (called the "regression line") that best represents the relationship between the variables like shown in the figure 1.

A simple linear regression model can be formulated as follows in eq.1:

$$\boldsymbol{Y = \beta 0 + \beta 1 X + \varepsilon}, \tag{1}$$

Where Y represents the dependent variable, $\beta 0$ and $\beta 1$ are the coefficients associated with the intercept and slope respectively, X is the independent variable, and ε is a random error.

The estimate of Y for each value of X is equivalent to the point on the straight line defined by the equation eq.2:

$$\boldsymbol{E(Y|X) = \beta 0 + \beta 1 X}. \tag{2}$$

The underlying hypothesis implies that, for each value of X, the expected error ε is zero $(E(\varepsilon) = 0)$ and the variance of the errors ε is constant $(V(\varepsilon) = \sigma^2)$. Moreover, the errors ε follow a normal distribution, $\varepsilon \sim N(0, \sigma^2)$, and are independent of each other.

The objectives of linear regression cover estimating the parameters $\beta 0$, $\beta 1$, and σ^2, while checking the model's suitability. Given n pairs of observations $(X1, Y1), (X2, Y2), \ldots, (Xn, Yn)$, the coefficients are determined by the least squares method, minimizing the sum of squared errors as calculated in eq.3:

$$L(\beta 0, \beta 1) = \sum(Yi - \beta 0 - \beta 1Xi)^2. \tag{3}$$

Solving the system of two equations with two unknowns, $\nabla L(\hat{\beta}_0, \hat{\beta}_1) = 0$, yields the optimal estimates of the coefficients β0 and β1.

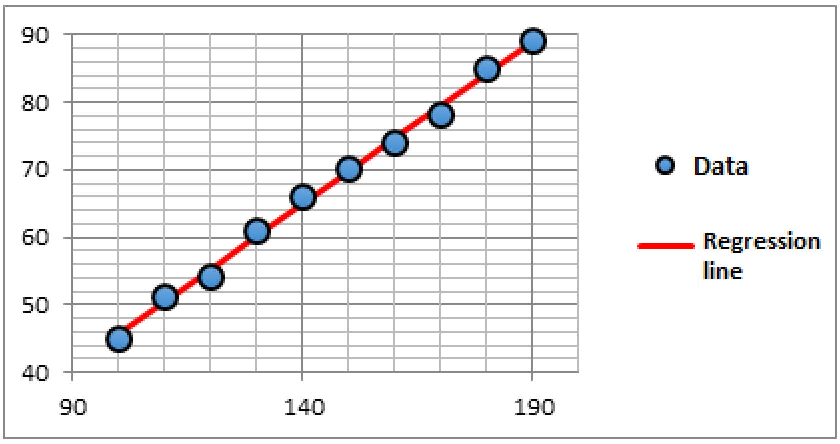


Figure 1. Example of a regression curve on data

Multiple linear regression

The multiple linear regression model is commonly used for the analysis of multidimensional data, constituting a natural extension of the classical linear model (Tranmer & Elliot, 2008). It applies specifically to the study of relationships between a quantitative variable Y, often referred to as the variable to be explained or response, and a set of p quantitative variables $X_1, ..., X_n$, designated as explanatory, control, endogenous or independent variables as shown in Figure 2. This statistical approach offers a powerful methodology for examining and modeling complex relationships involving several variables simultaneously.

The data are assumed to result from the observation of a statistical sample of size *n (n > p +1)* from eq.4:

$$R_{(p+1)} : (x_{1i}, \ldots, x_{ji}, \ldots, x_{pi}, y_i)_{i=1,\ldots,n.} \tag{4}$$

Where *i* varies from 1 to *n*.

In this context, the linear model assumes that the expectation of *Y* belongs to a subspace of R^n, generated by $\{1, X_1, \ldots, X_p\}$ where 1 denotes the vector of R_n consisting of 1. That is, the *(p + 1)* random variables verify in eq.5:

$$Y_i = \beta_0 + \beta 1_{x1i} + \beta 2_{x2i} + \cdots + \beta p_{xpi} + \varepsilon i \tag{5}$$

i = 1, 2,…, n

These relationships are subject to several hypothesis, as follows:

1. The ϵi represent error terms, unobserved, independent and identically distributed, with the equation eq.6:

$$E(\varepsilon i) = 0, Var(\varepsilon) = \sigma 2\, I. \tag{6}$$

2. The terms x j are considered deterministic (controlled factors) or the error ε is independent of the joint distribution of $X_1, \ldots, X_p$. In this case, this translates into the equations eq.7 and eq.8:

$$E\left(Y \mid X_{1,\ldots,Xp}\right) = \beta_0 + \beta 1_{X1} + \beta 2_{X2} + \cdots + \beta p_{Xp} \quad \textbf{(7)}$$

$$Var\left(Y \mid X_1, \ldots, X_p\right) = \sigma 2. \quad \textbf{(8)}$$

3. The unknown parameters $\beta_0, \ldots, \beta_p$ are considered constant.
4. Eventually, to specifically explore the laws of the estimators, a fourth assumption considers the normality of the error variable ε $(N(0, \sigma^2 I))$. The ε_i then follow a normal distribution $N(0, \sigma^2)$. The data are arranged in a matrix X $(n \times (p + 1))$ of general term x_{ji}, whose first column contains vector 1 $(x_{i=0}\ {}_1)$, and in a vector Y of general term y_i.

Writing the vectors $\varepsilon = [\varepsilon 1 - - - \varepsilon p]'$ and $\beta = [\beta 0 \beta 1 - - - \beta p]'$ the model is expressed in matrix form in eq.9:

$$y = X\beta + \varepsilon. \quad \textbf{(9)}$$

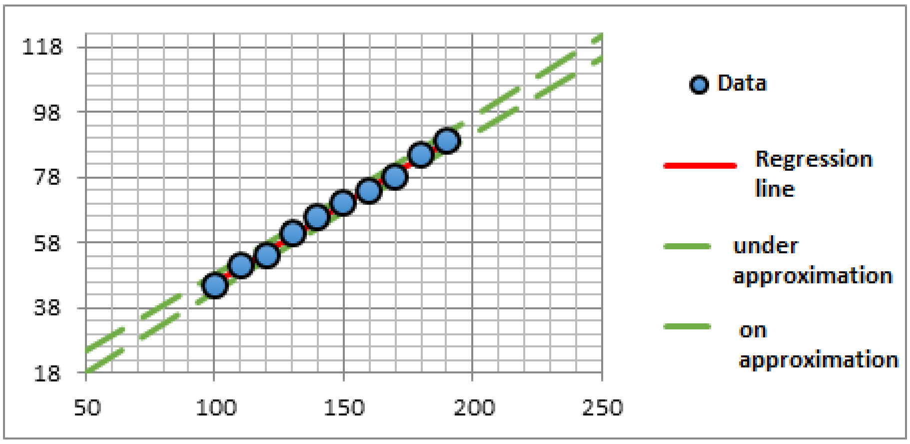


Figure 2. Prediction interval on data

1.2.2 Machine Learning

Machine learning is a field of artificial intelligence focused on creating techniques that enable computer systems to acquire knowledge from data. Instead of explicitly programming instructions, ML models use algorithms to detect patterns in data and perform specific tasks without the need for explicit programming. This is a valuable tool for assisting human activity, whether in the field of classification or regression.

Classification methods

Classification in machine learning is a fundamental domain where models are trained to assign labels or categories to unlabeled data based on labeled examples. The aim is to create a model capable of generalizing and correctly classifying new data.

There are two main types of classification: binary classification, where two distinct classes are defined (e.g. yes/no, cats/dogs), and multiclass classification, which involves more than two categories. Classification has a wide range of applications, including image recognition (identification of objects), fraud detection (identification of suspicious transactions), and medical diagnosis (classification of pathologies from medical images).

An additional field of machine learning is regression, which uses training data to predict continuous values as compared to classifications. It is particularly beneficial for estimating quantities. Regression has been applied in many fields such as financial forecasting (estimating stock prices), weather forecasting, and supply chain optimization by predicting product demand.

The main difference between regression and classification lies in the type of predicted outcome variable. In classification, the model anticipates a label or categorical class for each observation, whereas in regression, the aim is to predict a continuous numerical value. This fundamental difference guides the specific application of these two approaches in machine learning. In essence, classification is applied when the task is to assign specific categories or classes, while regression is employed to estimate numerical quantities.

Support Vector Machine

Support vector machines are a class of supervised learning algorithms used for both classification and regression. They were developed by Vladimir Vapnik and his team in the 1990s (Veisi, 2023). They represent a range of machine learning algorithms for resolving classification (Awad & al., 2015), regression (Honghai & al., 2005) and even anomaly detection problems (Hosseinzadeh & al. 2021). Their reputation is built on robust theoretical guarantees, extensive flexibility and ease of use, making their application accessible even to individuals without deep knowledge of data mining.

As shown in the Figure 3, the principle of support vector machines (SVMs) is elementary: the aim is to classify data into classes using a boundary that is as simple as possible. The aim is to maximize the distance between the different groups of data and the boundary separating them, known as the margin. SVMs are therefore described as wide-margin separators, and the data closest to this boundary are called support vectors.

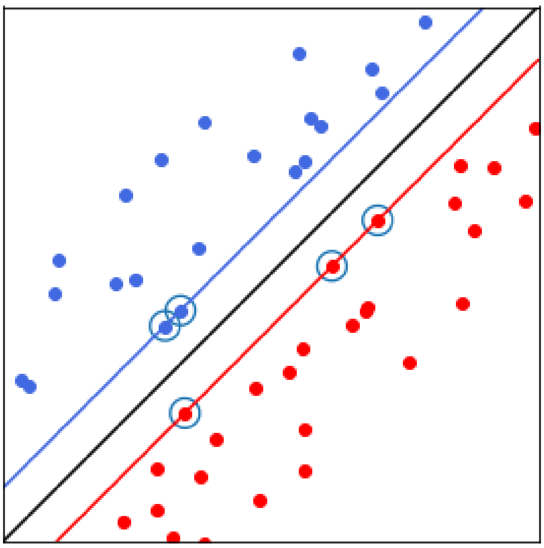


Figure 3. Example of SVM boundaries

In the two-dimensional plane of Figure 3, the black line represents the border, while the support vectors are identified as the circled points closest to the border. The margin is defined as the distance between the boundary and the blue and red lines.

In simple terms, a hyperplane is an n-1-dimensional subspace in an n-dimensional space. For a two-dimensional binary classification, this hyperplane takes the form of a line. In three dimensions, it takes the form of a plane, and so on.

These points are of critical importance in defining the hyperplane and the margin, as they have a considerable influence on the hyperplane's position. The margin, on the other hand, represents the distance between the hyperplane and the nearest data points in each class. The fundamental objective is to maximize this margin.

SVMs use a decision function to classify new data points, assigning a class according to the side of the hyperplane on which they are located. To deal with non-linear problems, SVMs utilize kernels, allowing the feature space to be transformed to make the data linearly separable in a higher-dimensional space.

They include a regularization parameter (C) that regulates the trade-off between maximizing margin and minimizing classification error. A higher C tolerates fewer classification errors, but may lead to overfitting.

The support vector classification method can be extended to solve regression problems. This method is called Support Vector Regression (SVR), which is a variant of SVM.

Support Vector Regression

SVRs are based on the same theory as SVMs used for classification, but are adapted to predict continuous values rather than classify data into discrete categories (Salcedo-Sanz & al., 2014). The margin in SVR is the area around the regression line where data points are tolerated, compared with classification SVMs, where the margin represents the distance between classes. The aim is to maximize this margin while keeping the prediction error within this tolerated limit as shown in the Figure 4.

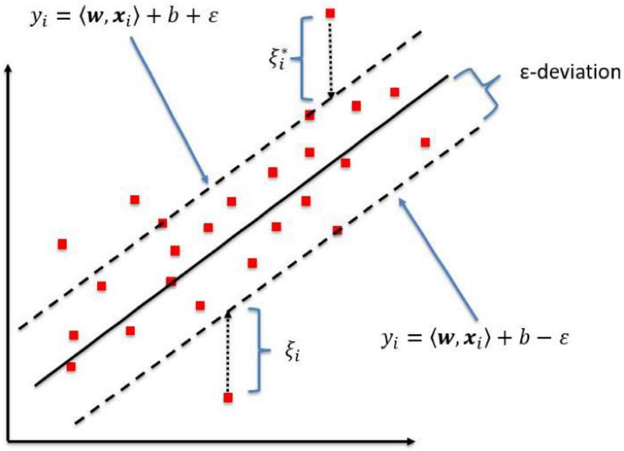


Figure 4. Support vector Regression (Kleynhaus & al., 2017)

The problem is formulated as the equation (eq.10):

$$\boldsymbol{min\, w, b, \xi, \xi *}\ \left(\mathbf{1/2}\ \| \boldsymbol{w} \|^{2} + \boldsymbol{C} \sum_{i=1}^{N} (\boldsymbol{\xi i} + \boldsymbol{\xi i *})\right) \tag{10}$$

Under constraints (eq. 11, 12) :

$$\boldsymbol{y_i} - \langle \boldsymbol{w}, \boldsymbol{x_i} \rangle - \boldsymbol{b} \leq \boldsymbol{\varepsilon} + \boldsymbol{\xi i} \tag{11}$$

$$\langle \boldsymbol{w}, \boldsymbol{x_i} \rangle + \boldsymbol{b} - \boldsymbol{y_i} \leq \boldsymbol{\varepsilon} + \boldsymbol{\xi * i} \tag{12}$$

$$\boldsymbol{\xi i}, \boldsymbol{\xi * i} \geq \mathbf{0}$$

Where,

N is the number of observations,

y_i is the target variable of observation i,

x_i is the input vector,

ε is the error tolerance, and

C is the regularization parameter.

The variables ξ_i and ξ^*_i are slack variables that allow individual errors to be tolerated.

This method offers a variety of kernel functions, each playing a crucial role in feature space transformation.

Linear kernel: Fundamental, it measures similarity by the scalar product between input vectors, i.e. $\langle x, x' \rangle$.

Polynomial kernel: Introducing a polynomial component, it is expressed as eq. 13:

$$(\gamma \langle x, x' \rangle + r)\ d, \tag{13}$$

Where *d* is determined by the degree parameter, And *r* by the coefficient parameter.

RBF (Radial Basis Function) kernel: Also known as the Gaussian kernel, it evaluates similarity as a function of Euclidean distance in eq.14

$$exp^{(-\gamma\|x-x\prime\|2)} \qquad \textbf{(14)}$$

Where γ set by the "gamma" parameter, which must be greater than 0.

Sigmoïde kernel: Using a sigmoid transformation, it measures similarity using eq.15:

$$\boldsymbol{tanh(\gamma\langle x, x'\rangle + r;} \qquad \textbf{(15)}$$

Where *r* is defined by the coefficient parameter.

Random forest

Several decision trees are combined into a single prediction using a machine learning technique called a random forest like in Figure 5. Their versatility, resilience, and capacity to manage a range of tasks, such as regression and classification, make them particularly effective and popular.

A decision tree is a node structure in the form of a tree, where each node represents a decision based on a particular feature. The leaves of the tree are the results, i.e. the predictions for a class (in the case of classification) or a numerical value (in the case of regression).

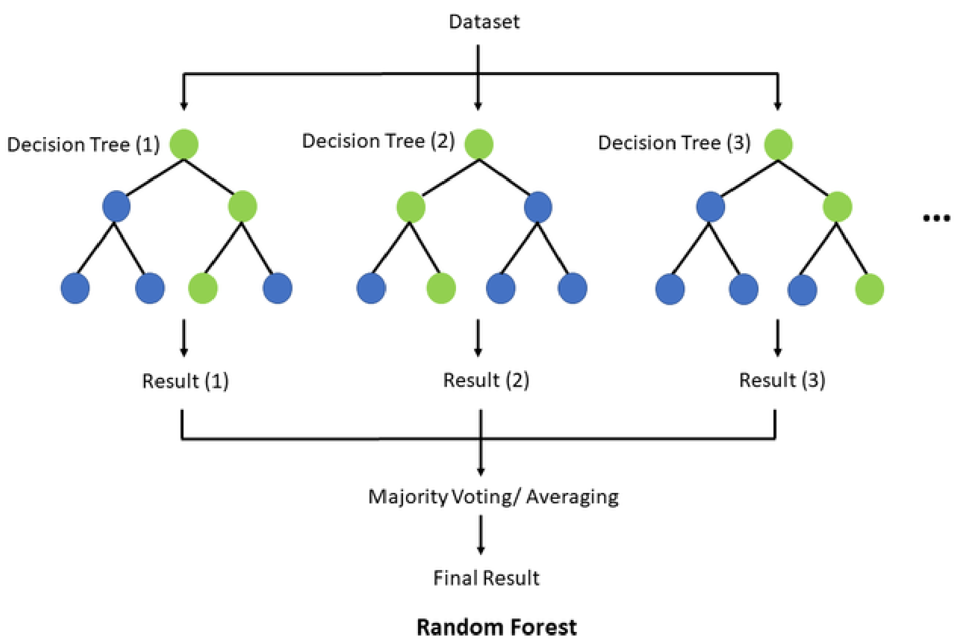


Figure 5. Decision tree vs Random forest

A number of randomly constructed independent decision trees form a random forest. Each tree is built on a random subsection of the training data, and at each decision node, a random subsection of the features is considered to determine the best split as show in Figure 6.

This randomization introduces diversity among the trees, making the forest robust and less susceptible to overfitting. The process of generating a prediction using a random forest requires the integration of each of the distinct responses given by each tree in the forest. This aggregation strategy makes overall predictions more stable and reliable.

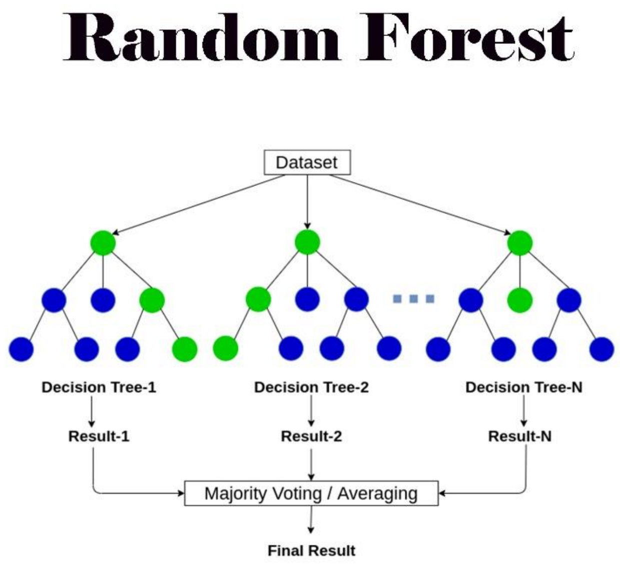


Figure 6. Prediction in Random forest (Roi Yehoshua, 2023)

Random forests work efficiently and stably on a variety of applications without the need for fine parameter optimization. When constructing trees, randomization adds variation, which facilitates the acquisition of various nuances and structures in the data. Their inherent ability to generalize to unknown data makes random forests particularly appropriate for complex issues for which simpler models might prove ineffective. Their ability to operate effectively "out of the box" makes them a valuable option, particularly in situations where time or resources for fine-tuning are in short supply.

Another strength of Random Forests resides in their ability to efficiently handle missing data without the need for intensive pre-processing (Vens & Costa, 2020). When an observation has missing data, a tree can still be built using the available features, and in the prediction phase, it will contribute according to its ability to generalize from the features present. This ability simplifies the process of handling missing data, which can be a practical advantage when applying the method to real datasets, often subject to information gaps.

1.2.3 Learning Methods

Many notable developments in the field of machine learning have changed the way computer systems learn and evolve in significant manners. Machine learning has experienced a renaissance in recent decades, driven by significant advancements in algorithms, increased computational power of computers, and access to vast amounts of data, all of which were initially based on the traditional notions of artificial intelligence.

Many approaches, including decision trees, neural networks, support vector machines, and many more, have influenced the development of machine learning. These advances have enabled computer systems to detect complex patterns, make accurate predictions and accomplish tasks once considered difficult or even impossible.

The development of machine learning has made a significant impact on many fields, from healthcare and finance to image recognition, natural language processing and many others (El Naqa & Murphy, 2015). Machine learning continues to make advancements, fueling the age of AI and helping to transform the way we approach the complex problems and challenges of the modern world.

Types of machines learning

The three main categories of machine learning are supervised learning, unsupervised learning and reinforcement learning.

These different machine learning methods contribute to an ever-evolving field, offering robust tools for solving a variety of complex problems in many fields.

The model is trained in supervised learning using an annotated dataset, where inputs are linked to known labels or outcomes. By learning to generalize these associations, the model can make predictions on new data using the patterns it has learned.

Unsupervised learning, on the other hand, focuses on discovering intrinsic structures in data without prior labels. Unsupervised algorithms group data into clusters or principal components to study similarities and patterns, revealing the underlying structure of the data.

Lastly, reinforcement learning is based on the idea of an agent interacting with its surroundings. The agent acts, examines how the environment reacts, and modifies its behavior to maximize a predetermined reward. This kind of learning is widely applied in the domains of robotics and games, where an agent must learn to make successive decisions in order to accomplish a given objective.

Supervised learning

One of the most important methods in ML is the supervised learning, in which a model is built on a set of labeled data means that the data is accompanied by appropriate outputs (Suthaharan, 2016). Through this process, the model learns to correlate inputs with desired outputs, producing a function that can precisely forecast the outcomes of fresh, unlabeled data.

Process of supervised learning: Inputs are associated with relevant classes to create a set of labelled data. An appropriate model, like random forest or linear regression, is chosen according on the type of task (classification or regression). Then, the model is trained on the training data, adjusting its parameters to minimize the error between predictions and actual labels. To reduce the error between its predictions and actual labels, the training process involves fitting the model to training data. To evaluate its performance, the model is trained on a test set, which is a set of unseen data. Its effectiveness is assessed using measures such as precision and recall. By analyzing the results, it is possible to adjust to the parameters of the model or consider an alternative to optimize its performance.

Supervised learning offers considerable flexibility and has applications in many fields, helping to solve complex problems through the effective use of labeled data.

There are many uses for supervised learning in industries like finance and health care. Algorithms used in supervised learning include specific techniques like Support Vector Regression (SVR), linear regression, and random forests (Nasteski, 2017). Numerous issues can be resolved with these techniques, ranging from financial trend analysis (De Prado, 2018) to medical diagnosis prediction (Uddin & al., 2019).

Unsupervised Learning

When a model is presented with unlabeled data, it must independently identify any inherent structures, patterns, or relationships in the data. This type of machine learning is known as unsupervised learning. In contrast to supervised learning, no labels or predetermined outcomes are given to the model to guide it. The main aim of unsupervised learning is to explore the structure inherent in the data and extract useful information without the need for explicit supervision.

There are several techniques commonly used in unsupervised learning, including the following:

Clustering: Clustering is a technique that aims to divide a dataset into homogeneous groups or clusters, where elements within the same cluster are similar to each other, while elements in different clusters are different. Popular clustering algorithms include K-means (Ding & he, 2004), Hierarchical Clustering (Kaufman, & Rousseeuw, 2009), and DBSCAN (Deng, 2020).

Dimensionality reduction: This technique aims to reduce the number of variables or dimensions in a dataset, while preserving as much information as possible. The aim is to simplify the representation of data while maintaining its important characteristics. Methods such as Principal Component Analysis (PCA) (Chen & Qian, 2008) and T-SNE (t-Distributed Stochastic Neighbor Embedding) (Gisbrecht & al., 2015) are often used.

Anomaly detection: Finding outliers or anomalies in a data set is another application for unsupervised learning. Observations that deviate markedly from the rest of the data are called anomalies. In this case, methods like One-Class SVM (Li & al., 2003) and Isolation Forest (Lesouple & al., 2021) are applied.

Association: Finding association links between various variables in a data set is the goal of this technique. To determine association rules from transactional data, the Apriori (Adewole & al., 2014) method is frequently used.

Process of Unsupervised learning: Unsupervised learning is a type of machine learning where the model is confronted with unlabeled data. Unsupervised learning aims to identify intrinsic structures or patterns in the data without prior knowledge of the desired results, unlike supervised learning, which trains the model using a set of labeled data. Initially, a collection of unlabeled data is assembled from a variety of sources, containing information on characteristics with no clear indication of the target variable. These data are then subjected to a cleaning and pre-processing process, aimed at eliminating outliers, normalizing scales, and correcting any irregularities present in the raw data. The choice of unsupervised learning algorithms depends on the specific objective of the analysis.

The model is then exposed to unlabeled data, and depending on the algorithm selected, it learns to discern meaningful structures, relationships or patterns within the data. The results obtained are then examined to extract relevant information. For example, when applying a clustering technique, the aim is to identify groups of similar data, while in the case of dimensionality reduction; the emphasis is on exploring the main components.

Reinforcement learning

Through interactions with a dynamic environment, an agent that uses reinforcement learning (RL) learns to make decisions that maximize a cumulative reward over time. The agent investigates the environment, acts, and modifies its plan in response to the rewards it receives. This process is founded on the idea of trial and error.

The main challenge of reinforcement learning is to find the right balance between exploration (Ding & Dong., 2020) (trying new actions to discover better policies) and exploitation (choosing the actions that seem to be the best based on current knowledge).

1. Agent and environment: An agent in reinforcement learning is an entity that operates and makes decisions inside an environment. The agent works in a certain environment, which reacts to the agent's actions.

2. Actions: The agent is capable of performing a variety of environmental actions. These operations can be continuous (e.g., setting a value within a continuous range) or discrete (e.g., selecting among multiple alternatives).

3. Reward: The environment provides a reward on the agent for completing a task. The quality of the action taken is determined by the reward. The agent wants to maximize the total reward that has accrued over time.

4. Policies: A policy defines the agent's strategy, in other words, the way it chooses its actions according to the current state of the environment. The agent's objective is to find an optimal policy that maximizes cumulative reward.

Process of RL: Initially, the agent interacts with the environment without any significant prior knowledge. During this first phase, the agent acts in the environment without receiving precise instructions. It then examines the state of the environment, including data on current circumstances, to make a decision based on its predefined policy. Once the decision has been made, the agent executes the action.

The environment evaluates the quality of the agent's choice and rewards it based on the action taken. The agent's policy is then updated using the reward. Essentially, this update aims to optimize expected future rewards by making adjustments to the policy.

This iterative process continues, with the agent iterating between observing the environment, taking actions, receiving rewards and updating its policy. The agent adapts continuously, refining its policy over time to improve its ability to maximize future gains.

Various RL algorithms have been documented in the literature, including methods such as Q-learning (Clifron & Laber, 2020), deep Q-learning (DQN) (Fan & al, 2020) and proximal policy optimization (PPO) (Engstrom & al., 2019).

1.2.4 Deep Learning Models Presentation

Deep learning is a branch of machine learning (Figure 7) that simulates the functions of the human brain to train computer models to perform tasks. More specifically, deep learning enables complex models to be learned from data by representing those using artificial neural networks, which are multi-layered structures.

They model complex data using artificial neural networks with multiple layers of hidden neurons. To accomplish tasks such as image recognition (Pak & Kim, 2017), speech recognition (Kamath & al., 2019), machine translation (Yang & al., 2020), time series prediction (Gamboa, 2017),…etc.. Deep neural networks can learn from huge amounts of unstructured data, including text, video and photos. Deep learning algorithms can find complex patterns in data and use them as predictive tools to generate accurate predictions.

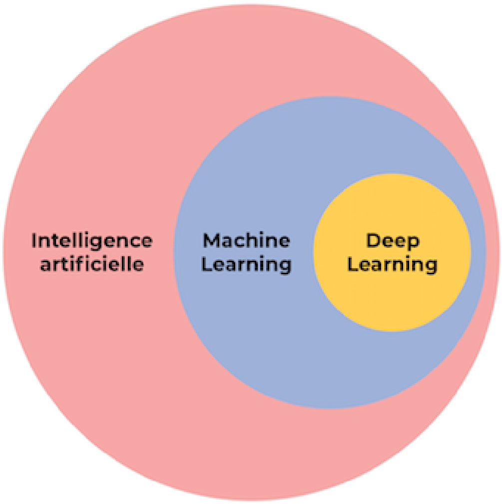


Figure 7. Deep Learning

The availability of large quantities of data, advancements in algorithms, and increased processing power have all contributed to the significant expansion of deep learning in recent years. It remains a very active field of research and application, which has led to major advances in many sectors.

Multi-Layer Perceptron

Inspired by the way real neurons operate; a neural network is like an algorithm in a computer context (Cox & Dean, 2017). Its unique quality is its ability to self-correct according to results, offering a means of learning and problem-solving without relying on traditional algorithms or programming.

Neural networks, as information assimilation systems, apply the principle of induction, suggesting experience-based learning. These networks derive an integrated decision-making system through the study of particular scenarios; its generic character depends on the quantity and complexity of the learning cases encountered in relation to the intrinsic difficulty of the problem to be solved.

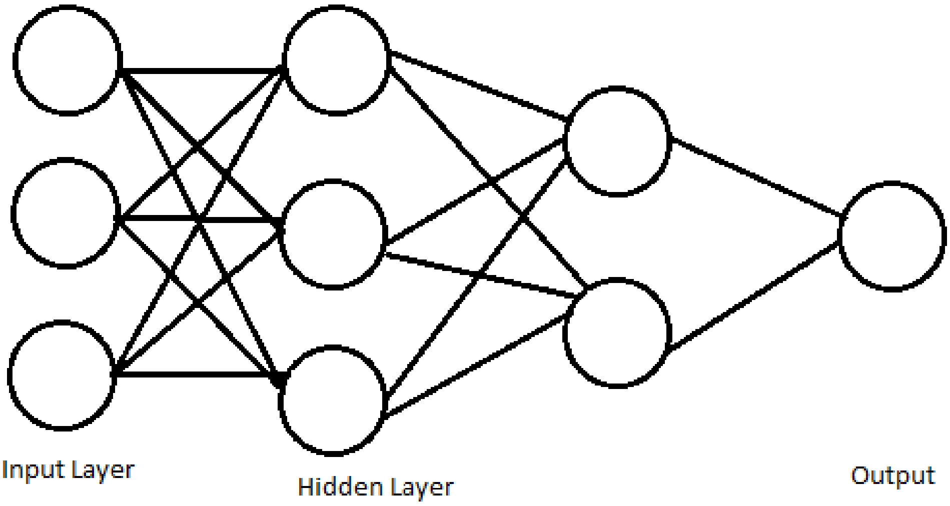


Figure 8. Multi Layers Perceptron

The Multi-Layer Perceptron (MLP) is usually structured in a sequence of layers, where each layer (i) receives its inputs from the outputs of the previous layer as in Figure 8. Each layer is composed of Ni neurons, taking their inputs from the N i-1 neurons of the previous layer. Each synaptic connection is associated with a synaptic weight (W ij[l]), so that the N i-1 input elements are weighted by these weights, then summed by the neurons of layer i. Mathematically, this can be expressed as follows in eq. 16:

$$\boldsymbol{Z_i^{[l]}} = \sum_{(j=1)}^{(Ni-1)} [\![\boldsymbol{W_{ij}^{[l]}} \boldsymbol{A}_{\boldsymbol{j}}^{[l-1]} + \boldsymbol{b}]\!]_{\boldsymbol{i}}^{[l]} \tag{16}$$

Where:

$Z^{[l]}_i$ is the resulted vector after the application of weights and bias *($b^{[l]}_i$)*,

$A^{[l-1]}_j$ represents the activation of the j-th neuron in the previous layer,

And *l* indicates the network level.

Multiplying several transformation matrices is like concatenating the layers of a neural network. From the point of view of modeling complex relationships, the network would be uninteresting if each layer applied a linear output function, as this would simplify the whole network into a single matrix that would be the product of the other matrices. This is where the importance of the activation function between layers comes into play.

The introduction of a non-linear function at the output of each layer *($A^{[l]}_i = g(Z^{[l]}_i)$)* is crucial to enable the network to capture complex patterns and avoid global linearity. Thus, the judicious choice of an output function becomes essential for the performance and relevance of the neural network.

Activation Function

Just as neurons in the brain respond to electrical signals, the activation function is crucial to the information processing carried out by an artificial neuron. Among the most popular and famous activation functions are Sigmoid, ReLU, the logistic function and the hyperbolic tangent.

Sigmoid

The sigmoid function is generally represented by the following equation eq.17:

$$\boldsymbol{\sigma(x) = 1/1 + e^{-x}} \tag{17}$$

Where *(x)* is the input to the sigmoid function.

Often used in network output layers for binary classification tasks, since it compresses values into a restricted range. It takes a process value as input and transforms it into an output process value between 0 and 1 (Figure9).

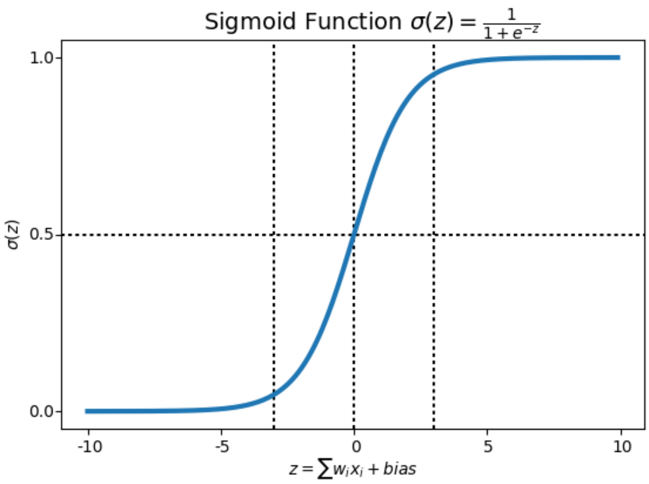


Figure 9. Function Sigmoid

Hyperbolic Tangent

The activation function that generally outperforms the logistic function is the hyperbolic tangent. This function transforms all real inputs into values between -1 and 1 (Figure 10), thus providing centralization around zero. However, it still presents the challenge of vanishing gradient, as it also reaches limit values where its gradient (derivative) becomes zero. Despite this, the hyperbolic tangent remains a frequently used option.

The hyperbolic tangent function is represented by the following equation eq.18:

$$\boldsymbol{tanh(z) = e^{z} - e^{-z}/e^{z} + e^{-z}} \tag{18}$$

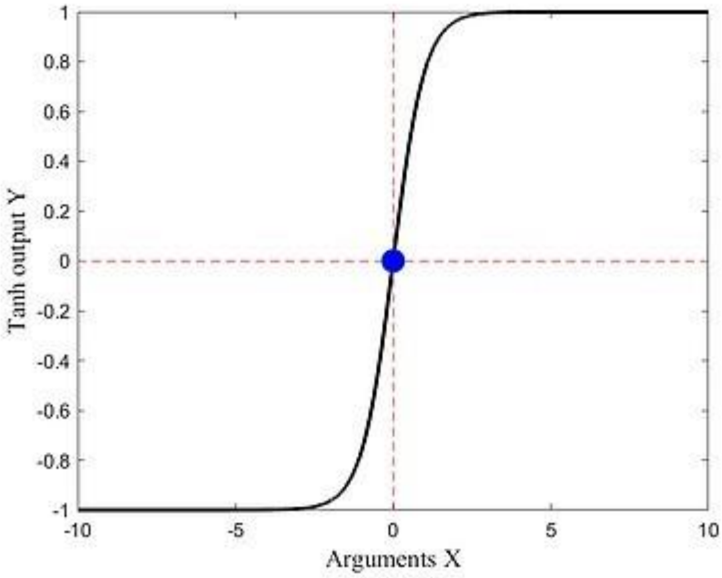


Figure 10. Function hyperbolic tangent

Rectified Linear Unit

The most commonly adopted activation function is the Rectified Linear Unit, characterized by its piecewise linear nature (Figure 11). Its major advantage lies in its ability to replace all negative input values with 0 and keep positive values unchanged, expressed mathematically as eq. 19:

$$\boldsymbol{max(0, x).} \qquad \textbf{(19)}$$

This function has a zero gradient for negative values and a gradient of 1 for positive values. This property is particularly beneficial during the learning phase, as it helps to overcome and rectify the vanishing gradient problem.

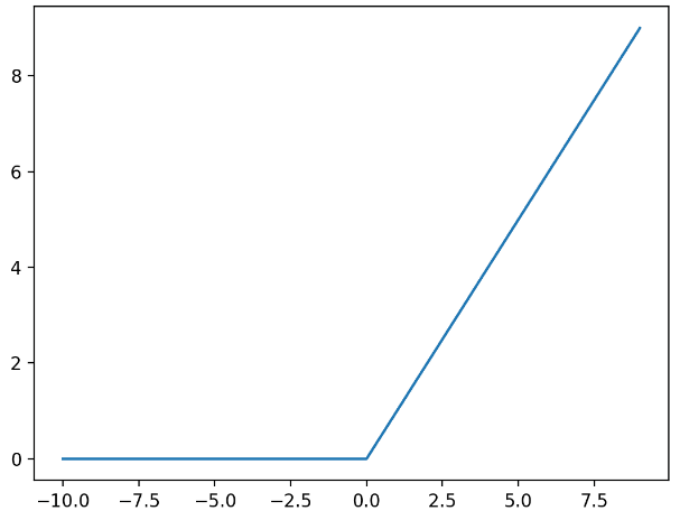


Figure 11. Function Relu

Convolutional Neural Network

Another type of neural network that excels at processing grid-structured data, such images, is the convolutional neural network (CNN). CNNs have been widely used in computer vision applications such as image classification, object identification and image segmentation. CNNs are designed to capture local and hierarchical patterns in data. Their innovative architecture makes them very useful for collecting spatial patterns in data.

In contrast to traditional neural networks (ANNs), CNNs are distinguished by their ability to extract complex features due to the integration of convolution and pooling layers, and a deep architecture (Namatevs, 2017). The effectiveness of CNNs in feature extraction increases with the depth of the architecture. Initially designed for image processing, CNNs are proving equally effective for processing two-dimensional or higher-dimensional inputs. Their process can be divided into three fundamental steps: convolution, pooling and the fully connected layer. These successive steps give CNNs the ability to apprehend complex, spatial and hierarchical information, positioning them as a powerful solution for a variety of tasks, from computer vision to other fields of application.

Convolution Layer

The convolution layer is a fundamental component of CNNs. It plays a crucial role in extracting features from input data.

The convolution layer uses filters, also known as kernels, to detect specific features in the input data. These filters are small matrices *(X, Y)* of weights that are applied locally to the image region like shown in Figure 12. Each filter is specialized in the detection of certain features. The filter is applied to local regions of the input by multiplying the filter elements with the corresponding values, then summing these products. This operation is repeated by moving the filter through the entire input, calling it stride. The result is a feature map that highlights the presence of the patterns detected by the filter.

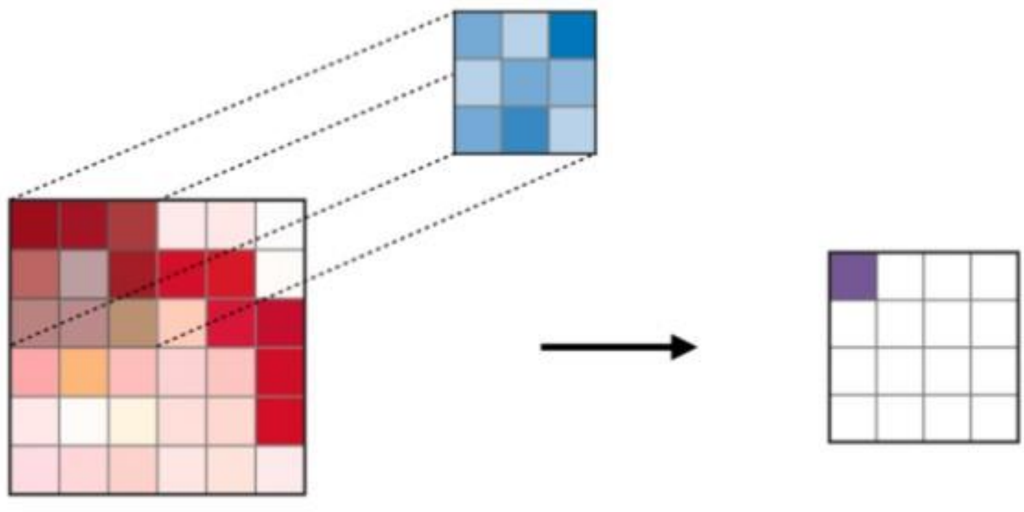


Figure 12. Convolution layer

Moving a filter over an input matrix create a new, smaller matrix called a "Feature Map". At each step, the product between the filter and the corresponding input submatrix contributes to progressively filling each cell of the Feature Map. The value of each feature map cell is the sum of the products of the filter elements and the corresponding matrix elements.

After the convolution operation, an activation function is usually applied to the resulting feature map to introduce non-linearity into the model. The ReLU function is frequently used, but other functions can also be chosen.

Pooling

The pooling layer reduces the spatial resolution of the extracted features, reducing the amount of computation required in subsequent layers while conserving the most important information. This makes the model more robust to minor variations in input data. The aim is to reduce the spatial dimension of extracted features while preserving their essential information.

The pooling layer works by applying a pooling operation (max pooling or average pooling) to local regions of the feature map resulting from convolution. Each region, often defined by a filter of size M is reduced to a single value.

The max pooling operation for a region *(i,j)* is defined as follows in eq.20:

$$\boldsymbol{MaxPooling}(\boldsymbol{I})_{\boldsymbol{ij}} = \boldsymbol{max}_{\boldsymbol{m=0}}^{\boldsymbol{P-1}} \boldsymbol{max}_{\boldsymbol{n=0}}^{\boldsymbol{P-1}} \boldsymbol{I}_{(\boldsymbol{i}\times\boldsymbol{P})+\boldsymbol{m},(\boldsymbol{j}\times\boldsymbol{P})+\boldsymbol{n}} \tag{20}$$

For the average pooling operation, each region is replaced by the average of its values in eq. 21:

$$\boldsymbol{AveragePooling}(\boldsymbol{I})_{\boldsymbol{ij}} = \boldsymbol{1}/\boldsymbol{P}^{2} \sum_{\boldsymbol{m=0}}^{\boldsymbol{P-1}} \sum_{\boldsymbol{n=0}}^{\boldsymbol{P-1}} \boldsymbol{I}_{(\boldsymbol{i}\times\boldsymbol{P})+\boldsymbol{m},(\boldsymbol{j}\times\boldsymbol{P})+\boldsymbol{n}} \tag{21}$$

The pooling layer reduces the spatial resolution of the extracted features, reducing the amount of computation required in subsequent layers while preserving the most important information. This makes the model more robust to minor variations in input data.

Fully Connected Layer

In order to connect the features discovered by the hidden layers to the output layer, fully connected layers are often placed at the end of the network. The model can produce scores for each class in the classification question thanks to this complete connection.

Each component of the flattened input vector is connected to each neuron in the fully connected layer (Figure 13). Weights are associated with each connection and will be adjusted during training. The operations of a fully connected layer are linear, meaning that the output of each neuron is a linear combination of its input values, each weighted by its corresponding weight. To add non-linearity to the model, a non-linear activation function is usually added after the linear processes.

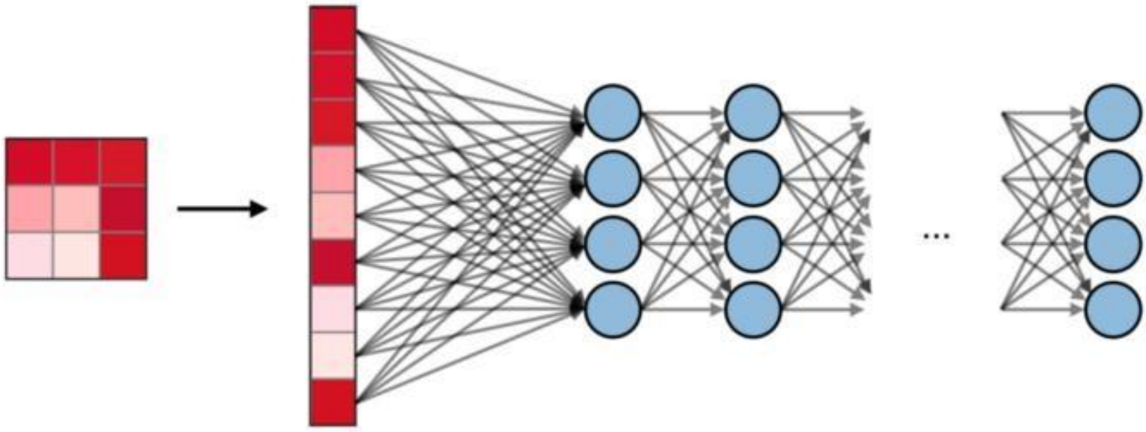


Figure 13. Fully Connected Layer

Transformer

The Transformer model is a revolutionary architecture introduced by Vaswani et al. in 2017 (Vaswani & al., 2017) for natural language processing tasks. Based models such as BERT, GPT, and others, the Transformer eliminated the need for recurrent or convolutional layers, introducing instead the attention mechanism.

Before the emergence of transformers, the implementation of neural networks required the use of vast annotated datasets, a costly and time-consuming process. By deducing mathematical patterns between elements, transformers have eliminated this necessity, opening the way to the exploration of billions of images and petabytes of textual data on the Internet as well as in corporate databases.

The Transformer model represents a deep learning architecture based on the multi-headed attention mechanism (Vaswani & al., 2017). What makes it stand out is its lack of recurrent units, resulting in reduced training times compared to previous recurrent neural architectures, such as Long-Term Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997). Its later version has been widely adopted for large-scale linguistic model training on large linguistic datasets, including the Wikipedia corpus and Common Crawl (OpenAI, 2019).

The Transformer architecture is a fundamental structure in the field of natural language processing (NLP) and has revolutionized many deep learning applications. It consists of two main parts: the encoder and the decoder. Each of these parts is made up of several stacked layers.

Encoder

The Transformer encoder is responsible for taking an input sequence and creating a vector representation of it. The self-attention layer and the feedforward layer are the two primary sub-layers that make up each of the multiple stacked layers that make up this structure.

Multi-Head Attention layer:

The Multi-Head Attention layer is the core of the Transformer. It enables the model to focus simultaneously on different parts of the input sequence, through multiple attention "heads". Each head generates a different projection of the input, and these projections are concatenated and linearly transformed.

The attention calculation can be expressed as a large matrix calculation using the Softmax function, which is useful for training due to computer optimizations of matrix operations that quickly compute matrix operations. The matrices *Q*, *K* and *V* are defined as where the *i* rows are vectors *q_{i}*, *k_{i}*, and *v_{i}* respectively. We can then represent attention as in eq. 22:

$$\boldsymbol{Attention(Q,K,V) = softmax\;(QK^T/\sqrt{d_{\{k\}}})\;V} \quad \textbf{(22)}$$

Where Softmax is shown on the horizontal axis.

The output of this layer is calculated for each position in the sequence, taking into account the relationships with all other positions.

Self-attention:

Self-attention allows the model to take account of long-term dependencies in a sequence by giving different weights to different elements in the sequence according to their relative importance for a specific task.

Consider *X* an input sequence, the self-attention output is calculated as the following equation eq.23:

$$\boldsymbol{MultiHead(X) = Concat(head_1, \ldots, head_h)W_O} \quad \textbf{(23)}$$

Where, $\boldsymbol{head_i = \;Attention(XW_{\,Qi}, XW_{\,Ki}\,, XW_{\,Vi})}$.

Feedforward layer:

After the self-attention layer, a feedforward layer is applied. It consists of two linear operations with an intermediate activation function, usually the ReLU as in eq. 24.

$$\boldsymbol{FFN(X) = \;ReLU(XW_1 + b_1)\,W_2 + b_2} \quad \textbf{(24)}$$

Where:

X represents the input of the feedforward layer, which is the output of an attention layer;

W are weight matrices;

b are the corresponding biases, one for each neuron in the feedforward layer;

XW_1+b_1 represents the linear operation, where each element of the input is multiplied by the corresponding weights and the bias is added. This is a linear transformation of the input.

The Transformer feedforward layer applies two linear transformations to the input, with a ReLU activation function between them. This enables the model to learn complex non-linear representations of the data. The weight matrices *W*, as well as the biases *b1* and *b2*, are learned during model training.

Decoder

The decoder, composed of several stacked layers, uses a multi-headed attention layer that takes into account the encoder and the decoded sequence. This approach allows the model to focus on different parts of the input and output during generation. The decoder, like the encoder, is responsible for generating an output sequence by taking into account the representation of the input sequence formed by the encoder. Each decoder layer incorporates an additional self-attention layer to take account of relationships with the already generated portion of the output sequence.

Multi-Head Target Auto-Attention layer:

This layer takes into account both the encoder and the decoded sequence up to the current position. It allows concentration on different parts of the input and output during generation with the calculated equation eq.25:

$$\boldsymbol{Attention(Q,K,V) = softmax(\frac{Q(K+MultiHead(X))}{\sqrt{(dk)}})\,V} \tag{25}$$

Where Q is calculated from the decoded sequence and K, V come from the encoder representation.

The target version of this specific operation means that, when generating an output sequence at a certain position, the model must focus on the already generated part of the output sequence. This helps to capture temporal dependencies when generating successive words in the output sequence.

Output Layer:

The last layer of the decoder generates the final output of the model. It projects the decoded representation into the output word space with the following equation eq. 26:

$$\boldsymbol{Output(X) = \ Softmax\ (XW_O + b_O)} \tag{26}$$

Where W_O and b_O are the weights and biases associated with the output layer,

and the Softmax function ill transform raw scores into normalized probabilities.

This process produces a probability distribution on the output vocabulary. Softmax ensures that the values obtained from the separate calculation of each position in the output sequence are reliable probabilities, in other words, that they range between 0 and 1, and that their sum is equal to 1.

The output generated at this stage of the output sequence is then selected according to the position with the highest probability. This process is repeated for each position in the sequence until the entire sequence has been created.

1.3 Energy forecasting and sustainability: Literature Review

The topic of energy planning and sustainability is constantly advancing due to the rich background that different works provide in comprehending the different approaches and challenges involved in anticipating energy as a time series.

Several of these studies focus on short-term forecasts, attempting to accurately predict short-term fluctuations in energy consumption (Divina & al., 2018). Others have taken a longer-term view, exploring medium- and long-term trends (Wang & al., 2020) to guide energy policy and investment. Energy prediction using time series reveals a diversity of research work encompassing fields from statistical modeling (Solyali D, 2020) to deep learning techniques (Kim & Cho 2019).

Reliable forecasting of energy requirements provides important guidance for the most effective possible distribution of energy resources (Liu & al., 2018) and development of power-saving techniques (Han & al., 2008). Furthermore, accurate predictions of energy consumption can serve as a stimulant for the efficient execution of market analyses, contributing to the advancement of economic growth (Pao, 2009). And its progress can be extended to other areas of temporal forecasting (Severiano & al., 2021), such as traffic flow projection (Yin & al., 2002), meteorological prediction (Carter & al., 1989), temperature estimation (Kršmanc& al., 2013), stock market performance forecasting (Vaisla & Bhatt, 2010), and solar radiation prediction (Deo & Şahin, 2017).

This field has enabled numerous researchers to propose innovative solutions, such as the use of techniques for predicting electrical energy while preserving user confidentiality (Tang & al., 2023). It has also been the driving force behind initiatives to improve the efficiency of solar distillers, through the application of machine learning to anticipate their energy production (Murugan & al., 2023).

The development of accurate predictive models for managing manufacturing procedures and optimizing the use of energy (Pelella & al., 2023) is another vital aspect of energy consumption prediction, and accurately forecast residential customers' energy consumption in order to optimize energy production based on demand or enhance the energy efficiency of public buildings within the framework of intelligent cities (Nazir & al., 2023).

Several methods have been tested and validated by the researchers for time series prediction and electric load prediction. These methods can be divided into 3 main categories (Wang & al., 2020):

1.3.1 Statistical Approaches

Traditional approaches such as ARIMA (Auto Regressive Integrated Moving Average) methods have been widely used to model energy time series. These methods provide a solid basis for understanding seasonal trends and short-term variations (Nepal & al., 2020). ARIMA models are known for their good performance in forecasting short-term electricity consumption (Mohamed & Bodger, 2005 and Sen & al., 2016). (Mohamed & al., 2010) Used the double-seasonal ARIMA model in the statistical analysis system, which performed well enough for a prediction with one-step ahead. The researchers in (Hor & al., 2006) also demonstrated the robustness of this model for predicting daily load patterns, while studying the climatic and meteorological changes that have a direct impact on electricity.

Among the most frequently, used forecasting techniques by energy experts (Musaylh & al, 2018) are ARIMA and multivariate adaptive regression Splines (MARS) (Sigauke & Chikobvu, 2010). The model SARIMA which is an extension of ARIMA that includes seasonal components, which makes it suitable for modeling time series with seasonal variations (Naim & al., 2018). These models performed best during periods of high volatility, such as holidays and periods of high thermal amplitude (Nowicka-Zagrajek & Weron, 2002).

The use of regression models, such as linear regression, can also be very effective for predicting electricity consumption (Yildiz & al., 2017). Linear regression models have been successfully applied in specific contexts for forecasting electrical energy consumption such as improve the accuracy of electricity consumption forecasts with google trends data (Hossain & Mahmud, 2018), comparing it with some other models (Samuel & al. 2017), or by using the phase-space reconstruction algorithm and the two-square kernel (Fan & al., 2018).

Other researchers have added exogenous variables to the regression, which becomes a multiple linear regression, such as humidity (Saber & Alam, 2017), temperature (Sun & al., 2017) and other meteorological factors (Wang, 2017). Other studies have shown that Sociological events (Yang & al., 2017) and economic factors (Zor & al., 2017) influence the short term load forecasting.

Exponential smoothing models (Abd Jalil & al., 2013) and Box-Jenkins (Sadaei & al., 2017) methods are another type of parametric statistical techniques utilized in this area of studies.

1.3.2 Machine Learning Models

Machine learning methods are widely used to predict energy consumption, due to their ability to model complex relationships between variables. One of the famous and successful models used is SVM (Türkey & Demren, 2011) for these excellent results in load forecasting (Guo & al., 2006). In order to anticipate electricity demand, other researchers, such (Kaytez & al., 2015), compared least squares SVMs (LS-SVM) with artificial neural networks (ANNs). The results showed that of all the models evaluated, the suggested model (LS-SVM) performed best and was the fastest. (Musaylh & al., 2018) Evaluated the comparison between Support Vector Regression (SVR), a branch of SVM, and statistical models such as Multivariate Adaptive Regression Spline (MARS) and ARIMA.

Support Vector Regression (SVR)-based method for short-term electric charge forecasting was developed by (Kavousi-Fard & al., 2014) and (Jiang & al., 2016), who combined the Modified Firefly Algorithm (MFA) and Particle Swarm Optimization (PSO) techniques, respectively, to optimize the SVR parameters. (Al-Misaylh & al., 2018) Tested another hybrid technique with SVR and PSO optimization with combining it with adaptive noise (ICEEMDAN) and empirical mode decomposition (EMDAN). Notably, a quantum-based dragonfly method in conjunction with an enhanced version of empirical mode decomposition with adaptive noise (CEEMDAN) (Zhang & Hong, 2019) has also shown excellent efficiency.

On the other hand, (Dong & al., 2018) introduced the cuckoo search (CS) algorithm into the SVR model, dealing specifically with seasonal data and incorporating a chaotic mapping function to diversify the exploration of the CS space. (Zhang & al., 2020) Similarly employed this chaotic mapping strategy, who incorporated variational mode decomposition (VMD) to mitigate data non-linearity. In addition, they incorporated a self-recursive mechanism, allowing the concept of memory to be taken into account in the model.

Similarly, (Son & Kim, 2017) adopted SVR to predict energy demand in the residential sector, achieving significant results using some meteorological and social factors, in addition to load data. A similar study by (Lusis & al., 2017) confirmed that, despite the use of a small data set, SVR models were effective when incorporating weather and other exogenous variables.

In addition, ensemble approaches have attracted considerable interest, such as the use of random forests (RF) by (Dudek, 2015) and (Lahouar & Slama, 2015) in load prediction, demonstrating their robust predictive capacities. However, (Mayrink & Hippert, 2016) explored regression trees with gradient boosting (GBRT), while (Papadopoulos & Karakatsanis, 2015) performed a comparative analysis between random forests (RF) and GBRT, which conclude that both algorithms are capable of producing satisfying results in the field of load prediction.

One of the most frequently used approaches to electrical load prediction for short term (Vemuri & al., 1993 and Sheikh & al., 2012), is artificial neural networks (ANN). The researchers distinguished themselves by their choice of architecture and activation functions for this model. (Tarmanini & al., 2023) This paper demonstrates how ANNs outperform ARIMA models in a real nonlinear electricity database located in Ireland.

Other work has shown that adding exogenous variables, such as weather conditions, time of day (Harrison & al., 2014), and the economic and demographic situation of the base (Hyde & Hodnett, 1997), can improve model performance.

Some researchers, such as (Henley & Peirson, 1997) have also pointed out that the relationship between electrical charge and temperature can vary depending on the time of day and the type of day. To address this complexity, models have been developed to represent this relationship for different time-of-day ranges and day types independently (Ramanthan & al., 2001).

A similar strategy was used by the authors (Chen & al., 2001 and Barbumescu & al., 2016) to predict future load values. They used a backpropagation technique and a multi-layer perceptron (MLP) architecture, incorporating historical load data as inputs to their models.

Furthermore, a proposal in (Khwaja & al., 2017) consists of a boosted neural network (BooNN), where several architectures of neural networks are trained iteratively, with the aim of improving the model error at each iteration.

Another proposal was made by (Hamid & Rahman, 2010 and Barbulescu & al., 2016 and Farfar & Khadir, 2019), who used the ANN model after grouping the days according to the similarity of their load profiles using clustering techniques.

ANNs are often combined with other techniques, as in (Dagdougui & al., 2019) for short-term modeling of building electricity consumption, which exploited an ANN whose evaluation was based on various optimization algorithms and building types. Similarly,(Al-Musaylh & al., 2019) contributed to prediction by proposing models for six-hour and daily periods in several regions of Queensland.

From a different perspective, (Cecati & al., 2015 and Xia & al., 2010) explored the possibilities of a radial-based ANN, while (Ribeiro & al., 2019) introduced a novel approach using an ensemble of wavelet neural networks to improve accuracy in short-term modeling. In parallel, (Ahmad & al., 2019) focused attention on the application of ANN-based forecasting to micro grids, offering a more focused perspective on the local scale.

1.3.3 Deep Learning Models

A significant amount of research has focused on exploring deep learning models for electrical load prediction, highlighting the adoption of various configurations of recurrent neural networks and thereby demonstrating the diversity of approaches in this field.

Recurrent neural networks (RNNs) and LSTM which are an alternative neural network architectures. They were developed to handle this type of time series data for their ability to capture long-term relationships between previous stages of the time series (Muzaffar & Afshari, 2019).

In a residential perspective (Marino & al., 2016) made a comparison between a standard LSTM and a sequence-to-sequence LSTM (S2S) for very short-term forecasts (one minute) while (Kong & al., 2017) implemented an LSTM model for load anticipation, while (Rahman & al., 2018) demonstrated the effectiveness of RNN models for commercial and residential buildings in medium- and long-term forecasts.

In addition, (Su & al., 2019) presented an architecture that uses a traditional LSTM to anticipate future gas use and a bi-directional multi-layer LSTM for feature extraction. In order to evaluate the LSTM and BI-LSTM models' attention mechanisms and further explore the possibilities of LSTMs, (Chitalia & al., 2020) tested the models on a variety of building types.

Another type of neural network that has gained significant popularity in the field of prediction is the convolutional neural network (CNN). The exploitation of convolutional neural network (CNN) architecture has become widespread due to its exceptional performance in feature extraction, as demonstrated in works such as (Levi & Hassner, 2015 and He, 2017). This approach uses several features to estimate electricity demand, including factors such as temperature (Deo & Sahin, 2017), weather conditions (Chow & Leung, 1996), as well as other relevant exogenous variables (Chow & Leung, 1996 and Jetcheva & al., 2014).

CNN was also combined with several other methods, such as recurrent neural networks (RNN) (Kim & al., 2019), fuzzy time series (Sadaei & al., 2019), SVR (Imani, 2021) and even K-means for clustering (Dong & al., 2017). Each approach has required a specific architecture for prediction, showing the diversity of strategies for integrating these techniques into the forecasting domain (Amarasighe & al., 2017). This study also compared the CNN 2D and CNN 1D architectures, two variations of the CNN model using a historical electrical load database (Bendaoud & Farah, 2020).

In electrical load prediction, generative techniques have garnered increasing attention in addition to conventional structures like convolutional neural networks (CNNs). Because of their capacity to produce realistic synthetic data, generative adversarial networks, or GANs, are emerging as a revolutionary solution. In order to create simulated electrical data, (Goodfellow et al., 2014) presented a strategy that sets two networks against one another: a generator and a discriminator. The goal of this approach is to attain a dynamic equilibrium.

Generative Adversarial Networks (GANs) represent an advanced deep learning technique (Zhang & Guo, 2020). Recent studies (Baasch & al., 2021 and Bendaoud & al., 2021) have also attested that GANs bring a novel approach to electric charge prediction. By generating novel electric charge time series in a realistic way, they have helped to improve predictive performance despite their complexity.

(Zhuang & al., 2023) Exploits the GAN to increase the size of load data sets, as well as to complete space load prediction.

(Liu & al., 2023) used a hybrid model combining TimeGAN with CNN-LSTM, the approach begins with the generation of synthetic data using TimeGAN to enrich the limited dataset. Next, the CNN network filters the input data to extract relevant features, while the LSTM network analyzes and predicts the temporal data. The experiments show a considerable improvement in the accuracy of short-term load estimates compared with models without employing synthetic data.

With the fast evolution of neural network architectures, transformers have emerged as key components in various fields of deep learning. Initially designed for natural language processing (Rahali & Akhloufi, 2023), transformers are now showing promise in extended applications.

Transformers have been used in medical imaging (Zhang & al., 2023), translation (Zhang & al., 2018) and prediction (Aksan & al., 2021). Furthermore, (Giuliari & al., 2021) recently used LSTM to forecast the human trajectory, and TRANSFORMERS performed better than LSTM-based techniques used alone. Using deep learning approaches, these writers (Wu & al., 2021) detected and categorized ambiguous and unexplained CEs of ICT patents in construction in 2021. Another study (Zhang & al., 2021) looked into the identification of suicide notes using a recurrent neural network based on transformers and BILSTM, which can be used in social networking sites.

Transformers have been utilized in a number of time series prediction applications (Ahmed & al., 2023). The researchers (L'heureux & al., 2022) has provided a method for adapting existing transformer architecture to perform electrical load prediction. The model was compared with an improved sequence-to-sequence GRU with adaptive time dependence, and proved to outperform the latter under various input and output parameters.

Another approach proposed in this article (Wang & al., 2022) is multitasking, where several decoders are associated with a single encoder to perform joint multi-energy load prediction. The proposed multi-decoder structure facilitates concentration at different levels of attention on the encoder output representation due to multi-headed attention.

Clustering was also combined with the transformer to obtain accurate daily load forecasts. In (Zhao & al., 2021) the approach incorporates a similar day selection method using Light GBM which is employed to determine the individual weight of each additional feature, while the k-means algorithm is implemented to group similar days into distinct clusters.

1.4 Conclusion

This section has mainly focused on energy-related literature, with particular emphasis on short-, medium- and long-term load forecasting methods. The literature review highlighted the evolution of approaches, initially focusing on statistical methods, then exploring the emerging fields of Machine Learning (ML) and Deep Learning (DL).

Forecasting methods were examined through the prism of classical statistical approaches, commonly used ML techniques, and significant advances in DL. Attention then turned to a detailed explanation of the most prevalent ML methods in the literature, offering a comprehensive overview of the approaches that have shaped research in the field of energy forecasting.

The section also explored different Deep Learning methods, diving into innovative architectures such as CNNs with various activation functions that were explained to provide a thorough understanding of these cutting-edge approaches.

Chapter 2: Load Charge: Advanced Data Analysis

2.1 Introduction

This chapter examines the significant research on electrical load with an actual database. In order to efficiently manage the power grid, it is important to understand how electricity is consumed. This study attempts to achieve this goal by looking at actual data. Trends and behaviors will be studied, helping to plan and adjust power generation to meet changing needs.

The next section of this chapter examines statistical analysis, a method for extracting significant information from all of this data. This practical approach will assist in identifying time trends, periods of high demand, and other significant aspects of electricity consumption. The integration of statistical accuracy and graphical readability should provide meaningful information for more effective planning and more accurate energy decision-making.

2.2 Load Charge

Our everyday existence depends significantly on electrical load, which powers our appliances, lights our houses, and runs our modern technologies. Understanding the electrical load is essential to ensure a reliable and efficient electricity supply.

Electrical charge refers to the amount of electricity transported by an electric circuit. The importance of electrical load is its capacity to provide energy to power electronic devices, lighting systems, industrial equipment and many other appliances.

Power plants, which generate electricity from a variety of energy sources such as coal, gas, hydro and renewables, are often the beginning of electrical load generation. Through grid distribution, the electrical load flow produced by this conversion can be distributed to households, companies and other infrastructures.

To optimize the planning and management of the power grid, it is essential to collect data on electrical load. Smart meters and other devices keep track of the electricity consumed by customers, providing useful information on usage trends, peak hours and seasonal variations. This data is then examined to plan electricity production, improve energy efficiency and forecast future needs.

2.2.1 Algerian Load Charge

Algeria, as a developing nation, is facing rapid industrial and demographic growth, resulting in a growing demand for electrical energy. To face this challenge, Sonelgaz, the Algerian national electricity and gas company, occupies a crucial position in the production and distribution of electricity.

Sonelgaz uses natural gas, an abundant resource in the country, to satisfy its growing energy needs. This efficient approach helps to ensure the stability and accessibility of energy supplies, while the diversification of sources, particularly natural gas, serves as a solution tailored to national growth.

The use of natural gas for power generation aligns Algerian objectives with satisfying national energy demand, while at the same time enhancing the energy self-sufficiency of the country.

Therefore, the management of electricity production and distribution by Sonelgaz, based on the use of natural gas, represents a strategic vision that aims to promote sustainable development and satisfy the dynamic energy needs of Algeria in this period of growth.

2.3 Electricity Production

The production and distribution of electricity involves several steps, from the generation of energy to its delivery to consumers.

Natural gas, extracted from underground storage tanks, is transported to the power plant via pipelines. Before its use, natural gas may be filtered to remove impurities and moisture, ensuring the preservation of the turbines. Once it arrives, it often undergoes a compression process to increase its pressure and improve combustion efficiency.

After entering the combustion chamber, the natural gas is burned, generating heat which is then used to heat water in a boiler, producing high-pressure steam. The hot combustion gases pass through a gas turbine, driving its rotation and generating electricity. At the same time, the steam generated in the boiler is directed to a steam turbine, contributing to the production of additional electricity.

The rotating action of gas and steam turbines drives a generator, converting the resulting kinetic energy into electricity. This integrated process, combining natural gas combustion, steam generation and turbine operation, provides an efficient mechanism for converting energy into electricity.

After passing through the steam turbine, the steam is condensed into water in a condenser, then cooled either by cooling towers or by a heat exchanger using water from a local source. The gases resulting from combustion are treated to eliminate pollutants such as nitrogen oxides and particulates, before being released into the atmosphere.

The electricity generated is converted to a higher voltage for transmission over long distances through the power grid. Before reaching consumers, it is then transformed to lower voltages, suitable for residential or commercial use.

2.4 Data Analysis

The exponential evolution of data in companies represents a major challenge for personnel, requiring an increased capacity to manage huge quantities of information. In this context, companies, including critical organizations such as Sonelgaz, the electricity production and distribution company, have made available large databases covering the period from 2000 to 2018. These data include hourly daily electricity consumption records.

In order to make short-term predictions about electricity consumption, it becomes imperative to perform an in-depth analysis of this vast database. Understanding these data and studying their trends is a crucial step before any modeling or prediction is attempted. Analysis of this data involves the application of descriptive statistics techniques to obtain measures of central tendency and dispersion, as well as to explore the distribution of variables.

Data visualization is also essential in this process, allowing information to be represented graphically in order to highlight trends, patterns and significant variations. This step facilitates visual perception of relationships between data, and helps identify recurring patterns.

2.4.1 Descriptive Statistics

Database Overview

To give an overall view and understand the distribution of the data, a number of statistical measures were calculated throughout the analysis of 18 consecutive years of electricity consumption data as shown in the Table Tab1.

Hours	Max	Min	Mean	Std
1h	11956	2842	5478.772913	1689.245760
2h	11422	2618	5182.804729	1611.402635
3h	11040	2590	5015.457269	1529.642495
4h	10632	2513	4935.584473	1471.294068

Tab 1. An overview for the first 4 hours of the day

To achieve the results done in the Table, three crucial measures were applied:

Average

The average has been calculated to determine the central value representative of electricity consumption over the period. It provides an indication of the average level of demand over the years with the following function eq. 27:

$$AVG = 1/N \sum_{i=1}^{N} xi \tag{27}$$

Where;

N is the total number of years, and x_i represents the power consumption for year i.

Maximum and minimum

The maximum and minimum values of electricity consumption have been identified with the equations eq. 28 and eq. 29. These extremes allow us to visualize peaks and troughs in demand over the study period.

$$Maximum = Max(x1, x2, \dots, xN) \tag{28}$$

$$Minimum = Min(x1, x2, \dots, xN) \tag{29}$$

Standard deviation

The standard deviation was calculated in eq. 30 to assess the dispersion of the data around the mean. A higher standard deviation indicates greater variability, while a lower one suggests greater consistency in consumption levels.

$$STD = \sqrt{1/N \sum_{i=1}^{N} (xi - AVG)^2} \quad \textbf{(30)}$$

Coefficient of Variation

The coefficient of variation is a relative measure of the dispersion of data in relation to its average. It is often expressed as a percentage, and is calculated as the ratio of the standard deviation to the average, then multiplied by 100 to obtain the percentage. The general formula is shown in eq. 30:

$$CV = \left({}^{AVG}/_{STD}\right) * 100 \quad \textbf{(31)}$$

The coefficient of variation (CV) gives an indication of the relative variability of the data compared to the average. A higher coefficient of variation suggests greater dispersion of the data in relation to the average, while a lower coefficient of variation indicates less dispersion. This measure is useful for comparing the relative dispersion between different data series.

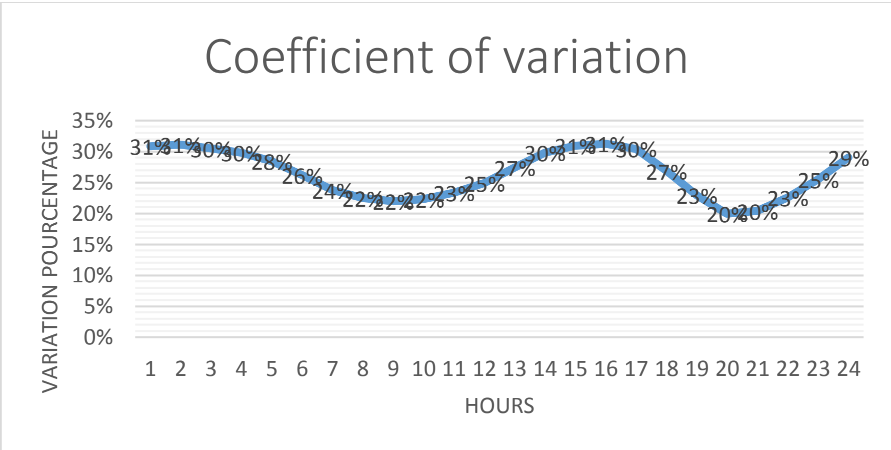


Figure 14. Coefficient of Variation Curve

A high coefficient of variation (CV) indicates increased dispersion or large fluctuations around the average. In the context of electrical energy consumption, a high CV indicates increased variability in consumption patterns. It is essential to note, however, that the curve illustrating the percentage of variation remains below 31%, as shown in the Figure 14, and that the average CV is around 26%. This observation suggests that the database has a relatively uniform distribution of values, with a moderate level of fluctuation around the mean.

Correlation

Pearson correlation is a statistical measure that evaluates the linear relationship between two continuous variables. It is calculated by dividing the covariance of the two variables by the product of their standard deviations. In the context of electrical energy consumption, the Pearson correlation could be used to evaluate the linear relationship between two time series of data, for example, the electrical consumption of an industrial sector in relation to total consumption.

The formula for the Pearson correlation between two variables X and Y is as follows in eq. 32:

$$Correlation = {cov(X,Y)} / {STD_x * STD_y} \quad (32)$$

If the Pearson correlation is close to 1, this indicates a strong positive correlation, while a value close to -1 indicates a strong negative correlation. A value close to 0 indicates a weak linear correlation.

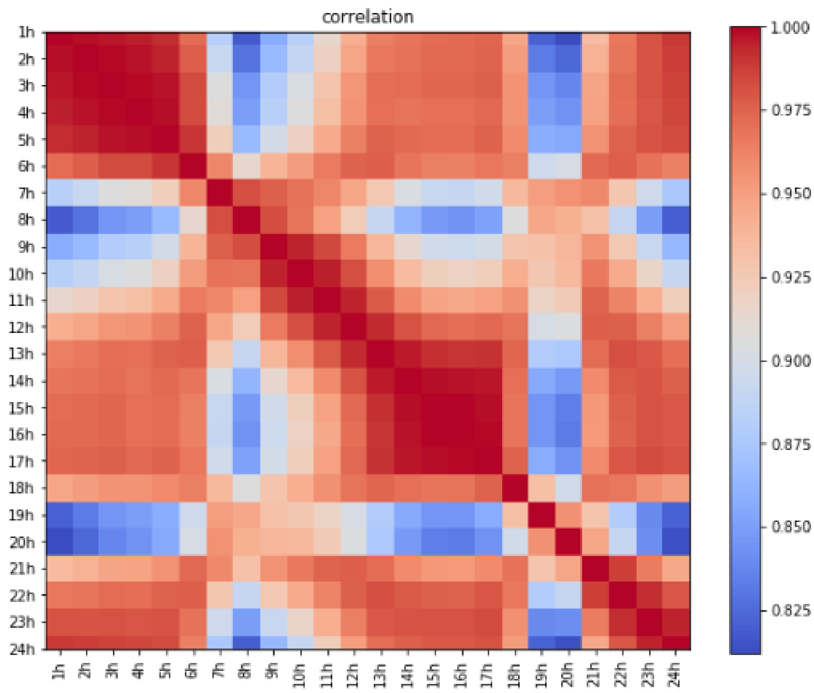


Figure 15. Correlation of the Data

In the Figure 15, horizontal examination highlights three well-defined time slots: from 1 am to 6am, from 12pm to 6pm, and from 10pm to midnight in this case (representing night, day and evening respectively). The observation also reveals the presence of the four seasons, with disturbances in the morning due to the daily movements of individuals (to work, school, etc.). Another disturbance occurs between 7pm and 9pm, probably due to the majority of people returning home.

2.4.2 Visualization

Data visualization provides an intuitive, visual understanding of temporal and seasonal trends in electricity consumption, making it easier to interpret the patterns and factors influencing energy demand.

Dispersion of the Data

The standard deviation of energy consumption per previous hour illustrates variations in energy demand throughout the day. The figure shows periods of high and low consumption, with a general tendency to increase over the years.

Analysis of the distribution shown in the Figure 16 shows a peak in consumption in August, corresponding to the summer period. This significant increase in electricity demand during this season can be attributed to the extreme heat, encouraging the intensive use of air conditioners. These observations underline the direct impact of climatic conditions on energy consumption patterns, in particular the high demand associated with the need for cooling during the hot months.

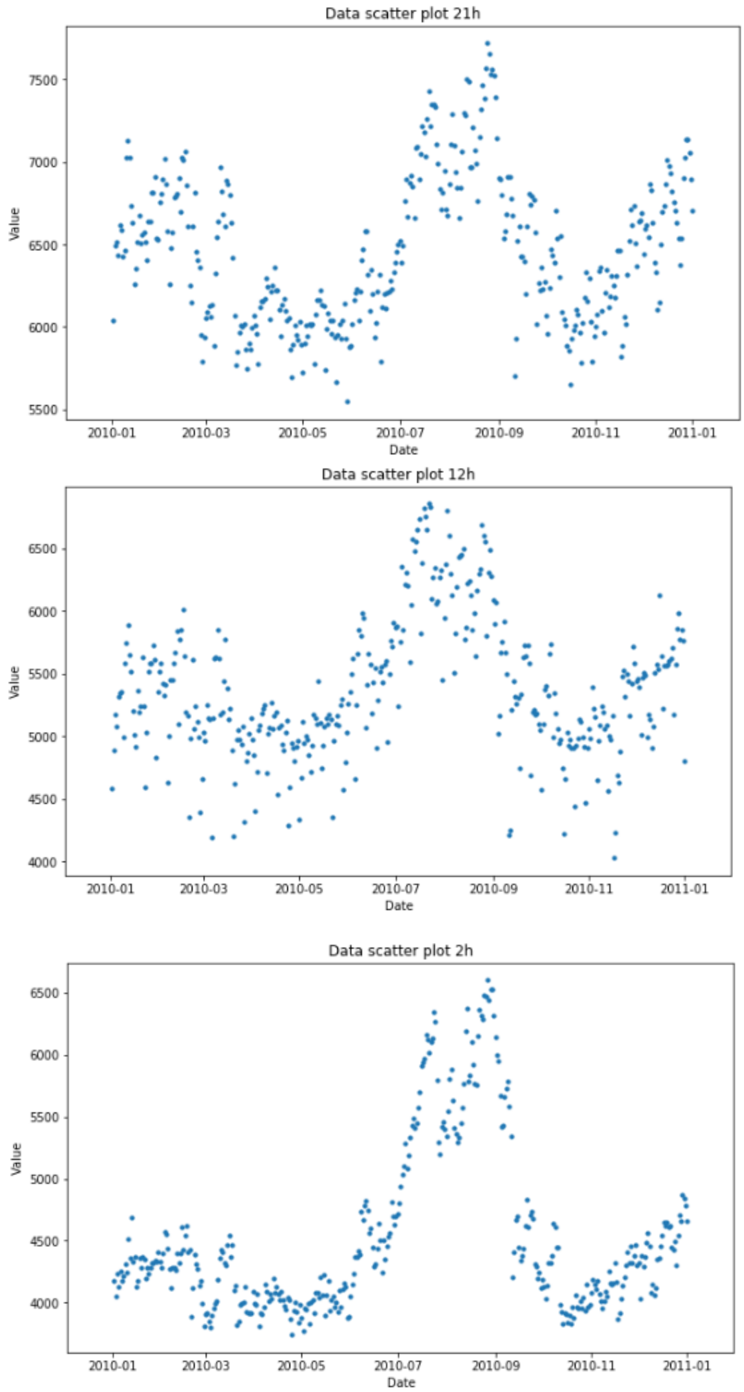


Figure 16. Dispersion of 2a.m, 12a.m and 9p.m

Clustering (K-means)

The use of the K-means algorithm for data visualization would make it easy to group different periods of power consumption into distinct clusters based on their similar characteristics. This could facilitate the identification of specific trends or behaviors in the data.

Applying K-means, different periods of the day could be grouped into clusters, offering a natural segmentation of power consumption patterns. This could be particularly useful for identifying seasonal trends, or unusual behavior.

K means has enabled researchers to observe how clusters evolve over the seasons, identifying significant variations in electricity consumption. In addition, the clusters could highlight atypical consumption periods, requiring special attention.

These clusters are formed by minimizing the intra-cluster variance, in other words the sum of the squares of the distances between the points in a cluster and its center (centroid).

How it works

Here is how the K-means algorithm works in detail:

Let X be the data set with n points, each point being a vector of dimension d.

Let C_i be the centroid of cluster i and S_i the set of points assigned to cluster i.

Centroid initialization: The process begins by randomly selecting k points in the dataset to form the initial centroids as shown in eq. 33.

K is the number of clusters formed.

$$C_i = X_{Random} \quad \textbf{(33)}$$

For i=1, 2,…, K.

Assigning points to clusters: Each data point is assigned to the cluster of which the centroid is closest in Euclidean distance as in eq. 34. This creates k initial clusters.

For j=1,…, n assign X_j to C_i where:

$$i = \arg min_k \left\| X_j - C_k \right\|^2 \quad \textbf{(34)}$$

Update Centroids: The centroids of each cluster are recalculated in eq. 35 by using the average of the coordinates of all the points assigned to it.

$$C_i = \frac{1}{|S_i|} \sum_{j \in S_i} X_j \quad \textbf{(35)}$$

For i=1,…, k.

Re-allocate Points to New Clusters: Data points are re-allocated to clusters according to the newly calculated centroids.

Repeat Steps 3 and 4: The centroid update and point reallocation steps are repeated until the centroids converge and the point allocations to clusters no longer change significantly.

The algorithm converges when the centroids do not change significantly, indicating that the clusters are stable.

The aim is to minimize the cost function, called "Inertia" (or WCSS - Within-Cluster Sum of Squares), defined as the sum of the squares of the distances from each point to its assigned centroid as shown in eq. 36:

$$\boldsymbol{Inertie} = \sum_{i=1}^{k} \sum_{j \in S_i} \left\| X_j - C_i \right\|^2 \tag{36}$$

K-means is sensitive to the initial choice of centroids, which can lead to different results. To minimize this effect, the algorithm is often run multiple times with different initializations, and the results are compared to select the best partition.

Applying K-means in the Data

Cluster analysis for one year shows three distinct groups, as illustrated in the Figure 17.

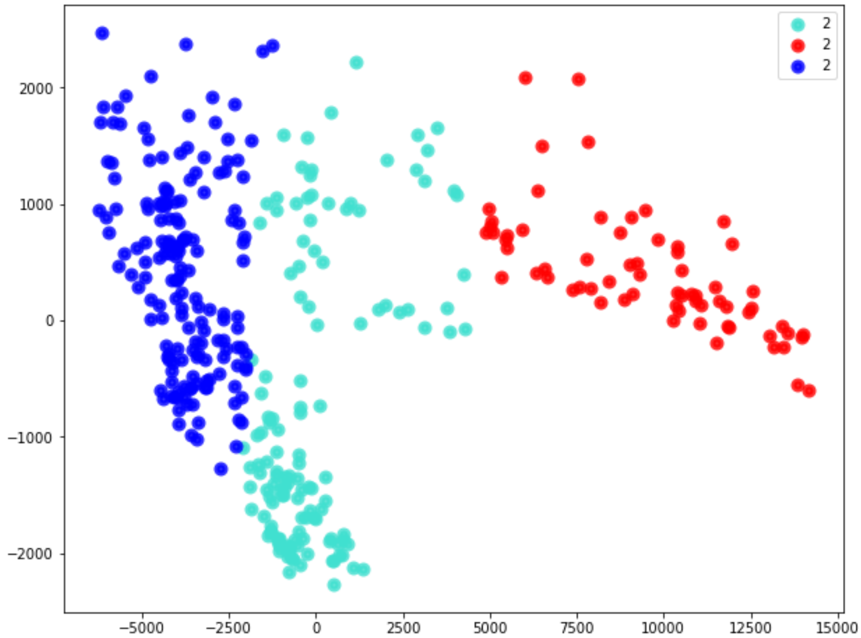


Figure 17. K-means results for one year

These clusters reflect the different seasons of the year, with one interesting observation: spring and autumn are grouped together in the same cluster due to their similar behaviors.

Winter

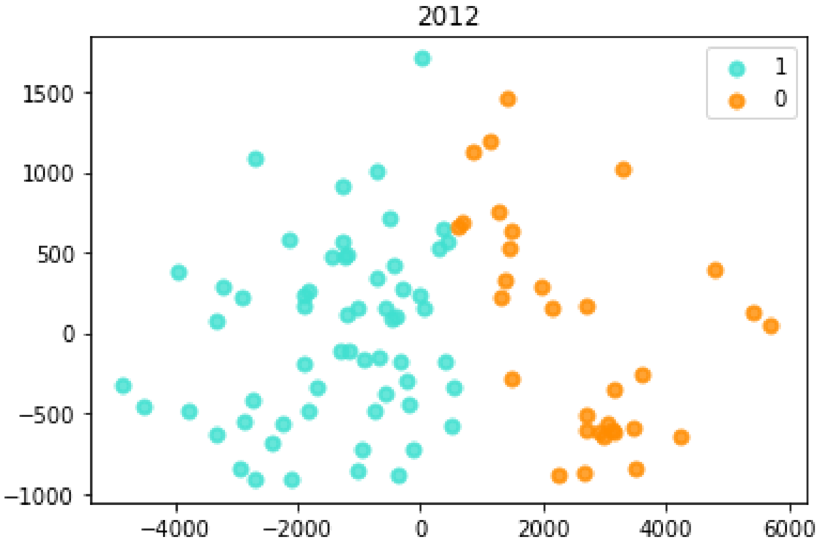


Figure 18. K-means results for winter

The observation in Figure 18 shows the two distinct groups present during the winter season, reflecting the variations between day and night. In winter, temperatures usually decrease at night and then warm up during the day. With longer nights in winter, the electricity consumption of the population increases at the end of the day and during the night. This tendency can be explained by the greater need for electricity due to climatic conditions and the longer duration of nighttime periods.

Summer

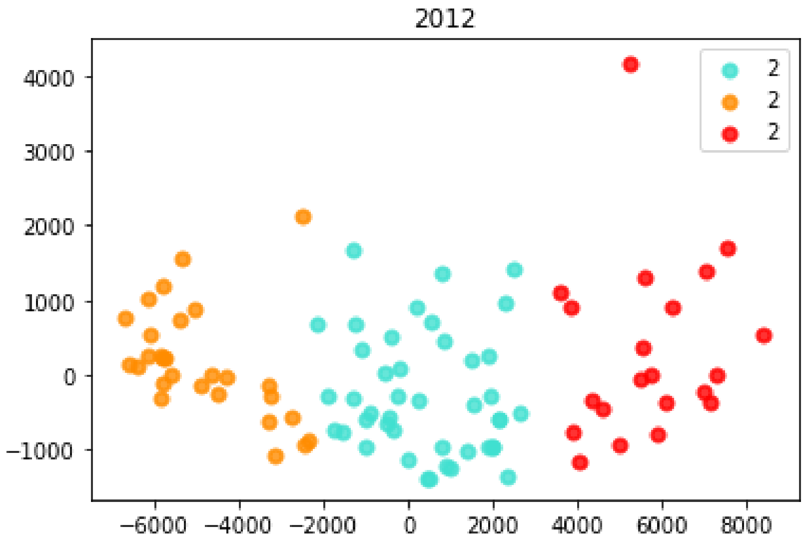


Figure 19. K-means results for Summer

In summer, a classification into three clusters can be seen in Figure 19. The first class corresponds to the end of spring, when temperatures progressively increase. The second class represents the summer season itself, and dominates the figure. Finally, the third class corresponds to the beginning of autumn, characterized by a progressive decrease in temperature. This division into classes reflects the distinct seasonal transitions and offers an insight into the variations in temperature and power consumption associated with these specific periods of the year.

Spring and Autumn

The two clusters identified for spring and autumn in Figure 20 and Figure 21can be explained by similar electricity consumption patterns during these transitional seasons.

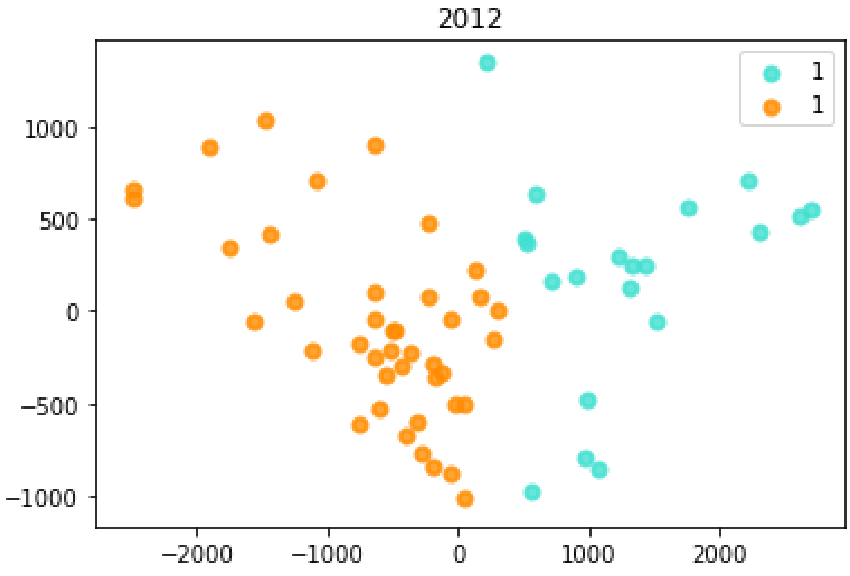


Figure 20. K-means result for Spring

During spring, temperatures begin to increase little by little, indicating the end of the winter period. The days get longer, and the need for heating decreases, but there may be an increased demand for electricity for seasonal activities such as gardening, outdoor work, and so on. The night-time hours become shorter, which can influence consumption patterns.

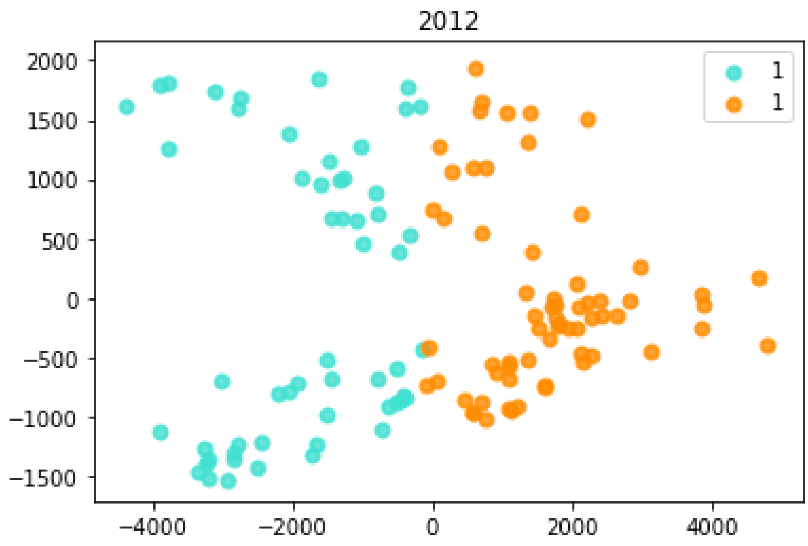


Figure 21. K-means result for Autumn

As autumn arrives, temperatures begin to fall, indicating the end of the summer period. Although the demand for air conditioning is reduced, the drop in temperature can lead to an increase in heating use. The days get shorter, causing changes in lighting habits and the use of electrical appliances.

2.4.3 Profile

The study of electricity consumption profiles consists in analyzing and understanding energy consumption patterns over a given period. This analysis enables the identification of trends, periods of high or low consumption, and the factors influencing variations in energy demand.

Consumption profile studies are also essential for electricity suppliers, network operators and energy researchers. It enables them to optimize production, manage load efficiently, and implement incentive measures for more sustainable energy use.

An overview of the whole year has been calculated to identify general trends, determine averages and understand annual fluctuations.

The annual analysis shows which months or seasons have the highest or lowest consumption.

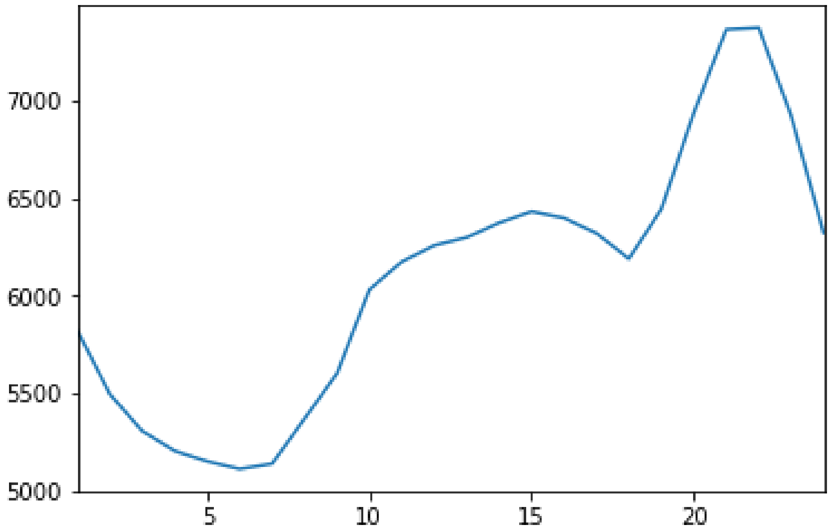


Figure 22. Profile curve of one year

Analysis of the electricity consumption profile over a full year in Figure 22 shows different temporal patterns. During the period from 12 a.m. to 6 a.m., a slight decrease in consumption is observed, indicating a period of night-time rest. At the start of the day, consumption progressively increases with the start of the activities such as work and school.

However, the most significant trend occurs in the evening. Between 7 p.m. and 8 p.m., consumption increases significantly, reaching a peak. This rise can be related to a variety of factors, such as evening activities, increased use of domestic lighting and appliances.

These temporal variations underline the importance of adjusting energy strategies to the specific needs of each period of the day.

Seasonal Profile

Winter

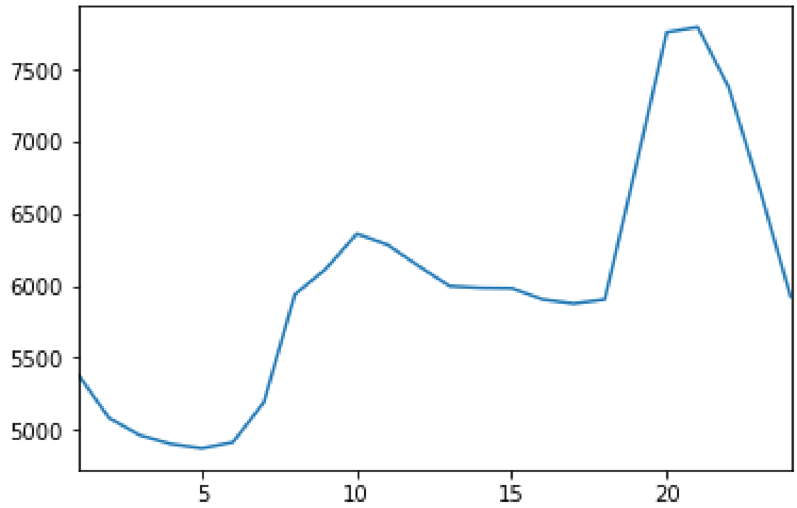


Figure 23. Profile curve of winter

In winter, peak electricity consumption occurs around 8 pm as shown in Figure 23. This observation is explained by the characteristics of winter, with its shorter days and longer nights. During this season, early and frequent use of electronic devices is more frequent, contributing to a high demand for electricity in the early evening. The need for earlier lighting increased heating and other indoor activities all influence electricity consumption during the winter months, causing a significant peak in the evening.

Summer

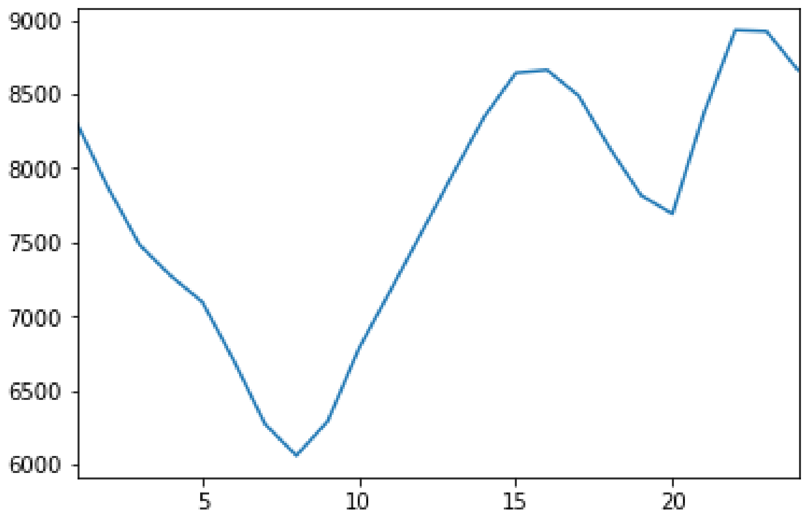


Figure 24. Profile curve of summer

In summer, the electricity consumption profile shows different trends in Figure 24. A significant decrease is observed early in the morning, followed by an increase at midday. However, the major difference resides in the shift in peak consumption to around 10 p.m. in the evening. This variation is related to the nature of summer, characterized by longer days and shorter nights. The morning decrease could be associated with lower use of lighting and electronic devices in the early hours of the day. The increase at midday could reflect increased activity during the day. Finally, the shift of the evening peak to 10 p.m. suggests prolonged use of electronic devices and higher electrical demand during summer evenings.

Spring

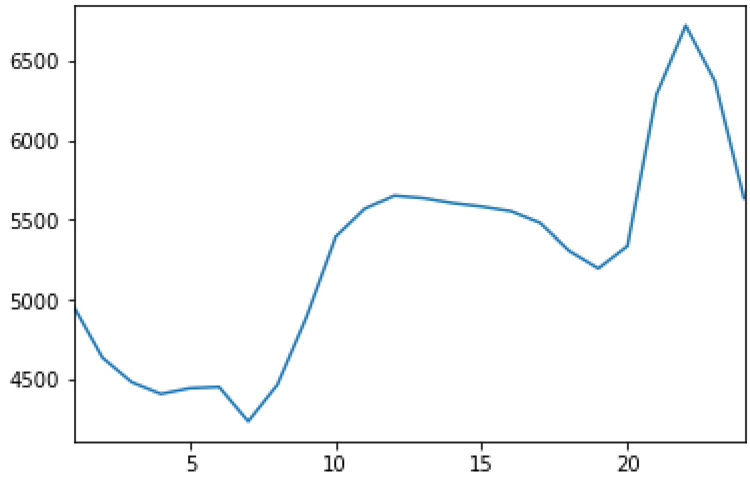


Figure 25. Profile curve of Spring

During the spring season, peak consumption remains unchanged from the summer, still occurring around 10 p.m. in the evening as shown in Figure 25. However, a slight increase in consumption is noted from 10 a.m. spring, like autumn, is a transitional period between the two main seasons of winter and summer. This particular pattern makes this season a little problematic, as it presents elements of each extreme, requiring adaptive management of energy demand.

Autumn:

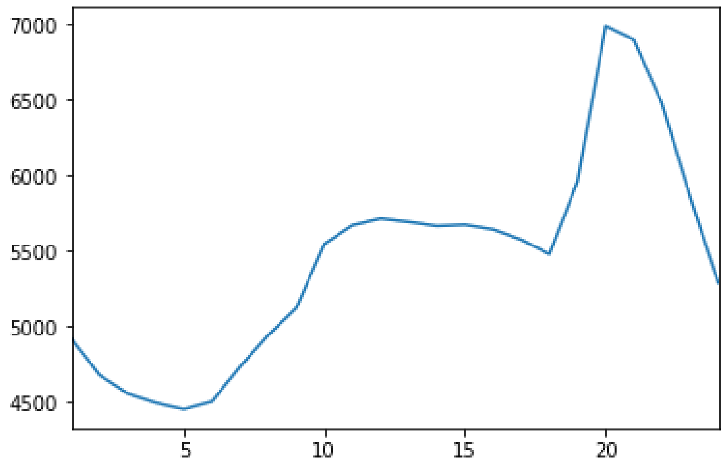


Figure 26. Profile curve of Autumn

In autumn, peak electricity consumption returns to its usual peak at 8 pm in Figure 26. This transition is explained by the fact that autumn days begin to behave more like winter days, with shorter days and longer nights.

Weekly profiles

Winter

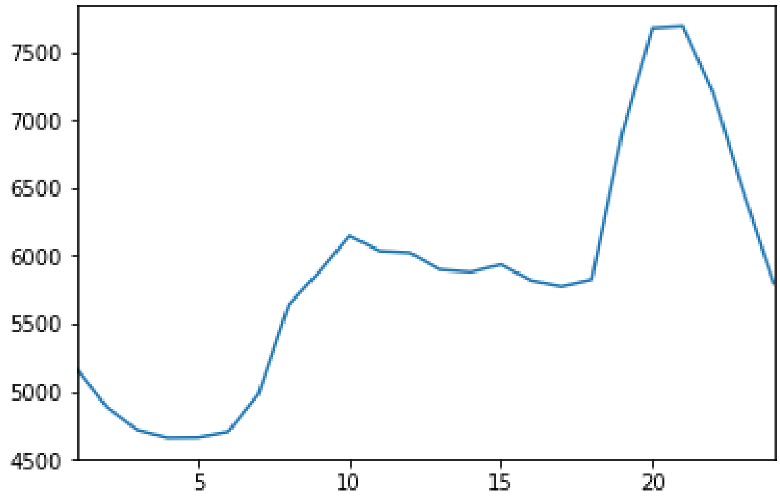


Figure 27. Weekly Profile in Winter

The analysis also looked at weekly patterns. In winter, the Figure 27 shows a decrease in consumption between 1 a.m. and 6 a.m., followed by an increase as people wake up, and so on. A disturbance is then observed between 10am and 5pm, leading to a consumption peak at 8pm.

Summer

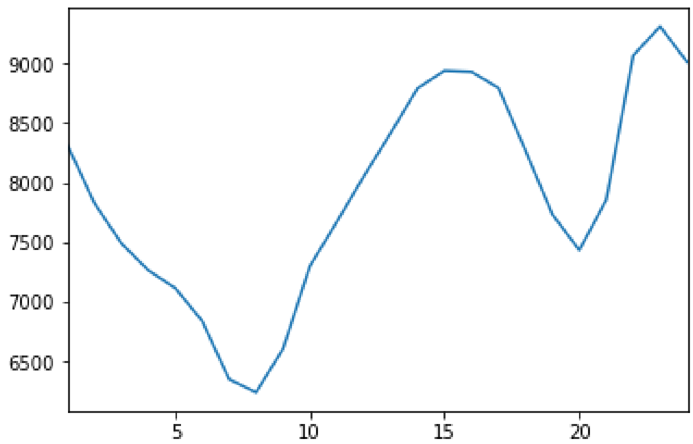


Figure 28. Weekly Profile in Summer

In summer, a slight difference between the seasonal profile and the weekly profile is noticeable in the Figure 28, particularly during the day between 10am and 5pm. It is notable that in the weekly profile, the increase in consumption is slightly more gradual.

Spring

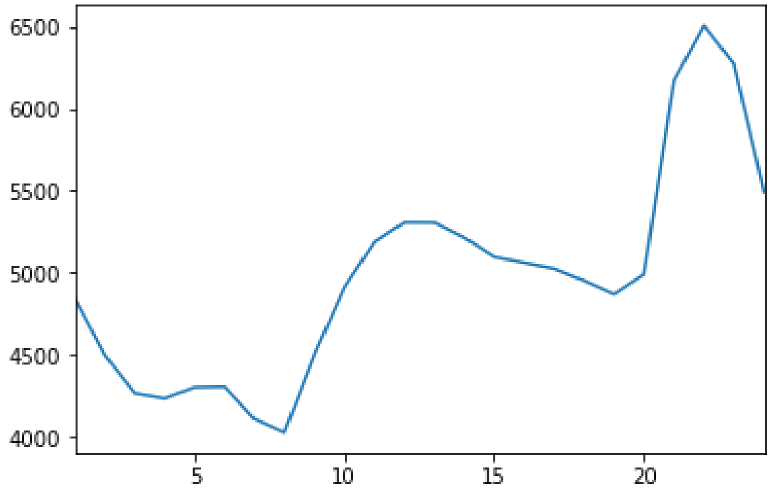


Figure 29. Weekly Profile in Spring

In spring, as the climate progressively warms up and the days get longer, peak consumption shifts to around 9pm as shown in Figure 29.

Autumn

In contrast to spring, often associated with beautiful days, autumn takes a reverse direction, with days becoming shorter and greyer. This explains the peak in consumption at 8pm shown in Figure 30, with the start of electronic appliance use a little earlier.

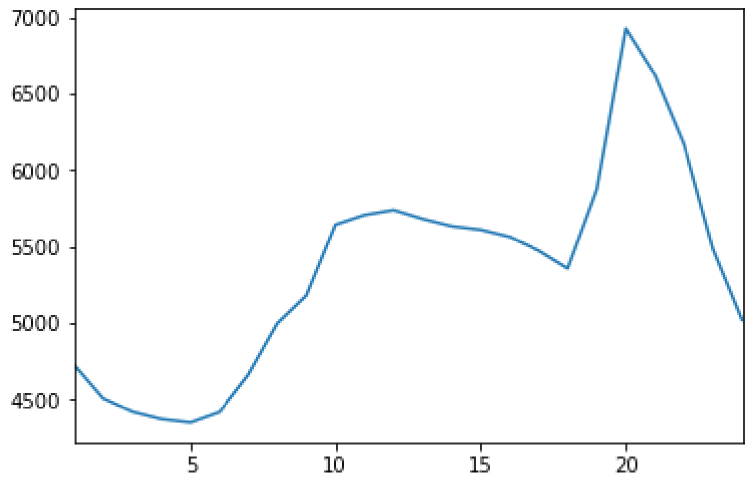
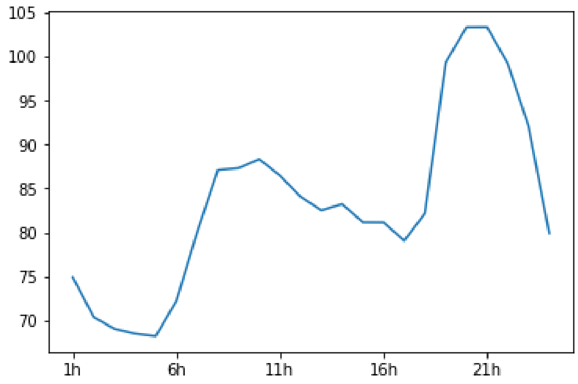
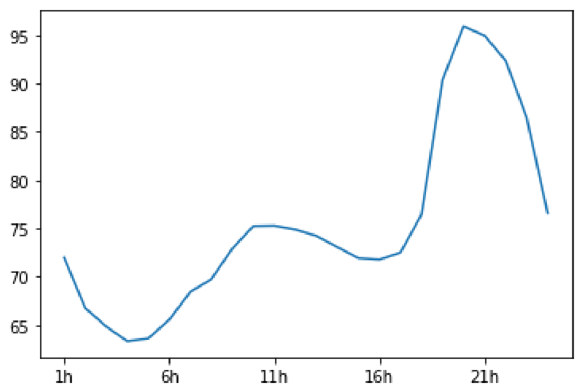


Figure 30. Weekly Profile in Autumn

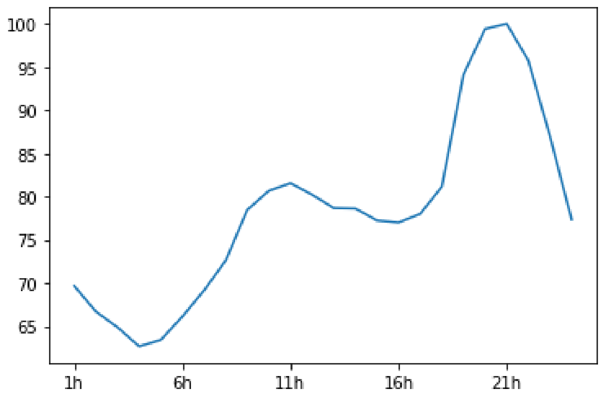
Daily Profiles



(A) Workday profile 2015



(B) Friday profile 2015



(C) Saturday profile 2015

Figure 31. Different daily profiles

For the daily profiles, the database has been subdivided into working days, Fridays as weekends and Saturdays, given that Saturday is a special day in Algeria when some people work while others in the public sector do not. The Figure 31 shows different profiles: the working day profile (A), the weekend profile (B) and the Saturday profile (C).

Disruptions observed on working days could be influenced by various factors related to the habits and daily activities of the population. At 3 p.m., this could correspond to a late afternoon period, when people may return from work or outdoor activities, leading to an increase in electricity consumption at home. At 4 p.m., there could be another peak in use, probably linked to domestic activities and the preparation of evening meals. These peak times may vary according to local habits, work schedules and cultural practices.

2.4.4 Trend and Seasonality

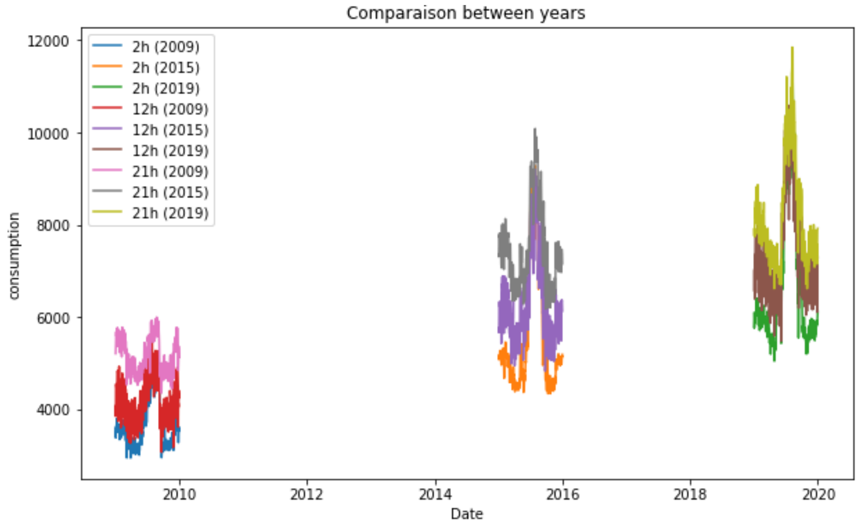


Figure 32. Comparison of different hours of the day in different years

The Figure 32 shows reduced electricity consumption at night, in contrast to higher consumption during the day, largely due to human activities. Moreover, an increasing tendency over the years is observed. These dynamics lead to take the study a step further by analyzing the frequency of temporal patterns, enabling to gain a better understanding of daily and seasonal variations in electricity consumption.

The study of frequency is crucial, as it looks at the regularity of time samples over the period of a day. This analysis, by focusing on daily intervals, helps to identify recurring trends, such as those observed in the power consumption shown in figure. This consumption varies on different time scales, such as daily, weekly, monthly and yearly, which can be understood by identifying the periodicity of trends.

Frequency is of particular importance for forecasting and modeling time series. By identifying the frequency of seasonal or cyclical patterns, we can select suitable forecasting models that capture these temporal patterns, thus improving forecast quality.

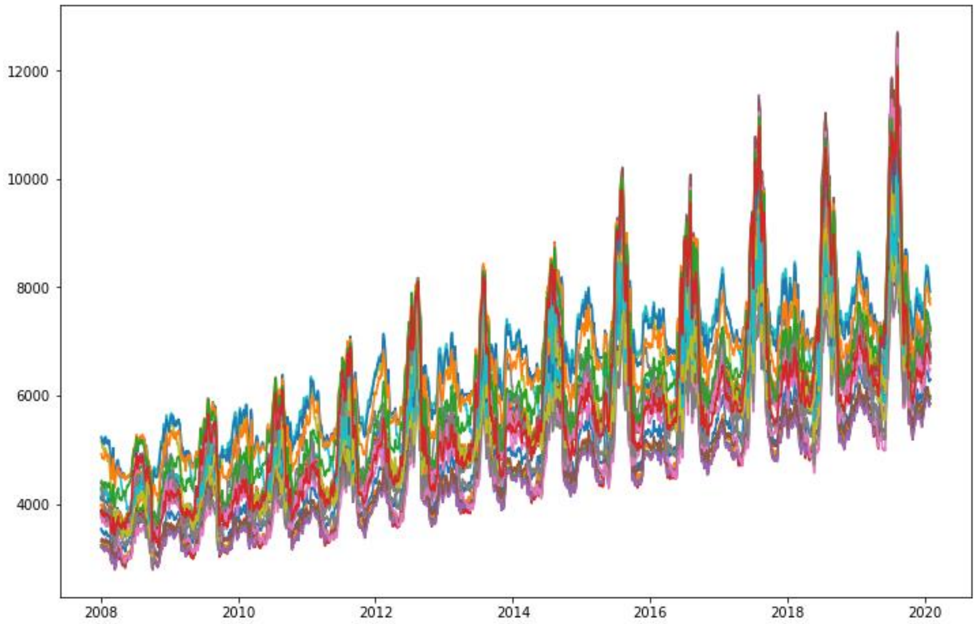


Figure 33. Trend cycle for each year

The Figure 33 illustrates visually a recurring cyclical pattern, with an ever-increasing amplitude over the years. To explore this dynamic further, it is essential to determine the periodicity of the cycle on an annual basis. In this context, periodicities are identified for one day, 217 days and 273 days.

The frequencies of 217 and 273 days, equivalent to around 9 months, suggest the possibility that they represent specific seasons and seasonal cycles. In other words, these periodicities could be associated with seasonal variations in electricity consumption. Daily frequency, on the other hand, captures the various occurrences that influence power consumption over the duration of a typical 24-hour day. In this way, the frequency analysis provides an in-depth perspective on the temporal and seasonal patterns present in the time series under study.

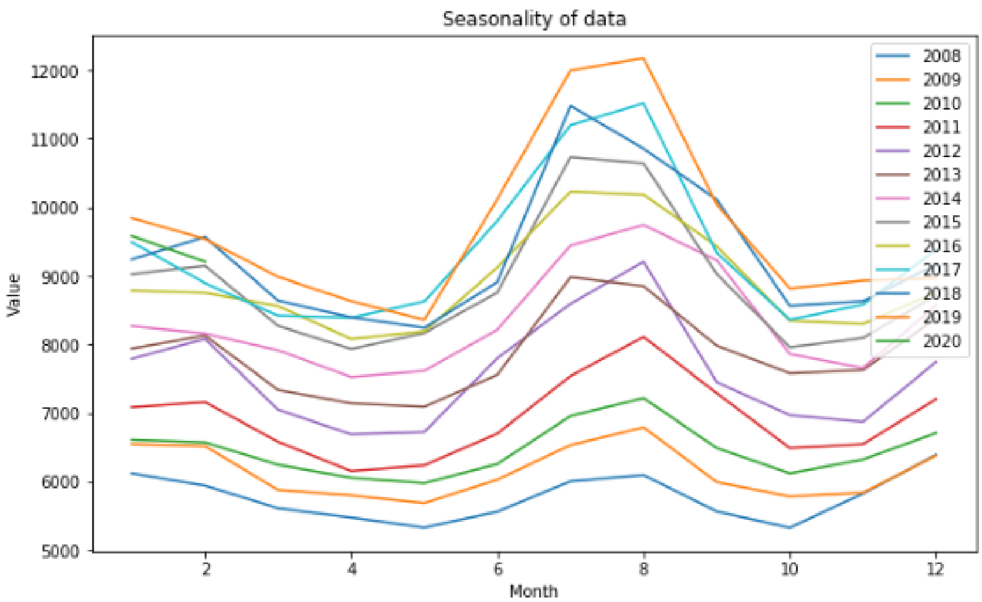


Figure 34. Profile for Seasonality

The Figure 34 shows that, despite variations from one year to the next in the amplitude of consumption, the temporal pattern representing the seasons remains relatively unchanged. Peak consumption remains constant in summer, a period generally associated with high temperatures. These findings highlight the regularity of seasonal consumption patterns, providing valuable information for energy supply planning and management.

2.4.5 Normalization and Metrics

Normalization and standardization

Data standardization and normalization are essential first steps to ensure model quality and performance. These techniques prepare the data to better adapt to the requirements of the algorithms, which often improves the performance and interpretability of the models.

Standardization transforms data in order to follow a common scale, while normalization adjusts them to fit within a specific range. These data analysis techniques ensure that the different variables are comparable. This is essential for algorithms that use distance measures or are sensitive to the scale of variables.

The exponential increase in data and the annual growth in electrical load, as demonstrated by the preceding analyses, highlight the imperative of normalizing and standardizing data. This step is essential to align the different scales and ensure that the models to be tested converge efficiently. Normalization and standardization facilitate the fair comparison of variables, accelerate model convergence, and contribute to the robustness and efficiency of subsequent analyses. The adoption of these preliminary processes enables massive data to be handled in a more consistent way, improving the quality of results in the context of electrical load analysis.

Normalization

Normalization, also known as feature scaling, adjusts feature values to a comparable scale. As a general rule, normalization resizes feature values to a specific scale range, either [0, 1] or [-1, 1]. The most commonly used normalization methods include Min-Max scaling and Max Abs scaling.

Mathematically, Min-Max scaling is defined as follows in eq. 37:

$$\boldsymbol{X_{sc}} = \frac{\boldsymbol{X} - \mathbf{min}(\boldsymbol{X})}{\mathbf{max}(\boldsymbol{X}) - \mathbf{min}(\boldsymbol{X})} \tag{37}$$

For Max Abs scaling, it is formulated as follows in eq. 38:

$$\boldsymbol{X_{sc}} = \frac{\boldsymbol{X}}{\mathbf{max}(|\boldsymbol{X}|)} \tag{38}$$

Given the constantly increasing demand for electricity, finding a reliable minimum load value for the normalization process is becoming complex. In order to avoid problems associated with minimum load values, such as loss of information, and to ensure the continued relevance of predictive models, the Max Abs Scaling technique is adopted for normalizing load data. This approach takes into account the evolution of electricity demand, thus ensuring that the proposed forecasting tool remains efficient and accurate in its future use.

Standardization

Standardization aims to give all characteristics a standard distribution, with a mean of zero and a standard deviation of one.

Commonly used methods for standardization are standard z-score (or normal standardization) expressed with the eq. 39:

$$X_{st} = \frac{X - Avg\ (X)}{STD} \quad \textbf{(39)}$$

Where x is the observation, $Avg\ (x)$ is the mean value of the observations,
And STD is the standard deviation of the observations.

In the context of electrical load forecasting, standardization is crucial to ensure that models are robust and perform effectively, particularly when data show significant variation. Using these normalization and standardization techniques, electrical load data can be optimally prepared for model training, improving the convergence of algorithms and their ability to generalize on new data.

Metrics

Accuracy metrics are indispensable tools for assessing the performance of forecasting models. Quantifying the accuracy of predictions, these measures provide crucial insights into the effectiveness of a model, enabling informed decisions to be made.

Analysis of models and error trends provides companies with valuable information on the factors contributing to inexactitudes. Metrics of accuracy play an essential role in enabling companies to compare different forecasting models with each other. Comparing the performance of several models using a consistent series of metrics, organizations are able to identify the most accurate and reliable model for their specific needs. This methodical approach optimizes the choice of forecasting model, enhancing the reliability of predictions for practical applications.

To evaluate the effectiveness of the electrical load prediction models, a number of metrics are used to quantify the accuracy of the predictions in comparison with actual values. These metrics offer a detailed insight into model performance, contributing to a more in-depth evaluation of the forecasting approaches.

Mean Absolute Error (MAE):

The Mean Absolute Error (MAE) represents a simple measure of accuracy, evaluating the average of the absolute differences between predicted and actual values. This metric offers an intuitive evaluation of the global error of the prediction. MAE is calculated by summing the absolute differences for each observation, then dividing this sum by the total number of observations.

To calculate the MAE, the first step is to obtain the absolute differences between the actual and predicted values for each observation. These absolute differences are then summarized and finally divided by the total number of observations.

The MAE formula can be represented as follows in eq. 40:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_p\right| \tag{40}$$

Where:

y_i is the actual value, y_p is the prediction, and n is the total number of observations.

Root Mean Square Error (RMSE):

The Root Mean Square Error (RMSE) represents a commonly used measure of accuracy, calculating the square root of the mean square difference between predicted and actual values. RMSE is particularly relevant when large errors have a significant impact on the evaluation. Differing from MAE, RMSE gives greater weight to major errors, making it more reactive to outliers. Mathematically, it is just the square root of the MSE as shown in eq. 41. It is frequently used to have a measurement in the same unit as the predicted variable.

$$RMSE = \sqrt{MSE} \tag{41}$$

Mean Square Error (MSE):

The Mean Square Error (MSE) is a measure of accuracy that calculates the average of the squares of the differences between predicted and actual values. The MSE, does not involve measuring the square root of this mean, which means that it gives greater weight to large errors without proportional adjustment. The MSE is used to assess the overall quality of predictions, but can be sensitive to the occurrence of outliers because of the absence of the square root in the calculation; .

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_p\right)^2 \tag{42}$$

Mean Absolute Percentage Error (MAPE):

The Mean Absolute Percentage Error (MAPE) is a measure of accuracy frequently used in forecasting. It quantifies the average percentage difference between predicted and actual values. This metric provides a relative evaluation of forecast accuracy, making it appropriate for comparing the performance of different time series or predictive models. MAPE is expressed as a percentage as in eq. 43, and lower values indicate greater forecast accuracy.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_p}{y_i}\right| * 100 \tag{43}$$

Correlation Coefficient (R):

The Correlation Coefficient (R) is a statistical measure evaluating the strength and direction of the linear relationship between two variables, in this case, predicted and actual values in the context of forecasting. It ranges from -1 to 1, where -1 indicates a perfect negative correlation, 1 a perfect positive correlation, and 0 an absence of linear correlation. The correlation coefficient can be used to determine the extent to which predicted values linearly follow actual values, offering crucial insights into the quality and nature of the prediction.

$$R = \frac{\sum_{i=1}^{n}\left(y_i - y_p\right)\left((mean(y_i)) - \left(mean(y_p)\right)\right)}{\sqrt{\sum_{i=1}^{n}\left(y_i - y_p\right)^2 \sum_{i=1}^{n}\left((mean(y_i)) - \left(mean(y_p)\right)\right)^2}} \tag{44}$$

Coefficient of Determination (R^2):

The Coefficient of Determination (R^2) is a measure that represents the proportion of variance in the dependent variable (actual values) explained by the independent variable (predicted values) as shown in eq. 45. In other words, it quantifies the accuracy of the model by evaluating the proportion of variation in the actual data that can be explained by the predictive model. The R^2 varies from 0 to 1, where 0 indicates that the model explains no variance, and 1 means that it explains all variance. A high R^2 indicates a better fit of the model to the actual data as shown in eq. 45. However, it is important to interpret this with caution, as a high R^2 does not necessarily guarantee high accuracy in predicting actual values.

$$\boldsymbol{R^2 = 1 - \frac{\sum_0^n (y_i - y_p)^2}{\sum_0^n (y_i - mean(y_i))^2}} \qquad \textbf{(45)}$$

R^2 is calculated as the proportion of total variation explained by the model. An R^2 of 1 indicates a perfect model fit, while an R^2 of 0 indicates that the model explains no variation.

Accuracy metrics enable to objectively evaluate the performance of forecasting models. They show how closely predictions correspond to actual values, eliminating the subjectivity associated with simple visual observation.

These measures help to compare different forecasting models, enabling to choose the most appropriate for a specific task. With common criteria, decision-making becomes simpler.

In addition, accuracy metrics are essential for monitoring the evolution of model performance over time. Evaluating them regularly enables to detect improvements or potential problems, giving the opportunity to take measures to improve the quality of the forecasts. These metrics are indispensable tools for guiding the choices we make and ensuring the success of any future forecasts.

2.4.6 Conclusion

This chapter is dedicated to an in-depth exploration of the Algerian electricity consumption database. The different variations and profiles present in the database were carefully examined, highlighting changes over time. This deep data analysis enabled a better understanding of consumption patterns, seasonal trends and significant changes in the electrical behavior of users. A close analysis of the data has enabled the extraction of crucial information for the development of accurate forecasting models adapted to the specific needs of the Algerian energy situation.

Chapter 3: Load Charge forecasting: Contributions

3.1 Introduction

The importance of short-term forecasting is a particular challenge in the context of electricity load management in Algeria. Due to the dynamic nature of daily habits, weather fluctuations and other intrinsic factors, an accurate understanding of short-term energy demand is crucial to ensuring stable and efficient power distribution.

Machine learning and deep learning models are becoming essential tools in this search, as they can accurately anticipate variations in energy demand. These models are capable of taking into account complex, non-linear parameters, such as seasonal behavior, public holidays and even exceptional events.

In the specific context of electricity load forecasting in Algeria, these models can be adapted to capture local features and seasonal trends. The increased accuracy of short-term forecasts helps to plan power generation and distribution more efficiently, thus helping to avoid network overloads and optimize the use of available resources.

Furthermore, in order to improve model performance, we are integrating the innovative concept of transfer learning into our approach. This method will improve the performance of models in learning and adapting to the various aspects of Algerian energy behavior, drawing on knowledge acquired in other similar fields. This increases the robustness of the models in the face of unforeseen variations and changes in consumption patterns.

In the context of this research, the test results are carefully analyzed, and the advantages and limitations of each model are discussed in order to provide a comprehensive perspective. The aim is to deploy a forecasting strategy that is not only accurate, but also adapted to the complexities of individual behavior in this particular context.

3.2 Importance of forecasting

The use of machine learning (ML) and deep learning (DL) techniques in power load prediction has gained crucial importance, providing significant benefits in terms of optimizing power generation and distribution.

ML and DL models are able to process large, complex data sets, featuring weather parameters, consumption patterns and other relevant variables. This enables more accurate electricity load forecasts, enabling better anticipation of variations in demand.

They can provide real-time forecasts, which are essential for dynamic load management. Being able to respond quickly to demand variations enables more efficient use of resources and helps avoid overload situations.

Accurate power load forecasts allow energy production to be optimized. Producers can adjust their production according to forecasts, minimizing costs and improving the overall efficiency of the energy supply network.

Moreover, by accurately anticipating demand, the company can adjust its production in a way that encourages the use of cleaner energy sources, thereby helping to reduce greenhouse gas emissions and promote more sustainable energy production.

In the modern era, where data occupies a predominant place, improving energy demand forecasting is at the heart of strategic challenges. For this study, Sonelgaz provided us with a dataset covering a period of 19 years, from 2000 to 2019, and including hourly electricity consumption data for a total of 24 hours of consumption per day. This extensive volume of data represents a valuable resource for the implementation of machine learning and deep learning models, which are particularly appropriate for this type of data to ensure reliable prediction.

To optimize model performance, the database was divided into two distinct parts. Eighty percent (80%) of the data was assigned to model training, enabling the algorithms to assimilate the underlying trends and patterns in power consumption over the defined period. The other 20% was reserved for the testing and validation phase, providing an objective evaluation of the predictive capacity of the models on unknown data.

The experimental phase was characterized by the implementation of several model architectures with different parameters. The aim of this diversity of approaches is to evaluate the performance of each model configuration, and to identify the optimum combination that will deliver the best possible model. This rigorous approach underlines this commitment to providing accurate and reliable electrical forecasts, making full advantage of the wealth of data available to management.

3.3 Application of Machine Learning Models

The use of ML models is increasingly emerging as an innovative solution for accurately anticipating complex fluctuations in demand. This part explores the many opportunities offered by ML in the energy sector, highlighting the technological advances that are reshaping the way companies approach the planning and management of energy resources.

3.3.1 Linear Regression

Linear regression is an essential approach, offering an analytical approach to modeling and anticipating trends in electrical load. This exploration focuses on the implementation of linear regression, in particular on how it can be used to capture variations in electricity consumption from peaks observed in the previous section over three consecutive days.

The algorithm implemented follows a methodical sequence.

Determining Peak Consumption: The process begins by identifying the maximum consumption peaks for each day in the database as shown in Table tab 2. These peaks are essential for detecting significant fluctuations in energy demand.

	maxi
Date	
2011-01-01	7019.25
2011-01-02	7182.75
2011-01-03	7245.00
2011-01-04	7200.50
2011-01-05	7143.25

Tab 2. Example of identifying the peaks

Characteristic preparation: Before starting the model training process, a crucial step is to prepare the data to capture temporal trends.

The fundamental idea here is to extend the dataset by creating additional columns, each representing peak consumption ("maxi" in the previous table) at different points in time. The practical result is the creation of columns such as "max-1", "max-2", and "max-3" (in the following table tab 3), where each value in these columns is derived from shifting the value of "max" down through a number of time intervals.

	maxi	max-1	max-2	max-3
Date				
2011-01-04	7200.50	7245.00	7182.75	7019.25
2011-01-05	7143.25	7200.50	7245.00	7182.75
2011-01-06	6830.50	7143.25	7200.50	7245.00
2011-01-07	6550.00	6830.50	7143.25	7200.50
2011-01-08	6823.50	6550.00	6830.50	7143.25

Tab 3. Example of the calculated three previous days

The "max-1" column captures the value of "maxi" at time t-1 (the previous day).

The "max-2" column captures the value of "maxi" at time t-2 (two days before).

The "max-3" column represents the value of "maxi" at time t-3 (three days before).

These new columns enrich the dataset by including information on energy consumption at different times in the past, giving the model a more extended temporal perspective.

Linear Regression Model training: The consumption peaks identified in this way are used to train a linear regression model. The modeled function takes the form of the eq. 46:

$$\boldsymbol{Y = a(t-1) + b(t-2) + c(t-3) + d,} \tag{46}$$

Where a, b, and c represent the coefficients to be determined,

And t-1, t-2, t-3 are the consumption peaks of the previous three days.

The database is then divided into two distinct parts: 80% for model training and 20% for testing and validation, and the percentage accuracy rate is calculated to assess the ability of the model to generalize on data not used during training.

Obtained Results

The resulting function is eq. 47:

$$\boldsymbol{Y = 1,16081242(day1) - 0,44436053(day2) + 0,25249535(day3) + 232,756074973} \quad \textbf{(47)}$$

This Function provides a model fitted to the shape of the data, enabling reliable forecasts based on the history of peak consumption. This approach offers an analytical and pragmatic method for improving electrical load prediction, underlining the central role of linear regression in this process.

metrics	Score
R^2	0.98
RMSE	165
MAE	127

Tab 4. Linear Regression Results

In the table Tab 4 the R2 (coefficient of determination), RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) metrics provide valuable insights into the performance of a linear regression model.

R2 measures the proportion of variance in the dependent variable (electric charge in this case) explained by the model. A coefficient of determination close to 1 indicates an excellent fit of the model to the data, suggesting a high ability to explain and predict variability in electrical charge. The obtained R2 result is 0.98, meaning 98% good prediction, which is an excellent result.

RMSE measures the mean square error between predicted and actual values, with a square root applied at the end. The lower the RMSE, the better the model's accuracy. An RMSE of 165 suggests a mean prediction error of 165 Megawatts in the electric charge measurement, which is a fairly acceptable result.

MAE measures the mean of the absolute error values between predicted and actual values. MAE is also a measure of model accuracy, and a lower value is preferable. The finding MAE is 127 indicates a mean absolute prediction error of 127 Megawatts in the measurement of electrical load.

With this algorithm, the score was 98%, a very satisfactory result. The resulting function provides a model that fits the shape of the used data perfectly. This will enable to predict future consumption with confidence, based on the peak consumption of the previous three days.

3.3.2 Multiple Regression

In statistics, as explained in previous chapters, multiple linear regression is an extended mathematical method of simple linear regression. Its aim is to model variations in an endogenous variable as a function of simultaneous variations in several exogenous variables. In the context of this study, a multiple regression analysis could show a positive correlation between daily electricity consumption and temperature fluctuations throughout the day.

However, before starting these analyses, it is imperative to verify the existence of a linear relationship between peak consumption and daily temperature.

The profile study made earlier showed that electricity consumption peaks during the summer. This finding has enabled to include the maximum temperature of each day in the database, thus enlarging the range of the analysis. The main objective is to explore and quantify the relationship between peak electricity consumption and temperature variations, thus enhancing the relevance of multiple linear regression in the analytical approach.

The following two figures (Figure 35 and 36) show scatter plots to verify linearity between power consumption and temperature, using the hours of 1am and 5pm as examples.

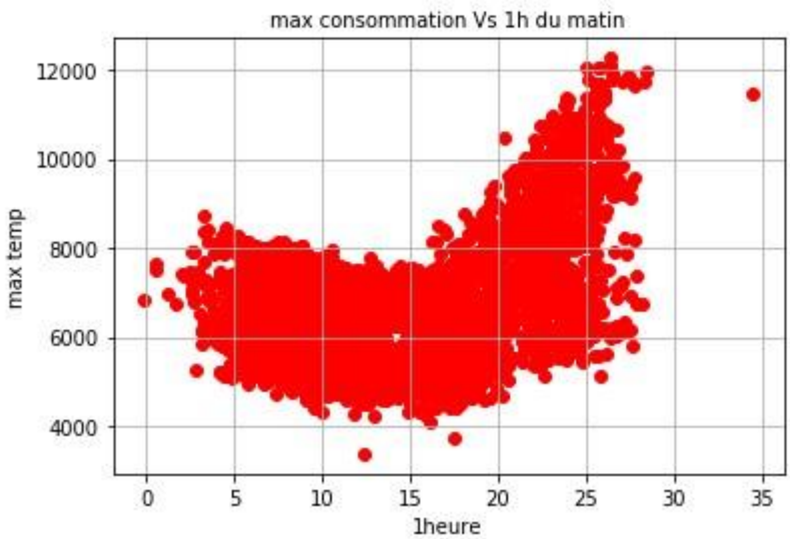


Figure 35. Scatter Plot for 1am

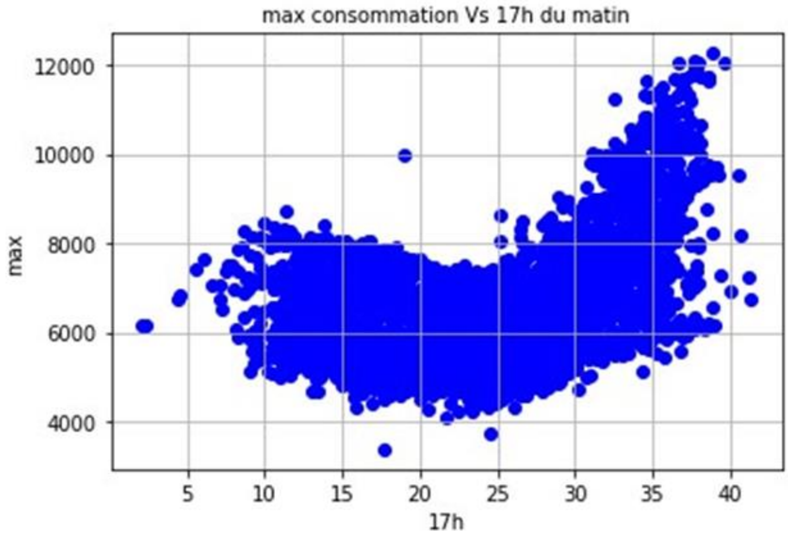


Figure 36. Scatter Plot for 5pm

It is observed that there is effectively a linear correlation between peak consumption and climatic variations over the day according to the scatters plot. More specifically, electricity consumption increases in proportion to temperature rise.

Once the algorithm has been tested, we obtain a function of four variables, similar to simple linear regression which is eq. 48:

$$Y = a(X) + b(Y) + c(Z) + d, \tag{48}$$

Where: *a, b, c* are the coefficients;

X represents the temperature at 1 a.m.;

Y represents the temperature at 10 p.m.;

Z represents the temperature at 8 p.m.

And *d* represents the peak electricity consumption for the day.

These times were chosen to capture different phases of the day, covering early morning, mid-day and early evening. This enables to model daily temperature variations and see how they affect electricity consumption at different times.

Obtained Results

Using the same algorithm as for the linear regression model, but replacing the maximum consumption data for the previous three days with the corresponding maximum temperatures, the results obtained are presented in the following table Tab 5:

metrics	Score
R^2	0.98
RMSE	110
MAE	97

Tab 5. Multiple Linear Regression Results

The results of the RMSE and MAE metrics for the model based on maximum temperatures show a slight improvement, probably due to the strong correlation between temperature and power consumption, compared with the simple linear regression model.

3.3.3 Support Vector Regression

SVR is particularly adept at modeling complex, non-linear relationships between variables. In the case of your study, where power consumption may be non-linear, SVR offers a robust approach to capturing these complex patterns.

SVR can also be powerful in its ability to perform generalization, allowing learned relationships to be generalized to unobserved data. This could be crucial for anticipating power consumption in climatic conditions that have not been specifically observed in your dataset.

In particular, it offers the possibility of adjusting various hyper parameters, including the choice of kernel and regularization parameters. This enables fine-tuning of the model for optimum performance.

The kernel parameter is one of the most critical components of SVR. It influences the transformation of input data and plays a central role in the ability of the model to capture complex relationships. Three commonly used kernel types are the linear kernel, the polynomial kernel and the Radial Basis Function (RBF) kernel. Each of these kernels has distinct characteristics that make them suitable for specific contexts.

The linear kernel is suitable for simple linear relationships between variables. The polynomial kernel is useful for modeling higher-degree non-linear relationships, requiring adjustment of the parameter d, which represents the order of the polynomial. The RBF, or Gaussian, kernel is frequently chosen for its flexibility in modeling complex nonlinear relationships, with the σ parameter controlling the width of the kernel.

The choice of kernel depends on the nature of the relationship to be modeled, and it is often necessary to experiment with different kernels and parameters to determine which provides optimal performance for a specific dataset. Kernel selection is therefore a crucial step in the SVR implementation process, and requires an iterative approach to achieve a well-fitting model.

Obtained Results

To refine the performance of the SVR model, a grid search approach was implemented. This approach, known as grid search, involves a systematic exploration of different combinations of two key hyper parameters: the kernel and the regularization parameter *C*.

Using this grid search approach shown in Tab 6, each possible combination of kernel and *C* was evaluated, training the model and assessing its performance at each iteration. The aim was to identify the optimal configuration that maximizes forecast accuracy for power consumption as a function of temperature.

Hyper parameters	values
Kernel	[Sigmoid, Linear, RBF]
C	[0.1, 1, 5, 10, 100]

Tab 6. Table of list of the selected hyper parameters

Kernel	C	MAPE	RMSE	MAE
RBF	5	5%	514	348
	10	4%	414	277
	100	1%	103	76.3
sigmoid	5	18%	1313	1094
	100	17%	1300	1095
linear	10	32%	2345	2219
	100	40%	2198	2145

Tab 7. Obtained Result SVR

After examining the SVR results with different kernels and parameters, it is clear that the RBF kernel outperforms the other kernels, particularly with high values of *C*, demonstrating a steady improvement in prediction accuracy in Tab 7. The RBF kernel is particularly powerful for modeling complex non-linear relationships between features and the target variable. It can capture subtle patterns in the data that might be difficult to represent with linear or sigmoid kernels. The sigmoid kernel shows higher errors, while the performance of the linear kernel lies between RBF and sigmoid, with a noticeable increase in errors as C increases.

3.3.4 Random Forest

The Random Forest approach, as a set of decision trees, offers a powerful perspective in exploring power consumption models. As opposed to linear models, Random Forests offer greater flexibility in modeling complex, non-linear relationships between variables. In this section, we will explore the efficiency of Random Forest as a predictive model, by evaluating its ability to anticipate power consumption based on a diverse set of features.

The basic concept is to construct a group or set of decision tree models to achieve a more robust and accurate prediction.

A Random Forest builds several independent decision trees during the training phase. Each tree is trained on a random subset of the training data, using a process called bootstrap sampling. This means that each tree is trained on a slightly different set of data, introducing diversity into the ensemble. Once the trees have been constructed, each tree gives a prediction for a given observation. The final Random Forest prediction is obtained by a majority voting process in the case of classification, or by averaging in the case of regression.

Random Forests have several relevant parameters that can influence model performance.

Number of trees (n_estimators): This is the number of decision trees included in the ensemble. A higher number of trees can improve model stability, but it is important to monitor model performance to avoid overfitting.

Maximum tree depth (max_depth): Defines the maximum depth of each decision tree. Deeper trees can capture more complex relationships, but can also lead to over-fitting. Setting this depth is crucial to balancing model complexity.

Minimum number of samples for a split (min_samples_split): Specifies the minimum number of samples required for a tree node to perform a division. Adjustment can help control tree growth and prevent overfitting.

Minimum number of samples in a leaf (min_samples_leaf): Indicates the minimum number of samples required in a leaf of the tree. This also helps to control leaf size and regulate model complexity.

Maximum number of features to consider for a split (max_features): Determines the maximum number of features to be considered when a tree performs a division. This can affect the diversity of trees in the ensemble.

Boostrap Sampling (bootstrap): Indicates whether bootstrap sampling is used when building trees. Bootstrap sampling can introduce variability, which can be beneficial for model robustness.

Obtained Results

Using a grid search to adjust hyper parameters is a very efficient and wise practice, as it is for SVR. This systematic approach makes it possible to explore different combinations of hyper parameters and select the one that maximizes model performance for a specific task.

Hyper Parameters	Values
n_estimators	[25, 50, 100]
max_depth	[5, 10, 25, 30]
min_samples_split	[2, 5, 10]
min_samples_leaf	[10, 20, 24]

Tab 8. Table values of hyper parameters

After fine-tuning the hyper parameters through a grid search of the table Tab 8, the best-performing models are listed in table Tab 9. The best-performing Random Forest model is configured with:

Max_depth: The maximum tree depth in the model is set to 5.

Min_samples_leaf: The minimum number of samples in tree leaves is set to 5.

Min_samples_split: The minimum number of samples required to split a node is specified as 24.

N_estimators: The model uses 100 estimators (trees) in the set.

N_estimator	Max_depth	Min_samples_split	Min_samples_leaf	MAE	RMSE	MAPE
100	5	5	24	62.15	93.92	2%
100	3	10	24	262.7	347.9	5%
100	4	3	2	197.2	260.9	3%

Tab 9. Obtained Results

These results indicate that the Random Forest model, with the specified hyper parameters, performed particularly well in the power consumption forecasting task, with acceptable mean and root mean square errors. The Figure 37 show an example of a decision tree from the RF done for the forecasting:

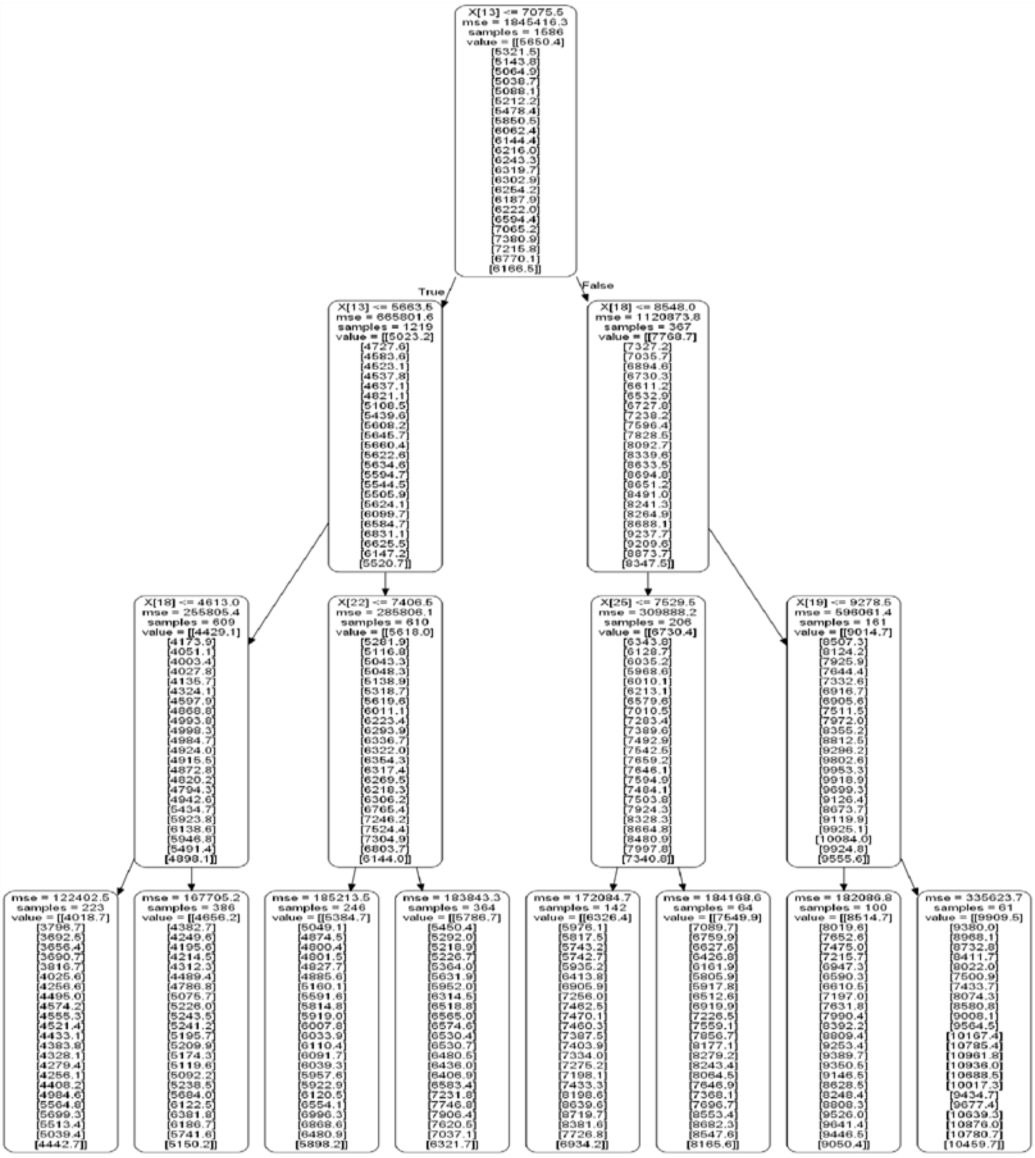


Figure 37. Example of the decision Tree

3.4 Application of Deep Learning Models

Deep Learning, a prominent area of machine learning, has established as a powerful method for exploring and understanding complex patterns in big data. In the field of power consumption forecasting, where relational dynamics are frequently characterized by non-linear relationships and multivariate influences, DL offers a promising approach to capture these subtle complexities. By leveraging deep neural network architectures, this approach goes beyond traditional models to allow a hierarchical representation of features, making it easier to capture the subtle patterns and optimize accurate predictions. In this section, we will explore the application of Deep Learning to the power consumption dataset, aiming to exploit the ability of these models to learn complex patterns and provide more sophisticated forecasts tailored to the dynamic characteristics of this context.

3.4.1 Artificial Neural Network

The introduction of the artificial neural network (ANN) in the context of power consumption forecasting can be based on its architecture, inspired by the human brain, and its ability to learn complex models.

ANNs have emerged as a powerful technology in the field of machine learning. In the context of power consumption forecasting, ANNs offer an innovative approach to capturing complex, non-linear relationships among data. Structured in layers of interconnected neurons, ANNs are designed to learn and extract patterns at different scales, enabling adaptive representation of temporal features and dynamic trends.

Finding the optimal configuration of a (RNA) is a crucial step in the modeling process. In this approach, various architectures have been explored by varying the number of neurons, the number of hidden layers, and the activation functions. The aim is to find the most appropriate model for the complexity of relationships present in the power consumption dataset.

Obtained Results

Hidden Layer	N° neurons	Activation Function	MAPE	MAE	RMSE
1	30	ReLu	2%	119	118
2	30, 20	ReLu	1%	54.40	101
2	30, 20	ReLu, Tanh	97%	5024	5200
1	25	Relu	1%	53.47	73.3
1	25	tanh	44%	3209	3189
1	25	Sigmoid	96%	5662	5510

Tab 10. Obtained Results of different architectures

The table Tab 10 shows the results of evaluating different Artificial Neural Network (ANN) architectures for predicting power consumption.

After an in-depth comparison of the results, a significant observation became apparent: architectures using the hyperbolic tangent and sigmoid activation functions performed significantly worse than those using ReLu. All architectures shared the same epoch time, and the data set was divided into 80% for training and 20% for testing and validation. Despite these similarities, the most encouraging result was associated with a single hidden layer made up of 25 neurons using the ReLu function, displaying a remarkable error rate of 1%, indicating 99% accuracy in predictions. This particular architecture competed closely with one featuring two hidden layers, each with 20 and 30 neurons, respectively. The main discrepancy lies in the MAE and RMSE measurements, where a marginal difference of 1MW is observed.

3.4.2 Convolutional Neural Network

In the context of power consumption forecasting, the use of advanced deep learning techniques, such as CNNs, is a promising approach. CNNs, which were originally designed for image analysis, have demonstrated their effectiveness in capturing complex patterns and modeling spatial relationships, making them relevant for temporal prediction in time series such as electricity consumption.

The fundamental characteristics of CNNs, such as the ability to learn relevant features and to perform convolution on data, make them particularly appropriate for detecting temporal patterns present in complex time sequences. The use of convolutional filters also enables efficient exploration of relationships between different parts of the data, making it easier to capture important temporal dependencies.

CNNs can be applied to time-series data using two main architectures: 1D CNN (Convolutional Neural Network 1 Dimension) and 2D CNN (Convolutional Neural Network 2 Dimensions). Although both share common concepts such as convolution, filter or pooling layers, and the fully connected layer, their specific applications are different depending on the nature of the data.

Common architecture of CNNs (1D and 2D):

Convolution layer: Used to extract local features from data through filters.

Filtering or Pooling Layer: Used to reduce the size of extracted features and provide invariance to transformations.

Fully Connected Layer: Involved in final classification or prediction, combining extracted features to make decisions.

Convolutional Neural Network 1D:

CNN are advanced architectures commonly used in such areas as computer vision. Their adaptability extends to the processing of one-dimensional sequences, such as time series. In the specific case of 1D CNN, the architecture is designed to capture local patterns in a sequence.

Before going straight on to training the different model architectures, the data base needs to be first prepared. Training and test data are initially one-dimensional, as the use of 1D convolution (Conv1D) in a CNN model is envisaged. However, convolution layers in convolutional neural networks expect inputs in matrix form (width, height, channels).

More precisely, for each data example, an additional dimension is added to represent channels, also known as chains or features. In the case of one-dimensional sequences, such as a time series processed with 1D convolution, the term "channel" is used to refer to a particular feature or dimension in the sequence.

This processing ensures that each channel is treated independently, enabling the network to detect specific patterns in each sequence efficiently. Once the data has been prepared, it is passed on to the CNN1D model architecture, by going directly into the first Conv1D layer.

Convolution filters are applied to the sequence, capturing local patterns specific to each channel. Each filter produces an "activation map" corresponding to a specific feature detected in the sequence. Next, the pooling layer reduces the spatial dimensionality of the activation maps, preserving the most important features while reducing model complexity.

Finally, the extracted features are flattened and connected to one or more fully connected layers, enabling the information to be combined for final decision-making.

Obtained Result

CONV Layer	Pooling	MLP	A.Function	Results
2	1	2	ReLu	1.37%
3	1	2	ReLu	10%
1	None	1	ReLu	16%
1	1	1	ReLu	15%

Tab 11. Obtained Results for different architecture of CNN1D

The preceding table Tab 11 illustrates some of the architectures tested for CNN 1D, among many others. The results of these architectures are presented in terms of MAPE.

The architecture configuration of the best result obtained is in Tab 12 as follows:

Convolutional Layer 1	Convolutional Layer 1	Pooling Layer	Dense Layer
30 kernels Size =1	16 kernels Size =1	Max Pooling Size=1	50 neurons ReLu

Tab 12. Configuration of the best architecture CNN 1D

The performance of this configuration is highly satisfactory for electric charge prediction. This architecture, which integrates two convolution layers with 30 and 16 filters respectively, generated a Mean Absolute Error (MAE) of 70.5 and a Root Mean Squared Error (RMSE) of 96.5. These impressive results demonstrate the power of this architecture to predict electrical charge with outstanding accuracy, keeping errors to a minimum.

Convolutional Neural Network 2D:

Electrical charge data are organized in a 2D grid, where each row represents 24 hours in time and each column represents a day of the week. Each grid element contains the electrical charge value at that particular hour.

Before applying the data to the CNN 2D model, they must be standardized and normalized to ensure faster model convergence. Then the data are divided into training and test sets to evaluate the performance of the model on unseen data as in each model.

In the context of 2D CNNs, as in 1D CNNs, an important step preceding data input to the model is the remodeling of the database.

This transformation is essential when adopting 2D CNN architectures. Its aim is to prepare the data in such a way as to make it compatible with the structure expected by a 2D CNN model.

The database is structured in 24-hour sequences initially, before reshaping. However, after this step, the reshaping result adopts a matrix form of (Total number of samples, Number of hours, Number of days, channel). This transformation enables the model to exploit the spatiotemporal features of the data optimally.

The CNN 2D model could have an architecture similar to the following architecture:

Input Layer: The 2D grid representing the time series.

2D Convolution layer: 2D convolution filters analyze spatiotemporal patterns in the grid. These filters can capture trends at different time scales and days of the week.

2D Pooling Layer: A 2D pooling layer reduces spatial dimensionality while preserving important patterns.

Additional convolution and pooling layers: Repetition of convolution and pooling layers to extract hierarchical features.

Flatten layer: Extracted features are flattened into a vector for use in fully connected layers.

Fully Connected layers: These layers combine the extracted features for the final prediction.

Output layer: The output layer gives the prediction of the electrical charge.

Once the model has been trained, it can be used to predict electrical charge for new instants in time.

This approach enables the 2D CNN to capture the complex spatiotemporal patterns present in electrical load time series, providing a more robust model adapted to the two-dimensional nature of temporal data.

Obtained Result

CONV Layer	Pooling	MLP	A.Function	Results
1	1	2	ReLu	4%
2	1	1	ReLu	1.29%
2	1	2	Tanh	70%
2	None	1	ReLu	5%
1	None	1	ReLu	2.45%

Tab 13. Obtained Result with CNN 2D

A low MAPE value indicates a high accuracy of the relationship between the values predicted by the model and the actual values as shown in the table Tab 13. The best architecture which gave the best result is the second architecture with the following configuration in Tab 14:

Convolutional Layer 1	Convolutional Layer 1	Pooling Layer	Dense Layer
16 kernels Size (1,1)	24 kernels Size (2,1)	Max Pooling Size(1,1)	24 neurons ReLu

Tab 14. Best Architecture Configuration

The results achieved with the selected CNN2D model configuration are very positive and show exceptional forecasting performance. With an accuracy of almost 99%, the model was able to predict 99% of its predictions correctly. This shows how effectively the model was able to represent the complex patterns identified in the electric charge data.

The model achieves a mean absolute error (MAE) value of 42.58. A low rating indicates great model accuracy. The MAE calculates the average amount of errors between predictions and actual values. The capacity of the model to produce accurate predictions is reinforced by the comparatively low MAE score.

The Root Mean Squared Error (RMSE), with a value of 59.71, is also remarkably low. The RMSE takes into account both the magnitude of the errors and their distribution, providing a robust measure of model accuracy. Such a low value of RMSE suggests that the model manages to minimize errors consistently.

3.4.3 Generative Adversial Network

Generative Adversarial Networks (GANs), introduced by Ian Goodfellow and colleagues in 2014 (Goodfellow & al., 2014), are a class of deep learning models that generate synthetic new data from an existing data set. GANs have found varied application in a variety of fields, including image generation (Mao & al., 2021), text summarization (Moravvej & al., 2021) and, in the context under discussion, time series forecasting (Brophy & al., 2023).

GAN is based on two competing networks in a framework, representing an innovative approach to the development of generative models capable of generating data independently. These two essential components are called the generator and the discriminator (Huang & al., 2022).

The generator, which takes the form of a convolutional neural network, aims to create new instances of a specific object. In parallel, the discriminator, in the form of a deconvolutive neural network, evaluates the authenticity of an object by determining whether or not it belongs to a given data set.

During the training process, these two entities enter into competition, generating a mutual improvement mechanism known as backpropagation. This competition helps to continuously adjust and optimize the performance of both the generator and the discriminator, reaching an equilibrium where the generator can produce increasingly realistic data, and the discriminator can better distinguish the generated data from that coming from the training set.

How it works

The generator works as a creator of data instances. It is implemented in the form of a convolutional neural network. The main aim of the generator is to generate new synthetic data that is as similar as possible to that existing in the training data set. More specifically, the generator learns from the existing training data set to produce data that is not discernible from the real data as shown in Figure 38.

The discriminator works like an expert, responsible for evaluating the authenticity of the data. It takes the form of a deconvolutive neural network. The discriminator must learn to discriminate between real data and data generated by the generator. Its role is to provide feedback to the generator on the quality of the data produced. More specifically, the discriminator is trained to give a high probability to real data and a low probability to generated data. His expertise is to precisely discriminate between the two types of data.

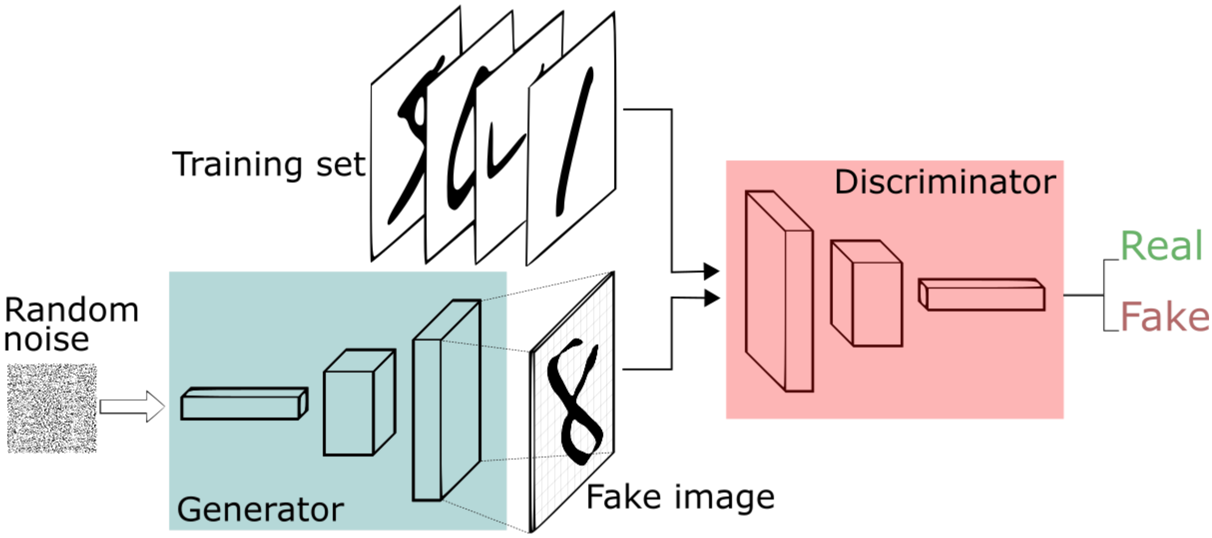


Figure 38. Architecture of GAN

The dynamic interaction between the generator and the discriminator within the GAN is essential to the learning process. Initially, the generator is generally unable to produce realistic data, and the discriminator is not capable of clearly discriminating between real and generated data. The weights and parameters of both networks are randomly initialized.

The generator uses these initial parameters to produce synthetic data. This data is then combined with real data to form a training dataset for the discriminator. The discriminator is then trained on this mixed training ensemble to improve its ability to discriminate between real and synthetic data. During this phase, the discriminator adjusts its parameters to minimize classification error, aiming to make discrimination between the two types of data increasingly accurate.

Once the discriminator has been trained, it evaluates the quality of the data generated by the generator. It provides feedback to the generator, indicating how accurate its data is. This action takes the form of an error signal, informing the generator how to adjust its parameters to improve the quality of its output.

The generator then adjusts its parameters to produce data that will further fool the discriminator in the next iteration. The purpose of the generator is to make its data increasingly difficult to differentiate from the real data.

This iterative process is repeated several times, with each cycle improving the performance of both the generator and the discriminator. The two networks evolve in parallel until an equilibrium is reached, where the generator is able to produce realistic data and the discriminator has difficulty to discriminate it from real data.

Obtained Results

The best-obtained result was an MAPE of 8%, meaning that 92% of predictions were correct. Looking at the previous graphs evaluating the results obtained, as seen in Figure 39 that the generator's prediction rate, shown in orange, is slightly higher than the actual values. There is also some disturbances noticed, indicating that the discriminator was able to distinguish between the generated and actual values, up to around 8500 epochs, where the error rate decreased.

For the second graph, we examine the performance rate of the generated values compared with the real values. The generator starts to improve after 4000 epochs and continues until around 8000 epochs, when we observe a stabilization. This is because the results of the generator become similar to the real values, and the discriminator can no longer distinguish them effectively.

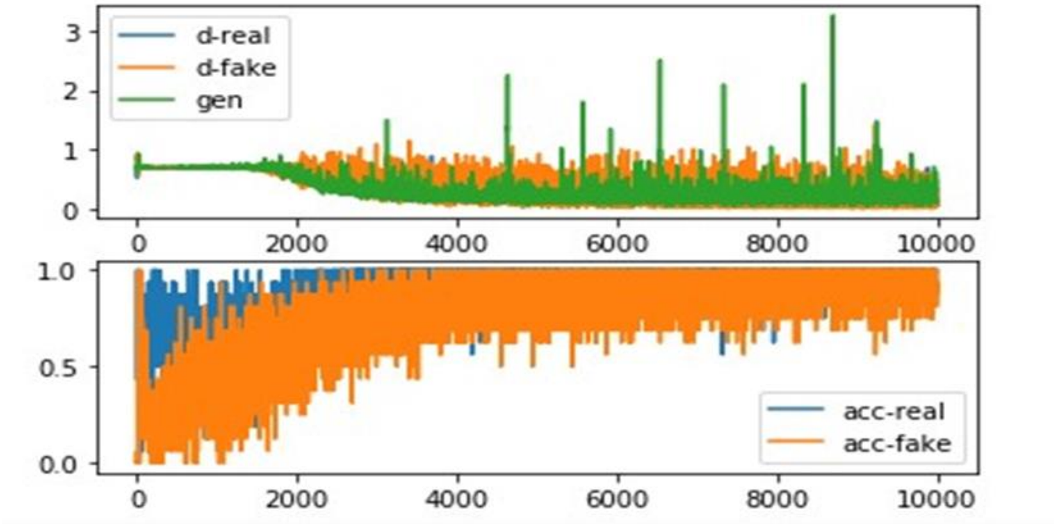


Figure 39. Obtained Results for GAN

3.5 Models Improvements: Transfer Learning

3.5.1 Definition

Transfer learning is characterized by its success in exploiting previously acquired knowledge to improve performance in a different but related task. This strategy aims to transfer representations learned in a source domain, which is frequently defined by a large volume of data (big data), to a specific target domain.

The first step in learning transfer is to pre-train a model on a specific task and data from the source domain. This pre-training process enables the model to capture general and hierarchical features that are inherent in the training data. These features can be extracted from the internal layers of the model and used as abstract representations, encapsulating relevant knowledge.

Once the model has been pre-trained, the weights are used as a starting point for training a model specifically for the target task.

During this phase, the upper layers of the model can be reset or adjusted to accommodate the new target task data. It should be noted that target task data can sometimes differ considerably from source domain data. In such cases, domain fitting techniques can be employed. These techniques aim to adjust the model so that it performs better on the target task, despite the substantial differences between the two domains.

In other words, transfer learning offers a strategic approach to effectively exploiting pre-existing knowledge, ensuring the adaptability of models to the specific challenges of a target task, even when data differ significantly between the source and target domains.

3.5.2 Motivations

Transfer learning, a major pillar of machine learning, is increasingly in demand in a variety of fields, including time series. Time series often feature complex characteristics such as trends, seasons and patterns of variation, which can make prediction difficult. Transfer learning offers a promising approach to these challenges, allowing models pre-trained on rich datasets from a specific domain to be adapted and reused in other contexts. This approach offers several advantages, including reducing the need for large training datasets, the ability to generalize knowledge from one domain to another, and accelerating the model development process. Allowing efficient reuse of knowledge already acquired, transfer learning represents a considerable saving in time and resources. In the field of time series, it can be a powerful tool for improving forecast accuracy, reducing the time needed to train models, and facilitating their adaptation to specific contexts. By leveraging representations learned from similar or related challenges, transfer learning is opening the door to significant advances in time series modeling and prediction, offering new perspectives in fields as varied as finance, meteorology, health and many others.

3.5.3 Traditional Learning VS Transfer Learning

Transfer learning is a concept that exploits the knowledge gained during the process of learning a model on specific data to improve the performance of that same model or another model on different data.

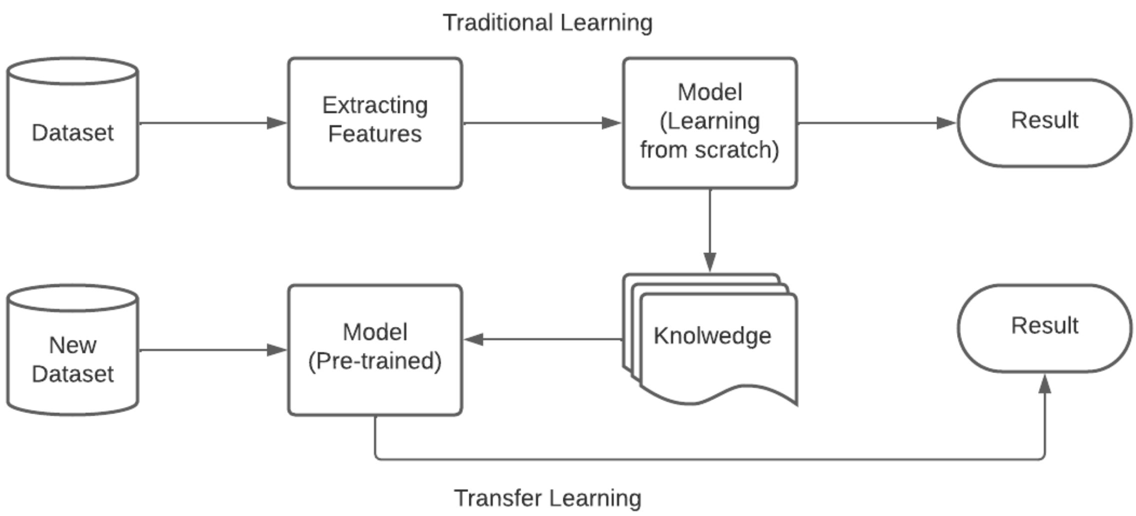


Figure 40. Diagram of the difference between Traditional and Transfer Learning

As the Figure 40 shows in traditional learning, a model is specifically designed and trained for a particular task, usually from scratch. Training a specific model for each new task can be time-consuming and often requires significant computational resources. This approach is based on the idea that each problem must be solved independently, without capitalizing on previous experience.

On the other hand, transfer learning offers a more cost-effective alternative by exploiting the knowledge gained from learning on a source task. Rather than starting from the beginning, the model is initially trained on a source dataset, usually large and diverse, to acquire general, abstract features. When the model is applied to a new task, only some parts of the model are adjusted. The upper layers can be retrained to adapt to the specific characteristics of the target task, while the inner layers preserve the general knowledge previously acquired.

3.5.4 Transfer Learning Approaches

There are different types of transfer learning:

Feature-based transfer learning: In this approach, the model pre-trained on a source task is used as a feature extractor. The layers of the source model, usually the inner layers, are used to extract generic features. These features are then used as inputs for a new model specific to the target task, whose upper layers are often retrained and adjusted to fit the new task.

Model-based transfer learning: In contrast to the first approach, transfer to complete models uses the entire pre-trained model as the starting point for the new model. All layers of the model are adjusted during training on the target task. This can be particularly effective when features learned in the source domain are directly applicable to the target task.

Task-specific transfer learning: This type of transfer learning involves the transfer of knowledge from a source task similar to the target task. For example, if a model is pre-trained on manuscript classification, it can be transferred to help with the alphabetic character recognition task.

Domain-specific transfer learning: The transfer learning process is performed between domains that share similar features. If a model is pre-trained on images of cars, it could be successfully transferred to help with the task of classifying other types of vehicles.

Homogeneous and heterogeneous transfer: Homogeneous transfer is used when the characteristics of the source and target tasks are similar, making adaptation simpler. Heterogeneous transfer, on the other hand, is used when the tasks are different, requiring a more complex adaptation of preexisting knowledge to the new task.

Horizontal and vertical transfer: Horizontal transfer concerns the transfer learning between similar tasks at the same hierarchical level. For example, if a model is pre-trained to recognize different bird species, it could be transferred to help with the task of recognizing different mammal species (same hierarchical level). Vertical transfer involves transferring between tasks of different hierarchical levels, such as from object recognition to scene recognition.

Sequential transfer learning: In sequential learning transfer, the model is continuously updated with new data over time. This enables the model to be gradually adapted to changes in data and tasks in a dynamic environment.

3.5.5 Transfer Learning for Time-Series

The ability of transfer learning, which was first studied in the field of computer vision, to improve model performance for new tasks by using knowledge previously acquired when performing similar tasks, has contributed to its growing popularity. Although the sequential nature of data presents particular challenges, this concept has a particularly interesting application in the field of time series.

Time series have distinctive features such as seasonal trends, fluctuating patterns and time dependency, as they record observations in chronological order. The aim of transposing transfer learning to this particular scenario is to improve the performance of prediction models applied to a target time series by leveraging the knowledge acquired from a source time series.

This method is particularly effective when there is a lack of training data. Transfer learning becomes a powerful strategy for overcoming the limits of data availability by allowing the model to be generalized to related or similar scenarios. However, this is not the only advantage.

In fact, companies continue to discover that the use of transfer learning in the time series domain is a considerable time-saver. This strategy accelerates the process of developing predictive models using pre-existing knowledge, giving an edge over competitors by reducing development times without sacrificing performance.

Time-series transfer learning is emerging as a powerful strategy, not only for improving model accuracy, but also for optimizing the operational efficiency of companies faced with data and time constraints.

3.5.6 Application of Transfer Learning for Load charge Forecasting

The inclusion of transfer learning in the prediction of electrical load consumption is of crucial importance to companies, offering significant advantages in terms of time efficiency and space optimization.

In the context of power consumption prediction, transfer learning translates into an enhanced ability to generalize models and better anticipate trends and patterns in temporal data. Consequently, companies can develop more accurate predictive models, even when power consumption-specific learning data is limited.

Moreover, the use of transfer learning offers a significant time saving advantage. Models can be easily adjusted and fine-tuned to meet specific electricity consumption requirements, based on current information. As a result, the model creation process is faster, reducing implementation times and enabling the company to respond more quickly to variations in demand.

Transfer learning enables the development of more efficient and compact models in terms of space optimization. Computing resources can be used more efficiently, reducing the need for complete relearning and requiring less processing and storage power. This leads to financial and functional benefits, such as reduced expenses for the necessary data infrastructure to implement the models.

Proposed Approach

Having benefited from a vast database of historical data provided by Sonalgaz, which was carefully studied in the previous chapter, where frequency analysis revealed a cyclical nature to consumption that is repeated nearly every year, a strategic approach was developed to optimize the power consumption prediction process.

In this approach, the database is split into two distinct parts. In the first part, the best architecture of the CNN1D model is carefully selected and trained on this initial section of the historical database. This process enables the model to acquire a deep understanding of the characteristics and patterns of power consumption present in this first section of the data.

The next phase of this strategy is based on transfer learning. The CNN1D model, previously trained on the first part of the historical consumption database, is then transferred and applied to the second part of the database. This approach is based on the observation that the annual profile of electricity consumption remains essentially the same from one year to the next, showing only an increase in the amplitude of consumption.

The fundamental motivation behind this transfer learning approach is to significantly reduce model learning time. According to frequency analysis shown in Figure 41, the amount of electricity consumed annually follows a predictable cycle each year as shown in the Figure 41. Therefore, fluctuations in the amplitude of electricity consumption can be effectively adjusted by the model by exploiting the knowledge acquired during initial learning, thus avoiding complete re-training.

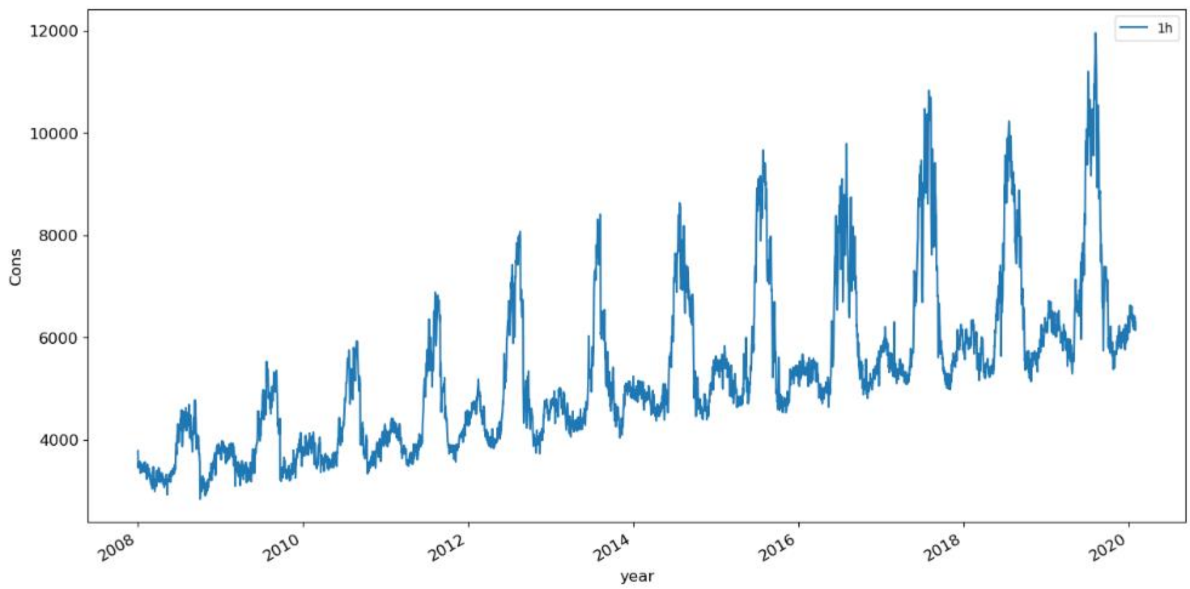


Figure 41. Progression of one consumption hour over the years

Moreover, this method offers an interesting advantage for memory optimization. The company is able to reduce the amount of historical energy consumption data that needs to be retained for future forecasting, avoiding the need to entirely retrain the model. The result is a significant reduction in costs through more efficient management of storage resources.

Model Creation

The development of a new model will heavily rely on CNN, which is widely recognized for its effectiveness in extracting relevant information and features from databases. This method transfers this important knowledge by reusing the layers of a pre-trained model. One of the most crucial steps in adapting this model to a new database is to remove the two MLP layers from the original design that performed as the output of the model.

This adaptation is performed by transferring the pre-trained weights to the new model, with the particular ability to deactivate the training of these layers. This ingenious approach enables the new model to benefit from the knowledge acquired by the pre-trained model during its previous training.

In this way, the new model preserves the feature extraction capabilities of the CNN, while adjusting its parameters to the specific features of the new database. To improve performance on the target task, a new output layer is then added to the simplified MLP model, specifically designed for electric charge prediction. This new layer will then be trained on the data for this new task, while the weights transferred from the CNN will be used as a favorable starting point for this new training phase.

3.5.7 Results and Discussion

Results	MAPE	MAE	RMSE
CNN1D	1.37%	70.5	96.5
Transfer Learning	1.0%	47.89	65.25

Tab 15. The Obtained Results for Transfer Learning

According to the results presented in the following table Tab 15, transfer learning outperformed traditional CNN1D learning, providing superior performance in considerably less time. It is important to note that the CNN1D results were achieved after 100 epochs, whereas with transfer learning, comparable results were achieved in just 20 epochs. This finding underlines the significant time saving, reinforcing the effectiveness of learning transfer in this particular context.

The results obtained underline the potential of transfer learning to improve the accuracy of time series forecasts. This approach takes advantage of the knowledge and experience accumulated from an extended dataset, enhancing model performance on new data.

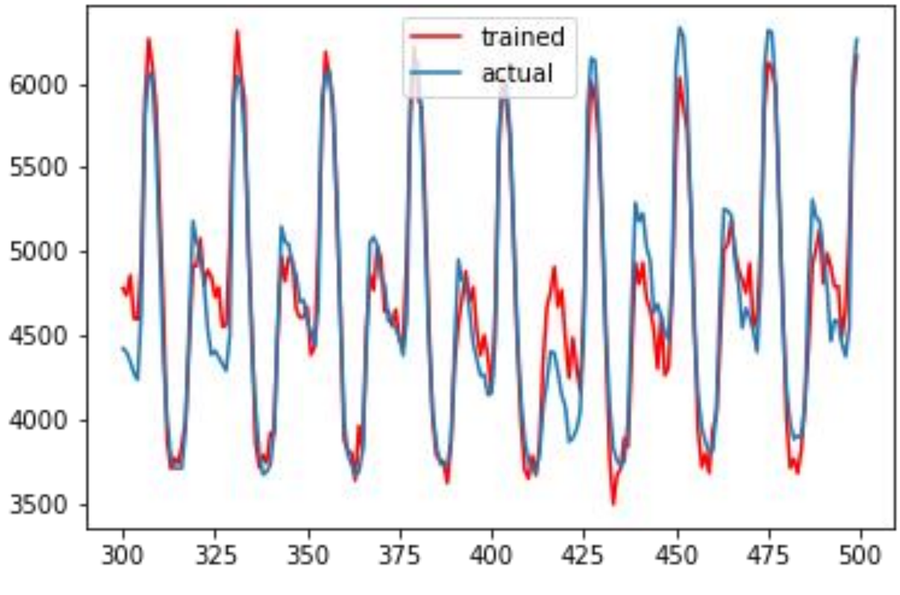


Figure 42. Evaluated curve of trained and actual model

As shown in Figure 42, which provides a visual representation of the performance of the trained model. This representation closely follows the actual curve, demonstrating the success of this approach. Once this validation process is complete, the next analytical step involves in-depth validation to ensure that this model is reliable and efficient.

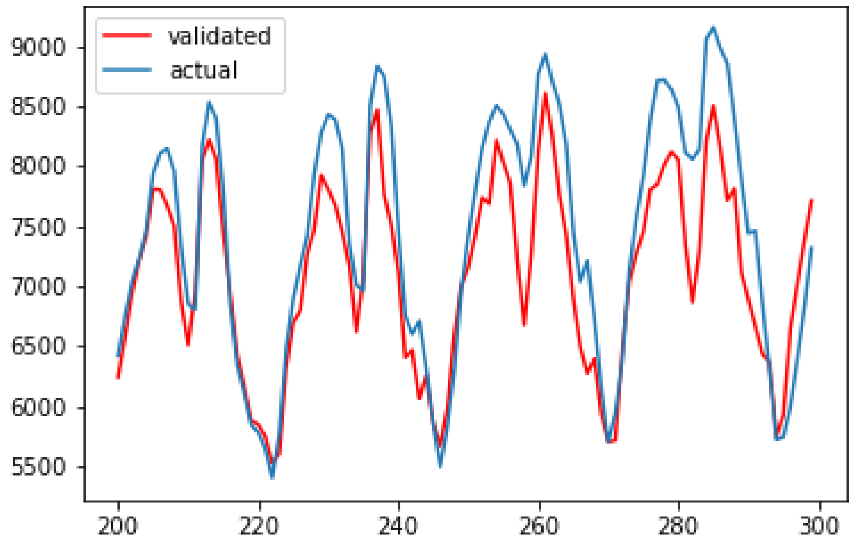


Figure 43. Evaluated curve of the actual model that has been tested and validated

The testing and validation phase represents a crucial step in evaluating the performance of the model on new data, while identifying potential areas for improvement. Testing the model through a series of tests and measurements gives greater confidence in its ability to accurately predict the load charge. Figure 43 illustrates that the predicted data matches the shape of the actual values, reinforcing the validity of this approach.

The success of this testing and validation phase has determined the relevance of this model in real-life scenarios, as well as its potential to induce significant change and generate impact in various fields and applications.

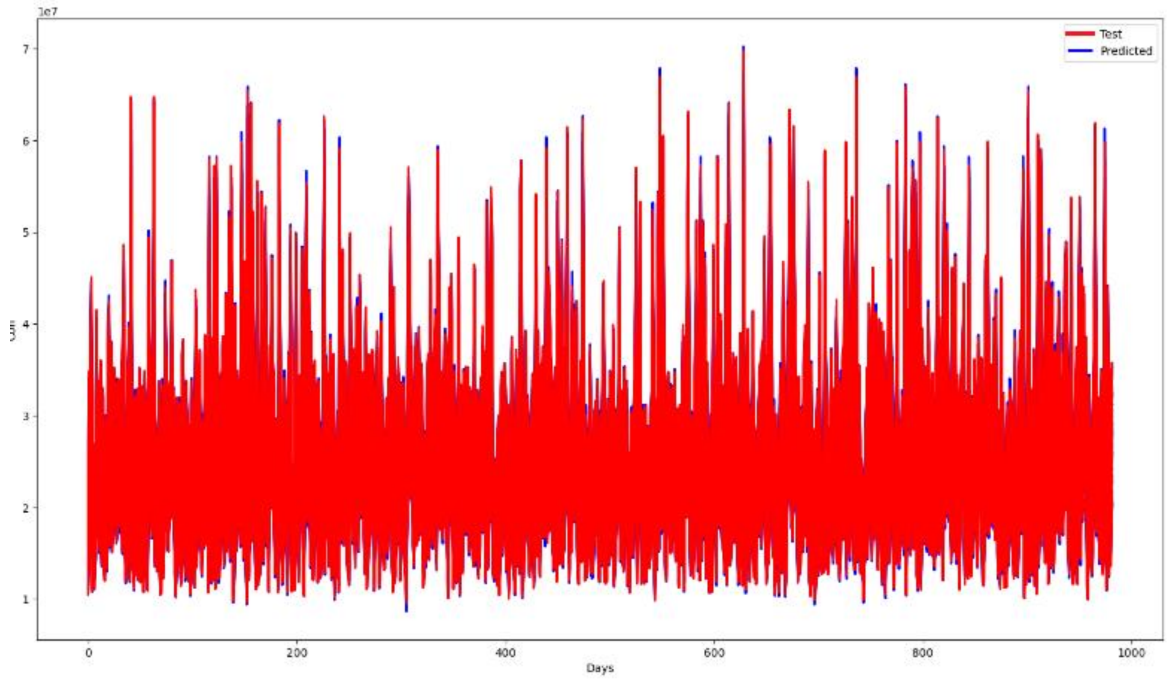


Figure 44. Plot of the actual and the tested data

The new model transferred has a MAPE of 1.0% and acceptable RMSE and MAE losses, making the results very encouraging. The attached Figure 44 clearly shows that these results are well above what is considered satisfying. They provide persuasive evidence of the ability of transfer learning to leverage previous learning and expertise to significantly improve model performance when confronted with new data.

3.6 Comparative Study

Models / Metric	Accuracy	RMSE	MAE
Linear Regression	98	165	127
Multiple Regression	98	110	97
Support Vector Regression	99	103	76.3
Random Forest	98	93.3	62.15
Artificial Neural Network	99	53.47	73.3
Convolutional Neural Network 1D	98.7	96.5	70.5

Convolutional Neural Network 2D	98.8	59.71	42.58
Generative Adversial Network	92	198	239
Transfer Learning	99	65.25	47.89

Tab 16. Comparative Table of the Obtained Result

The evaluation results in the table Tab 16 showed the diversity of performance observed among the different models examined.

LR and MR produced a respectable result in terms of accuracy, but also presented relatively high RMSE and MAE errors. This finding may indicate that the linear approach may have limitations in capturing the complexity of relationships between variables, which may be changing over time.

The SVR showed high accuracy, with acceptable RMSE and MAE errors. However, this performance could reveal sensitivity to seasonal variations, or inaccurate model complexity due to the sensitivity of hyper parameters.

RF model performance was promising, with competitive RMSE and MAE errors, reflecting its adaptability to data complexity.

The ANN showed excellent accuracy, associated with very low RMSE and MAE errors, demonstrating its strong ability to learn complex relationships in the data.

The 1D and 2D CNNs demonstrated robust performance, particularly the 2D CNN, which posted particularly low RMSE and MAE errors, demonstrating its ability to capture spatial features.

On the other hand, GAN performed relatively poorly, with higher RMSE and MAE errors, indicating persistent difficulties in generating accurate data even after many iterations.

The transfer learning approach was distinguished by high accuracy and low RMSE and MAE errors. These results clearly demonstrate the ability of transfer learning to exploit existing knowledge to significantly improve performance on new data while significantly reducing training time.

3.7 Conclusion

Comparative analysis of the various models tested in the prediction of electrical consumption reveals a significant diversity in their performance.

More advanced methods, such as neural networks (ANN, CNN 1D, CNN 2D) and transfer learning, have shown higher performance and a more powerful capacity to learn complex patterns. They also emerge as robust choices, demonstrating an impressive ability to model complex temporal and spatial relationships.

Chapter 4: The Impact of Load Charge Forecasting in Environment

4.1 Introduction

The dynamic evolution of our modern societies towards increasing dependence on electricity has driven a significant rise in energy consumption over the last few decades. This steady growth, while crucial to satisfy the growing needs of the current population, has simultaneously generated considerable environmental challenges. Among these, pollution caused by energy production occupies a prominent place, contributing significantly to greenhouse gas emissions and other forms of environmental deterioration.

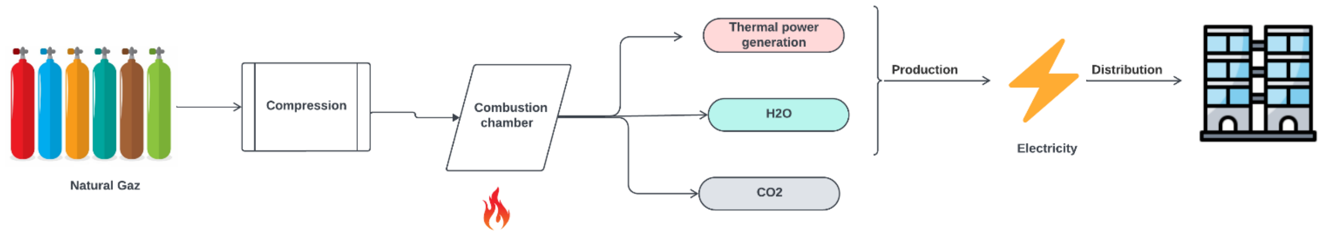


Figure 45. Schema of the Generation of Electricity

The graphic diagram in Figure 45 illustrates the various steps involved in producing electricity from natural gas. Natural gas is first put through a compression and filtering process. This step is crucial to ensure that the gas is of the right quality and at the right pressure before being used to generate electricity.

After being compressed and filtered, the natural gas is then sent to a combustion chamber. The gas is burned in this space, releasing a large quantity of thermal energy. The steam is then produced by heating the water using this thermal energy.

The steam generated in the process is directed to a turbine connected to an electrical generator. As the turbine rotates under the effect of the steam, it drives the generator, producing electricity. This electricity is then ready for distribution to the power grid.

However, it is essential to note that a major inconvenience of this process is the production of carbon dioxide (CO2) during the combustion of natural gas. Managing these CO2 emissions is therefore a major environmental challenge when it comes to generating electricity from natural gas.

Adjusting electricity production according to load forecasts allows the company to avoid the need for back-up fossil fuel power plants. This practice contributes directly to the reduction of greenhouse gas emissions, and therefore decreases the overall carbon emissions footprint of electricity production.

In particular, by avoiding the frequent use of fossil fuel power plants, emissions of carbon dioxide (CO2), sulphur dioxide (SO2) and nitrogen oxides (NOx) are significantly reduced.

These gases are known for their impact on global warming, air acidification and smog formation. Thus, accurate planning based on load forecasts not only optimizes energy efficiency, but also promotes more sustainable and ecofriendly practices in the energy sector.

In this challenging context, accurate forecasting of electricity consumption is emerging as an essential and strategic tool in the responsible management of supply and demand. This rigorous anticipation of future energy needs offers the possibility of implementing proactive measures to adjust electricity production efficiently, therefore minimizing unnecessary excess and limiting harmful emissions.

This contribution aims to explore in depth the strong correlation between accurate forecasting of electricity consumption, the fight against pollution generated by energy production, and the imperative need to adopt more sustainable energy technologies and practices. The implications of this innovative approach in reducing CO2 emissions, its alignment with global environmental objectives as defined in international agreements, and its role in building a more resilient, equitable and ecologically responsible energy future will be examined. With these crucial aspects in consideration, this exploration aims to highlight potential solutions for reconciling energy needs with the essential preservation of our planet.

4.2 CO2 Emissions

Greenhouse gas emissions of CO2, or carbon dioxide, refer to the release of this natural gas into the atmosphere, from any source. CO2 is a colorless, unscented gas naturally present in the air, but its level has risen significantly since the industrial revolution due to human activities (Raupach & al, 2013). CO2 is a greenhouse gas, which means it contributes to the Earth's natural greenhouse effect by absorbing and emitting heat radiation. However, the increase in CO2 emissions resulting mainly from human activities has major implications for climate and the environment.

The main sources of CO2 emissions include Fossil fuel combustion (Gregg & Andres, 2008). When we burn fossil fuels such as coal, oil and natural gas to produce energy, whether for electricity, heating or transport, this releases CO2 into the atmosphere. The industrial processes, such as the production of cement and steel, can also generate significant CO2 emissions as it shows the Figure 46.

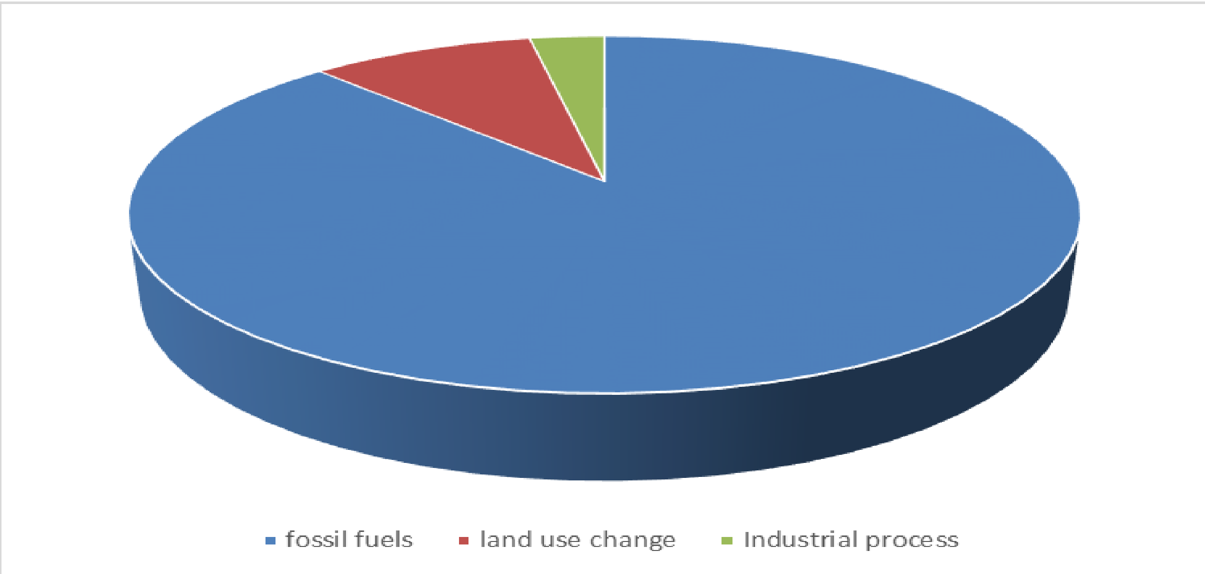


Figure 46. Human sources of Carbone Dioxide (Quéré & al., 2013)

Anthropogenic emissions of greenhouse gases, mainly carbon dioxide (CO2), were estimated at 25 Gt per year in 2006 (Le Quéré & al., 2013) These emissions rose to 37.1 Gt in 2018 (Le Quéré & al., 2018) and continued to increase in 2019 to reach 42.2 ± 3.3 Gt (Friedlingstein & al., 2020). Global economic sectors such as industry, transport and energy are mainly responsible for these emissions (Fletcher, 2018), according to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, published in 2014.

When it comes to examining the causes of CO2 emissions in more detail, Algeria stands out for its specific characteristics, particularly when it comes to power generation. The emission factor is strongly influenced by the choice of energy source, and in Algeria, natural gas is the most common energy source.

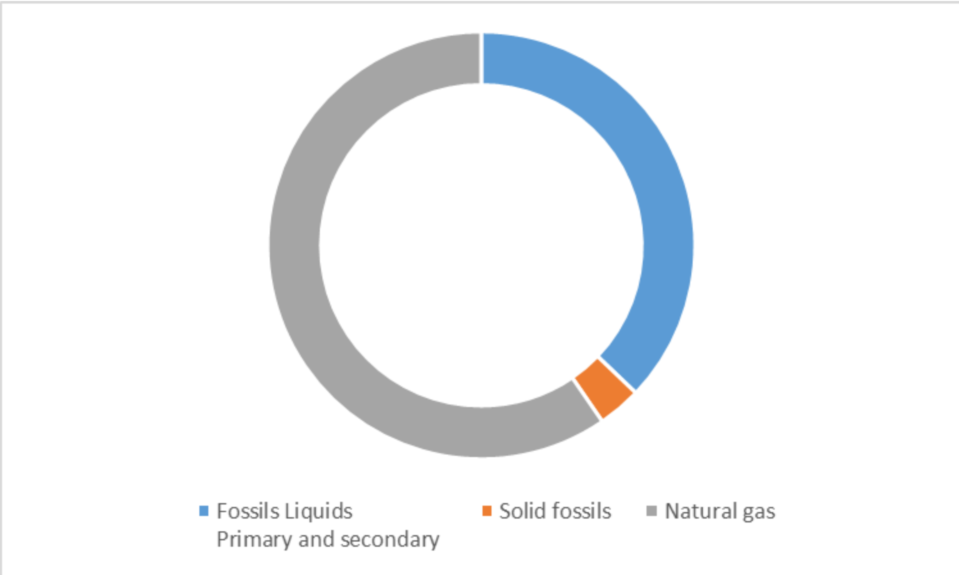


Figure 47. CO2 emissions in the energy sector by fuel type

It is remarkable in Figure 47 that natural gas occupies a dominant position, accounting for around two-thirds of the Algerian energy consumption. This trend underlines the strategic importance of natural gas in the country's energy mix. In addition to its significant contribution to energy consumption, natural gas plays an essential role in power generation, industrial applications and the domestic sector.

The reason for this focus on gas is partly due to the vast reserves of natural gas in Algeria, making the country one of the most important energy producers in the world. Algerian energy policies have also evolved to promote the sustainable use of natural resources, and natural gas is often seen as a cleaner option to liquid and solid fossil fuels.

At the same time, this predominance is opposed to primary liquid fossil fuels, which include crude oil, condensate and LPG (liquefied petroleum gas) extracted on the production site, as well as solid fuels such as coal and wood.

The estimated CO2 emission factor for Algeria, with natural gas as the main source of electricity generation, is 548 g/KWh (Convention des Maires). This value reflects the considerable weight of fossil fuels, even though natural gas is often considered a less polluting option than other fossil energy sources.

In comparison, the world average CO2 emission factor for electricity generation is significantly lower, at around 6 g/KWh. This disparity highlights the need for Algeria to diversify its energy mix and explore more sustainable alternatives in order to reduce its impact on the environment.

The adoption of cleaner technologies, such as renewable energies, could play a crucial role in improving the country's emission factor. Harnessing the potential of solar, wind and other clean energy sources can help not only to reduce CO2 emissions, but also to strengthen energy resilience and promote sustainable development.

4.3 Energy vs CO2 Emissions

Electricity consumption has been growing steadily over the years, and to satisfy the growing needs of consumers, energy suppliers are redoubling their efforts to increase electricity production and compensate for the shortfalls in society. However, it is crucial to note that this increasing production of electricity is not without consequences for the environment.

Accurate forecasting of electricity consumption is becoming an essential tool in the preventive fight against pollution associated with energy production, given the strong correlation between these two factors and the continuing escalation of environmental pollution, which is characterized by a gradual increase from one year to the next.

Sonalgaz is a good illustration of this process, using fossil fuels, in particular natural gas, to generate electricity. This practice has a significant impact on the ecosystem. The high combustion rate of natural gas in the production of electricity leads to the significant release of carbon dioxide (CO2) into the atmosphere. These emissions contribute directly to air pollution, which is inhaled by citizens, with adverse consequences for air quality and public health.

The problem of global warming is exacerbated by the fast conversion of natural gas into CO2 emissions, which increases the atmospheric concentration of greenhouse gases. This is why it is becoming increasingly important to reconsider the energy sources used to generate electricity, and to choose more environmentally-friendly alternatives.

In mathematics, the relationship between electricity consumption (E), carbon dioxide (CO2) emissions and greenhouse gases can be simplified by the following equation Eq. 49:

$$\boldsymbol{CO2 = Electricity\ C\ x\ Emission\ Factor} \qquad \textbf{(49)}$$

Where,

Electricity *C* is the electricity consumption;

Emission Factor is the emission factor, representing the amount of CO2 emitted per unit of electricity.

This equation Eq.49 demonstrates the direct correlation between the amount of electricity consumed and the associated CO2 emissions. The emission factor plays a crucial role in determining the total environmental impact of energy production. However, electricity network providers can positively influence this equation by adjusting production to anticipated needs, thereby making a significant contribution to reducing greenhouse gas emissions. This mathematical approach offers a valuable analytical tool for estimating and optimizing the environmental impact of electricity consumption, aligning energy practices with sustainability objectives.

Electricity demand management offers a significant opportunity to reduce dependence on fossil fuel power plants. This leads to a significant reduction in carbon dioxide (CO_2) emissions, making a crucial contribution to the fight against climate change and its harmful impacts.

Reducing the use of power plants fueled by fossil fuels, such as coal, oil and natural gas, contributes directly to the reduction of CO_2 emissions into the atmosphere. These greenhouse gas emissions are one of the main causes of global warming, with serious consequences such as melting polar ice caps, rising sea levels and extreme weather events.

Therefore, by optimizing electricity demand management, the company will not only encourage a more efficient use of energy resources, but also actively participate in the reduction of environmental impacts. This approach represents a concrete and immediate step towards climate change mitigation and the creation of a more sustainable and environmentally friendly energy future.

4.4 Proposed approach

It is imperative to reduce carbon dioxide (CO_2) emissions from power generation in order to fight climate change and ensure the transition to a more sustainable energy future. The combustion of fossil fuels in power plants releases greenhouse gases into the atmosphere, with devastating effects on the global ecosystem.

One of the most effective ways of reducing these emissions is to integrate renewable energies such as solar, wind and hydro power into the energy mix. These renewable energy sources produce electricity without emitting CO_2 during the generation process. By harnessing these resources, dependence on fossil fuels can be reduced, lowering the associated CO_2 emissions.

The role of smart grids in reducing CO_2 emissions is also crucial. These electrical distribution systems use advanced technologies to optimize electricity management. They enable more efficient integration of renewable energies by adjusting production according to the availability of these sources, boosting a more sustainable use of energy.

The use of energy storage technologies also helps to mitigate the fluctuations inherent in intermittent renewable energies. By storing electricity generated during periods of high production, it can be used when demand is high, reducing the need for back-up fossil fuel power plants.

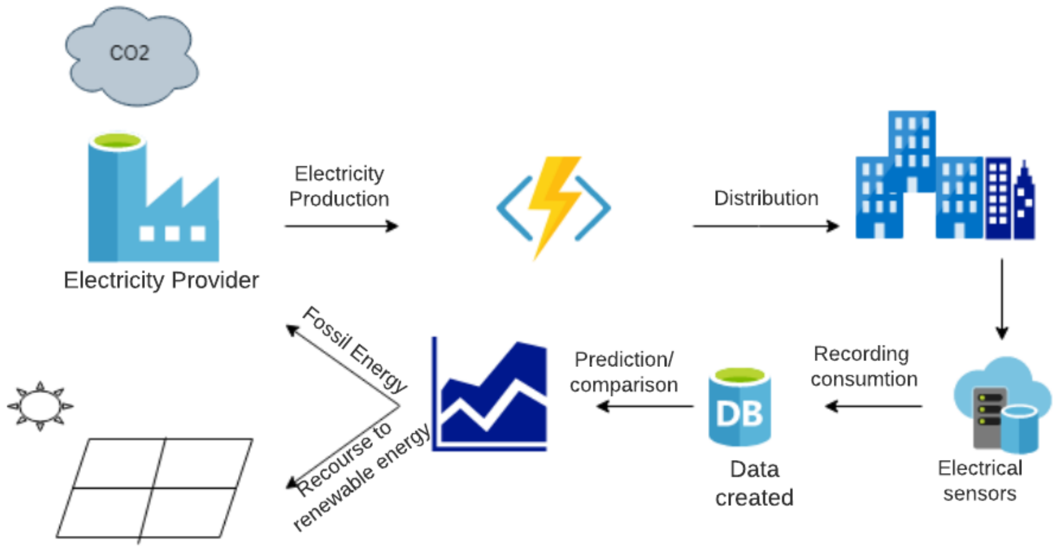


Figure 48. Diagram of the Proposed Approach

Electricity is generated at the plant, a process which results in the emission of CO2 into the atmosphere. This electricity is then transmitted and distributed through the national grid like demonstrated in Figure 48. Once individuals consume this electricity, their consumption is accurately measured by electrical sensors integrated into smart grids, forming a detailed historical database of electricity consumption.

This database is used to perform an advanced prediction, implemented using a revolutionary new model called the transformer. This prediction precisely estimates the CO2 emissions generated by power generation. If this estimate reaches a limit that exceeds the regulatory limits for CO2 emissions, this underlines the crucial importance of integrating renewable energies.

The integration of renewable energies, such as solar and wind power, is becoming essential to help minimize CO2 emissions. These renewable energy sources have the advantage of an emission factor close to zero, making a significant contribution to reducing the carbon footprint of power generation.

In order to maximize energy efficiency, electricity produced during periods of high demand can be stored in batteries. This storage enables the electricity to be used later during periods when renewable energies are unable to satisfy consumer energy demand. This strategy guarantees a stable supply while minimizing dependence on fossil fuels, leading to a significant reduction in CO2 emissions associated with electricity production.

Revolutionary technological innovations in the energy sector are changing the way electricity is produced, distributed and consumed. Smart power grids, powered by artificial intelligence algorithms, optimize demand management and facilitate the more efficient integration of renewable energies. Advanced storage batteries guarantee more flexible use of electricity, reducing dependence on conventional energy sources.

In this approach, the model used to predict power consumption is based on the transformer's encoder, known for its efficiency in natural language processing through the attention mechanism. This advanced technology enables accurate prediction by exploiting historical trends and complex patterns of power consumption.

Today, technological innovation is not only a powerful source of transformation, but also an important tool for solving global problems such as climate change, food security, public health and many others. Emerging technologies, such as that used to predict electricity consumption, offer considerable potential for improving quality of life and promoting sustainable development. Combining the effectiveness of advanced predictive models with the benefits of technological innovation, we are in a position to forge a future where intelligent electricity management makes a significant contribution to solving global challenges.

4.5 Inspiration

Transformers, in the context of natural language processing, are revolutionary neural model architectures that have transformed the machine learning landscape. Transformers were first introduced by Vaswani et al. in 2017 (Vaswani & al., 2017) and have since proved to be a major breakthrough in the field of natural language processing (NLP) and beyond.

Transformers were originally designed to solve the limitations of recurrent sequence models and short-term memory models, which had difficulty in capturing long-term relationships in sequential data. The architecture of transformers is based on an attention structure that allows the model to consider all parts of the input sequence simultaneously, effectively removing strict sequential dependency.

Previous versions of ML models used a similar approach on a larger scale, as illustrated by the n-gram model (Sidorov & al., 2014). They performed frequency mapping of relationships between different pairs of words or groups of words in their training dataset, trying to predict the next word. However, a critical limitation of these early technologies was that they could not retain context beyond a certain input length. For example, an early ML model had difficulty generating a meaningful paragraph, as it could not maintain context between the first and last sentences of a paragraph.

At the heart of the architecture of the transformers is the attention mechanism, an evolutionary innovation that revolutionizes the way the model understands sequences. This mechanism allows the model to assign weights to different parts of the input sequence, according to their relevance to the task in progress. The result is an enhanced understanding of the relationships between elements, even when they are distant from one another. This ability to assess the importance of these connections offers a significant improvement in contextual understanding, enabling more effective capture of long-term dependencies.

Transformers became particularly famous with the introduction of Google's "BERT" (Bidirectional Encoder Representations from Transformers) model in 2018 (Devlin & al., 2018), which achieved record performance in a variety of NLP tasks. Since then, transformers have been extended to many others fields.

4.5.1 Architecture of Transformer

The Transformer architecture has been used in many fields, successfully proving its adaptability and efficiency in modeling complex sequences.

The Transformer architecture consists of an encoder and a decoder, both with their own components and processes as shown in the Figure 49.

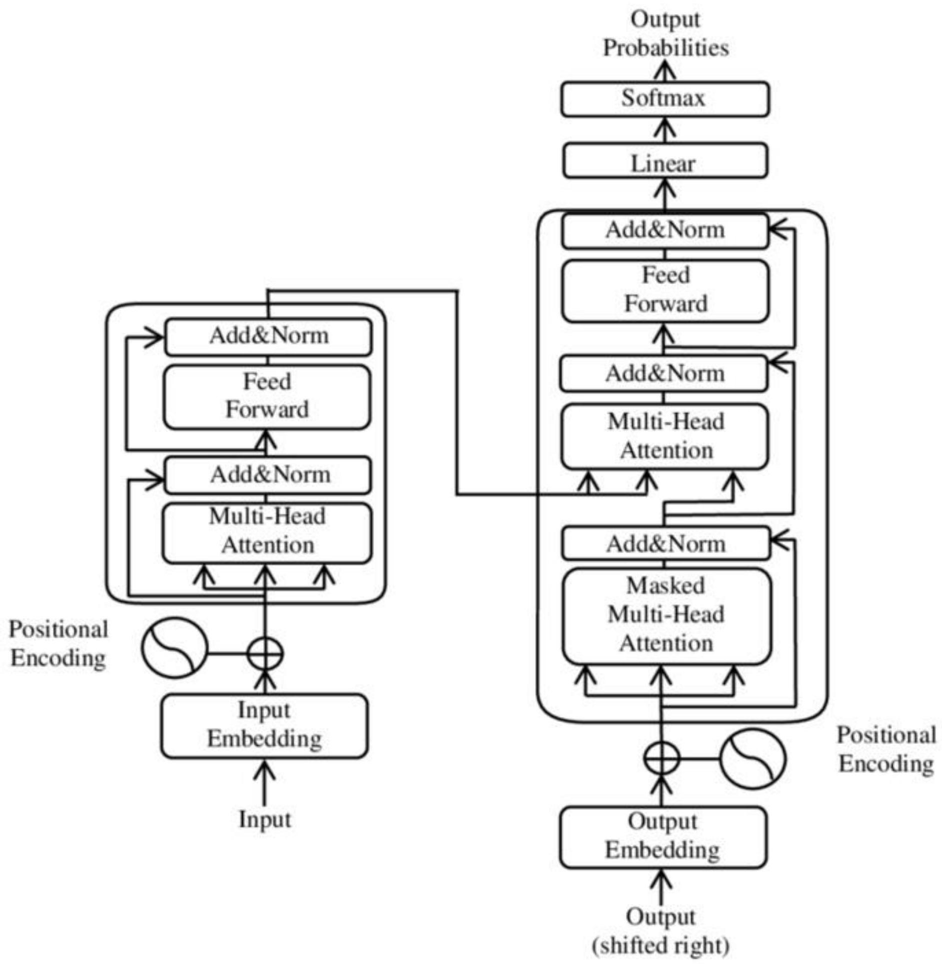


Figure 49. Transformer Architecture

The Transformer uses separate blocks for encoding and decoding. The encoders are responsible for processing the input sequence, while the decoders generate the output sequence. Each encoder and encoder block is composed of several sub-layers.

The Transformer is characterized by the absence of convolution or recursion layers. However, it does incorporate a key method called positional coding, designed to assign each data unit in a sequence a representation of its relative position. This technique is crucial to enable the model to capture the sequentially of input data, an essential aspect for adequate context understanding, particularly in natural language processing.

Encoder Part:

The encoder in the Transformer architecture has a central role in processing the input sequence. Each encoder layer aims to extract features from the sequence, while each decoder layer uses these features to generate an output sequence. Composed of several uniform blocks, generally referred to as "encoder layers", each encoder layer incorporates two essential sub-layers.

Multi-Head Self-Attention:

This sublayer enables the model to capture long-term dependencies between elements of the sequence, allowing the model to learn different attention representations for different parts of the sequence. Each attention head operates as a separate information pathway, enabling the capture of complex patterns and relationships.

The mathematical formula for calculating attention weights is defined by Eq. 50 :

$$\boldsymbol{attention}\ (\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \ \boldsymbol{softmax}(\frac{\boldsymbol{QK^T}}{\sqrt{\boldsymbol{dk}}})\boldsymbol{V} \tag{50}$$

Here, *Q* represents the matrix containing the queries,

K the set of keys,

And *V* the values, representing all the parts of the sequence.

For the multi-head attention modules of the encoder and decoder, *V* is composed of the same sequence of elements as *Q*. These matrices are generated from the input sequence. The softmax function normalizes the attention weights to create a probability distribution. These weights are then applied to the values, producing a weighted output. The *Q*, *K* and *V* parameters enable the model to learn how each part of the sequence is related to the others. The parameter *dk* normalizes the scale of attention scores.

However, in the attention module that takes encoder and decoder sequences into account, *V* is different from the sequence represented by *Q*.

For simplicity, the values of *V* are weighted and summed with attention weights *a*, determined by eq. 51:

$$\boldsymbol{a} = \ \boldsymbol{softmax}(\frac{\boldsymbol{QK^T}}{\sqrt{\boldsymbol{dk}}}) \tag{51}$$

This function indicates that the weights *a* are defined by the way in which every element of the sequence (*Q*) is influenced by every other element of the sequence (*K*). The attention weights calculated determine the influence of each element in the sequence on the others. These weights are then applied to the corresponding values, generating a weighted output. This allows the model to focus on specific elements of the sequence according to their importance.

The Softmax function is then applied to the weights a to obtain a distribution between 0 and 1. This creates a probability distribution, indicating the relative importance of each element. These weights are then applied to all the elements of the sequence introduced into V, which represent the same vectors as Q for the encoder and decoder, but which are different for the module with encoder and decoder inputs.

The attention mechanism is divided into several "heads". Each head can be seen as a different way of paying attention to different parts of the sequence. This enables the model to learn different and complementary aspects of the data.

These heads of attention enable the model to learn from different representations of *Q*, *K* and *V*, offering a diversity and wealth of information beneficial to learning.

Each head uses distinct linear projections of the matrices *Q*, *K* and *V*, represented by the weight matrices *WQ*, *WK* and *WV* respectively. These weights are learned during model training.

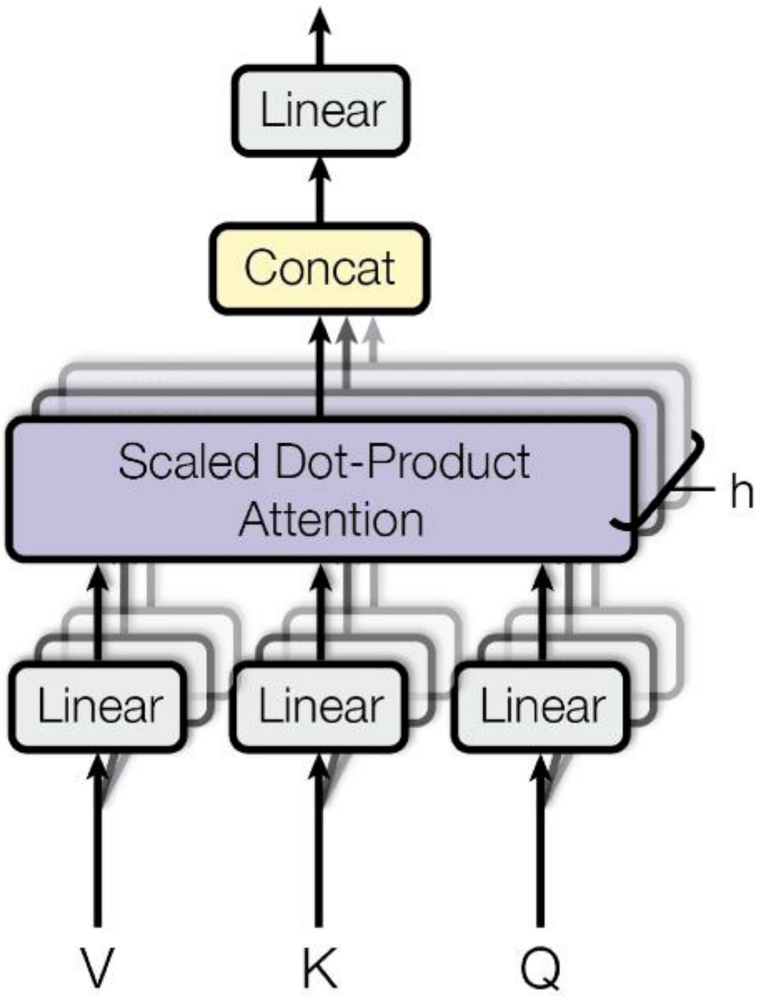


Figure 50. Heads Attention

These parallel projections shown in Figure 50 allow the model to explore different ways of interpreting information.

The single attention mechanism is then applied to each of these sets of projections Q_i, K_i and V_i in parallel like in the previous equation.

The results of the different attention heads are concatenated along the feature dimension then projected again using another weight matrix to produce the result with the equation eq. 52.

$$\boldsymbol{Multihead(Q, K, V) = concat\ (head_1, \ldots, head_h)W^0} \quad \textbf{(52)}$$

This approach enhances the capacity of the model to learn complex, non-linear relationships in the data.

Feed Forward Layer:

The second sublayer is a fully connected reaction network consisting of two linear transformations with activation of the ReLU between the two.

After the attention mechanism in the transformer, the output usually passes through a linear layer in the feedforward network. This layer is responsible for transforming the representations learned by the attention mechanism into a feature space better suited to the specific task of the model.

This step consists of matrix multiplication of the output of the attention mechanism by a weight matrix *W1* . This allows the learned features to be projected into a latent space.

$$\boldsymbol{Intermediate\ Output = \ Attention\ Output\ x\ W1 + \ b1} \quad \textbf{(53)}$$

Where *W1* is the weight matrix,

X matrix multiplication,

And *b1* is the bias.

The intermediate output is then subjected to an activation function, usually a ReLU, which introduces a non-linearity into the network.

The activated output is transformed again by matrix multiplication with a second weight matrix producing the final output of the feedforward network.

The aim of this linear layer is to project the information learned by the attention mechanism into a feature space more suited to the final task. Linear and non-linear transformations enable the model to learn complex, non-linear representations of data, improving its ability to perform specific tasks.

Each sublayer in the Transformer model is followed by a normalization layer. This normalization layer normalizes the sum of the inputs of the sublayer (*x*) with the output generated by the sublayer itself sublayer (*x*). This normalization helps to keep the model training stable through the control of activation values, which can be crucial for efficient learning and network convergence.

Decoder Part:

The decoder has a similar structure to the encoder, with a few differences specific to its output sequence generation function.

At the beginning of the process of sequence generation, the decoder starts with a special symbol marking the initial part of the sequence. Before diving into self-attention, the decoder considers the context of the input sequence using encoder-decoder attention. Then, the decoder uses multi-headed self-attention to focus on different parts of the output sequence generated up to now.

To guarantee independent generation and avoid any use of future position information, all future positions in the output sequence are hidden from the start of the process. In other words, the model cannot take into account information beyond the current position during the self-attention stage, thus ensuring consistency in generation.

Through auto-attention and encoder-decoder attention, the decoder generates a new symbol for the output sequence. This newly generated symbol is then integrated into the input of the decoder, updating the context for subsequent generation steps. This iterative process continues until a special end-of-sequence symbol is generated or the predefined maximum length is reached. These steps take place iteratively, enabling the decoder to generate the complete output sequence.

4.5.2 Transformer for Time-Series

The Transformer model was originally developed for natural language processing, but has since been successfully extended to other fields, including time series analysis.

The model offers a number of advantages for time series analysis, particularly in terms of capturing long-term temporal dependencies, parallelism, adaptability and applicability to different contexts. However, it is important to note that the choice of model will always depend on the specific nature of the data and the objectives of the task.

Time series can contain complex temporal dependencies over long periods. The Transformer, as opposed to recurrent models, has no sequential dependency constraints and can therefore capture long-term relationships more effectively. The attention mechanism enables the model to identify and weight important elements in the sequence, improving its ability to capture complex temporal trends.

This mechanism enables the Transformer to automatically adjust to the optimal size of the temporal window for a specified challenge. It can give more weight to relevant elements in the sequence, adapting its temporal perception according to the characteristics of the data.

4.5.3 Transformer Based Encoder (System Architecture)

Using data divided into 24-hour intervals, the basic Transformer architecture and some additional models were integrated and compared to predict load data for the following day. The customized encoder in this research work is based on the Transformer architecture and will be compared with other models.

The Transformer, a neural network widely used in the field of NLP, is showing its extended versatility beyond linguistic tasks. One of its outstanding applications concerns the prediction of electrical charge, where it can be used to great effect.

The Transformer Encoder architecture is formally characterized by a piling of residual encoding blocks. During the training process, this transformer encoder links the input sequence into a contextualized encoding sequence.

The encoder block initially consists of a bidirectional self-attention layer, followed by two feedback layers. Various implementations have been explored to optimize performance on time-series data, each customizing its methods to suit specific needs.

In order to efficiently capture short-term patterns, a crucial modification was made to the attention map. This adaptation incorporates a window that limits backward attention, allowing analysis to focus on closer time sequences.

A normalization step is performed on the input sequence, followed by a transformation into a fixed-dimension vector for each element. This prepares the data uniformly for consistent analysis.

Each element in the electric charge prediction instance then represents a specific temporal value of electric charge, creating a meaningful representation for a better understanding of temporal patterns.

Positional encoding is essential to maintain the chronological order of power consumption times. Positional encoding assigns a specific value to each element of the time sequence according to its relative position in the series. This feature is crucial if the model is to capture temporal dynamics and correctly interpret variations in power consumption over time.

Maintaining positional encoding, the Transformer model can appropriately contextualize temporal information, preserving the chronological sequence of hours. In this way, it is able to recognize and exploit trends, seasonal patterns, and other important temporal features inherent in power consumption data. This approach ensures that the model can make accurate predictions while taking into account the temporal evolution of the data, which is crucial in the field of time series.

$$\boldsymbol{Position_{Vector}}(\boldsymbol{pos},\boldsymbol{i}) = \begin{cases} \mathbf{sin}(\boldsymbol{pos}/\boldsymbol{n}^{2i}/_{d}) \\ \mathbf{cos}(\boldsymbol{pos}/\boldsymbol{n}^{2i}/_{d}) \end{cases} \quad \textbf{(54)}$$

Where *d* is the dimension of the position vector;

And *n* is the frequency of oscillations in sinusoidal and cosinusoidal representations, thus influencing the way positions are encoded in token embedding. This ensures adequate frequency variation to efficiently encode token positions in the sequence, depending on the dimension of the model.

The sine function is used for even-index dimensions (0, 2, 4, etc.) in position embedding.

The cosine function is used for odd-index dimensions (1, 3, 5, etc.) in position embedding.

This means that each dimension of the position embedding is alternately modified according to one or other of these functions, providing a rich representation of the relative position of tokens in the sequence.

Once the position vectors have been generated, they are combined with their respective elements in the sequence. This step involves the concatenation or addition of position vectors to existing records. The aim of this operation is to combine position information with the characteristics of each element, using the following equation.

$$\textbf{Sequence with position} = \boldsymbol{X} + \boldsymbol{P} \tag{55}$$

The input sequence *X* (represented by feature vectors) is combined with the position vectors *P* (generated in the previous step).

The insertion of these position vectors gives the model the ability to differentiate between elements according to their relative position in the sequence.

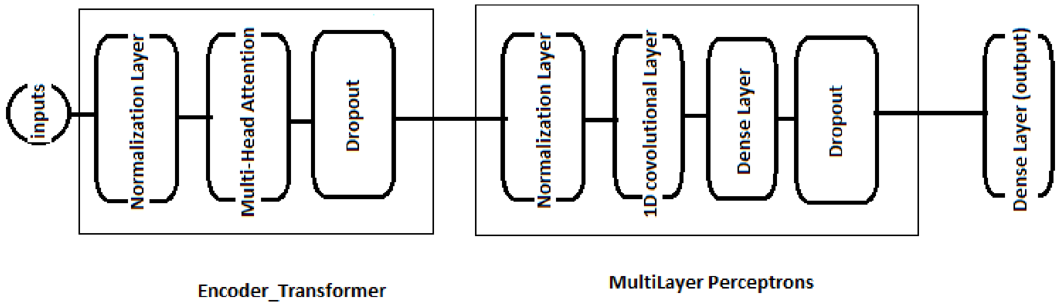


Figure 51. Transformer Based Encoder Architecture

Once the time vectors have been encoded with their positions, the layer normalization procedure is implemented, as illustrated in the Figure 51. This normalization guarantees the stability of the data stream by normalizing the activations. The output of this normalization is then forwarded to a multi-headed attention layer. This layer enables the model to focus on different parts of the sequence, capturing the complex relationships between elements.

For regularization, an output filter is applied to the outputs of the attention layer. This involves randomly deactivating certain neurons to prevent over-fitting of the model. The resulting outputs are then added to the original inputs, creating a residual connection. This connection facilitates the transfer of crucial information across the network and serves as the output for the encoder part, simultaneously becoming the input for a new normalization layer.

Next, a 1D convolution layer is used to introduce convolution operations on the resulting attention sequence. A dense layer is applied, and the output of this layer is again added to the residual connection. A dropout is applied after the 1D convolution layer to regularize the model. Finally, an output layer is applied to produce the electric charge predictions. This global process offers a robust architecture, combining attention, convolution and residual connections for accurate prediction of electric charge time series.

Good performance of this architecture will have a significant impact on the quality of electrical load predictions. This will result in greater accuracy in predicting CO2 emissions, which will be essential in determining whether permitted emissions thresholds have been exceeded. This assessment will be crucial in deciding whether or not to integrate renewable energy sources, contributing to the efficient and sustainable management of energy consumption.

4.6 Results and discussion

The Transformer-Encoder model is compared with the RNN and LSTM models because of their similarities in approach to assess its performance, particularly in predicting power consumption. The models exploit historical data to learn trends and temporal patterns, contributing to more accurate predictions. The ability to take into account the influence of historical data is a feature common to all of them. The Transformer-Encoder model, like RNN and LSTM, is capable of handling long-term temporal dependencies. This means they can capture complex relationships and patterns that extend over multiple time periods.

All the three architectures can adapt to fluctuating temporal characteristics, which is crucial in predicting power consumption subject to variation. Both the Transformer-Encoder and the RNN and LSTM are qualified to encode temporal information, each in its own specific way. The Transformer-Encoder uses self-attention mechanisms, while the RNN and LSTM use hidden states to encode sequential information.

The models were evaluated using parameters frequently employed in time series analysis, including MAE, MSE, RMSE, as well as MAPE accuracy. These parameters act as standard measures for assessing the ability of models to capture and predict data patterns in time series.

MAPE quantifies the average percentage deviation between predicted and actual values. By subtracting MAPE from 100, we obtain the measure of accuracy, representing the proportion of accurate predictions in percent. This formulation improves the interpretation of accuracy.

The results obtained are listed in the table Tab 17, providing a visual representation of the relative performance of the models evaluated.

Models	MAPE	MAE	MSE	RMSE
Transformer-Encoder	2%	82	36	71
RNN	11%	234	354	152
LSTM	2%	110	120	80

Tab 17. Table Obtained Results

Using four Transformer encoder blocks to simplify the model while retaining the ability to capture complex temporal patterns.

In addition, to simplify the model, we opted for more focused attention. This choice is appropriate for the electric charge prediction task, where local patterns are often more meaningful. In this Transformer-based electric charge prediction model, several parameters have been carefully configured to ensure effective representation of complex temporal patterns.

For this model, the head size determines the dimension of the key, query and value vectors in the attention layer. Specifically, in the multi-head attention layer, the number of dimensions of the scalar product between the query and key vectors is equal to the head size, initialized to 12. A larger head size enables the attention layer to capture more complex relationships between elements in the sequence. However, it also increases the complexity of the model.

The number of heads specifies how many independent attention heads are used simultaneously. Each attention head learns different patterns in the data. In the proposed model, this parameter has been initialized to two per encoder block, helping to simplify the model while preserving the ability to detect local patterns. A higher number of heads, while facilitating the learning of more patterns, requires increased computational resources.

The size of the feed-forward layer determines the size of the output after the attention layer. In this model, the size of the feed-forward layer is set to 24. This means that the feed-forward layer produces output of dimension 24 for each position in the sequence, enabling the capture of complex patterns.

The Transformer-Encoder model's outstanding performance in the training phase demonstrates its significant potential for solving the challenges inherent in prediction. This ability to achieve a high level of accuracy enhances its relevance in the time series field, particularly for the prediction of electricity consumption.

When comparing RNNs to transformers, a crucial distinction lies in the presence of parallelism in RNNs. Unlike RNNs, transformers have the ability to process multiple input sequences simultaneously, due to their inherent capacity for parallelism. This feature enables transformers to recognize and represent complex relationships more efficiently. In RNNs, which process sequences sequentially, the input sequence representation must be compressed into a single state vector before processing subsequent sequences. This compression can lead to a loss of important information and limit the model's ability to accurately capture long-term dependencies.

RNNs also have limitations, such as potential gradient explosion problems, which can restrict their ability to effectively handle long-term dependencies, even when using sophisticated mechanisms such as LSTMs. In summary, the comparison between RNNs and transformers highlights the intrinsic advantages of transformers in time series modeling, highlighting their ability to overcome the limitations associated with RNNs.

In the context of reducing CO2 emissions associated with electricity generation, this capability becomes crucial. Accurate forecasting of electricity consumption enables more precise planning of energy activities. Electricity generation plays a significant role in the relationship between electricity consumption forecasts and CO2 emissions. In other words, accurate forecasting of energy needs has a direct impact on carbon dioxide emissions by improving the production, delivery and use of electricity.

The table Tab 18 shows the results for CO_2 emissions, highlighting the values predicted by the transformer over the period of a few hours in a day, for illustrative purposes. A significant observation emerges regarding the direct correlation between calculated CO_2 emissions and predicted values: an increase in consumption is directly proportional to an increase in CO_2 emissions.

Hours	1H	6H	12H	17H	21H
Predicted Values	7250	6250	6877	7145	7840
CO2 Emissions	3973	3768	3916	3739	4296

Tab 18. Results of CO2 Emissions

The emission factor is closely linked to the source of electricity used to generate power. If, for example, electricity generation depends mainly on the use of gas, the emission factor will be high due to the high CO_2 emissions generated by gas combustion. On the other hand, when electricity is generated from renewable sources, the emission factor tends to approach zero.

In the case of Algeria, the emission factor associated with gas use in the country has been set at 548 gCO_2/kWh for the specific purposes of this study. It is important to note that this value differs significantly from the world average, which stands at 6 gCO_2/kWh. This disparity highlights the importance of taking national specificities into account when assessing the environmental impact of electricity generation.

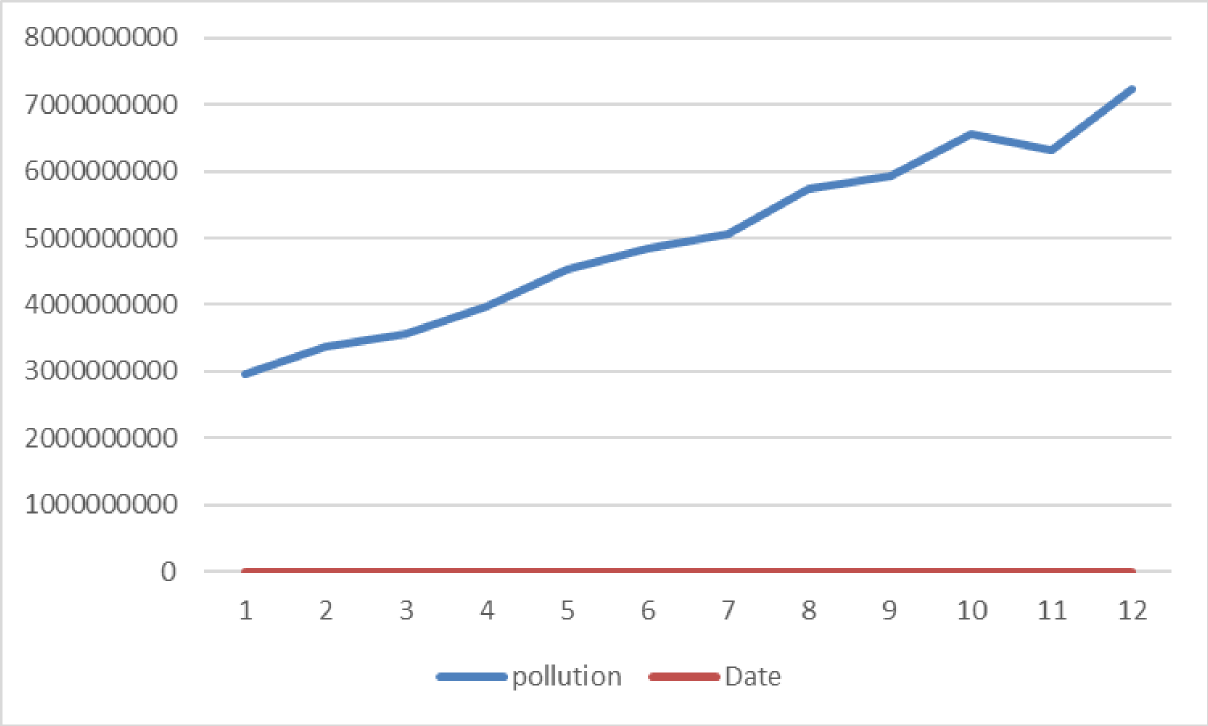


Figure 52. Pollution curve in Algeria

It is crucial to realize that all forms of production have an impact on the global environment, as the International Energy Agency (IEA) predicts that global pollution will reach 32,252 million tons by 2020. Although the Figure 52 are expressed in grams, commitment to reducing greenhouse gas emissions remains essential, particularly through greater integration of renewable energies.

Without the exceptional growth in renewable energies, electric vehicles, heat pumps and energy efficiency technologies, the increase in CO2 emissions would have been almost three times greater. These advances therefore play a decisive role in the fight against rising global emissions, highlighting the importance of sustainable initiatives.

Understanding these emissions specific to Algeria provides an essential basis for the development of more sustainable energy policies, aimed at minimizing environmental impact while improving health on a national scale.

4.7 Conclusion

The efficiency of the transformer encoder in predicting electrical load offers a significant breakthrough for specialists. Through accurate modeling of electricity consumption trends, experts can precisely anticipate when the CO2 production threshold is likely to be exceeded. This predictive capability enables energy professionals to adjust their energy mix in real time, promoting a transition to more sustainable sources, such as renewable energies. This approach makes a tangible contribution to reducing CO2 emissions, by optimizing the management of electricity production, thus supporting global efforts to promote environmental sustainability.

Conclusion and Perspectives

Energy is a fundamental concept that governs the universe and life on Earth. Its relevance to modern society is hard to overestimate. It is essential to many aspects of daily life, from industry and technology to households, influencing the economy, the environment and safety.

Without energy, modern society could not survive. Energy is an essential component of the global economy. Fluctuations in energy prices have a direct impact on production costs, prices of goods and services, and economic growth.

The research and development in the field of energy is leading to technological advances that not only improve energy efficiency and reduce emissions, but also open the way to new industries and economic opportunities. Worldwide access to energy is not identical. Millions of people in developing countries lack access to reliable, affordable energy, limiting their economic, educational and development opportunities.

Electricity, one of the most versatile and widespread forms of energy, which plays a crucial role in present-day civilization. It is based on the notion of electrical charge, a fundamental characteristic of matter. In our homes, electricity lights up our rooms, powers our electronics, heats our homes and runs our appliances. In industry, it powers the machines that make our consumer goods, and keeps production lines running.

Power system operators must maintain a balance between electricity supply and demand to avoid blackouts and power failures. Accurate prediction of demand enables them to optimize grid operation, avoiding overloads and minimizing energy losses when transmitting electricity over long distances.

After an in-depth study, including both statistical and visual analyses of the database, various machine learning and deep learning techniques and models were explored in the context of electrical load forecasting. This exploration led to the identification of promising models that generated very satisfactory results.

These models were evaluated using a variety of metrics such as RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error), which are crucial indicators of the performance of prediction models. The results obtained demonstrated a significant reduction in prediction error compared with traditional methods, attesting to the effectiveness of machine and deep learning approaches in this specific field.

In addition, a visual analysis of the predictions against real data also confirmed the accuracy of the models, highlighting their ability to accurately capture the complex trends and patterns present in electrical load data. These promising results open up new perspectives for the optimization of energy operations, offering suppliers a better understanding of consumption trends and more accurate forecasts for the efficient management of power generation and distribution.

Each model presents its own advantages and challenges. For example, linear regression offers simple interpretability but can be limited in capturing non-linear relationships, while SVRs are effective at modeling complex relationships but require careful optimization of hyper parameters. Random Forests are robust and can handle heterogeneous data, but may suffer from over-fitting. CNNs are excellent for capturing spatial or temporal patterns in data, while GANs offer the ability to generate realistic new data.

The Algerian dataset covers 18 years of electricity demand data, with measurements taken every 24 hours. Due to its considerable size, this dataset offers a wealth of information on electricity demand. It also reflects the country's socio-economic context, making it an interesting dataset requiring in-depth analysis.

In this specific study, different architectures and parameters were explored for each of these models, highlighting performance metrics such as RMSE and MAE. Each model, we found, has its own strengths and weaknesses, and the choice of model often depends on the specific characteristics of the data and the objectives of the task.

The results of the machine learning models were compared to highlight the ability of deep learning models to capture consumer characteristics and trends on a large historical database. Regression, while offering a simple interpretability, showed limited performance in terms of prediction accuracy. RMSE and MAE values were relatively high, indicating a significant deviation between predicted and actual values, with an accuracy of 98% of correct prediction.

Compared with the SVR, the results showed an improvement over the SVR, with lower RMSE and MAE values and a MAPE of 1%. Performance was influenced by the choice of kernel and hyper parameters, with the C parameter and RBF kernel giving the best results. Random Forests showed robust performance in electric charge prediction. With appropriate optimization of hyper parameters, notably the number of estimators and maximum tree depth, RMSE and MAE values were reduced, indicating a better match between predicted and actual values.

With the right architecture, CNNs can detect local and global patterns in temporal data, enabling them to better model complex relationships between variables and improve predictive performance. The convolutional layers of CNNs, with their ability to identify recurring or seasonal patterns in time series, while pooling layers can reduce data dimensionality while preserving essential information.

In terms of performance, CNNs have been observed to significantly reduce prediction errors, measured by metrics. In addition, MAPE, a relative measure of error, is often very low, approximately 1%, indicating high forecast accuracy.

The use of CNNs in short-term electric charge prediction has been particularly successful. Because of their ability to capture fine temporal features and model complex relationships between variables, CNNs have proved effective in predicting electricity demand over short periods of time.

In this study, an extensive historical database provided by Sonelgaz was scrutinized, revealing a cyclical pattern in electricity consumption repeating itself almost every year. In this contribution, the strategic approach of learning transfer was developed to optimize the power consumption prediction process.

Initially, the CNN1D model was trained on the first part of the historical database. Subsequently, this previously trained model was transferred and applied to the second part of the database, specifically focused on electricity consumption. This strategy is based on the study that the annual profile of electricity consumption remains essentially the same from one year to the next, with variations mainly in the amplitude of consumption.

The major advantage of this approach lies in its efficiency in terms of memory optimization. By avoiding a complete re-training of the model, the company can reduce the amount of historical power consumption data to be stored for future forecasts. This approach has given excellent results, with very low MAE and RMSE values, demonstrating the robustness and reliability of the method.

Electricity generation can have a significant impact on the environment and public health, particularly when fossil fuels such as natural gas are used. Emissions of carbon dioxide (CO2) and other air pollutants from fossil fuel combustion contribute to air pollution, climate change and a range of associated health problems.

The transition to a cleaner electricity generation is essential to reduce this harmful environmental impact. Through the use of renewable energy sources such as solar, wind and hydro, as well as cleaner technologies such as carbon capture and storage, we can reduce CO2 emissions and improve air quality.

Electricity load prediction can play a crucial role in this transition to greener production. With accurate forecasts of electricity demand, power system operators can optimize the use of renewable energy sources, reduce dependence on fossil fuels and minimize the CO2 emissions associated with power generation.

The second contribution focuses on the accurate prediction of electricity consumption. The aim is to assess whether electricity suppliers have exceeded the average threshold for CO2 emissions into the atmosphere, a crucial indicator for assessing their environmental impact. This approach uses sophisticated forecasting model to anticipate future electricity demand and calculate the future CO2 emissions, helping suppliers to adjust their energy mix accordingly.

Analyzing electricity consumption forecasts and comparing them with acceptable CO2 emission thresholds, suppliers can determine if they need to integrate more renewable energies into their production. Renewable energies such as solar, wind and hydropower are cleaner alternatives to fossil fuels, reducing CO2 emissions and helping to combat climate change.

The Transformer attention system has proven its effectiveness in this field, through its ability to capture long-term dependencies between data. Unlike recurrent architectures such as recurrent neural networks (RNNs) and short- and long-term memories (LSTMs), the Transformer uses attention mechanisms that enable it to focus on specific parts of the input sequence during processing.

This attention feature enables the Transformer model to efficiently manage long-term relationships between different parts of the input data, which is particularly beneficial in the field of power load prediction. Power consumption patterns can be complex and influenced by many interdependent factors, requiring in-depth analysis of long-term temporal dependencies.

Compared with RNNs and LSTMs, which can encounter problems of vanishing gradient or sequence forgetting when learning long sequences, the Transformer is better equipped to handle these long-term dependencies with its attention mechanism. This translates into improved performance in power load prediction, offering more accurate and reliable results to support decision-making in power grid management.

This approach enables electricity suppliers to make informed decisions about their energy mix, promoting a transition to greener, more sustainable power generation. By reducing the CO2 emissions associated with electricity generation, this strategy helps preserve air quality and mitigate adverse impacts on the environment and human health.

Perspectives

The prediction models studied achieved excellent results, but it would make enormous progress if they could be further enriched by incorporating a wider variety of data. This would include meteorological, economic and social data, to capture more holistically and accurately the multiple factors influencing electricity consumption.

In addition, exploring new energy management methods, capitalizing on advances in artificial intelligence and optimization, would maximize the use of renewable energies while reducing CO2 emissions. This would also guarantee the reliability of the power grid.

Promoting the deployment of smart grid and advanced metering technologies would have a critical contribution to make in providing real-time data on electricity consumption. This would enable more responsive grid management and more informed decision-making.

Encouraging the research and development of innovative technologies such as advanced energy storage and transport electrification could accelerate the energy transition to a more sustainable future.

Finally, raising awareness among consumers, businesses and political decision-makers of the importance of reducing energy consumption and adopting eco-responsible practices would be essential to encourage greener, more sustainable electricity production.

At the same time, exploring the use of federated learning to guarantee data confidentiality in real time would be a promising avenue to consider.

References

- *Abd Jalil, N. A., Ahmad, M. H., & Mohamed, N. (2013). Electricity load demand forecasting using exponential smoothing methods. World Applied Sciences Journal, 22(11), 1540-1543.*

- *Abiyev, R. H., & Ma'aitaH, M. K. S. (2018). Deep convolutional neural networks for chest diseases detection. Journal of healthcare engineering, 2018.*

- *Adewole, K. S., Akintola, A. G., Abdulsalam, S. O., & Ajiboye, A. R. (2014). Frequent pattern and association rule mining from inventory database using apriori algorithm.*

- *Aditya Ramesh et al. Zero-Shot Text-to-Image Generation. 2021. arXiv: 2102. 12092 [cs.CV].*

- *Ahmad, A., Javaid, N., Mateen, A., Awais, M., & Khan, Z. A. (2019). Short-term load forecasting in smart grids: An intelligent modular approach. Energies, 12(1), 164.*

- *Ahmed, S., Nielsen, I. E., Tripathi, A., Siddiqui, S., Ramachandran, R. P., & Rasool, G. (2023). Transformers in time-series analysis: A tutorial. Circuits, Systems, and Signal Processing, 42(12), 7433-7466.*

- *Aksan, E., Kaufmann, M., Cao, P., & Hilliges, O. (2021, December). A spatio-temporal transformer for 3d human motion prediction. In 2021 International Conference on 3D Vision (3DV) (pp. 565-574). IEEE.*

- *Al-Musaylh, M. S., Deo, R. C., Adamowski, J. F., & Li, Y. (2018). Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. Advanced Engineering Informatics, 35, 1-16.*

- *Al-Musaylh, M. S., Deo, R. C., Adamowski, J. F., & Li, Y. (2019). Short-term electricity demand forecasting using machine learning methods enriched with ground-based climate and ECMWF Reanalysis atmospheric predictors in southeast Queensland, Australia. Renewable and Sustainable Energy Reviews, 113, 109293.*

- *Al-Musaylh, M. S., Deo, R. C., Li, Y., & Adamowski, J. F. (2018). Two-phase particle swarm optimized-support vector regression hybrid model integrated with improved empirical mode decomposition with adaptive noise for multiple-horizon electricity demand forecasting. Applied energy, 217, 422-439.*

- *Amarasinghe, K., Marino, D. L., & Manic, M. (2017, June). Deep neural networks for energy load forecasting. In 2017 IEEE 26th international symposium on industrial electronics (ISIE) (pp. 1483-1488). IEEE.*

- *Awad, M., Khanna, R., Awad, M., & Khanna, R. (2015). Support vector machines for classification. Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers, 39-66.*

• *Baasch, G., Rousseau, G., & Evins, R. (2021). A Conditional Generative adversarial Network for energy use in multiple buildings using scarce data. Energy and AI, 5, 100087.*

• *Barbulescu, C., Kilyeni, S., Deacu, A., Turi, G. M., & Moga, M. (2016, May). Artificial neural network based monthly load curves forecasting. In 2016 IEEE 11th International Symposium on Applied Computational Intelligence and Informatics (SACI) (pp. 237-242). IEEE.*

• *Barbulescu, C., Kilyeni, S., Deacu, A., Turi, G. M., & Moga, M. (2016, May). Artificial neural network based monthly load curves forecasting. In 2016 IEEE 11th International Symposium on Applied Computational Intelligence and Informatics (SACI) (pp. 237-242). IEEE.*

• *Bendaoud, N. M. M., & Farah, N. (2020). Using deep learning for short-term load forecasting. Neural computing and applications, 32, 15029-15041.*

• *Bendaoud, N. M. M., Farah, N., & Ahmed, S. B. (2021). Comparing Generative Adversarial Networks architectures for electricity demand forecasting. Energy and Buildings, 247, 111152.*

• *Berner, C., Brockman, G., Chan, B., Cheung, V., Dębiak, P., Dennison, C., ... & Zhang, S. (2019). Dota 2 with large scale deep reinforcement learning. arXiv preprint arXiv:1912.06680.*

• *Brophy, E., Wang, Z., She, Q., & Ward, T. (2023). Generative adversarial networks in time series: A systematic literature review. ACM Computing Surveys, 55(10), 1-31.*

• *Carter, G. M., Dallavalle, J. P., & Glahn, H. R. (1989). Statistical forecasts based on the National Meteorological Center's numerical weather prediction system. Weather and Forecasting, 4(3), 401-412.*

• *Cecati, C., Kolbusz, J., Różycki, P., Siano, P., & Wilamowski, B. M. (2015). A novel RBF training algorithm for short-term electric load forecasting and comparative studies. IEEE Transactions on industrial Electronics, 62(10), 6519-6529.*

• *Chen, G., & Qian, S. E. (2008). Simultaneous dimensionality reduction and denoising of hyperspectral imagery using bivariate wavelet shrinking and principal component analysis. Canadian Journal of Remote Sensing, 34(5), 447-454.*

• *Chen, H., Canizares, C. A., & Singh, A. (2001, January). ANN-based short-term load forecasting in electricity markets. In 2001 IEEE power engineering society winter meeting. Conference proceedings (Cat. No. 01CH37194) (Vol. 2, pp. 411-415). IEEE.*

• *Chitalia, G., Pipattanasomporn, M., Garg, V., & Rahman, S. (2020). Robust short-term electrical load forecasting framework for commercial buildings using deep recurrent neural networks. Applied Energy, 278, 115410.*

• *Chow, T. W., & Leung, C. T. (1996). Neural network based short-term load forecasting using weather compensation. IEEE transactions on power systems, 11(4), 1736-1742.*

• *Chow, T. W., & Leung, C. T. (1996). Neural network based short-term load forecasting using weather compensation. IEEE transactions on power systems, 11(4), 1736-1742.*

• *Clifton, J., & Laber, E. (2020). Q-learning: Theory and applications. Annual Review of Statistics and Its Application, 7, 279-301.*

• *Convention des Maires, “Les facteurs d’émission.”*

• *Cox, D. D., & Dean, T. (2014). Neural networks and neuroscience-inspired computer vision. Current Biology, 24(18), R921-R929.*

• *Dagdougui, H., Bagheri, F., Le, H., & Dessaint, L. (2019). Neural network model for short-term and very-short-term load forecasting in district buildings. Energy and Buildings, 203, 109408.*

• *De Prado, M. L. (2018). Advances in financial machine learning. John Wiley & Sons.*

• *Deng, D. (2020, September). DBSCAN clustering algorithm based on density. In 2020 7th international forum on electrical engineering and automation (IFEEA) (pp. 949-953). IEEE.*

• *Deo, R. C., & Şahin, M. (2017). Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland. Renewable and Sustainable Energy Reviews, 72, 828-848.*

• *Deo, R. C., & Şahin, M. (2017). Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland. Renewable and Sustainable Energy Reviews, 72, 828-848.*

• *Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.*

• *Ding, C., & He, X. (2004, July). K-means clustering via principal component analysis. In Proceedings of the twenty-first international conference on Machine learning (p. 29).*

• *Ding, Z., & Dong, H. (2020). Challenges of reinforcement learning. Deep Reinforcement Learning: Fundamentals, Research and Applications, 249-272.*

• *Divina, F., Gilson, A., Goméz-Vela, F., García Torres, M., & Torres, J. F. (2018). Stacking ensemble learning for short-term electricity consumption forecasting. Energies, 11(4), 949.*

• *Dong, X., Qian, L., & Huang, L. (2017, February). Short-term load forecasting in smart grid: A combined CNN and K-means clustering approach. In 2017 IEEE international conference on big data and smart computing (BigComp) (pp. 119-125). IEEE.*

• *Dong, Y., Zhang, Z., & Hong, W. C. (2018). A hybrid seasonal mechanism with a chaotic cuckoo search algorithm with a support vector regression model for electric load forecasting. Energies, 11(4), 1009.*

• *Dudek, G. (2015). Short-term load forecasting using random forests. In Intelligent Systems' 2014: Proceedings of the 7th IEEE International Conference Intelligent Systems IS’2014, September 24-26, 2014, Warsaw, Poland, Volume 2: Tools, Architectures, Systems, Applications (pp. 821-828). Springer International Publishing.*

• *El Naqa, I., & Murphy, M. J. (2015). What is machine learning? (pp. 3-11). Springer International Publishing.*

• *Engstrom, L., Ilyas, A., Santurkar, S., Tsipras, D., Janoos, F., Rudolph, L., & Madry, A. (2019, September). Implementation matters in deep rl: A case study on ppo and trpo. In International conference on learning representations.*

• *Fan, G. F., Peng, L. L., & Hong, W. C. (2018). Short term load forecasting based on phase space reconstruction algorithm and bi-square kernel regression model. Applied energy, 224, 13-33.*

• *Fan, J., Wang, Z., Xie, Y., & Yang, Z. (2020, July). A theoretical analysis of deep Q-learning. In Learning for dynamics and control (pp. 486-489). PMLR.*

• *Farfar, K. E., & Khadir, M. T. (2019). A two-stage short-term load forecasting approach using temperature daily profiles estimation. Neural Computing and Applications, 31, 3909-3919.*

• *Fletcher, C. (2018). Climate change: what the science tells us. John Wiley & Sons.*

• *Friedlingstein, P., O'sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., ... & Zaehle, S. (2020). Global carbon budget 2020. Earth System Science Data Discussions, 2020, 1-3.*

• *Gamboa, J. C. B. (2017). Deep learning for time-series analysis. arXiv preprint arXiv:1701.01887.*

• *Gisbrecht, A., Schulz, A., & Hammer, B. (2015). Parametric nonlinear dimensionality reduction using kernel t-SNE. Neurocomputing, 147, 71-82.*

• *Giuliari, F., Hasan, I., Cristani, M., & Galasso, F. (2021, January). Transformer networks for trajectory forecasting. In 2020 25th international conference on pattern recognition (ICPR) (pp. 10335-10342). IEEE.*

• *Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.*

• *Gregg, J. S., & Andres, R. J. (2008). A method for estimating the temporal and spatial patterns of carbon dioxide emissions from national fossil-fuel consumption. Tellus B: Chemical and Physical Meteorology, 60(1), 1-10.*

• *Guo, Y. C., Niu, D. X., & Chen, Y. X. (2006, August). Support vector machine model in electricity load forecasting. In 2006 International Conference on Machine Learning and Cybernetics (pp. 2892-2896). IEEE.*

• *Hamid, M. A., & Rahman, T. A. (2010, March). Short term load forecasting using an artificial neural network trained by artificial immune system learning algorithm. In 2010 12th International Conference on Computer Modelling and Simulation (pp. 408-413). IEEE.*

• *Han, B., Zhang, D., & Yang, T. (2008, June). Energy consumption analysis and energy management strategy for sensor node. In 2008 International Conference on Information and Automation (pp. 211-214). IEEE.*

• *Harrison, I. O., Danisa, A., & Ishaku, B. (2014). Short Term Electric Load Forecasting of 132/33KV Maiduguri Transmission Substation using Adaptive Neuro-Fuzzy Inference System (ANFIS). International Journal of Computer Applications, 107(11).*

• *Hasan, M., & Boris, F. (2006). Svm: Machines à vecteurs de support ou séparateurs à vastes marges. Rapport technique, Versailles St Quentin, France. Cité, 64.*

• *He, W. (2017). Load forecasting via deep neural networks. Procedia Computer Science, 122, 308-314.*

• *Henley, A., & Peirson, J. (1997). Non-linearities in electricity demand and temperature: parametric versus non-parametric methods. Oxford bulletin of economics and statistics, 59(1), 149-162.*

• *Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.*

• *Honghai, F., Guoshun, C., Cheng, Y., Bingru, Y., & Yumei, C. (2005, September). A SVM regression based approach to filling in missing values. In International Conference on Knowledge-Based and Intelligent Information and Engineering Systems (pp. 581-587). Berlin, Heidelberg: Springer Berlin Heidelberg.*

• *Hor, C. L., Watson, S. J., & Majithia, S. (2006, June). Daily load forecasting and maximum demand estimation using ARIMA and GARCH. In 2006 International conference on probabilistic methods applied to power systems (pp. 1-6). IEEE.*

• *Hossain, M. S., & Mahmud, M. Z. (2018). "Electricity consumption forecasting using linear regression with Google Trends data: A case study of New York State." Energy Reports, 4, 17-24.*

• *Hosseinzadeh, M., Rahmani, A. M., Vo, B., Bidaki, M., Masdari, M., & Zangakani, M. (2021). Improving security using SVM-based anomaly detection: issues and challenges. Soft Computing, 25, 3195-3223.*

• *Huang, Z., Zhang, W., Wang, D., & Yin, Y. (2022). A GAN framework-based dynamic multi-graph convolutional network for origin–destination-based ride-hailing demand prediction. Information Sciences, 601, 129-146.*

• *Hussein, A., Sallam, M. E., & Abdalla, M. Y. A. (2023). Exploring New Horizons: Surgical Robots Supported by Artificial Intelligence. Mesopotamian Journal of Artificial Intelligence in Healthcare, 2023, 40-44.*

• *Hyde, O., & Hodnett, P. F. (1997). An adaptable automated procedure for short-term electricity load forecasting. IEEE Transactions on Power Systems, 12(1), 84-94.*

• Imani, M. (2021). Electrical load-temperature CNN for residential load forecasting. Energy, 227, 120480.

• Jetcheva, J. G., Majidpour, M., & Chen, W. P. (2014). Neural network model ensembles for building-level electricity load forecasts. Energy and Buildings, 84, 214-223.

• Jiang, H., Zhang, Y., Muljadi, E., Zhang, J. J., & Gao, D. W. (2016). A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization. IEEE Transactions on Smart Grid, 9(4), 3341-3350.

• Jumper, J., & Hassabis, D. (2022). Protein structure predictions to atomic accuracy with AlphaFold. Nature methods, 19(1), 11-12.

• Kamath, U., Liu, J., & Whitaker, J. (2019). Deep learning for NLP and speech recognition (Vol. 84). Cham, Switzerland: Springer.

• Kaufman, L., & Rousseeuw, P. J. (2009). Finding groups in data: an introduction to cluster analysis. John Wiley & Sons.

• Kavousi-Fard, A., Samet, H., & Marzbani, F. (2014). A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting. Expert systems with applications, 41(13), 6047-6056.

• Kaytez, F., Taplamacioglu, M. C., Cam, E., & Hardalac, F. (2015). Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. International Journal of Electrical Power & Energy Systems, 67, 431-438.

• Khwaja, A. S., Zhang, X., Anpalagan, A., & Venkatesh, B. (2017). Boosted neural networks for improved short-term electric load forecasting. Electric Power Systems Research, 143, 431-437.

• Kim, J., Moon, J., Hwang, E., & Kang, P. (2019). Recurrent inception convolution neural network for multi short-term load forecasting. Energy and buildings, 194, 328-341.

• Kim, T. Y., & Cho, S. B. (2019). Predicting residential energy consumption using CNN-LSTM neural networks. Energy, 182, 72-81.

• Kleynhans, T., Montanaro, M., Gerace, A., & Kanan, C. (2017). Predicting top-of-atmosphere thermal radiance using merra-2 atmospheric data with deep learning. Remote Sensing, 9(11), 1133.

• Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2017). Short-term residential load forecasting based on LSTM recurrent neural network. IEEE transactions on smart grid, 10(1), 841-851.Rahman, A., Srikumar, V., & Smith, A. D. (2018). Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Applied energy, 212, 372-385.

• Kršmanc, R., Slak, A. Š., & Demšar, J. (2013). Statistical approach for forecasting road surface temperature. Meteorological applications, 20(4), 439-446.

• *L'Heureux, A., Grolinger, K., & Capretz, M. A. (2022). Transformer-based model for electrical load forecasting. Energies, 15(14), 4993.*

• *Lahouar, A., & Slama, J. B. H. (2015). Day-ahead load forecast using random forest and expert input selection. Energy Conversion and Management, 103, 1040-1051.*

• *Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., ... & Zaehle, S. (2018). Global carbon budget 2018. Earth System Science Data Discussions, 2018, 1-3.*

• *Lesouple, J., Baudoin, C., Spigai, M., & Tourneret, J. Y. (2021). Generalized isolation forest for anomaly detection. Pattern Recognition Letters, 149, 109-119.*

• *Levi, G., & Hassner, T. (2015). Age and gender classification using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 34-42).*

• *Li, K. L., Huang, H. K., Tian, S. F., & Xu, W. (2003, November). Improving one-class SVM for anomaly detection. In Proceedings of the 2003 international conference on machine learning and cybernetics (IEEE Cat. No. 03EX693) (Vol. 5, pp. 3077-3081). IEEE.*

• *Liu, J., Chen, Y., Zhan, J., & Shang, F. (2018). An on-line energy management strategy based on trip condition prediction for commuter plug-in hybrid electric vehicles. IEEE Transactions on Vehicular Technology, 67(5), 3767-3781.*

• *Liu, Y., Liang, Z., Li, X., & Bakeer, A. (2023, June). Generative Adversarial Network and CNN-LSTM Based Short-Term Power Load Forecasting. In 2023 IEEE 17th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG) (pp. 1-6). IEEE.*

• *Lusis, P., Khalilpour, K. R., Andrew, L., & Liebman, A. (2017). Short-term residential load forecasting: Impact of calendar effects and forecast granularity. Applied energy, 205, 654-669.*

• *Makridakis, S., & Hyndman, R. J. (1974). Forecasting: methods and Applications. Operational Research Quarterly, 25(4), 648-649.*

• *Mao, X., & Li, Q. (2021). Generative adversarial networks for image generation. Springer.*

• *Marino, D. L., Amarasinghe, K., & Manic, M. (2016, October). Building energy load forecasting using deep neural networks. In IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society (pp. 7046-7051). IEEE.*

• *Mayrink, V., & Hippert, H. S. (2016, November). A hybrid method using Exponential Smoothing and Gradient Boosting for electrical short-term load forecasting. In 2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI) (pp. 1-6). IEEE.*

• *Mohamed, N., Ahmad, M. H., Ismail, Z., & Suhartono, S. (2010, July). Short term load forecasting using double seasonal ARIMA model. In Proceedings of the regional conference on statistical sciences (Vol. 10, pp. 57-73).*

- *Mohamed, Z., & Bodger, P. (2005). Forecasting electricity consumption in New Zealand using economic and demographic variables. Energy, 30(10), 1833-1843.*
- *Moravvej, S. V., Mirzaei, A., & Safayani, M. (2021). Biomedical text summarization using conditional generative adversarial network (CGAN). arXiv preprint arXiv:2110.11870.*
- *Murugan, D. K., Said, Z., Panchal, H., Gupta, N. K., Subramani, S., Kumar, A., & Sadasivuni, K. K. (2023). Machine learning approaches for real-time forecasting of solar still distillate output. Environmental Challenges, 13, 100779.*
- *Muzaffar, S., & Afshari, A. (2019). Short-term load forecasts using LSTM networks. Energy Procedia, 158, 2922-2927.*
- *Naim, I., Mahara, T., & Idrisi, A. R. (2018). Effective short-term forecasting for daily time series with complex seasonal patterns. Procedia computer science, 132, 1832-1841.*
- *Namatēvs, I. (2017). Deep convolutional neural networks: Structure, feature extraction and training. Information Technology and Management Science, 20(1), 40-47.*
- *Nasteski, V. (2017). An overview of the supervised machine learning methods. Horizons. b, 4, 51-62.*

- *Nasteski, V. (2017). An overview of the supervised machine learning methods. Horizons. b, 4, 51-62.*

- *Nazir, A., Shaikh, A. K., Shah, A. S., & Khalil, A. (2023). Forecasting energy consumption demand of customers in smart grid using Temporal Fusion Transformer (TFT). Results in Engineering, 17, 100888.*
- *Nepal, B., Yamaha, M., Yokoe, A., & Yamaji, T. (2020). Electricity load forecasting using clustering and ARIMA model for energy management in buildings. Japan Architectural Review, 3(1), 62-76.*
- *Nowicka-Zagrajek, J., & Weron, R. (2002). Modeling electricity loads in California: ARMA models with hyperbolic noise. Signal Processing, 82(12), 1903-1915.*
- *Open, A. I. (2019). Better language models and their implications. San Francisco: OpenAI.*
- *OpenAI. GPT-4 Technical Report. 2023. arXiv: 2303.08774 [cs.CL].*
- *Pak, M., & Kim, S. (2017, August). A review of deep learning in image recognition. In 2017 4th international conference on computer applications and information processing technology (CAIPT) (pp. 1-3). IEEE.*
- *Pao, H. T. (2009). Forecast of electricity consumption and economic growth in Taiwan by state space modeling. Energy, 34(11), 1779-1791.*
- *Papadopoulos, S., & Karakatsanis, I. (2015, February). Short-term electricity load forecasting using time series and ensemble learning methods. In 2015 IEEE Power and Energy Conference at Illinois (PECI) (pp. 1-6). IEEE.*

• *Pelella, F., Viscito, L., Magnea, F., Zanella, A., Patalano, S., Mauro, A. W., & (Bianco, N. 2023). Comparison between Physics-Based Approaches and Neural Networks for the Energy Consumption Optimization of an Automotive Production Industrial Process. Energies, 16(19), 6916.*

• *Quéré, C. et al. (2013). The global carbone budget 1959-2011. [Le bilan mondial du carbone 1959-2011.]*

• *Rahali, A., & Akhloufi, M. A. (2023). End-to-end transformer-based models in textual-based NLP. AI, 4(1), 54-110.*

• *Ramanathan, R., Engle, R., Granger, C. W., Vahid-Araghi, F., & Brace, C. (2001). Short-run forecasts of electricity loads and peaks. In Essays in econometrics: Collected Papers of Clive WJ Granger Volume 1, Spectral Analysis, Seasonality, Nonlinearity, Methodology, and Forecasting (pp. 497-516).*

• *Raupach, M. R., Quéré, C. L., Peters, G. P., & Canadell, J. G. (2013). Anthropogenic CO2 emissions. Nature Climate Change, 3(7), 603-604.*

• *Ribeiro, G. T., Mariani, V. C., & dos Santos Coelho, L. (2019). Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting. Engineering Applications of Artificial Intelligence, 82, 272-281.*

• *Roi Yehoshua, (2023). Random Forest. Medium.com,(https://medium.com/@roiyeho/random-forests-98892261dc49) March 2023*

• *S. Vemuri et al. 1993. Neural network based short term load forecasting. IEEE Transactions on Power Systems. 8: 336-342.*

• *Saber, A. Y., & Alam, A. R. (2017, November). Short term load forecasting using multiple linear regression for big data. In 2017 IEEE symposium series on computational intelligence (SSCI) (pp. 1-6). IEEE.*

• *Sadaei, H. J., e Silva, P. C. D. L., Guimaraes, F. G., & Lee, M. H. (2019). Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series. Energy, 175, 365-377.*

• *Sadaei, H. J., Guimaraes, F. G., da Silva, C. J., Lee, M. H., & Eslami, T. (2017). Short-term load forecasting method based on fuzzy time series, seasonality and long memory process. International Journal of Approximate Reasoning, 83, 196-217.*

• *Salcedo-Sanz, S., Rojo-Álvarez, J. L., Martínez-Ramón, M., & Camps-Valls, G. (2014). Support vector machines in engineering: an overview. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 4(3), 234-267.*

• *Samuel, I. A., Emmanuel, A., Odigwe, I. A., & Felly-Njoku, F. C. (2017). A comparative study of regression analysis and artificial neural network methods for medium-term load forecasting. Indian Journal of Science and Technology, 10(10), 1-7.*

- *Sen P, Roy M, Pal P. Application of ARIMA for forecasting energy consumption and GHG emission:A case study of an Indian pig iron manufacturing organization. Energy 2016;116:1031e8*

- *Severiano, C. A., e Silva, P. C. D. L., Cohen, M. W., & Guimarães, F. G. (2021). Evolving fuzzy time series for spatio-temporal forecasting in renewable energy systems. Renewable Energy, 171, 764-783.*

- *Sheikh, S. K., & Unde, M. G. (2012). Short term load forecasting using ANN technique. International Journal of Engineering Sciences & Emerging Technologies, 1(2), 97-107.*

- *Sidorov, G., Velasquez, F., Stamatatos, E., Gelbukh, A., & Chanona-Hernández, L. (2014). Syntactic n-grams as machine learning features for natural language processing. Expert Systems with Applications, 41(3), 853-860.*

- *Sigauke, C., & Chikobvu, D. (2010). Daily peak electricity load forecasting in South Africa using a multivariate non-parametric regression approach. ORiON, 26(2).*

- *Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... & Hassabis, D. (2017). Mastering the game of go without human knowledge. nature, 550(7676), 354-359.*

- *Sim, Y., Chung, M. J., Kotter, E., Yune, S., Kim, M., Do, S., ... & Choi, B. W. (2020). Deep convolutional neural network–based software improves radiologist detection of malignant lung nodules on chest radiographs. Radiology, 294(1), 199-209.*

- *Solyali, D. (2020). A comparative analysis of machine learning approaches for short-/long-term electricity load forecasting in Cyprus. Sustainability, 12(9), 3612.*

- *Son, H., & Kim, C. (2017). Short-term forecasting of electricity demand for the residential sector using weather and social variables. Resources, conservation and recycling, 123, 200-207.*

- *Su, H., Zio, E., Zhang, J., Xu, M., Li, X., & Zhang, Z. (2019). A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced Deep-RNN model. Energy, 178, 585-597.*

- *Sun, X., Ouyang, Z., & Yue, D. (2017, November). Short-term load forecasting based on multivariate linear regression. In 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2) (pp. 1-5). IEEE.*

- *Suthaharan, S. (2016). Machine learning models and algorithms for big data classification. Integr. Ser. Inf. Syst, 36, 1-12.*

- *Tang, L., Xie, H., Wang, X., & Bie, Z. (2023). Privacy-preserving knowledge sharing for few-shot building energy prediction: A federated learning approach. Applied Energy, 337, 120860.*

- *Tarmanini, C., Sarma, N., Gezegin, C., & Ozgonenel, O. (2023). Short term load forecasting based on ARIMA and ANN approaches. Energy Reports, 9, 550-557.*

- *Testagrose, C., Shabbir, M., Weaver, B., & Liu, X. (2023, May). Comparative Study Between Vision Transformer and EfficientNet on Marsh Grass Classification. In The International FLAIRS Conference Proceedings (Vol. 36).*

- *Thai, M. T., Phan, P. T., Hoang, T. T., Wong, S., Lovell, N. H., & Do, T. N. (2020). Advanced intelligent systems for surgical robotics. Advanced Intelligent Systems, 2(8), 1900138.*

- *TomBrownet al. "Language Models are Few-Shot Learners". In: Advances inNeural Information Processing Systems. Ed. by H. Larochelle et al. Vol. 33. Curran Associates, Inc., 2020, pp. 1877–1901. url: https://proceedings.neurips.cc/ paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf.*

- *Tranmer, M., & Elliot, M. (2008). Multiple linear regression. The Cathie Marsh Centre for Census and Survey Research (CCSR), 5(5), 1-5.*

- *Türkay, B. E., & Demren, D. (2011, December). Electrical load forecasting using support vector machines. In 2011 7th International Conference on Electrical and Electronics Engineering (ELECO) (pp. I-49). IEEE.*

- *Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. BMC medical informatics and decision making, 19(1), 1-16.*

- *Vaisla, K. S., & Bhatt, A. K. (2010). An analysis of the performance of artificial neural network technique for stock market forecasting. International Journal on Computer Science and Engineering, 2(6), 2104-2109.*

- *Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.*

- *Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.*

- *Veisi, H. (2023). Introduction to SVM. In Learning with Fractional Orthogonal Kernel Classifiers in Support Vector Machines: Theory, Algorithms and Applications (pp. 3-18). Singapore: Springer Nature Singapore.*

- *Vens, C., & Costa, F. (2011, December). Random forest based feature induction. In 2011 IEEE 11th international conference on data mining (pp. 744-753). IEEE.*

- *Wang, C., Wang, Y., Ding, Z., Zheng, T., Hu, J., & Zhang, K. (2022). A transformer-based method of multienergy load forecasting in integrated energy system. IEEE Transactions on Smart Grid, 13(4), 2703-2714.*

• *Wang, J. (2017, April). Power system short-term load forecasting. In 2017 5th International Conference on Machinery, Materials and Computing Technology (ICMMCT 2017) (pp. 250-253). Atlantis Press.*

• *Wang, J. Q., Du, Y., & Wang, J. (2020). LSTM based long-term energy consumption prediction with periodicity. Energy, 197, 117197.*

• *Wu, H., Shen, G. Q., Lin, X., Li, M., & Li, C. Z. (2021). A transformer-based deep learning model for recognizing communication-oriented entities from patents of ICT in construction. Automation in Construction, 125, 103608.*

• *Xia, C., Wang, J., & McMenemy, K. (2010). Short, medium and long term load forecasting model and virtual load forecaster based on radial basis function neural networks. International Journal of Electrical Power & Energy Systems, 32(7), 743-750.*

• *Yang, S., Wang, Y., & Chu, X. (2020). A survey of deep learning techniques for neural machine translation. arXiv preprint arXiv:2002.07526.*

• *Yang, W., Wang, J., & Wang, R. (2017). Research and application of a novel hybrid model based on data selection and artificial intelligence algorithm for short term load forecasting. Entropy, 19(2), 52.*

• *Yildiz, B., Bilbao, J. I., & Sproul, A. B. (2017). A review and analysis of regression and machine learning models on commercial building electricity load forecasting. Renewable and Sustainable Energy Reviews, 73, 1104-1122.*

• *Yin, H., Wong, S., Xu, J., & Wong, C. K. (2002). Urban traffic flow prediction using a fuzzy-neural approach. Transportation Research Part C: Emerging Technologies, 10(2), 85-98.*

• *Zhang, G., & Guo, J. (2020). A novel ensemble method for residential electricity demand forecasting based on a novel sample simulation strategy. Energy, 207, 118265.*

• *Zhang, J., Luan, H., Sun, M., Zhai, F., Xu, J., Zhang, M., & Liu, Y. (2018). Improving the transformer translation model with document-level context. arXiv preprint arXiv:1810.03581.*

• *Zhang, T., Schoene, A. M., & Ananiadou, S. (2021). Automatic identification of suicide notes with a transformer-based deep learning model. Internet interventions, 25, 100422.*

• *Zhang, Y., Wang, J., Gorriz, J. M., & Wang, S. (2023). Deep learning and vision transformer for medical image analysis. Journal of Imaging, 9(7), 147.*

• *Zhang, Z., & Hong, W. C. (2019). Electric load forecasting by complete ensemble empirical mode decomposition adaptive noise and support vector regression with quantum-based dragonfly algorithm. Nonlinear dynamics, 98, 1107-1136.*

• *Zhang, Z., Hong, W. C., & Li, J. (2020). Electric load forecasting by hybrid self-recurrent support vector regression model with variational mode decomposition and improved cuckoo search algorithm. IEEE Access, 8, 14642-14658.*

• *Zhao, Z., Xia, C., Chi, L., Chang, X., Li, W., Yang, T., & Zomaya, A. Y. (2021). Short-term load forecasting based on the transformer model. information, 12(12), 516.*

• *Zhuang, J., Guo, Z., Gong, X., Hu, Y., Lian, H., Xu, W., & Zhao, S. (2023, April). Spatial Load Forecasting Method Based on GIS and Generative Adversarial Network. In 2023 8th Asia Conference on Power and Electrical Engineering (ACPEE) (pp. 1965-1969). IEEE.*

• *Zor, K., Timur, O., & Teke, A. (2017, June). A state-of-the-art review of artificial intelligence techniques for short-term electric load forecasting. In 2017 6th international youth conference on energy (IYCE) (pp. 1-7). IEEE.*