PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
Ministry of Higher Education and Scientific Research

UNIVERSITY ECHAHID HAMMA LAKHDAR - EL OUED
FACULTY OF EXACT SCIENCES
Computer Science department

End of study memory
Presented for the Diploma of

ACADEMIC MASTER

Domain: **Mathematics and Computer Science**
Industry: **Computer Science**
Specialty: **Distributed Systems and Artificial Intelligence**

Theme

Forcasting the Energy consumption at El-Oued by Using Artificial Intelligence Techniques.

Presented by:

Boughezala Safa

Djouadi Imane

Defended on 08-june-2023

Jury members:

Mr. Ammar Boucherit	Univ. El Oued	Supervisor
Mr.Meftah Charaf Eddine	Univ. El Oued	President
Mr. Yagoub Mohammed Amine	Univ. El Oued	Examinator

Academic Year 2022/2023

شكر و تقدير

قال رسول الله صلى الله عليه و سلم:

"من لم يشكر الناس لم يشكر الله"

صدق رسول الله صلى الله عليه و سلم

الحمد الله على إحسانه و الشكر له على توفيقه و امتنانه و نشهد أن لا إله إلا الله وحده لا شريك له تعظيما لشأنه و نشهد أن سيدنا و نبينا محمد عبده و رسوله الداعي إلى رضوانه صلى الله عليه و على آله و أصحابه و أتباعه و سلم.

بعد شكر الله سبحانه و تعالى على توفيقه لنا لإتمام هذا البحث المتواضع نتقدم بجزيل الشكر إلى الوالدين العزيزين الذين أعانونا و شجعونا على الاستمرار

في مسيرة العلم و النجاح، و إكمال الدراسة الجامعية و البحث؛ كما نتوجه بالشكر الجزيل إلى من شرفنا بإشرافه على مذكرة تخثنا الأستاذ الدكتور "**بوشريط عمار**" الذي لن تكفي حروف هذه المذكرة لإيفائه حقه بصبره الكبير علينا، ولتوجيهاته العلمية التي لا تقدر بثمن؛ و التي ساهمت بشكل كبير في إتمام و إستكمال هذا العمل؛ إلى كل أساتذة قسم الرياضيات و الاعلام الالي؛ كما نتوجه بخالص شكرنا و تقديرنا إلى كل من ساعدنا من قريب أو من بعيد على إنجاز و إتمام هذا العمل.

" رب أوزعني أن أشكر نعمتك التي أنعمت علي و على والدي و أن أعمل صالحاً ترضاه و أدخلني برحمتك في عبادك الصالحين "

إهداء

أحمد الله عز وجل على منه و عونه لإتمام هذا البحث.

إلى الذي وهبني كل ما يملك حتى أحقق له آماله، إلى من كان يدفعني قدما نحو الأمام لنيل المبتغى، إلى الإنسان الذي امتلك الإنسانية بكل قوة، إلى مدرستي الأولى في الحياة

أبي الغالي على قلبي أطال الله في عمره؛

إلى التي وهبت فلذة كبدها كل العطاء و الحنان، إلى التي صبرت على كل شيء، التي رعتني حق الرّعاية و كانت سندي في الشدائد، و كانت دعواها لي بالتوفيق ، إلى من إرتحت كلما تذكرت إبتسامتها في وجهي نبع الحنان **أمي** أعز ملاك على القلب و العين جزاها الله عني خير الجزاء في الدارين؛

إليهما أهدي هذا العمل المتواضع لكي أُدخل على قلبهما شيئًا من السعادة

إلى **إخوتي**(عماد و زوجته، عبد المومن وزوجته، يحي، عبد الباري و يعقوب)و **أخواتي**(هاجر، خولة، أسماء، مروة و نسيمة) الذي تقاسموا معي عبء الحياة و إلى عصافير و أزهار العائلة(عبيدة، عبد المولى، عائشة، علياء، عبد الغفور،عفراء، تيماء، رملة، ميرا، محمد خالد ،يزيد ،ياسر ،غسان،احمد ياسين، ادم، نوح)و الى جدتي واعمامي و اخوالي و كل عائلة بوغزاله؛

إلى أصدقائي ومعارفي الذين أُجلُّهم وأحترمهم؛

الى عائلتي المستقبلية عائلة "جار الله" حفظهم الله و رعاهم؛

و إلى كل من يؤمن بأن بذور نجاح التغيير هي في ذواتنا و في أنفسنا قبل أن تكون في أشياء أخرى...

قال الله تعالى " : إن الله لا يغير ما بقوم حتى يغيروا ما بأنفسهم" ......

الآية 11 من سورة الرعد

الطالبة: صفاء بوغزاله

إهداء

إلى من أُفَضِّلها على نفسي، ولِمَ لا؛ فلقد ضحَّت من أجلي

ولم تدَّخر جُهدًا في سبيل إسعادي على الدَّوام

(أُمِّي الحبيبة)

نسير في دروب الحياة، ويبقى من يُسيطر على أذهاننا في كل مسلك نسلكه

صاحب الوجه الطيب، والأفعال الحسنة.

فلم يبخل عليَّ طيلة حياته

(والدي العزيز)

إلى إخوتي. . . . سندي وعضدي ومشاطري أفراحي وأحزاني

إلى أصدقائي، وجميع من وقفوا بجواري وساعدوني بكل ما يملكون، وفي أصعدة كثيرة

أُقدِّم لكم هذا البحث، وأتمنَّى أن يحوز على رضاكم.

الطالبة: إيمان جوادي

Abstract

The problem of forecasting energy consumption has become the subject of numerous studies. Locally, the same issue has become of great importance for SONELGAZ institution since it has noticed the recurrence of the problems of a large increase in consumption at specific times (summer and winter).

On the other hand, since the forecast of energy consumption requires the use of reliable Dataset and techniques for the forecasting of the evolution of consumers' energy needs, we aim in collaboration with SONELGAZ through this research to apply artificial intelligence techniques to solve this problem in order to providing an opportunity for SONELGAZ to control the crisis of the increasing demand for energy at critical times.

Key words: Energy Consumption, Forecasting, AI techniques

Résumé:

La problématique de la prévision de la consommation d'énergie fait l'objet de nombreuses études. Localement, le même problème est devenu d'une grande importance pour l'institution SONELGAZ depuis qu'elle a constaté la récurrence des problèmes d'augmentation importante de la consommation à des moments précis (été et hiver).

Néanmoins, étant donné que la prévision de la consommation d'énergie nécessite l'utilisation de données fiables et de techniques de prévision de l'évolution des besoins énergétiques des consommateurs, nous visons en collaboration avec SONELGAZ à travers cette recherche à appliquer des techniques d'intelligence artificielle pour résoudre ce problème afin de fournir une opportunité pour SONELGAZ de maîtriser la crise de la demande croissante en énergie à des moments critiques.

Mots-clés : Consommation d'énergie; Prévisions; Techniques d'IA

ملخص

أصبحت مشكلة التنبؤ باستهلاك الطاقة موضوع دراسات عديدة. محليًا، أصبحت نفس المشكلة ذات أهمية كبيرة لمؤسسة SONELGAZ لأنها لاحظت تكرار مشاكل الزيادة الكبيرة في الاستهلاك في أوقات محددة (الصيف والشتاء).

ومن جهة لأخرى، ونظرًا لأن التنبؤ باستهلاك الطاقة يتطلب استخدام مجموعة بيانات وتقنيات موثوقة للتنبؤ بتطور احتياجات الطاقة للمستهلكين، فإننا نهدف بالتعاون مع SONELGAZ من خلال هذا البحث إلى تطبيق تقنيات الذكاء الاصطناعي لحل هذه المشكلة من أجل توفير فرصة لشركة سونلغاز للسيطرة على أزمة الطلب المتزايد على الطاقة في الأوقات الحرجة.

الكلمات المفتاحية: استهلاك الطاقة، التنبؤ، تقنيات الذكاء الاصطناعي.

contents

List of figures..ix

List of tables...x

General introduction...1

Part 01 Theoretical part...3

chapter 01 Energy consumption...4

Introduction...5

1.Definition of electrical energy..5

2.Methods of generating electrical energy...5

2.1.Wind power plants:...5

2.2.Solar power plants :...5

2.3.Geothermal power plants:...6

2.4.Hydroelectricity power plants:... 6

3.Sonelgaz Company...6

3.1.Definition:..6

3.2.Sonelgaz products..6

3.2.1.Natural gas production..7

3.2.2.Production of electrical energy..7

3.2.3.Renewable energy…………………………………………………………..7

3.3.The goals of the company..7

4.Definition of electrical energy consumption……………………………………..8

5.Electrical energy consumption sectors..8

6.The evolution of electrical energy consumption..9

7.Factors of electrical energy consumption...9

Conclusion...10

chapter 02 forecast techniques...11

Introduction...12

1.Definition of forecast models:..12

2.Type of forecast models:..12

2.1.machine learning models...12

2.2.Neural Networks and Deep Learning..14

2.3.statistical models:……………………………………………………………….. 16
3.Evaluation metrics of Machine Learning models..18
4.Litterature review:..20
4.1.Electricity demand estimation method..20
4.2.Previous studies on using AI techniques to predict electricity consumption.........20
Conclusion...22
Part02 Practical part.. 23
chapter 03 Implementation and results……………………………...…………24
Introduction.. 25
1.Proposed approach for electricity consumption prediction......................................25
1.1.Data Collection and Preprocessing..25
1.2.Techniques used for the prediction...28
1.3.Results and Analysis...29
1.3.1.Overview of the results...29
Part One: municipality of El-Oued..29
Part Two: El-Oued city.. 33
1.3.2.Results analysis...36
Part one: municipality of El-Oued:... 36
Part two: El-Oued city:..38
1.4.Results and discussion...39
Conclusion...40
General Conclusion...41
*B*ibliograph*y*.. 42

List of figures

Figure 2.1	Forecast models……………………………………………………….	12
Figure 2.2	Linear regression model predictions……………………………….	14
Figure 3.1	Machine learning Workflow……………………………………….	25
Figure 3.2	Data format after restructure……………………………………….	27
Figure 3.3	Data format after organizing and merge……………………………	27
Figure 3.4	Prediction by DT…………………………………………………...	36
Figure 3.5	Prediction by RF……………………………………………………	36
Figure 3.6	Prediction by LSTM………………………………………………..	37
Figure 3.7	Prediction by Prophet…………………………………………….....	37
Figure 3.8	Prediction by SARIMA…………………………………………….	37
Figure 3.9	Prediction by DT…………………………………………………...	38
Figure 3.10	Prediction by RF……………………………………………………	38
Figure 3.11	Prediction by LSTM………………………………………………..	38
Figure 3.12	Prediction by Prophet………………………………………………	39
Figure 3.13	Prediction by SARIMA…………………………………………….	39

List of tables

Table 2.1	exemples of results of related works…………………………...	22
Table 3.1	Variables of the dataset……………………………………...	25
Table 3.2	File of first season…………………………………………...	26
Table 3.3	The features and label of dataset (case1)………………………	26
Table 3.4	The features and label of dataset(case2)…………………….....	27
Table 3.5	The DT results for the testing and training phase……………...	29
Table 3.6	The RF results for the testing and training phase……………...	30
Table 3.7	The LSTM results for the testing and training phase……….....	30
Table 3.8	The Prophet results for the testing and training phase……...….	31
Table 3.9	The SARIMA results for the testing and training phase…….....	31
Table 3.10	The performance evaluation for different techniques………....	32
Table 3.11	The DT results for the testing and training phase……………...	33
Table 3.12	The RF results for the testing and training phase……………...	33
Table 3.13	The LSTM results for the testing and training phase……….....	34
Table 3.14	The Prophet results for the testing and training phase……...….	34
Table 3.15	The SARIMA results for the testing and training phase…….....	34
Table 3.16	The performance evaluation for different techniques………....	35

General introduction

Electricity is one of the most significant and prevalent sources of energy as well as one of the primary and essential resources for people, since they are related on it for the performance of all of their economic and social activities in everyday life. In other words, electricity is a crucial pillar of the national economy and its sustainable services are the basis of people's daily lives.

In this context, and given the importance of electricity for humans because of the benefit it provides, there has been a significant increase in its consumption, whether for domestic or for industrial use. Even though there are a number of contributing elements to this increase, including population growth and technological advancement, it is crucial for electricity generation and distribution companies to establish a short and long-term strategies to predict the rate of consumption.

Research problem

Due to the overt growth in electricity consumption, it is so vital to investigate the feasibility of forecasting energy consumption in order to accurately estimate demand, save energy, and prevent any problem due to the deficit.

To do this, the use of artificial techniques to predict electrical energy consumption using historical data for El-Oued city over a period of 10 years, will be very advantageous.

In this study, we will address an important issue represented by the application of artificial intelligence techniques in order to predict the future consumption of electricity and to give the opportunity to the SONELGAZ institution to prepare for the future demand, by careful monitoring of medium voltage electricity sales chains intended for domestic consumption at of El-Oued city in the period between the first quarter of 2012 and the last quarter of the year 2022.

Therefore, we can summarize our problematic in the following questions:

- Is it possible to forecast the consumption of electricity intended for household use in El-Oued city ?

- To what extent are we able to predict with high accuracy the amount of electricity consumption of households in El Oued ?

- What is the best machine learning model that can predict household consumption in El-Oued city with high accuracy ?

Importance of the subject:

Research on predicting electricity usage is crucial for a number of reasons. Therefore, the interest of this study lies in brief in the following three points:

1. It enables companies that produce and distribute electricity to plan their energy supply and generation based on forecast demand, which can help ensure a reliable power supply and prevent power outages.
2. An accurate prediction of electricity consumption enables consumers to better control their energy consumption and reduce costs. If consumers know ahead of time that electricity demand will be high at a certain time, they can take measures to decrease their electricity usage during that time, such as turning off lights or unnecessary equipment and therefore, avoid electricity cut-off.
3. An electricity consumption prediction study is very important for SONELGAZ company, in order to plan investments in energy infrastructure and identify the needs for renewable energy sources.

Research goals

We aim to achieve a number of goals through our research on this subject. The most crucial of these goals is to:

- Find the most accurate model to predict household electricity consumption in El-Oued city by comparing various machine learning models.

- Provide a trained model to the SONELGAZ institution so that its technicians can use machine learning and artificial intelligence to predict the amount of electricity will be used in a specific period/time.

Document structure

This work is composed of four chapters, where we have devoted two chapters for the theoretical part and only one chapter for the practical part. In the first chapter, we discussed the concepts related to the problem of electricity consumption as well as the methods used to solve this problem. Then, in the second chapter, we described how machine learning models work and their importance in the case of our study. In the third chapter, we presented some existing works in the literature in order to benefit of their advantages.

Finally, in the last chapter we focused on the presentation and analysis of results obtained from our approach with the machine learning models used for the prediction of the amount of electricity consumption.

Part 01
Theoretical part

chapter 01
Energy consumption

Introduction

Energy is considered one of the most important pillars of development in any country, especially electric energy, because it is the pillar of the global economy, and it is also a criterion that explains the country's progress or backwardness, which causes the occurrence of a negative energy. Consumption and its demand for continuous increase, whether by individuals or economic institutions. In this chapter, we present some general concepts of electrical energy and the sectors that consume it, and then we move on to presentation introduction to the Algerian National Electricity and Gas Company.

1. Definition of electrical energy

Electrical energy is indeed derived from the movement of charged particles, typically electrons, through a medium. This movement of charged particles generates an electric current, which can be harnessed and used to power various devices and systems. The energy carried by the electric current can be measured in terms of voltage, current, and power, and is typically expressed in units such as volts, amperes, and watts. Electrical energy is a versatile and widely used form of energy, powering everything from small electronic devices to large industrial machines and power grids [1].

2. Methods of generating electrical energy

2.1. Wind power plants:

Wind turbines are machines that convert the kinetic energy from the wind into electrical energy. Wind power plants consist of multiple wind turbines connected to the power grid.

Wind Power in Algeria, The availability of wind energy typically varies from one topographic region to another and also depends on the climate. The climate of Algeria varies substantially between its northern and southern regions. The northern part is special since it boasts a prime Mediterranean location, the Atlas Mountains, and other high plains. However, the southern winds are more powerful than the ones in the north. The majority of southern lands are lower in latitude than the northern region, and more than 70% of Algeria's total surface area is covered by desert. Southern winds have speeds between 4 and 6 m/s. The best location is thought to be Adrar because of its well-known operational and strong wind characteristics. Strong winds can cause damage to a high hill or ridge [2].

2.2. Solar power plants :

Solar power plants use photovoltaic cells to convert solar radiation into electrical energy. They can be built on a large scale or as smaller, decentralized systems.

In 1988, Algeria began using solar energy in the Southern project. Algeria began equipping bigger towns, like Skikda and Oran, with the necessary tools to maximize the potential of solar energy overall. Either a CSP (concentrated solar power plant) system or a PV (photovoltaic) system can be installed to produce solar energy [2].

2.3.Geothermal power plants:

Geothermal power plants generate electricity by utilizing heat from the earth's core. Water is pumped into the ground where it is heated by the earth's natural heat and then returned to the surface as steam, which powers a turbine connected to a generator [3].

Algeria is known for its collection of hot springs, which number more than 200 and are dispersed around the country [2].

2.4.Hydroelectricity power plants:

Hydroelectricity is a type of renewable energy that produces electricity by harnessing the force of moving water [4].

Hydroelectricity in algerian, Despite being significant and projected to be 65 billion m3, the overall flows through Algerian territory are of little use to the nation because of the following factors: fewer days with rain, concentration in fewer places, high evaporation, and swift evacuation to the sea.

From a schematic perspective, the surface resources get scarcer as you move south. Currently, there are estimated to be 25 billion cubic meters of useable and renewable energy, of which roughly two thirds come from surface resources. With 265GWh produced in 2003, hydraulic electricity made up less than 1% of all power output [2].

3. Sonelgaz Company

3.1. Definition:

SONELGAZ is one of the oldest known institutions in Algeria, a public electricity and gas company that actively contributes to the economic and industrial development of the country. In order to provide more about this company, we will discuss the products it offers to its customers, and its tasks and functions. At the end, we present the goals for which the company goes above and beyond [5].

3.2.Sonelgaz products

Sonelgaz, like any other company, provides products to its customers and The products offered by Sonelgaz are as follows [5]:

3.2.1. Natural gas production

Sonelgaz produces natural gas, which is transported and distributed throughout the national territory. The following statistics represent information about the gas network :

Natural Gas penetration rate: **65%**

Transport network: **23,194 km**

Distribution network: **139,322 km**

Length of the gas network: **162,516 km**

Number of customers: **6,886,407**

3.2.2. Production of electrical energy

Where is electricity produced, transported and distributed on the national territory, The following figures provide information on the electricity network:

Electrification rate: **98%**

Installed capacity: **24,561 MW**

Transport network: **32,720 km**

Distribution network: **367,573 km**

Length of the electrical network: **400,293 km**

Number of customers: **10,983,538**

3.2.3. Renewable energy

This energy is produced by power plants of the following types:

Photovoltaic plant: **356.1 Mwp**
Wind power plant**: 10.2 Mw**

3.3.The goals of the company

The Sonelgaz company strives, through the missions and functions it exercises, to achieve a certain number of objectives and results which have been set as follows:

• Act to meet the ever-increasing demand for electrical energy through the optimal use of basic resources and take advantage of them while preserving the environment.

• Sonelgaz Group companies plan to develop electricity production, transmission and distribution activities, as well as gas transmission and distribution activities.

• Development and improvement in the field of energy services to develop and diversify its products.

• The continued development of maintenance and operation equipment and management and operation modernization projects to participate in industrial and commercial achievements around the world to reach the end customer.

• In general, Sonelgaz Electricity and Gas Company is an industrial and commercial company that always seeks to participate in global competitiveness to obtain a share of the global market [5].

4. Definition of electrical energy consumption

Electric energy consumption is the total amount of power utilized for different purposes, including transportation, residential, commercial, industrial, and other ad hoc uses. The actual need for electricity from the current electricity supply is to run electrical equipment like machines, lighting, and appliances. Kilowatt-hours (kWh) are a common unit of measurement for electric energy consumption, and they are a crucial indicator for comprehending patterns of energy use and spotting chances to increase energy efficiency [6].

5. Electrical energy consumption sectors

Several sectors fall into the field of electrical energy consumption, and this is due to its various and different uses. The most prominent sectors in which electrical energy appears are [6]:

5.1. Agricultural sector

Electrical energy is used as fuel for means and machines such as tractors, water pumps, etc. as direct uses, while it is also used indirectly in the production of livestock feed and fertilizers, among other uses.

5.2. Residential sector

Electricity is available, especially to citizens in general, to meet their particular needs for lighting and heating, and for household chores in general (such as cooking).

5.3. Public service sector

Its share of electrical energy is limited to the perimeter of public services such as commercial buildings, hospitals, educational establishments and the rest of the sectors (transmission sector, public works sector).

5.4. Industrial and economic sector

Where electricity is used in all industrial projects and products of plastic, rubber, textile and other industries.

6. The evolution of electrical energy consumption

Indeed, technological advancements and the increasing access to electricity have played a significant role in the rise of electrical energy consumption. The invention of new electrical appliances and devices, coupled with the increasing use of electricity in different sectors, has led to a substantial increase in global electricity consumption.

For instance, the widespread use of air conditioners and refrigerators in households and the use of electric-powered machinery in industries has significantly contributed to the rise of electrical energy consumption. Additionally, the growth of the population and the expansion of urban areas have also played a role in the increase of electricity demand.

As a result of these factors, the global electrical energy consumption has been on a steady rise since the late 19th century, and it is projected to continue increasing in the coming years. Therefore, it is essential to find ways to balance the increasing demand for electricity with sustainable energy solutions to mitigate the adverse effects of over-reliance on non-renewable sources of energy [7].

7. Factors of electrical energy consumption

We note that the frequency of electricity consumption increases every year, and the variables and factors differ for economic, technological, cultural, social, and demographic reasons. The following points summarize the most important factors driving energy consumption:

• Climate change: the increase in electricity consumption coincides during the hot months (summer) and decreases during the cold months (winter).

• Population growth: as new neighborhoods are built and constructed, which generates an increase in the number of subscribers to networks, and thus increases the demand for electricity.

• Technological development: including high-level and efficient technology in social life and in various sectors such as industry, economy, etc.

• Elevation and improvement of the standard of living: which results in an increase in the use of luxury products such as electrical appliances and others.

• Economic growth: it reflects the increasing uses of electrical energy by economic agents and the rest of the world sectors (transport sector, public works, etc.) [6].

Conclusion

In this chapter, we conducted a general study on the consumption of electrical energy and its direction, which require searching for other modern methods that are efficient and accurate, such as machine learning models, of which we will present some models and which we will try Explain how it works in the next chapter.

chapter 02
forecast techniques

Introduction

The study of forecast models plays a major role in the planning process. It is very important in the process of evaluating the growth and development of some variables related to the field studied. Our main goal is to come up with a prediction that provides future information based on previous studies. In this chapter we will introduce the definition of forecast models, discuss the study of the most used prediction models and identify the most important algorithms, with the aim of selecting the most suitable one for our study.

1. Definition of forecast models:

A forecasting model is a statistical tool designed to predict future trends and outcomes based on historical data. It involves analyzing past patterns and trends to make informed predictions about future events, sales, demand, or inventory levels [8].

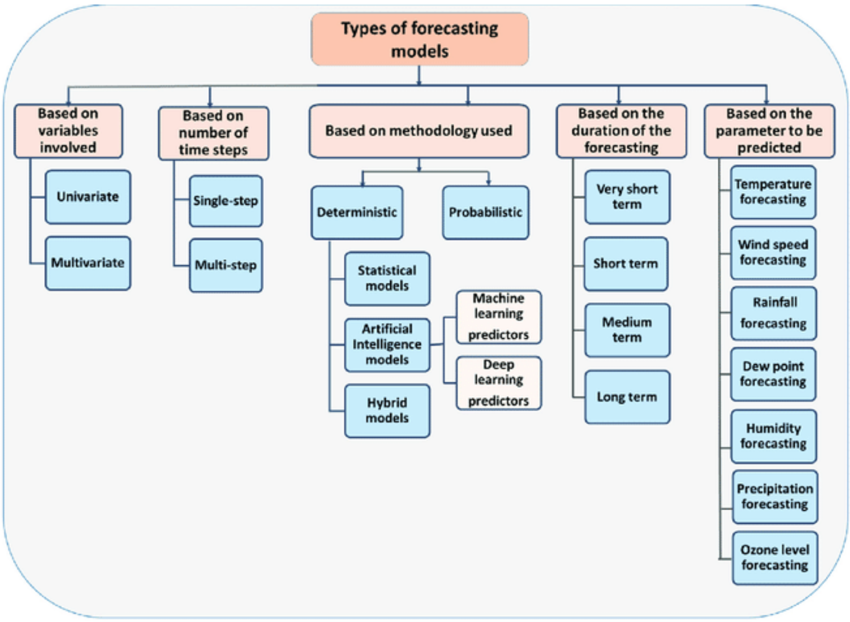

Figure 2.1: Forecast models.

2. Type of forecast models:

There are many types of Forecasting, including:

2.1. machine learning models

Machine learning is a subset of artificial intelligence that uses statistical techniques to allow computer systems to automatically learn and improve from experience, without being explicitly programmed. It involves training models on large amounts of data, identifying patterns and relationships, and using those patterns to make predictions or take actions on new data. Machine learning has become increasingly popular due to the explosion of digital data and the need for more efficient and accurate analysis methods. It has been successfully applied to a wide range of fields, including image and speech recognition, natural language processing, recommendation systems, and predictive analytics [6].

2.1.1. Types of machine learning models:

There are many machine learning models used in forcasting, including:

- **Regression model**

Regression are a type of supervised learning.it is used when predicting a continuous variable, which can therefore take any value. The classes representing the note change have therefore were considered constant and the regression algorithm allowed to obtain decimal values that we finally rounded to get a vector of predicted level [9].

The algorithms used for the regression are

- Linear regression

Linear regression is a type of regression analysis in which a linear relationship is established between the dependent variable and one or more independent variables. The goal is to find the best linear relationship that describes the variation in the dependent variable based on the independent variables. The equation of a simple linear regression model is of the form y = b0 + b1x, where y is the dependent variable, x is the independent variable, b0 is the y-intercept, and b1 is the slope of the line.

In multiple linear regression, there are multiple independent variables, and the equation becomes y = b0 + b1x1 + b2x2 + ... + bnxn. The goal is still to find the best linear relationship that describes the variation in the dependent variable based on the independent variables.

Linear regression is commonly used in data analysis and machine learning for tasks such as predicting a continuous variable or determining the strength of the relationship between two variables [10].

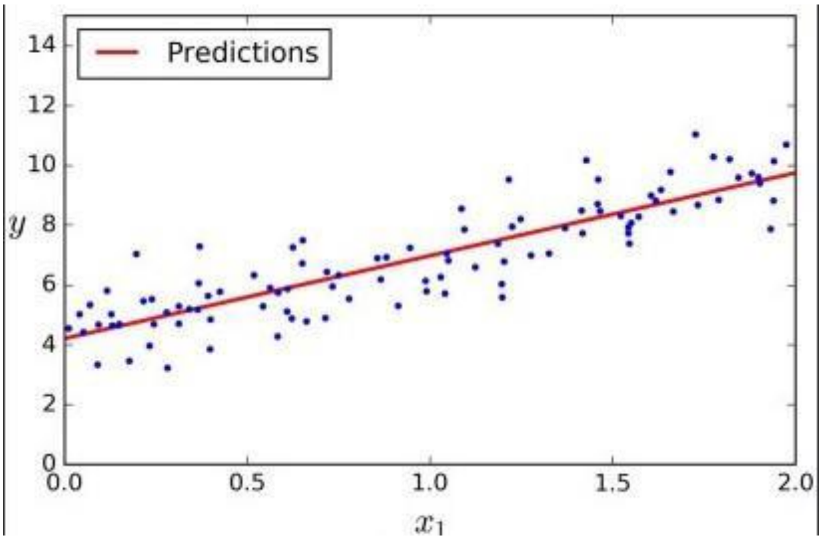

Figure 2.2: Linear regression model predictions [6].

- **Regression Tree**

Decision trees for regression work by recursively splitting the data into smaller subsets based on the values of the predictor variables, until a stopping criterion is met. The splitting criterion is typically based on the reduction in the variance of the response variable in the subsets. Once the tree is built, the predicted response for a new observation is obtained by traversing the tree from the root to the leaf node that corresponds to the new observation and using the mean response value of the training data in that leaf node as the predicted response. Decision trees for regression are popular because they are easy to interpret and can handle nonlinear relationships between the predictors and the response [10].

- **Random forest**

In random forests, each tree in the ensemble is built using a bootstrap sample of the training data, which means that some samples may appear multiple times and others may be left out. This is known as bagging, and it helps to reduce the variance of the model by reducing overfitting.

When building each tree, instead of considering all input features at each split, only a random subset of features is considered. This helps to introduce some randomness into the model and reduces correlation among the trees in the ensemble. The parameter max_features controls the size of the feature subset to be considered at each split. By default, max_features is set to the square root of the total number of features.

2.2. Neural Networks and Deep Learning

Artificial Neural Networks (ANN) or Simulated Neural Networks (SNN) are modeled after the structure and function of the human brain. The basic building block of an ANN is a neuron, which receives inputs from other neurons, processes them, and produces an output signal. The output signal of one neuron can be the input signal of another neuron, allowing information to be transmitted and processed throughout the network.

The input layer of an ANN receives data from the outside world and passes it to the hidden layers, which perform intermediate computations on the data. The output layer of the network produces the final output of the network based on the computations performed by the hidden layers. During the training process, the weights and thresholds of the neurons in the network are adjusted so that the network produces the desired output for a given input.

ANNs are commonly used in machine learning tasks such as classification, regression, and pattern recognition, and have achieved state-of-the-art performance on a wide range of tasks. However, they can be computationally expensive to train and may require large amounts of data to achieve good performance.There are many models of neural networks are [11]:

2.2.1. Recurrent Neural Networks (RNNs)

RNNs are particularly suited for tasks that involve sequential data, where each element in the sequence is dependent on the previous ones. One of the main features of RNNs is the use of a hidden state that allows the network to maintain a memory of the past inputs. This hidden state is updated at each step of the sequence and serves as an input to the next step, allowing the network to make predictions based on its previous state.

However, RNNs can suffer from the problem of vanishing gradients, which occurs when the gradients used to update the weights during training become very small, leading to slow convergence or even stagnation of the training process. This issue is particularly pronounced in long sequences, where the effect of the gradients can quickly become diluted over time. To address this issue, a variation of RNNs called Long Short-Term Memory (LSTM) networks was introduced. LSTMs have a more complex architecture than traditional RNNs, with additional gating mechanisms that allow them to selectively forget or remember information from the past. This makes them particularly effective at modeling long-term dependencies in sequences.

The recurrent neural network algorithms are:

- **LSTM (Long Short Term Memory)**

LSTM is an extension of RNN, introduced by Hochreiter and Schmidhuber in 1997, designed to avoid the long-term dependency issue, unlike RNN, LSTM can remember data for long periods. In RNN architecture , hidden layers have a simple structure (e.g. single tanh layer), while the LSTM architecture is more complex, It is constituted of 4 hidden layers.

The principal component of LSTM is the cell state. To add or remove information from the cell state, the gates are used to protect it, using sigmoid function (one means allows the modification, while a value of zero means denies the modification.). We can identify three different gates :

Forget gate layer : Looks at the input data, and the data received from the previously hidden layer, then decides which information LSTM is going to delete from the cell state, using a sigmoid function (One means keeps it, 0 means delete it). It is calculated as: $f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f)$

Input/Update gate layer : Decides which information LSTM is going to store in the cell state. At first, input gate layer decides which information will be updated using a sigmoid function, then a Tanh layer proposes a new vector to add to the cell state. Then the LSTM update the cell state, by forgetting the information that we decided to forget, and updating it with the new vector values. It is calculated as $i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i)$

and $\tilde{C}_t = \tan h\,(W_c.[h_{t-1}, x_t] + b_C)$

Output Layer : decides what will be our output by executing a sigmoid function that decides which part of the cell LSTM is going to output, the result is passed through a Tanh layer (value between − 1 and 1) to output only the information we decide to pass to the next neuron. It is calculated as: $O_t = \sigma(W_0[h_{t-1}, x_t] + b_0)$ and $h_t = o_t * \tan h\,(C_t)$ [12].

- Facebook prophet

Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well [13].

2.3. statistical models:

- **time series:**

Utilizing historical data to make predictions about the future is known as time-series forecasting. Time-series forecasting models are statistical models that employ previous data points ordered in chronological order to predict future values. These models extrapolate data trends and patterns after analysis in order to forecast future values. Future events are predicted using time-series forecasting models, a specific type of predictive modeling [14].

- **Autoregressive (AR) models:**

An autoregressive (AR) model is a type of statistical model used to describe a time series data. In an AR model, the value of the dependent variable at a given time point is modeled as a linear combination of its own past values. The order of the AR model, denoted as 'p', represents the number of lagged values used in the model.

Mathematically, an AR(p) model can be expressed as:

$$Y_t = c + \phi_p Y_{t-p} + \varepsilon_t$$

Where:

Y_t:represents the dependent variable at time t

c:is a constant term.

ϕ_p:the coefficient.

Y_{t-p}:represent the lagged value of the dependent variable.

ε_t:the error term at time t [14].

- **moving average (MA):**

Moving Average (MA) is another type of statistical model commonly used in time series analysis. MA models focus on the past error terms or residuals.

In an MA model, the value of the dependent variable at a given time point is represented as a linear combination of error terms from previous time points. The order of the MA model, denoted as 'q', indicates the number of lagged error terms included in the model.

The general form of an MA model can be represented as:

$$Y(t) = \theta_1 * \varepsilon(t-1) + \theta_2 * \varepsilon(t-2) + \cdots + \theta_q * \varepsilon(t-q) + \varepsilon(t)$$

$Y(t)$ represents the dependent variable at time t.

μ is the constant term.

$\varepsilon(t), \varepsilon(t-1), \dots, \varepsilon(t-q)$are the error terms or residuals at different time points.

$\theta_1, \theta_2, \dots, \theta_q$are the coefficients that measure the influence of past error terms on the current value [15].

- **Autoregressive Moving Average(ARMA):**

The ARMA model serves as a foundation for various forecasting models in different fields. It consists of a constant term added to the combination of autoregressive (AR) lags and their coefficients, along with the combination of moving average (MA) lags and their coefficients, and white noise. This equation forms the underlying structure for subsequent models and provides a framework for a wide range of forecasting approaches in diverse domains [16].

The general form of an ARMA model can be represented as:

$$Y_t = c + \phi_p Y_{t-p} + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

- **Autoregressive integrated moving average (ARIMA) model:**

ARIMA stands for Autoregressive Integrated Moving Average. It is a popular and powerful time series forecasting model that combines autoregressive (AR), moving average (MA), and differencing techniques to capture both short-term and long-term dependencies in a time series.

The order of an ARIMA model is denoted as (p, d, q), where:

- p represents the order of the AR component.
- d represents the degree of differencing required to achieve stationarity.
- q represents the order of the MA component.

The parameters (p, d, q) are determined through statistical methods, such as analyzing autocorrelation and partial autocorrelation plots [17].

- **Seasonal Autoregressive Integrated Moving Average(SARIMA):**

is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component. It is specifically designed to capture and forecast data with recurring seasonal fluctuations or patterns.

The notation for SARIMA models is (p, d, q)(P, D, Q, s), where:

(p, d, q) represents the non-seasonal AR, differencing, and MA components, as in ARIMA.

(P, D, Q) represents the seasonal AR, seasonal differencing, and seasonal MA components.

s denotes the length of the seasonal cycle, such as the number of observations in a season.

The parameters (p, d, q, P, D, Q) are determined by analyzing autocorrelation and partial autocorrelation plots, similar to ARIMA [18].

3. Evaluation metrics of Machine Learning models

- **Mean Absolute Error (MAE)**

MAE stands for Mean Absolute Error and it measures the average magnitude of the errors in a set of predictions, without considering their direction (positive or negative). It is defined as the average of the absolute differences between the predicted values and the actual values. MAE is a common metric used in regression analysis to evaluate the performance of a model, particularly when the dataset contains outliers or when we want to have a more balanced view of the errors.

$$MAE(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} |y_i - \hat{y}_i|$$

where ŷ is the predicted label, y is the true label and n is the number of samples used. It helps in determining dissimilarity between predicted outcomes and actual outcomes.To determine the average error, it is a more natural technique [19].

- **Mean Square Error (MSE)**

When comparing various estimators, the mean square error is useful, especially when one of them is biased [9].

$$MSE = \frac{1}{n}\sum_{j=1}^{n}(y - y_i)^2$$

- n= The number of errors
- y =value of consumption
- y_i = predicted value.

- **Root Mean Square Error (RMSE)**

Root Mean Squared Error (RMSE) is also a measure of calculating the variances between the estimated value and the true value. It is calculated by the square root of the Mean Square Error (MSE) [20].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y - y_i)^2}$$

- **Normalized Root Mean Square Error (NRMSE)**

The RMSE makes it easier to compare models with different scales. the normalized RMSE(NRMSE), which connects the RMSE to the variable's observed range [9].

$$NRMSE = \frac{RMSE}{\bar{y}}$$

- **R2 Score**

The R2 score, also known as the coefficient of determination, is a statistical measure that represents the proportion of the variance in the dependent variable (target) that is predictable from the independent variables (features) in a regression model. In other words, it measures how well the model fits the data by comparing the variability of the actual target values with the variability of the predicted target values.

The R2 score ranges from 0 to 1, with 1 indicating a perfect fit and 0 indicating that the model does not explain any of the variability in the target variable. A negative R2 score indicates that the model is worse than a horizontal line (i.e., predicting the mean of the target variable for all samples). The R2 score can be used to evaluate the performance of regression models and compare different models [19].

$$R - Squared = 1 - \frac{First\ Sum\ of\ Errors}{Second\ Sum\ of\ Errors}$$

where First Sum of Errors is the sum of squared residuals (the difference between the predicted and actual values), and Second Sum of Errors is the total sum of squares (the total variability in the dependent variable).

4. Litterature review:

4.1. Electricity demand estimation method

Large businesses and economic organizations use a variety of tools and techniques to estimate the demand for electrical energy consumption because they depend on the following factors when estimating expected electricity consumption:

The method based on referencing the date of electricity consumption on the same day from the same date in the previous years to make a prediction for the present and calculate the rate of increase, which is calculated in this way

Consumption next year = Consumption last year + x. Consumption last year.

Where x: the Rate of increase.

x. Last year's consumption: the value of last year's load.

The increase in the growth rate which was originally calculated for the next year, and if it turns out that the consumption increases or decreases, it is modified according to the situation [6].

4.2. Previous studies on using AI techniques to predict electricity consumption

Paper 01: Using Support Vector Machine to Predict Next Day Electricity Load of Public Buildings with Sub-metering Devices

Accurate short-term electrical load forecasting is essential for enabling demand side management in the construction industry. It is feasible to forecast both the overall electrical load and the loads of particular building service systems (such as air conditioning, lighting, power, and other equipment) for buildings that have electricity sub-metering systems installed. In this article, a Support Vector Machine (SVM)-based approach is suggested to forecast system-level loads. 24 SVM models (one model per hour) were developed and used to forecast the hourly electricity load for each type of system. Simple weather predictions and hourly electricity usage from the preceding two days serve as the prediction method's inputs. The proposed method beats three other common data mining techniques, according to a case study.Both CV_RMSE and N_MBE use (ARIMAX, Decision Tree, and Artificial Neural Network). In order to forecast system level electrical consumption for public buildings, the SVM approach is recommended [21]

Paper 02: Research on electricity consumption forecasting model based on wavelet transform and multi-layer LSTM model

This research suggests a forecasting model based on the combination of wavelet transform and multiple LSTM that takes into account the time series properties of electricity consumption data. The experimental results directly demonstrate that the method in this paper has significantly improved the prediction accuracy of daily electricity consumption, and the noise reduction processing of WT can be used to some extent through training and prediction of sample data, as well as horizontal comparison with traditional LSTM and Bi-LSTM algorithms. Increase the model's stability and precision. With the use of this technology, it is possible to more accurately anticipate daily electricity consumption, create practical plans for the generation and transmission of power, and efficiently reduce the waste of electricity resources [22].

Paper 03: Electricity Consumption Prediction in an Electronic System Using Artificial Neural Networks

In this paper, a method for electricity consumption prediction based on artificial neural networks is proposed. The electricity consumption dataset is obtained from a cold storage facility, which generates data in hourly intervals. The data obtained are measured for a period of over 2 years and then separated to four seasons, so different models are developed for each season. Five different network structures (ordinary RNN, LSTM, GRU, bidirectional LSTM, bidirectional GRU) for five different values of horizon, i.e., input data (one day, two days, four days, one week, two weeks) are examined. Performance indices, such as mean absolute percentage error (MAPE), root mean square error (RMSE), mean absolute error (MAE) and mean square error(MSE), are used in order to obtain qualitative and quantitative comparisons among the obtained models. The results show that the modifications of recurrent neural networks perform much better than ordinary recurrent neural networks. GRU and LSTMB structures with horizons of 168h and 336h are found to have the best performances [23].

Paper 04: Energy consumption prediction by using machine learning for smart building: Case study in Malaysia

This research aims to address the problems by developing a predictive model for energy consumption in Microsoft Azure cloud-based machine learning platform. Three methodologies which are Support Vector Machine, Artificial Neural Network, and k-Nearest Neighbour are proposed for the algorithm of the predictive model. Focusing on real-life application in Malaysia, two tenants from a commercial building are taken as a case study. The data collected is analyzed and pre-processed before it is used for model training and testing. The performance of each of the methods is compared based on RMSE, NRMSE, and MAPE metrics. The experimentation shows that each tenant's energy consumption has different distribution characteristics [24].

We identified several study publications related to the forecasting of electric energy consumption. After submitting them, we classify them according to the year of publication, the technology used, the data set, and performance measures. We summarize them in the following table:

Paper	Year	Method	Data used	Performance matrics
[21]	2015	ARIMAX REPTree ANN SVM	electricity data of Shanghai during 2013	NMBE :12.4% /CV_RMSE :22.4% NMBE :12.1% / CV_RMSE :27.2% NMBE :11.9% / CV_RMSE :22.1% NMBE :7.7% / CV_RMSE :15.2%
[22]	2022	LSTM	The electricity consumption in the area controlled by the U.S	RMSE :0.041 MAPE :0.946% R^2:0.988
[23]	2022	LSTM GRU	The electricity consumption of cold storage facility	MAE :1.075810 / MSE :1.968030 MAE :0.540190 / MSE :0.523995
[24]	2021	KNN ANN SVM	energy consumption of commercial building located in Klang Valley Malaysia	RMSE:5.0025748/ MAPE:3.02 RMSE:8.874015 / MAPE:5.02 RMSE:4.750689 / MAPE:2.76

Table 2.1: exemples of results of related works

Conclusion

In this chapter, we conduct a theoretical study of forecast models. Where we discussed the most important predictive methods and how they work. We have shown the great importance of these methods in knowing the future values. Now that we have defined the machine learning models, we will apply some models to the problem of electricity consumption after going through some steps that we will discuss in the next chapter.

Part02
Practical part

chapter 03

Implementation and results

Introduction

in this chapter, first we prepare the data before applying regression models to it, which is the electricity consumption in megawatts for El-Oued city and the municipality (El-Oued) over an 11-year period. The performance of automatic learning algorithms for prediction is then evaluated, represented by regression and classification models, and we compare these models to determine which one is the most effective at assisting in making the best prediction decision. A classification or regression algorithm's predictive ability is typically measured by its predictive accuracy or error rate, in test examples. Our display of the results will be in the form of tables and graphs.

1. Proposed approach for electricity consumption prediction

The fundamental stages of the machine learning process consist of data collection and preparation, model selection, model training using training data, performance evaluation of the model, and finally the visualization of the obtained results. Such steps can be summarized as follows:

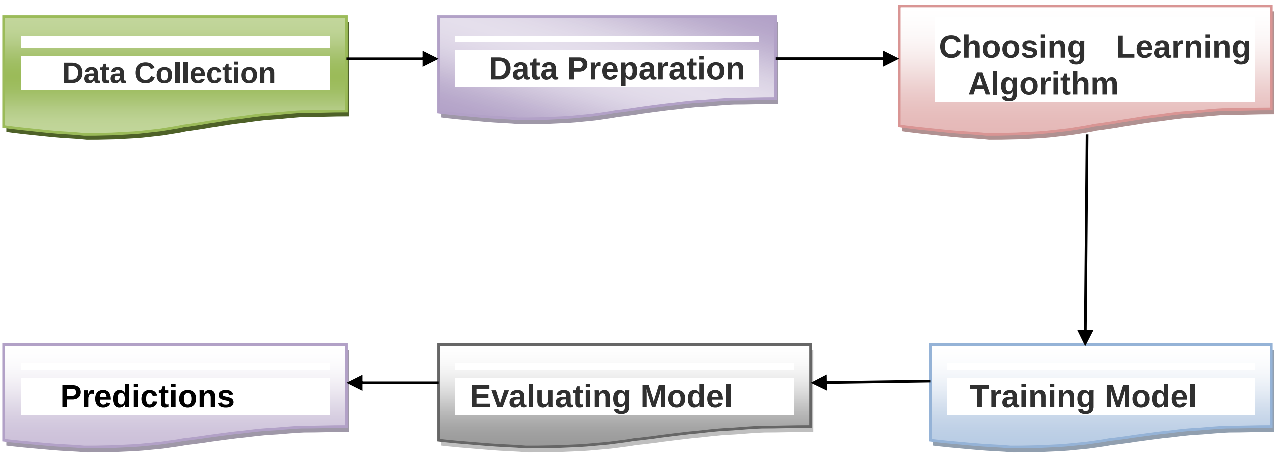

Figure 3.1: Machine Learning Workflow

1.1. Data Collection and Preprocessing

- **Overview of the dataset**

The data obtained from Sonelgaz company which were collected over the 10 years (from 2012 to 2022).the data comes in the form txt file.

Here are the most important points and characteristics of the collected data:

Each client's information is collected seasonally each year:

client	T1_2012	T2_2012	T3_2012	T4_2012	....	T1_2022	T2_2022	T3_2022	T4_2022

Table 3.1: Variables of the dataset

- **Data preprocessing and cleaning**

Data preprocessing is a data mining approach that involves putting intelligible format into raw data. Data preprocessing is a tried-and-true way for resolving issues when the real data is often insufficient, either due to missing attribute values or missing particular qualities of interest.

Step one: convert the data file

Since the gathered data was originally a text document, we transformed it into an Excel file in CSV format. This will enable us to process the data more easily when we come to the coding stage.

Step two: data cleaning

After conducting a data analysis, we discovered that there is missing data, so we must solve this problem and clean the data by removing incompleted data (deleting the rows that contain empty values). We keep only the essential and useful information from which the model learns. At the end of this step, we will have a total of 11,042 rows for municipality of El-Oued.

Step three: processing negative data

Since the obtained data includes data with negative values, we have changed the negative values to positive values in order to prevent any issues or flaws that these values might introduce.

Step four: Calculate the sum and Splitting the data

Calculate the sum of each column and then separate each season in the csv file,so we get 4 file ,and each file contains 11 columns and one row.

First file:

Season	t1_2012	t1_2013	t1_2014	t1_2015	t1_2016	t1_2017	t1_2018	t1_2019	t1_2020	t1_2021	t1_2022
Sum	12978630	14098068	14452512	15427490	14817060	14627272	15433716	15541748	15769255	13417632	16260966

Table 3.2: file of first season

Step five: Data re-organization

Finding the ideal combination of characteristics that we require for the prediction process is the goal of the characteristic selection procedure. We choose the entry and exit characteristics in an effort to increase the model's effectiveness.

Therefore, and because we will use different techniques of AI, we have prepared two datasets for:

- **First case :**

Such case has been for the machine learning models ads also for LSTM. Here, we have restructured the data in a CSV file containing 6 columns as follows:

Features					**Label**
Y	S	3Y	2Y	1Y	Target

Table 3.3: The features and label of dataset (case 1)

As one can see, the first column represents the year, the second, represents the season number, the third represents the consumption value of the third previous year, the fourth represents the consumption value of the second previous year, the fifth represents the consumption value for the previous year, and the last column represents the consumption value for the year.

	Year	T	3Y	2Y	1Y	target
0	2015	1.0	12978630.0	14098068.0	14452512.0	15427490.0
1	2016	1.0	14098068.0	14452512.0	15427490.0	14817060.0
2	2017	1.0	14452512.0	15427490.0	14817060.0	14627272.0
3	2018	1.0	15427490.0	14817060.0	14627272.0	15433716.0
4	2019	1.0	14817060.0	14627272.0	15433716.0	15541748.0

Figure 3.2: Data format after restructure

- **Second case:**
  In fact, the dataset of such characteristics had been exploited for SARIMA and Prophet. In this case, we have restructured the data in a CSV file containing 2 columns to be used in the time series models as follows:

Features	Label
Date	Value

Table 3.4: The features and label of dataset (case 2)

Step six: merge the data

After organizing the data of the four files, we will merge them all into one file for that. We will end up with one large file that contains all the variables and data.

	Year	T	3Y	2Y	1Y	target
0	2015	1.0	12978630.0	14098068.0	14452512.0	15427490.0
1	2015	2.0	12447482.0	11767980.0	12850339.0	15008678.0
2	2015	3.0	31089885.0	26245084.0	32542742.0	34858734.0
3	2015	4.0	24349374.0	22785614.0	25318993.0	23986772.0
4	2016	1.0	14098068.0	14452512.0	15427490.0	14817060.0
5	2016	2.0	11767980.0	12850339.0	15008678.0	14904585.0
6	2016	3.0	26245084.0	32542742.0	34858734.0	37261293.0
7	2016	4.0	22785614.0	25318993.0	23986772.0	22192289.0
8	2017	1.0	14452512.0	15427490.0	14817060.0	14627272.0
9	2017	2.0	12850339.0	15008678.0	14904585.0	14896826.0

Figure 3.3: Data format after organizing and merge.

Step seven: splitting the data

In this step, we will split the data into 2 part: one will be used for training and the other will be used for testing. The data was divided into 67% for training and 33% for testing.

1.2. Techniques used for the prediction

1.2.1. Simple Regression Techniques (Implementation)

- **Decision tree**

We start by importing the libraries, DecisionTreeRegressor from sklearn.tree for decision tree regression and train_test_split from sklearn.model_selection, then the target is set after reading the data, and then the train_test_split function is used to split the data into training and test sets. Then a DecisionTreeRegressor object is created and set the parameter (min_samples_leaf=10 and min_samples_split=2), then the fit approach is used to train the decision tree Regressor using the training data, and then the predict method is used to make predictions on the test set using the trained model.

- **Random forest**

A group of decision-making tree algorithms makes up an alphabetic forest. We create a random forest algorithm to solve a regression problem using the **RandomForestRegressor** class from the **Sklearn.ensemble** library. The parameter **n_estimators** is the most crucial component of the **RandomForestRegressor** class. This parameter specifies the total number of trees in the random forest; we start with **n_estimators = 100** and set the maximum tree depth to 10 and random_state=44 min_samples_leaf=5 min_samples_split=5. next, using a method that is fit to the training data, we introduce the random forest model .Then employ the trained model to perform predictions on the entire test set using the method predict.

1.2.2. AI Time Series Analysis

- **LSTM (Implementation)**

We start by importing the libraries, LSTM,Dense,Input from keras.layers and train_test_split from sklearn.model_selection, then the target is set after reading the data, and then the train_test_split function is used to split the data into training and test sets. Then, we initialize the model as Sequential(),than we put LSTM layer and 11 dense layers,then we have an output layer to prediction of a value.and we used tha activation function 'linear'.Second, we add the compiler function 'adam' as an improved scale to calculate the loss in order to train the model on the weights of the model and use the mean absolute error.

Finally, we will train our model at iteration of 100 times.

- **prophet**

Here, We start by importing the class Prophet from prophet and read the file of sum and convert the T column to datetime ,then Convert the name of the two columns to 'ds' and 'y'. We fit the model by instantiating a new Prophet object. Then we call its fit method and pass in the historical dataframe. And then we use the helper method Prophet.make_future_dataframe to predict the next five years (By default it will also include the dates from the history).

1.2.3. Statistical Time Series Analysis (Implementation)

- **SARIMA**

Import the module statsmodels.api and read the file of sum ,convert the T column to datetime.then,splitting the data and we create a SARIMAX object from sm.tsa and we put the parameter(endog=train, order(1,1,3), seasonal_order(1,1,3,4)).then we use get_forecast method to obtain forecasts of test values and we put parameters step= len(test).

1.3. Results and Analysis

1.3.1. Overview of the results

Part One: municipality of El-Oued

a) Decision tree

We applied DT model:

Training					Testing				
MAE	MSE	RMSE	NRMSE	R2_score	MAE	MSE	RMSE	NRMSE	R2_score
4424288.67	28653381054909.31	5352885.30	0.173016	0.686	4184312.50	20796868752581.75	4560358.40	0.151571	0.70

Table 3.5: The DT results for the testing and training phase**.**

As shown in Table(3.5) the DT model achieved (R^2 =0.70, RMSE=4560358.40, MAE=4184312.50). The plots of the results are provided down below (3.4) .

b) Random Forest

We applied RF model:

Training					testing				
MAE	MSE	RMSE	NRMSE	R2_score	MAE	MSE	RMSE	NRMSE	R2_score
2658140.67	11796339944206.7	3434580.02	0.106793	0.896	1817856.21	4558281096691.36	2135013.137	0.105529	0.91

Table 3.6: The RF results for the testing and training phase.

As shown in Table(3.6) the RF model achieved (R^2 =0.91, RMSE=2135013.137, MAE=1817856.21). The plots of the results are provided down below (3.5) .

c) LSTM

We applied LSTM model:

Training					Testing				
MAE	MSE	RMSE	NRMSE	R2_score	MAE	MSE	RMSE	NRMSE	R2_score
2327274.09	8621452412254.57	2936230.987	0.094905	0.905	855693.81	1457649719140.54	1207331.65	0.040128	0.97

Table 3.7: The LSTM results for the testing and training phase.

As shown in Table(3.7) the LSTM model achieved (R^2 =0.97, RMSE=1207331.65, MAE=855693.81). The plots of the results are provided down below (3.6).

d) **Prophet:**

We applied Prophet model:

MAE	MSE	RMSE	NRMSE	R2_score
1556387.56	5015041459776.558	2239428.8244497874	0.06623	0.943

Table 3.8: The prophet results for the testing and training phase.

As shown in Table(3.8) the Prophet model achieved (R^2 =0.94,RMSE=2239428.82, MAE=1556387.56). The plots of the results are provided down below (3.7).

e) **SARIMA:**

We applied SARIMA model:

training					testing				
MAE	MSE	RMSE	NRMSE	R2_score	MAE	MSE	RMSE	NRMSE	R2_score
10403907.1296	195931933040702.6	13997568.826074	0.124824	-1.8230999	2103577.338603	7826282325329.015	2797549.342787	0.08698582	0.931

Table 3.9: The SARIMA results for the testing and training phase.

As shown in Table(3.9)the SARIMA model achieved (R^2=0.93,RMSE=2797549.34, MAE=2103577.3). The plots of the results are provided down below (3.8).

Comparison between results:

The accuracy of the Training set and the Accuracy of the Testing set using the Municipality of El-Oued Dataset are compared in the table that follows (3.10).

	Training					testing				
	MAE	MSE	RMSE	NRMSE	R2_score	MAE	MSE	RMSE	NRMSE	R2_score
DT	4424288.67	28653381054909.31	5352885.30	0.173016	0.686	4184312.50	20796868752581.75	4560358.40	0.151571	0.70
RF	2658140.67	11796339944206.7	3434580.02	0.106793	0.896	1817856.21	4558281096691.36	2135013.137	0.105529	0.91
LSTM	2327274.09	8621452412254.57	2936230.987	0.094905	0.905	**855693.81**	1457649719140.54	**1207331.65**	0.040128	0.97
SARIMA	10403907.1296	195931933040702.6	13997568.826074	0.124824	-1.8230999	2103577.338603	7826282325329.015	2797549.342787	0.08698582	0.931

Table 3.10: the performance evaluation for different techniques.

through the results obtained and shown in the previous Table (3.10), we find that:

We note that the LSTM model was the most accurate among the models, it has the lowest error value, RMSE =1207331.65, then the RF model shows good prediction. with an error rate of RMSE =2135013.137. then the Prophet Achieved results acceptable,RMSE=2239428.82.The SARIMA model gave results were good in the testing phase with RMSE =2797549.342787 and DT model It gave the error value being slightly high with RMSE =4560358.40.

Part Two: El-Oued city

a) Decision tree

Training					Testing				
MAE	MSE	RMSE	NRMSE	R2_score	MAE	MSE	RMSE	NRMSE	R2_score
5774164.96	53008633962847.36	7280702.84	0.56563	0.68	5393820.21	37716404058280.35	6141368.25	0.536596	0.712

Table 3.11: The DT results for the testing and training phase**.**

As shown in Table(3.11) the DT model achieved (R^2 =0.71,RMSE=6141368.25, MAE=5393820.2). The plots of the results are provided down below (3.9).

b)Random Forest:

Training					Testing				
MAE	MSE	RMSE	NRMSE	R2_score	MAE	MSE	RMSE	NRMSE	R2_score
3582408.53	24173113344498.8	4916616.04	0.337766	0.885	2524096.50	9910072101854.17	3148026.69	0.341237	0.88

Table 3.12: The RF results for the testing and training phase**.**

As shown in Table(3.12) the RF model achieved (R^2 =0.88,RMSE=3148026.69, MAE=2524096.5). The plots of the results are provided down below (3.10).

d)LSTM:

Training					Testing				
MAE	MSE	RMSE	NRMSE	R2_score	MAE	MSE	RMSE	NRMSE	R2_score
2869992.95	14160354824144.6	3763024.69	0.292345	0.914	2246013.36	7143113500125.36	2672660.37	0.233521	0.945

Table 3.13: The LSTM results for the testing and training phase**.**

As shown in Table(3.13) the LSTM model achieved (R^2 =0.94, RMSE=2672660.37, MAE=2246013.3). The plots of the results are provided down below (3.11).

e)Prophet:

MAE	MSE	RMSE	NRMSE	R2_score
3606782.97850	25790444295803.36	5078429.3138531875	0.0695	0.92

Table 3.14: The Prophet results for the testing and training phase.

As shown in Table(3.14) the Prophet model achieved (R^2 =0.92, RMSE=5078429.31, MAE=3606782.9). The plots of the results are provided down below (3.12).

f) SARIMA:

Training					Testing				
MAE	MSE	RMSE	NRMSE	R2_score	MAE	MSE	RMSE	NRMSE	R2_score
17658324.93040979	780742063592461.9	27941761.99870835	0.0603	-2.354	5031787.960924175	42415544182473.64	6512721.718488641	0.10212	0.902

Table 3.15: The SARIMA results for the testing and training phase.

As shown in Table(3.15) the SARIMA model achieved (R^2 =0.90, RMSE=6512721.71, MAE=5031787.9). The plots of the results are provided down below (3.13).

Comparison between results:

	Training					Testing				
	MAE	MSE	RMSE	NRMSE	R2_score	MAE	MSE	RMSE	NRMSE	R2_score
DT	5774164.96	53008633962847.36	7280702.84	0.56563	0.68	5393820.21	37716404058280.35	6141368.25	0.536596	0.712
RF	3582408.53	24173113344498.85	4916616.04	0.337766	0.885	2524096.50	9910072101854.176	3148026.69	0.341237	0.88
LSTM	2869992.95	14160354824144.6	3763024.69	0.292345	0.914	**2246013.36**	7143113500125.36	**2672660.37**	0.233521	0.945
SARIMA	17658324.93040979	780742063592461.9	27941761.99870835	0.0603	-2.354	5031787.960924175	42415544182473.64	6512721.718488641	0.10212	0.902

Table 3.16:the performance evaluation for different techniques.

Through the results obtained, we find the following:

- The DT model gave a slightly higher error value with RMSE = 6141368.25.

- The RF model achieved similar results in both the learning and testing phases, the results were good. Where the error value as follows: RMSE=3148026.69.
- The LSTM model achieved the best results in the testing phase, and it has the lowest error value, RMSE=2672660.37.
- The Prophet model had a acceptable performance.the results were good,where error value RMSE=5078429.31.
- The SARIMA model had a Unacceptable performance, so the results were much unsatisfactory. With a error value RMSE = 6512721.71.

So in this part we can say that the LSTM model is the most satisfactory models in terms of performance.

1.3.2. Results analysis

Part one: municipality of El-Oued:

We present in this section the results of the prediction of the value of the electricity consumption for data of municipality of El-oued.

a) Decision tree

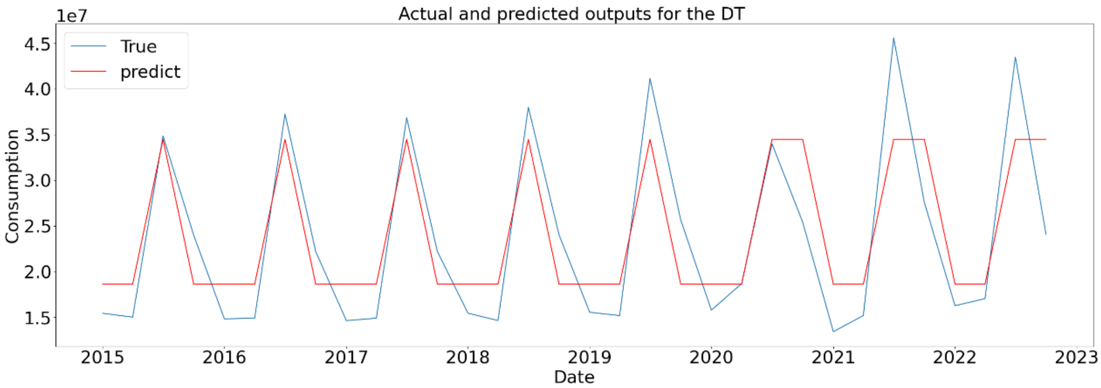

Figure 3.4: prediction by DT.

b) Random Forest:

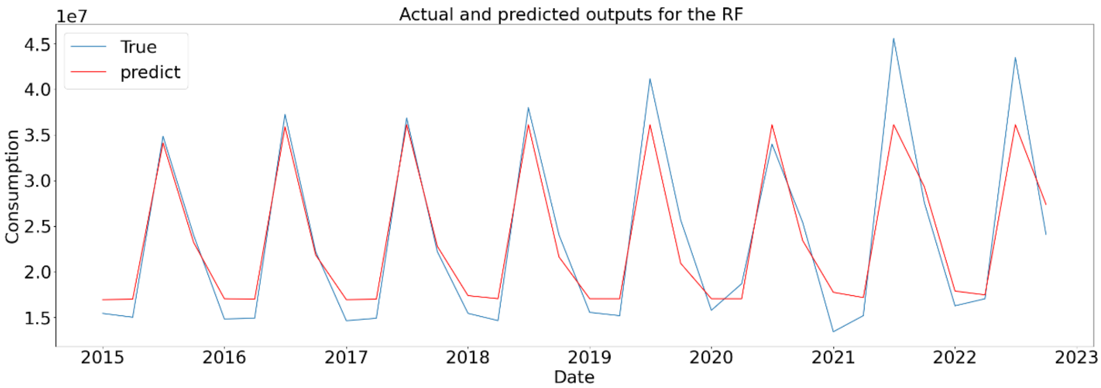

Figure 3.5: prediction by RF.

c) LSTM:

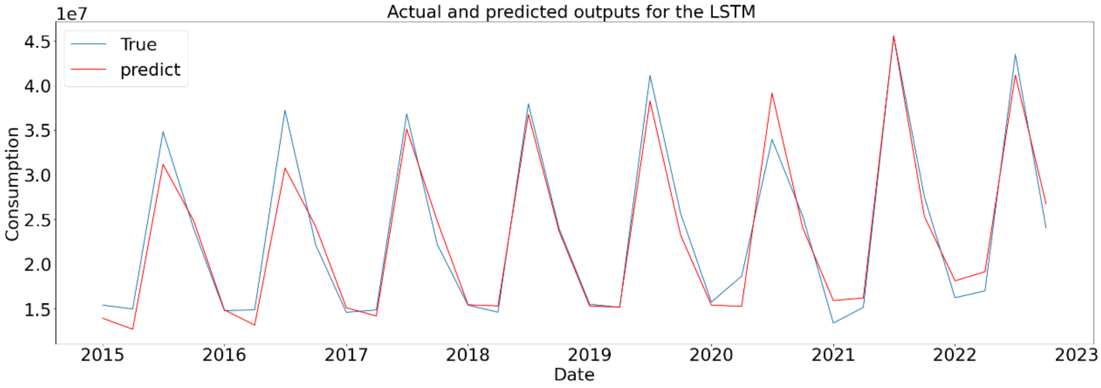

Figure 3.6: prediction by LSTM.

d) Prophet:

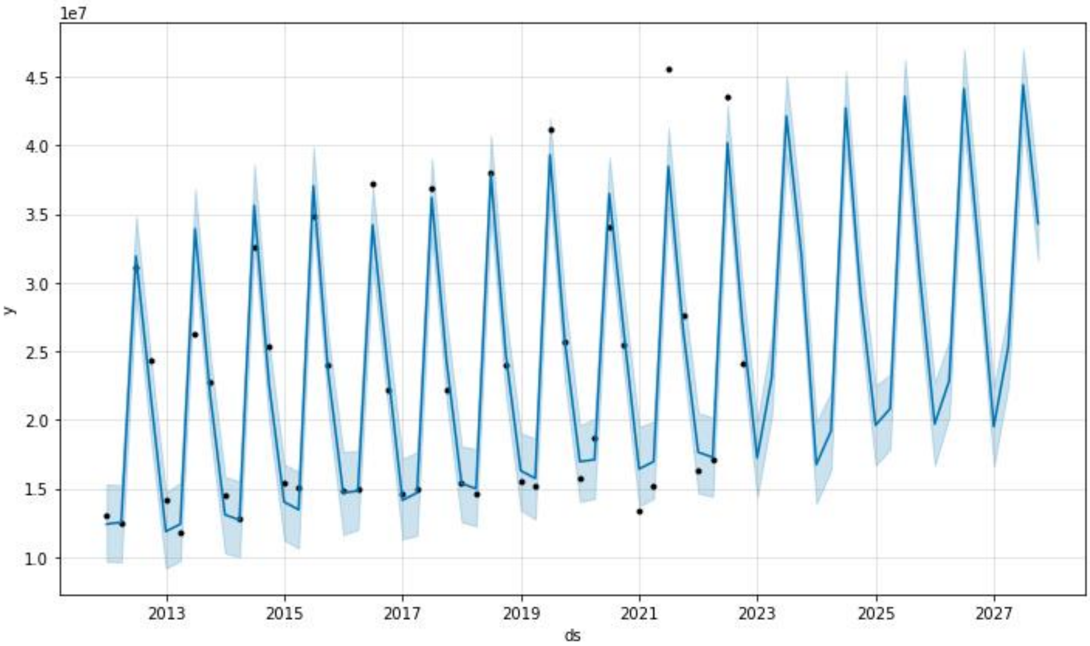

Figure 3.7: prediction by Prophet.

e) SARIMA:

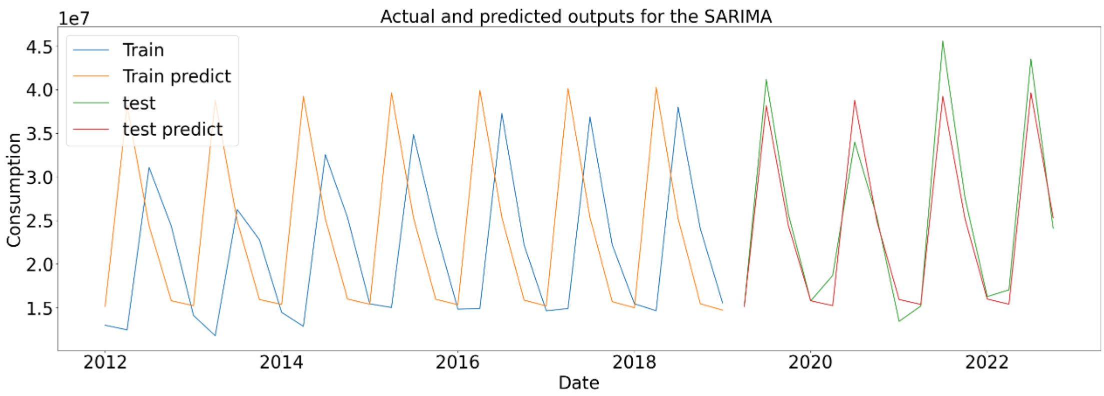

Figure 3.8: prediction by SARIMA**.**

Part two: El-Oued city:

We present here the results of the prediction of the value of the electricity consumption for data of El-oued city.

a) Decision tree

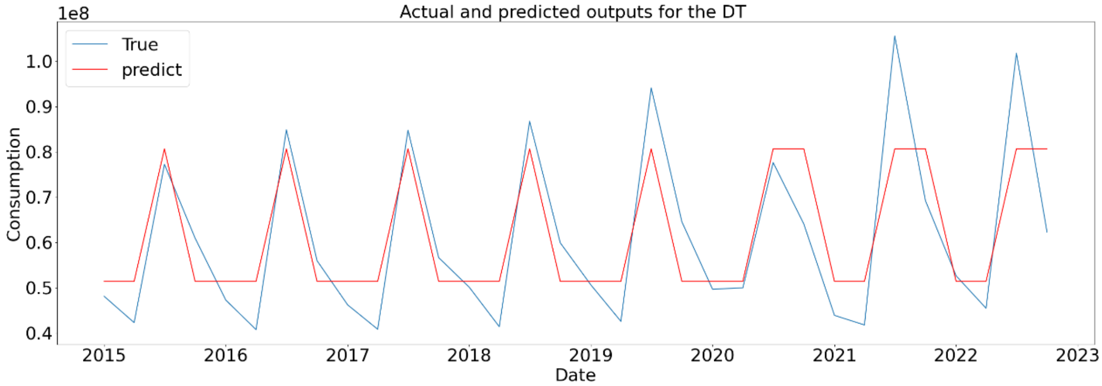

Figure 3.9: prediction by DT.

b) Random Forest:

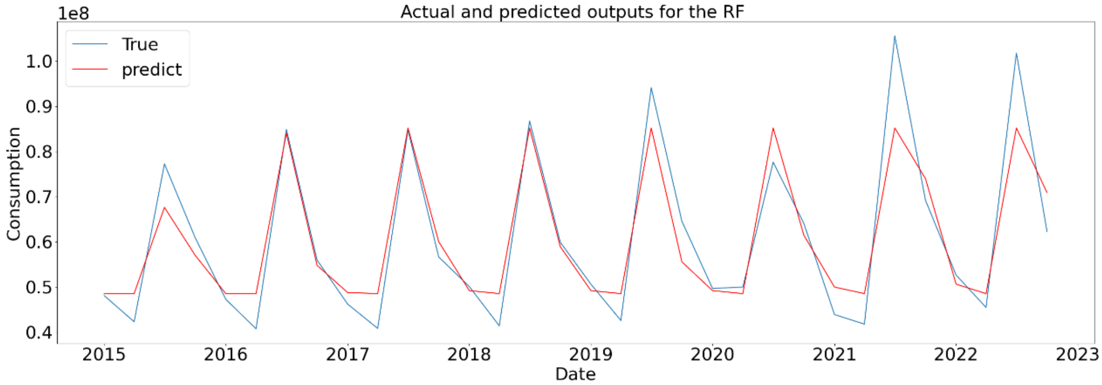

Figure 3.10: prediction by RF.

c) LSTM:

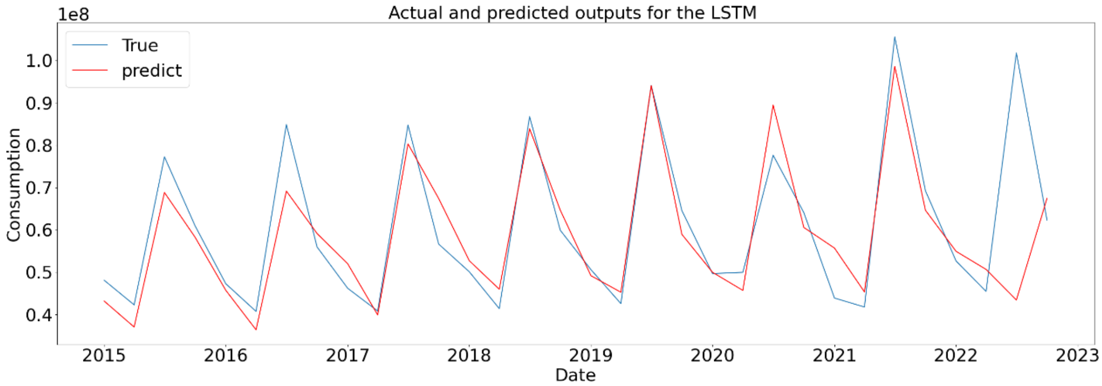

Figure 3.11: prediction by LSTM.

d) Prophet:

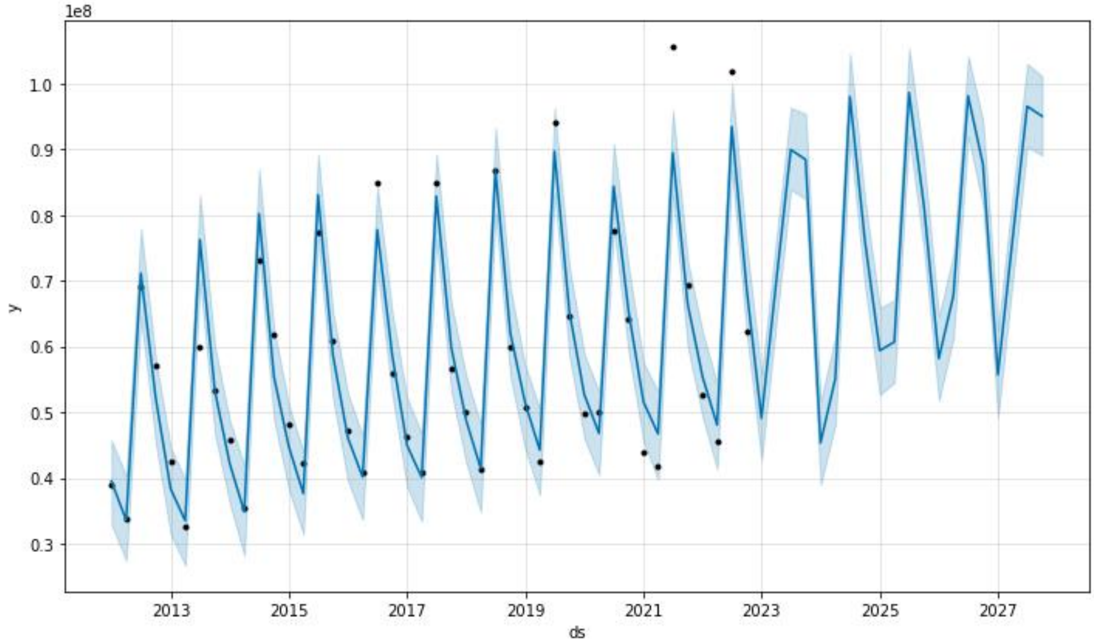

Figure 3.12: prediction by Prophet**.**

e) SARIMA:

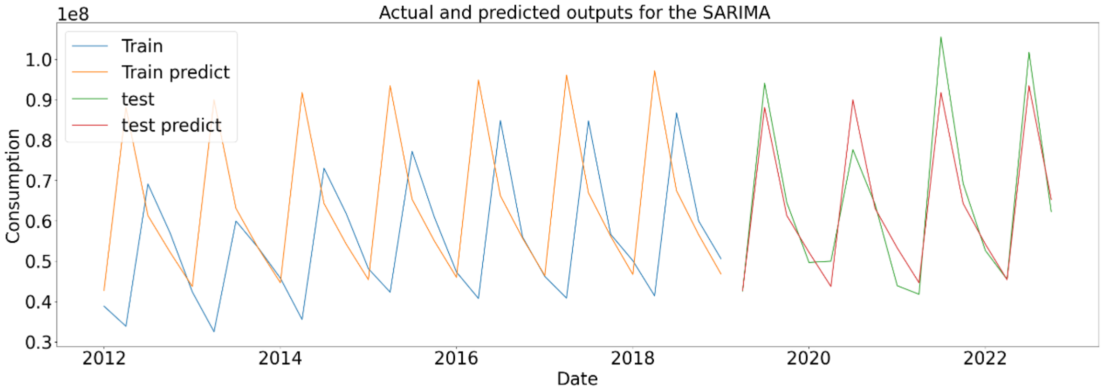

Figure 3.13: prediction by SARIMA**.**

1.4. Results and discussion

In our thesis, we used different algorithms for the forecast of electricity consumption in El-Oued. Concretely, we have analyzed the electricity consumption of the city and the municipality of El-Oued in both short and long term.

As a result, we concluded that the best algorithm was the LSTM and the accuracy was very encouraging for the short term case.

Finally, we concluded that the electricity consumption forecast is very useful for Sonelgaz and we intend to extend our work with other functionalities to better help the institutions producing and distributing electricity.

Conclusion

In this chapter, we collected and processed the data, then presented detailed results for each model and presented a comparative analysis of the results of the different models for predicting electricity consumption on two sets of data, which are the municipal data of El-oued, and El-oued.city Finally, some conclusions are drawn from our analysis of the results obtained.

General Conclusion

Due to the fact that electricity controls every aspect of life, as demand rises in tandem with technology and urbanization's rapid development, there will be a shortage of energy that cannot be produced in the required quantities. This will result in a crisis for both producers and consumers. Therefore, accurate demand forecasting for electrical energy is a key decision-making challenge since it enables better energy resource planning and management.

In this study, we discussed the potential for highly accurate predictions of electricity consumption by applying certain automatic learning models, embodied in regression models to predict consumption value, such as LSTM, random forest, decision tree, SARIMA, and SVR.

All prediction models have been applied to two different sets of data: data from the municipality of El-oued and data from El-oued city.

After comparing the results of regression models, we found that the model based on neural networks (LSTM) is significantly better than the other models for all data, but it is also possible to rely on model RF to forecast future electricity consumption values.

In terms of research views, we may say that LSTM, RF can all be used to predict value-based consumer behavior.

Over the course of working on this project, we learned a lot about artificial intelligence approaches, particularly automatic learning models, and how to apply these models to predicting electrical energy consumption.

B*ibliograph*y

[1] "Electrical Energy and Power," [Online]. Available: https://byjus.com/physics/electrical-energy-and-power/.

[2] T. E. Boukelia and . M. S. Mecibah, "Solid waste as renewable source of energy," *International Journal of Energy and Environmental Engineering*, 2012.

[3] Mansoor-ul-Hassan, "POWER GENERATION METHODS,TECHNIQUES AND ECONOMICAL STRATEGY," *International Technical Sciences Journal* , 2014.

[4] «Hydroelectric Energy,» [En ligne]. Available: https://education.nationalgeographic.org/resource/hydroelectric-energy/. [Accès le 3 2023].

[5] [En ligne]. Available: https://www.sonelgaz.dz/fr.

[6] . S. ALLAL et . H. ARIOUA, «Modèles de machine Learning pour la prédiction de la consommation en énergie électrique Cas d'étude : La société Sonelgaz,» M'sila, 2022.

[7] «Energy Production and Consumption,» Our World in Data, [En ligne]. Available: https://ourworldindata.org/energy-production-consumption. [Accès le 15 03 2023].

[8] «Forecasting Model,» highradius, [En ligne]. Available: https://www.highradius.com/resources/glossary/treasury/forecasting-model/. [Accès le 04 2023].

[9] B. Blouzi et N. M. H. Lalmi, «An Automatic Prediction of Solar Radiation for Renewable Energy using Machine Learning models,» El- OUED, 2022.

[10] «Machine Learning Models,» [En ligne]. Available: https://www.mathworks.com/discovery/machine-learning-models.html.

[11] A. Graves, «Generating Sequences With Recurrent Neural Networks,» 2014.

[12] F. Laghrissi, S. Douzi, K. Douzi et B. Hssina, «Intrusion detection systems using long short-term memory (LSTM),» *Big Data*, 2021.

[13] «forecasting and scale,» prophet, [En ligne]. Available: https://facebook.github.io/prophet/?fbclid=IwAR2aVY1pd4NHjlQzwkp5W8V11MRfNvj3RK1IRjnGKo7DrgWY0JQG5zZe9is. [Accès le 2023 03 20].

[14] A. Kumar, «Different types of Time-series Forecasting Models,» Data Analytics, 31 03 2023. [En ligne]. Available: https://vitalflux.com/different-types-of-time-series-forecasting-models/. [Accès le 05 2023].

[15] A. LaBarr, «What are Moving Average Models,» [En ligne]. Available: https://www.youtube.com/watch?v=zNLG8tsA_Go. [Accès le 05 2023].

[16] «Time Series Models,» Towards Data Science, 22 09 2020. [En ligne]. Available: https://towardsdatascience.com/time-series-models-d9266f8ac7b0. [Accès le 05 2023].

[17] A. HAYES, «Autoregressive Integrated Moving Average (ARIMA) Prediction Model,» investopedia, 18 12 2022. [En ligne]. Available: https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp. [Accès le 05 2023].

[18] J. Brownlee, «A Gentle Introduction to SARIMA for Time Series Forecasting in Python,» machinelearningmastery, 17 08 2018. [En ligne]. Available: https://machinelearningmastery.com/sarima-for-time-series-forecasting-in-python/. [Accès le 05 2023].

[19] P. Meel, «Predicting Flight Delays with Error Calculation using Machine Learned Classifiers,» *ResearchGate*, 2020.

[20] Razak Olu-Ajayi et Hafiz Alaka, «Building energy consumption prediction using deep learning».

[21] Yangyang Fu, Zhengwei Li, Hao Zhang et Peng Xu, «Using Support Vector Machine to Predict Next Day Electricity Load of Public Buildings with Sub-metering Devices,» *ELSEVIER*, 2015.

[22] C. Dianwei, «Research on electricity consumption forecasting model based on wavelet transform and multi-layer LSTM model,» *ELSEVIER*, 2022.

[23] Miona Andrejevi´c Stošovi, Novak Radivojevi´ et Malinka Ivanova, «Electricity Consumption Prediction in an Electronic System Using Artificial Neural Networks,» *electronics*, 2022.

[24] Mel Keytingan M. Shapi , Nor Azuana Ramli et Lilik J. Awalin, «Energy consumption prediction by using machine learning for smart building: Case study in Malaysia,» *ELSEVIER*, 2021.

[25] B. Mahesh, «Machine Learning Algorithms -A Review,» *ResearcherGate*, 2019.

[26] . L. Seunghye, N. Ngoc-Hien , K. Armagan et L. Jaehong , «Super learner

machine-learning algorithms for compressive strength prediction of high performance concrete,» *fib*, 2022.