**PEOPLE’S DEMOCRATIC REPUBLIC OF ALGERIA**

**MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH**

**FACULTY OF SCIENCE**

**DEPARTEMENT OF COMPUTER SCIENCE**

Submitted in partial fulfilment of the requirements for the Master degree in Computer science

Option: Advanced and applied information system

# DESIGN AND IMPLEMENTATION OF HOME ENERGY MANAGEMENT SYSTEM ADVISOR

**THESIS COMMITTEE:**

SUPERVISOR: Dr. Adel LAHSASNA

REVIEWERS: Dr. I. BOULENMOUR and Dr. M. CHERIBAT

**AUTHOR:**

Abderraouf BOUAKKAZ

Faculty of Science

University 20 Aout 1955-Skikda

SKIKDA CITY, JUNE 2023

# Statement of Original Authorship

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Signature: ______________________________

Date: ____June 26, 2023__________

# Acknowledgements

First of all I would thank almighty GOD for giving me the courage to carry out this project.

I would like to express my gratitude to my thesis supervisor professor Lahsasna Adel for his advice, encouragements and aids.

I send my thanks to

- The President and members of the Jury who have agreed to evaluate this thesis.
- All my teachers who taught me during the cycle of studies.
- All administration staff at the University of Skikda.
- All those who contributed from near or far to the realization of this project, for their advice, their precious help and their moral support.

# Dedication

I dedicate this work to:

- The dearest people in my life: my mother "Zohra" and my father "Hocine", for their love, care and encouragement.
- My dear wife.
- My dear brothers.
- My dear sisters.
- The whole family from near and far.
- My colleagues and my friends.

# Abstract

In the face of global energy challenges, the ability to adapt and innovate is crucial for humanity's survival and progress. Effective energy management plays a pivotal role in optimizing energy consumption patterns and enhancing efficiency. This involves implementing strategies such as energy conservation practices, adopting energy-efficient technologies, and leveraging smart grid systems for improved monitoring and control.

To address these goals, this study proposes a decision support system with an integrated information system for energy management in residential homes. By utilizing various machine learning algorithms, the system aims to predict energy consumption and cost, ultimately optimizing energy performance. The performance and effectiveness of these algorithms are thoroughly evaluated, highlighting the potential benefits of machine learning in achieving energy efficiency and cost reduction within homes.

**Keywords:** *Information system, machine learning, renewable energy, housing, energy management systems, optimization.*

# Résumé


Face aux défis énergétiques mondiaux, la capacité d'adaptation et d'innovation est cruciale pour la survie et le progrès de l'humanité. Une gestion efficace de l'énergie joue un rôle central dans l'optimisation des schémas de consommation d'énergie et l'amélioration de l'efficacité. Cela implique la mise en œuvre de stratégies telles que des pratiques d'économie d'énergie, l'adoption de technologies économes en énergie et l'utilisation de systèmes de réseaux intelligents pour une surveillance et un contrôle améliorés.

Pour répondre à ces objectifs, cette étude propose un système d'aide à la décision avec un système d'information intégré pour la gestion de l'énergie dans les résidences. En utilisant divers algorithmes d'apprentissage automatique, le système vise à prédire la consommation et les coûts d'énergie, optimisant ainsi les performances énergétiques. Les performances et l'efficacité de ces algorithmes sont soigneusement évaluées, mettant en évidence les avantages potentiels de l'apprentissage automatique pour atteindre l'efficacité énergétique et la réduction des coûts dans les maisons.



**Mots-clés :** Système d'information, machine learning, énergies renouvelables, habitat, systèmes de gestion de l'énergie, optimisation.

# الملخص

في مواجهة تحديات الطاقة العالمية ، تعد القدرة على التكيف والابتكار أمرًا بالغ الأهمية لبقاء البشرية وتقدمها. تلعب الإدارة الفعالة للطاقة دورًا محوريًا في تحسين أنماط استهلاك الطاقة وتعزيز الكفاءة. يتضمن ذلك تنفيذ استراتيجيات مثل ممارسات الحفاظ على الطاقة ، واعتماد تقنيات موفرة للطاقة ، والاستفادة من أنظمة الشبكة الذكية لتحسين المراقبة والتحكم

لتحقيق هذه الأهداف ، تقترح هذه الدراسة نظام دعم القرار مع نظام معلومات متكامل لإدارة الطاقة في المنازل السكنية. من خلال استخدام خوارزميات التعلم الآلي المختلفة ، يهدف النظام إلى التنبؤ باستهلاك الطاقة والتكلفة ، وفي النهاية تحسين أداء الطاقة. يتم تقييم أداء وفعالية هذه الخوارزميات بدقة ، مما يسلط الضوء على الفوائد المحتملة للتعلم الآلي في تحقيق كفاءة الطاقة وخفض التكلفة داخل المنازل

**الكلمات المفتاحية: نظام المعلومات ، التعلم الآلي ، الطاقة المتجددة ، الإسكان ، أنظمة إدارة الطاقة ، التحسين**

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

ACU: Anticipative control unit

ADS: Acquisition data system

AI: Artificiel intelligence

AIS:  The automated information system

ANN: Artificial neural networks;

EMS: Energy management system;

IS: Information system;

LR: Linear regression;

ML: Machine learning

MSE: Mean squared error

MPC: Model Predictive Control;

RCU:  Reactive control unit

RF: Random forest;

SVM: Support vector machine.

# Chapter 1: Introduction

The current chapter serves as an introduction of our dissertation, aiming to highlight on the core issue of our study. It begins by providing a background in section 1.1, followed by an exploration of the research context and its objectives in section 1.2 and its purposes in section 1.3. Section 1.4 elaborates on the significance and scope of this study. Lastly, section 1.5 offers a comprehensive outline of the subsequent chapters that will be covered in this dissertation.

## 1.1.Background

In recent years, there has been a notable surge in energy consumption at the residential level due to the rapid expansion of highly populated new cities in Algeria. The conventional energy sources used in these areas, primarily fossil fuels, contribute significantly to greenhouse gas emissions. Consequently, the consequences of this trend have become increasingly apparent. To mitigate these effects, it is now imperative to prioritize the reduction of energy consumption in residential buildings and transition towards new and renewable energy sources, thereby replacing conventional and polluting energy systems. In light of this necessity, it becomes crucial to develop an optimal strategy for efficient energy management that incorporates novel methods, approaches, and innovative design and renovation practices within the housing sector.

## 1.2. Context

Humanity is currently navigating a challenging period in terms of energy, and the ability to adapt and innovate will be crucial and decisive for the survival of the human race within a highly developed environment. While making accurate forecasts about the future of energy production and consumption is exceptionally difficult due to the complex energy landscape at both national and global levels, it is evident that energy demand continues to rise while production undergoes a complex and uncertain transitional phase. In this context, it is important to address the following questions:

1. How can the residential sector contribute to energy conservation by optimizing final energy usage, particularly in terms of electrical energy?

2. How can we ensure local energy supply by integrating alternative sources of energy?

By exploring these inquiries, we aim to uncover strategies and solutions that enable energy-saving practices and promote diverse and reliable energy sources at the residential level.

### 1.2.1. Energy production and consumption issues

Alongside environmental concerns like global warming, effectively managing and controlling energy demand presents a significant challenge. This challenge is particularly pertinent to the production and consumption of electrical energy. To gain a deeper understanding of these issues, it is essential to delve into the fundamental principles underlying energy consumption and production problems.

Energy consumption challenges arise from various factors, including population growth, urbanization, and industrialization. As societies expand, the demand for energy, especially electrical energy, escalates. This surge in demand places strains on existing energy infrastructure and necessitates the development of new energy sources and technologies.

On the production side, meeting the growing energy demand requires a careful balance between conventional and alternative energy sources. Conventional energy sources, such as fossil fuels, face concerns regarding their limited availability, environmental impact, and carbon emissions. Transitioning to alternative and renewable energy sources, such as solar, wind, hydroelectric, and geothermal power, is crucial to achieving sustainable energy production.

Effective energy management involves strategies to optimize energy consumption patterns and enhance energy efficiency. This includes promoting energy conservation practices, implementing energy-efficient technologies, and adopting smart grid systems for better monitoring and control of energy usage. Additionally, demand response programs and incentives can incentivize consumers to adjust their energy consumption patterns during peak periods, thus reducing strain on the energy grid.

To address these challenges, energy companies, and individuals must collaborate to foster a comprehensive energy management approach. This approach involves investing in research and development to advance clean energy technologies, establishing supportive policies and regulations, promoting energy education and awareness, and encouraging sustainable practices across sectors.

By understanding the underlying principles and challenges associated with energy consumption and production, stakeholders can work towards implementing effective strategies that ensure a reliable, affordable, and sustainable energy future while mitigating the environmental impact of energy.

### 1.2.2. Peak power demand problems

Maintaining a stable balance between electricity supply and demand is a primary challenge in operating the electricity grid. However, one significant issue that electricity operators face is the occurrence of peak power demand. These peaks in demand often arise from the use of energy-intensive devices and equipment, primarily in the industrial and residential sectors.

Several factors influence electricity consumption, with economic activity and meteorological conditions being the primary parameters. Economic activity directly affects electricity demand, as increased industrial production and commercial activities lead to higher energy consumption. Similarly, during periods of extreme temperatures, such as heatwaves or cold spells, the use of heating or cooling systems intensifies, driving up electricity demand in the residential and commercial sectors.

The interplay between economic activity and meteorology creates variability in electricity consumption patterns. Understanding and predicting these variations are crucial for effective grid management. Electricity operators employ forecasting models that consider historical data, economic indicators, weather forecasts, and seasonal patterns to anticipate and plan for peak demand periods.

Efficient energy management strategies can help mitigate peak power demand challenges. Demand response programs, for example, encourage consumers to adjust their electricity usage during peak periods through incentives or time-based pricing. Smart grid technologies and advanced metering systems enable real-time monitoring and control of electricity consumption, allowing for better demand management and load balancing.

### 1.2.3. Position of information system in energy management

Information systems serve as a critical infrastructure for energy management, providing essential tools and capabilities to monitor, analyze, optimize, and report on energy-related activities. They facilitate data-driven decision-making, improve energy efficiency, and support sustainability goals in the context of energy management. Here are some key positions of information systems in energy management:

- Data Acquisition and Monitoring: Information systems enable the collection and aggregation of real-time data on energy consumption, production, and other relevant variables. This data can be sourced from smart meters, sensors, IoT devices, and energy management systems. By acquiring and monitoring this data, information systems provide a comprehensive view of energy usage patterns and help identify areas for improvement.
- Energy Analytics and Optimization: Information systems leverage advanced analytics techniques to analyze energy data and extract valuable insights. These insights can be used to identify energy inefficiencies, detect anomalies, and optimize energy usage across various systems and processes. By applying data-driven analytics, information systems facilitate informed decision-making for energy conservation and efficiency measures.
- Energy Planning and Forecasting: Information systems assist in energy planning by integrating historical and real-time energy data with predictive modeling techniques. They enable the development of accurate energy demand forecasts, which are crucial for efficient energy procurement and resource allocation. With reliable forecasts, organizations can optimize their energy generation, distribution, and consumption strategies.
- Energy Reporting and Compliance: Information systems facilitate energy reporting by automating the collection, analysis, and reporting of energy-related data required for regulatory compliance. These systems ensure accurate and timely reporting of energy consumption, emissions, and other sustainability metrics, helping organizations meet regulatory requirements and demonstrate their commitment to environmental goals.

- Energy Decision Support: Information systems provide decision support tools and dashboards that enable stakeholders to visualize and interpret energy-related data effectively. They assist in identifying energy-saving opportunities, evaluating the impact of energy management initiatives, and assessing the return on investment for energy efficiency projects. These decision support capabilities empower organizations to make data-driven choices for optimizing energy management strategies.

## 1.3.Purposes

The main objective of this study is to propose an approach that utilizes information systems to achieve optimal performance by balancing power production and consumption in real-time. This approach aims to strike a balance between meeting the energy needs of consumers, ensuring occupant comfort, and respecting the environment, all while minimizing costs. Specifically, the study focuses on minimizing energy costs and enhancing the system's reliability by reducing peak power consumption.

To achieve these goals, the proposed approach will leverage information systems to gather and analyze data related to power production, consumption patterns, and environmental factors. By integrating advanced algorithms and optimization techniques such as machine learning algorithms, the approach will enable intelligent decision-making to optimize the operation of the power system.

The information system will provide a centralized platform for monitoring and controlling power production and consumption, allowing for effective load management. It will consider factors such as energy demand, available renewable energy sources, electricity pricing, and grid conditions to make informed decisions in real-time.

This approach will prioritize occupant comfort by considering factors such as temperature, lighting, and other environmental parameters. It will aim to achieve energy efficiency without compromising the comfort and well-being of the occupants.

By dynamically balancing power production and consumption, the approach will help reduce peak power demand, thereby improving the reliability and stability of the power system. This can lead to reduced energy costs and a more sustainable and environmentally friendly energy grid.

### 1.4.Significance and scope

Automated energy control systems, known as energy management systems (EMS) or building automation systems (BAS), have become prevalent in both residential and non-residential buildings worldwide. These systems utilize computers as central processors to automate and optimize energy usage. The concept of EMS was first introduced in 1979 with the paper "Solar energy management system" by Men [1]. The introduction of personal computers in 1980 further improved the performance of EMS [2].

Several studies have been conducted to explore different approaches and strategies for energy management in various settings. In 1986, an optimization algorithm was proposed to manage energy by reducing electrical costs through demand reduction and usage time management [3]. Strategies for controlling and managing electronic devices were developed using home automation communication network systems [4]. In 1997, the concept of energy management systems in buildings was presented, comprising equipment equipped with microcontrollers for communication, a centralized control system, and a human-machine interface (HMI) for control optimization and energy consumption monitoring [5]. The authors [6] proposed an energy management system for buildings that utilized a multi-agent integrated control system to regulate subsystems such as electricity, air conditioning, and security. In [7], the authors focused on studying a control system based on PID neural networks for energy management in buildings. This approach used the Proportional, Integral, and Derivative control mechanism to optimize energy consumption. Rim's work [8] aimed to develop an energy management system (EMS) specifically for the residential sector. Their proposed system employed a multi-layer architecture known as Smart Grid Equipment (SGE), comprising anticipation, reactive, and local layers.

H. Zhang et al. [9] proposed a logic-based energy management strategy for commercial buildings, aiming to reduce electricity bills and $CO_2$ emissions by incorporating renewable energy sources. The authors in [10] implemented an energy management strategy using reinforcement learning algorithms in a microgrid system.

Their study focused on optimizing battery scheduling to minimize external electricity purchases and increase the utilization of local renewable energy sources. A novel architecture for energy management in microgrids based on reinforcement learning is proposed [11], this approach aimed to optimize system flexibility by considering energy demand, external grid conditions, and wind resources. In [12] an AI-based arbitrage strategy to maximize operating profit in electricity markets involving grid operators, energy storage systems, and customers is proposed. The authors [13] proposed an advanced satisfaction-based Home Energy Management (HEM) system using Deep Reinforcement Learning (DQN). This system programmed the operation of controllable devices and shiftable loads to optimize energy consumption. A supervised learning strategy for a Home Energy Management System (HEMS) with a demand response program is proposed [14]. This approach aimed to minimize daily energy costs by learning resident behavior and discretely controlling Energy Storage Systems (ESS) and Renewable Energy Sources (RES). Authors in [15] proposed an energy management strategy for residential microgrid systems using reinforcement learning based on Model Predictive Control (MPC). The objective was to find an optimal energy exchange policy considering fluctuating spot market prices, uncertain user demand and renewable energy generation, and collective peak power penalties.

These studies showcase the diverse range of approaches, including neural networks, reinforcement learning, and control systems, being employed to address energy management challenges in different contexts. The aim is to optimize energy consumption, reduce costs, and enhance the overall efficiency of energy systems.

### 1.5.Reading plan outline

The thesis is structured in five chapters as follows:

- In the first chapter as an introduction of the study, a comprehensive analysis of the energy sector context is presented. This analysis highlights the importance of energy management, particularly in the residential sector. The challenges and complexities associated with energy consumption and production in residential buildings are addressed, emphasizing the need for efficient energy management strategies. This chapter also introduces the purpose of the study, which focuses on proposing an approach for optimizing energy performance in the residential sector.

- The second chapter of the study provides a comprehensive overview of the information system (IS) in relation to energy management. This chapter focuses on explaining the fundamental concepts associated with IS and presents the general architecture of an IS designed for energy management purposes. The chapter begins by introducing the basic concepts related to IS, such as data, information, and knowledge. It elucidates the role of IS in collecting, storing, processing, and disseminating data to support decision-making processes in energy management. Next, the general architecture of the IS is described. This includes the components and subsystems that comprise the IS, such as data collection systems, data storage and management systems, data processing and analysis tools, and user interfaces. The chapter explores how these components work together to facilitate the flow of information and support various functions within the energy management domain. The main functions of the IS are then presented. These functions encompass data acquisition and monitoring, data processing and analysis, decision support, and reporting. The chapter highlights how these functions enable energy managers to access real-time data, analyze energy consumption patterns, make informed decisions, and generate reports to track and evaluate energy management performance.
- The third chapter is a general introduction to machine learning (ML). This chapter provides a comprehensive overview of the basic concepts, approaches, algorithms, and applications of machine learning. The chapter begins by introducing the fundamental concepts of machine learning, including the definition of ML, its relation to artificial intelligence (AI), and the role of data in the learning process. It covers key terminology such as training data, features, labels, and predictions. Next, the various approaches to machine learning are discussed. This includes supervised learning, unsupervised learning, reinforcement learning and deep learning. The chapter explores how each approach differs in terms of the available data, learning objectives, and the types of problems they are suited for. The chapter then delves into the different algorithms used in machine learning. It provides an overview of popular algorithms such as decision trees, support vector machines, neural networks, and clustering algorithms. Each algorithm is explained in terms of its underlying principles and how it can be applied to solve different types of problems. Additionally, the chapter highlights the wide range of applications

where machine learning is employed. The chapter illustrates how machine learning techniques are utilized to extract insights from data and make predictions or decisions in real-world scenarios.

- The fourth chapter of the study focuses on the implementation of the proposed decision support approach for energy management in homes. This chapter details the development and deployment of an information system specifically designed for this purpose. It outlines how the system integrates machine learning algorithms to predict energy consumption and energy costs. Next, the chapter explores the selection and implementation of machine learning algorithms for energy consumption and cost prediction. It discusses various algorithms, such as linear regression models, random forests, support vector machines and artificial neural networks that can effectively model and predict energy usage patterns. The advantages and limitations of each algorithm are considered in the context of energy management in homes. The chapter elaborates on the steps taken to train and fine-tune the machine learning models using historical data (solar irradiation, temperature and energy consumption). Furthermore, the chapter discusses the integration of the developed models into the information system. It explains how the prediction models are utilized within the system to provide decision support for optimizing energy consumption and cost. The system may offer recommendations on energy-efficient practices, optimal appliance usage, or energy-saving strategies based on the predictions generated by the machine learning algorithms.
- The final chapter of our study focuses on discussing the obtained results from the implementation of the machine learning algorithms in our energy management system. In this chapter, we evaluate the performance and effectiveness of the algorithms by assessing the accuracy of the energy consumption and cost predictions. We begin by presenting the evaluation metrics used to measure the accuracy of the predictions, such as mean squared error (MSE). This metric provide quantitative measures of the prediction quality and allow us to compare the performance of different algorithms. Next, we analyze the accuracy of the energy consumption and cost predictions generated by the implemented machine learning algorithms. We compare the predicted values with the actual values to determine the level of accuracy achieved. Furthermore, we discuss the impact of

the implemented system on energy efficiency and cost reduction in homes. We analyze how the use of accurate predictions and the integration of the decision support system into daily energy management practices have influenced energy consumption patterns and cost savings. We explore any observed improvements in energy efficiency, such as optimized appliance usage, reduced energy waste, and increased reliance on renewable energy sources.

# Chapter 2: Information system: overview

Faced with technological developments worldwide and the need to keep this rhythm of development in various fields and sectors, particularly in the energy sector. The interest in information technology has become very crucial and essential.

The information technology (IT) is characterized by great scalability of needs and technologies to respond better and faster to the needs of end users by providing them with the best solutions to facilitate their daily lives and give them more comfortable with reasonable control of costs and time.

This chapter is considered as a general presentation of information systems (IS) which is devoted to the description of the basic concepts related to the IS and the presentation of the general architecture of the IS and their main functions.

## 2.1.Generalities and definitions

In order to better understand the information system, this section is dedicated for giving some basic definitions and basic knowledge concerning the information system.

a. **Data:** A data is what is known and which serves as a starting point for reasoning aimed at determining a solution to a problem in relation to this datum. It can be an elementary description of a reality, the result of a comparison between two events of the same order (measurement) or in other words an observation or a measurement [16].In computer science, data is the representation of information in a program: either in the text of the program (source code), or in memory during execution. Data can be stored and classified in different forms: textual (string), numerical, images, sounds, etc. The variable data that makes a program flexible is usually read from a device user input (keyboard, mouse, etc.), a file, or from a network [17].
b. **Information:** Information is a concept of the discipline of information and communication sciences. It comes from the Latin verb "*informare*", which means "to give form to" or "to form an idea".

Information designates both the message to be communicated and the symbols used to write it. It uses a code of meaningful signs such as an alphabet of letters, a base of numbers, ideograms or pictograms [18].

c. **Process:** It is a word that comes from the Latin “pro” which means “forward” and from “cessus, cedere” which means “to go or walk”, which therefore means to go forward, to advance. This word is also at the origin of the word procedure which designates rather the method of organization, the strategy of change [19]. According to ISO 9000, a process is defined as “a set of interrelated or interactive activities which transforms inputs into outputs”.

A process (in computing): is a program executed or being executed by a computer. More precisely, it can be defined as [20]:

- ✓ a set of instructions to be executed, which may be in ROM, but most often loaded from mass memory to RAM;
- ✓ RAM address space, data, etc.; resources that allow data I/O, such as network ports.

### 2.2.System concept

Any human organization (a company, the State, etc.) can be perceived as a system. There are several definitions of “system”; and among them, we have chosen two that bring out the essential qualities of this concept [21]:

According to Joël De Rosnay, a system can be defined as a “set of elements in dynamic interaction, organized according to a goal”.

F. de SAUSSUR defined a system as follows: “Organized totality, made up of interdependent elements that can be only defined in relation to each other according to their place in this totality”.

A system has the following properties [21]:

- **the organization:** defined by the arrangement of the relationships between the elements that make up the system;
- **the totality**: is to say that a system is more than the sum of its elements, that it has properties that its components do not have;
- **the interaction between its elements:** which goes beyond the relations of the

cause-effect type;

- **the complexity:** which it is necessary to preserve even if we are unable to grasp its full richness.

### 2.2.1. Information system

The concept of information system (IS) was created in the early 1970s, to distinguish it from the computer system.

Reix R defined an information system (IS) as an organized set of resources (hardware, software, personnel, data, procedures) making it possible to acquire, process, store, communicate information (in the form of data, texts, images, sounds, etc.) in organizations [22],

The concept of information system (IS) covers two notions: on the one hand the reality of the system that transforms, undertakes, communicates and records information and on the other hand the computer system or the digital information system, which is an artificial object designed by man takes into account the potential of information and communication technologies (ICT) to help them acquire, process, store, transmit and retrieve the information that allows them to carry out their activities within the system [23].

### 2.2.2. Computer system

The computer system is a subset of the information system (IS), which ensures the automatable part of the functions of the IS. Originally very focused on operating system assistance, the digital information system (DIS), from 1980, extended to decision support systems, which constitute a specific sector of activity, economic intelligence, and dedicated tools, decision support systems.

In order for a digital information system to function effectively, it must provide relevant information to each decision-making center, regardless of its hierarchical level, so that it may control, decide, and act as necessary [24].

## 2.3.The information system structure

Computerized systems, i.e., the implementation of an information system (IS) of the company, can only be effective if they are directed from the start to the end.

The interaction between the system and its environment is possible thanks to information flows. These flows also circulate inside the system, which allows it to analyze its own functioning.

The elements of the system are themselves systems (or subsystems) as shown in Figure I.1, the automated system consists of three subsystems: the control system or decision system, information system and operating system [25] [26].

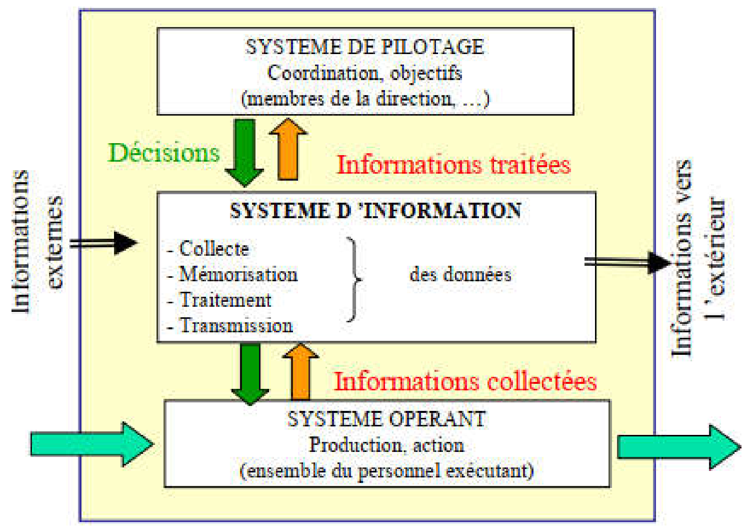


Fig.2. 1 systematic representation of an information system.

### 2.3.1. The control system

The control system also called a decision system or management system, decides on the actions to be carried out on the operating system according to the targeted objectives. It analyzes the both internal (internal functioning of the system) and external (environment) of the system. He ensures the control of tasks and ensures the regulation of the system. It is connected to the other subsystems by internal information flows (the information system and the operating system).

Information circulated in the system is exploited by the control system, which is used by the decision system to carry out the following activities:

– Think: adaptation to the environment, design.

- Decide: forecast, allocation, planning…
- Control.

### 2.3.2. Information system

The information system ensures the exchange of information between those who “think and organize” (control system) and those who “perform” (operating system). It has a central role since it supplies the system (company) with information. For this, it memorizes the information, processing and communications to the two other subsystems to which it is connected. All system information, of external or internal origin, therefore passes through the Information System.

Information is then sent to the operating system, which is responsible for carrying out the tasks entrusted to it. It in turn generates information for the decision-making system, which can thus control deviations and act accordingly.

The information system performs the following activities:

- Information gathering (data collecting).
- Memorization of information ;
- Information processing ;
- And dissemination of information (broadcasting).

### 2.3.3. Operating system

It is also called a production system and carries out the physical production of goods and services. Its activity is controlled by the decision system. It is linked to the environment by external physical flows and to the other subsystems by internal information flows.

It is responsible for performing all execution operations. It includes all functions related to the company's own activities such as customer invoicing, employee payroll, and stock management.

## 2.4. Information system functions

An information system with four main roles shown in the figure. 2.2: collect, store, process and disseminate information [25] [26]:

### 2.4.1. Data collection

This function actually corresponds to different types of tasks: first, it involves collecting information (so-called listening task).

Secondly, it is necessary to retain, among the information collected, those that are relevant with regard to the company's activities (analysis task). Finally, as a last step, it is necessary to enter the information retained in the Information System (input task).

The information comes from different sources includes:

a. **External sources:** The information comes from the system environment. These are generally flows from the system partners (customers, suppliers, administrations...). More and more, the company must listen to its environment to anticipate changes and adapt its operation. The development of communication means (internet in particular) makes it easier to find information but its exploitation remains delicate (quality and reliability of information).
b. **Internal sources:** the information system must be fed by the flows generated by the different actors of the system. These flows result from the activity of the system: supplies, production, employee management, accounting, sales...

### 2.4.2. Memorization of information

It implements technical and organizational means (archiving methods for example) for storing information in computers in a durable and stable manner (in the form of files or organized as databases for easier access).

### 2.4.3. Data processing

This means that it must be able to perform a certain number of treatments on the stored information (data). Here again, the treatment can be manual or automatic (carried out by a computer). The main types of processing consist of searching and extracting information, consulting, comparing the information with each other, modifying, organizing, updating, deleting information and producing (based on calculation rules).

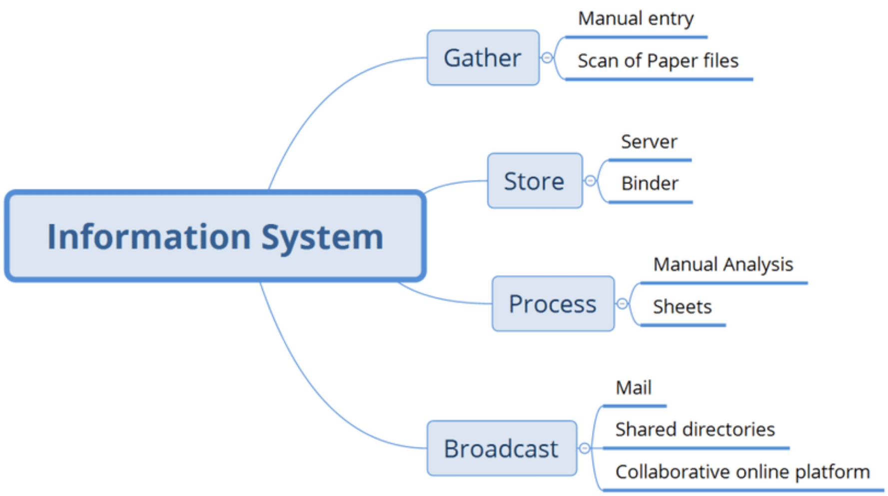


Fig.2. 2 representation of main functions of the Information System [27].

#### 2.4.4. Dissemination of information

It is about making information available to those who need it at the time it is necessary, in a directly exploitable form.

Many means of disseminating information are available: paper, oral forms, and increasingly digital media which can guarantee optimal transmission speeds and reach a maximum number of people. The Internet and interconnection of information systems make this even more evident.

### 2.5.Information system objectives

- **A communication tool:**
  - An internal communication tool: the IS plays the role of an intermediary between the operating system (OS) and the decision system (DS).
  - An external communication tool: the IS plays the role of an intermediary between the company and its environment.
- **A decision support tool:**
  - The IS provides information necessary for decision-making to decision-makers. It makes it possible to study the predictable consequences of decisions and to automate certain decisions.
  - The IS makes it possible to control the evolution of the organization (It makes it possible to detect internal malfunctions and abnormal situations).

- The IS allow to coordinate the activity of the various components of the company and in particular those of the OS.

### 2.5. Elements of an SI

The SI varies from one system to another, it may contain all or part of the following elements [25]:

- All system databases;
- Download the integrated management software package;
- Use the customer relationship management tool;
- The tool of the management of the logistics chain ;
- The business applications ;
- Manage the network infrastructure ;
- Manage data and application servers ;
- Check the safety devices.

### 2.6.Qualities of an IS

- Speed and ease of access to information: thanks to the use of high-performance software and hardware platforms and user-friendly interfaces.
- Reliability of information: the IS must provide safe, reliable and up-to-date information.
- Relevance of information: information must be filtered according to the user
- Information security: through the use of effective security software (filtering routers, anti-virus, firewalls, intrusion detectors, etc.).

### 2.7.The automated information system (AIS)

An automated information system (AIS) is an assembly of computer hardware, software, firmware, designed to perform a specific information processing operations, such as communication, calculation, dissemination, processing and storage of information (automated collection), manipulation, management, control, display, switching, exchange, transmission or reception of data, and includes computer software, firmware and hardware. Computers, word processing systems, networks or other electronic information processing systems, and associated equipment.

Management information systems are a common example of automated information systems [28].

## 2.8.Conclusion

In this chapter, we have seen the general basics of information systems which play a critical role in an interconnected world. They are essential for organizations to efficiently gather, process, store, and distribute information, enabling informed decision-making and facilitating effective communication. As technology continues to advance, the strategic implementation and continuous improvement of information systems will remain fundamental for success in the digital age.

# Chapter 3: Machine learning and its applications

Since their evolution, humans have used many types of tools and techniques to accomplish various tasks in an easier and more efficient way. The creativity of the human brain has led to the invention of different machines. These machines have made human life easy and fulfil various needs of his life including transportation, industries and computing. And artificial intelligence (AI) is one of them.

The Today, artificial intelligence (AI) has far exceeded the hype of quantum computing. One of the most widespread forms of artificial intelligence to date is Machine Learning (ML), Machine Learning relies on mathematical reasoning, translated into algorithms able to digest a large amount of information and data in order to acquire new knowledge or to understand a behavior.

The topic of this chapter is a general introduction to Machine learning (ML), which examines the basic concepts, approaches, algorithms and applications of machine learning.

## 3.1.Machine learning definition

According to Arthur Samuel, machine learning is defined as the field of study that gives computers the ability to learn without being explicitly programmed. Arthur Samuel was famous for his checkers program. Machine learning is used to teach machines to handle data more efficiently [29].

Machine learning generally refers to changes in systems that perform tasks associated with artificial intelligence (AI). Such tasks involve recognition, diagnosis, planning, control, prediction, etc. Changes can either be improvements to already performing systems or synthesis of new systems [30].

## 3.2. History of Machine Learning

Since previous years, the subject of thinking machines has preoccupied the minds of researchers and scientists. This concept is the basis of the thoughts of what will later be known as artificial intelligence (AI), as well as one of its sub-branches: machine learning. However, AI and machine learning algorithms are nothing new, dating back to the 1950s.

The realization of this idea is mainly due to the mathematician Alan Turing and his concept of the "universal machine" in 1936[31], which is the basis of today's computers. He will continue to lay the foundations of machine learning, with his article on "The computer and intelligence" in 1950, in which he develops, among other things, the Turing test.

After that, the American computer scientist Arthur Lee Samuels who is an IBM researcher developed one of the first machine learning programs: a self-learning program for playing checkers. In fact, he invented the term “Machine learning”. His approach to machine learning was explained in an article published in the IBM Journal of Research and Development in 1959[32].

The major advance in the field of artificial intelligence is due to the success of the computer developed by IBM, "Deep Blue"[33], which was the first to defeat the world chess champion Garry Kasparov in 1997. The Deep Blue project will inspire many others in the field of artificial intelligence, including another great challenge: IBM Watson [34], the computer whose goal is to win the game Jeopardy. This goal was achieved in 2011, when Watson won at Jeopardy! by answering questions using natural language processing.

During the following years, the mediated applications of machine learning follow one another much more quickly than before:

- In 2012, a neural network developed by Google manages to recognize human faces as well as cats in YouTube videos.
- In 2014, 64 years after Alan Turing's prediction, the dialoguer "Eugene Goostman"[35] was the first to pass the Turing test by managing to convince 33% of human judges after five minutes of conversation that he is not a computer, but a

13-year-old Ukrainian boy.

- In 2015, a new milestone was reached when Google's "AlphaGo"[36] computer won against one of the best players in the game of Go, the board game considered to be the toughest in the world.
- In 2022, OpenAI launched ChatGPT on November 30, 2022, Chat GT (Chat Generative Pre-trained Transformer) is a prototype of a conversational agent using artificial intelligence, developed by OpenAI and specialized in dialogue, and is continuously refined through the use of supervised learning and reinforcement learning techniques [37].

## 3.3.How Machine Learning Works

Machine Learning allows a computer or computer-assisted system such as an AI program or a robot to adapt its responses or behaviors to the situations encountered, based on the analysis of past empirical data from databases, sensors, or even from the web.

Machine Learning is based on mathematical reasoning, translated into algorithms, capable of digesting large amounts of information and data to acquire new knowledge or even understand behavior. The principle of Machine Learning is precisely to be able to learn independently from this data and to constantly evolve recursively.

These functions of machine learning detect key patterns and adjust their operation accordingly. Concretely, with each interaction with the application, a click (on the links), making a like "like" (for example the applications: Youtube, Facebook, twitter...), or making requests on the engines research and the application is able to deduce a behavior and adjust the proposed content and its interface accordingly [38].

Training a machine learning algorithm to create a model can be divided into three steps:

a. **Representation:** The algorithm creates a pattern to transform the entered data into the desired results. As the learning algorithm is exposed to more data, it will begin to learn the relationship between the raw data and which data points are strong predictors of the desired outcome.

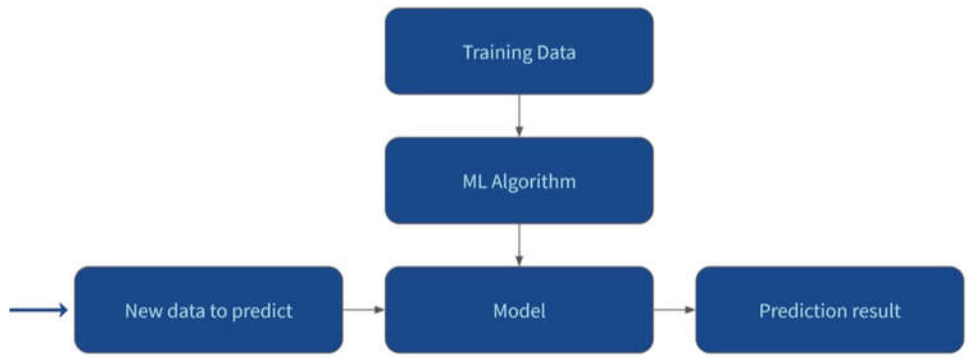


Fig.3. 1 the training a machine learning model.

**b. Evaluation:** As the algorithm creates multiple models, the algorithm will need to evaluate and score the models based on which model produces the most accurate predictions. It is important to remember that once the model is operationalized, it will be exposed to unknown data. Therefore, make sure the model is generalized and does not fit the training data.

**c. Optimization:** Once the algorithm has created and scored several patterns, select the best-performing algorithm. As you expose the algorithm to more diverse input data sets, select the most generalized model.

### 3.4.Feature (machine learning)

The term "feature" is used in machine learning and pattern recognition to describe a measurable property or characteristic of phenomena. Effective algorithms in pattern recognition, classification, and regression must choose informative, discriminating, and independent features. The features of syntactic pattern recognition are typically numerical, but they also include structural features such as strings and graphs. As with linear regression, the concept of "feature" is related to the concept of an explanatory variable [39].

#### 3.4.1. Feature types

Two types of features are commonly used in feature engineering: numerical and categorical [39]:

a. Numerical features are continuous values that can be measured on a scale and used directly by machine learning algorithms. Examples of numerical features include age, height, weight, and income.

b. Categorical features are discrete values that can be categorized in groups. Categorical features typically need to be converted to numerical features before they can be used in machine learning algorithms. Examples of categorical features include gender, color, and zip code.

The type of feature used in feature engineering depends on the specific machine learning algorithm that is being used. Some machine learning algorithms, such as decision trees, can handle both numerical and categorical features. Other machine learning algorithms, such as linear regression, can only handle numerical features.

### 3.4.2. Feature engineering

To extract meaningful information from different types of data, you need to figure out how to represent them. Feature engineering, feature extraction, or feature discovery involves using domain knowledge to extract features (characteristics, properties, attributes) from raw data. Using these extra features will improve the quality of the results from a machine learning process compared to supplying only raw data [40].

## 3.5.Types and approaches of Machine Learning

Machine learning techniques are needed to improve the accuracy of predictive models. Depending on the nature of the problem being addressed, there are different approaches depending on the type and volume of data. In this section, we discuss the categories of machine learning:

### 3.5.1. Supervised learning

Supervised learning, or supervised machine learning, is one of the most popular and successful types of machine learning.

In supervised learning, a labelled datasets are used to train algorithms, which can classify data or predict outcomes accurately.

Supervised learning is intended to find patterns in the data that can be applied to an analysis process. A labelled datasets are used to train algorithms, which can classify data or predict outcomes accurately. These data have labelled characteristics that define the meaning of the data.

In this type of learning, it is sought to define a prediction rule given by the relationship $R : \mathcal{X} \to \mathcal{Y}$ of a variable to be predicted $Y$ as a function of predictive variables $X$ [41].

**Example:** a medical diagnosis is a supervised learning rule ($X$ are the symptoms, $Y$ the diagnosis).

Supervised learning uses a training set to teach models to yield the desired output. This training dataset includes inputs and correct outputs, which allow the model to learn over time. The algorithm measures its accuracy through the loss function, adjusting until the error has been sufficiently minimized.

Supervised learning can be separated into two types of problems when data mining—classification and regression [42]:

a. **Classification:** uses an algorithm to accurately assign test data into specific categories in order to predict a class label, which is a choice from a predefined list of possibilities. It recognizes specific entities within the dataset and attempts to draw some conclusions on how those entities should be labelled or defined. Common classification algorithms are linear classifiers, support vector machines (SVM), decision trees, k-nearest neighbour, and random forest.
b. **Regression:** is used to understand the relationship between dependent and independent variables. It is commonly used to predict a continuous number, and make projections, such as for sales revenue for a given business. Linear regression, logistical regression, and polynomial regression are popular regression algorithms.

#### 3.5.1.1.Supervised learning algorithms

In supervised machine learning processes, various algorithms and computational techniques are used. Here are some of the most commonly used learning methods, typically calculated through the using programs like R or Python [42]:

a. **Linear regression:** Linear regression is used to determine a relationship between a dependent variable and one or more independent variables in order to predict future outcomes. However, linear regression can be a simple or multiple: in simple linear regression, there is only one independent variable and one dependent variable. As the number of independent variables increases, it is referred to as multiple linear regression. For each type of linear regression, it seeks to plot a line of best fit, which is calculated through the method of least squares. However, unlike other regression models, this line is straight when plotted on a graph.

b. **Logistic regression:** While linear regression is leveraged when dependent variables are continuous, logistic regression is selected when the dependent variable is categorical, meaning they have binary outputs, such as "true" and "false" or "yes" and "no." While both regression models seek to understand relationships between data inputs, logistic regression is mainly used to estimate the probability of an event occurring, such as voted or didn't vote, based on a given dataset of independent variables. Since the outcome is a probability, the dependent variable is bounded between 0 and 1.

c. **Support vector machines (SVM):** Support Vector Machines (SVM) are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis. SVMs have been widely applied to model classification problems (Patterns) and nonlinear regressions. Once SVM classifiers have been trained, they can be used to predict future trends.

d. **Neural networks (NN):** Neural interconnections in the brain are abstracted and implemented on digital computers as models of neural networks. Neural networks process training data by simulating the interconnectivity of the human brain through layers of nodes. New applications and new neural network (NN) architectures are being used and studied in more detail in research institutes to control costs and generate revenue in the market. Neural networks learn the mapping function through supervised learning, adjusting based on the loss function through the process of gradient descent. As the cost function is at or near zero, the model's accuracy became confident to yield the correct answer.

e. **Naive Bayes (NB):** Based on Bayes' theorem, Naive Bayes is a classification approach that adopts conditional independence of classes. The presence of one characteristic does not affect the presence of another in the probability of a given outcome, and each predictor has an equal impact. Naive Bayes classifiers come in three types: Multinomial, Bernoulli, and Gaussian. It is mainly used in text classification, spam detection, and recommendation systems.

f. **K-nearest neighbour:** The K-nearest neighbour, also known as the KNN algorithm, is a nonparametric algorithm that classifies data points according to their proximity and their association with other available data. This algorithm assumes that similar data points can be found in close proximity to each other. As a result, it seeks to calculate the distance between the data points, usually through the Euclidean distance, and then it assigns a category according to the most frequent category or average. Its ease of use and low calculation time make it a favorite algorithm of data scientists, but as the test dataset increases, the processing time lengthens, which makes it less attractive for classification tasks. KNN is usually used for recommendation engines and image recognition.

g. **Random forest:** A random forest is a supervised machine learning algorithm that can be used for both classification and regression. The "forest" references a collection of uncorrelated decision trees, which are then merged together to reduce variance and create more accurate data predictions.

#### 3.5.1.2.Supervised learning applications

- Computer vision;
- Pattern recognition;
- Recognition of handwriting;
- Speech Recognition ;
- Automatic language processing;
- Bioinformatics.
- Weather Forecasting
- Supply and demand analysis

### 3.5.2. Unsupervised learning

Unlike supervised learning, unsupervised learning uses unlabelled data. From that data, it discovers patterns that help solve for clustering or association problems.

Unsupervised learning is best suited when the problem requires a massive amount of unlabelled data. Facebook Instagram, Snapchat, Facebook, etc. all contain large amounts of unlabelled data. For example, social media applications, such as

Twitter, Instagram, Snapchat, Facebook etc. all contain large amounts of unlabelled data. Which makes understanding the meaning of these data very complicated and requires algorithms that can begin to understand the meaning by being able to classify the data according to the patterns or clusters [43] that they find.

In this type of learning, the system targets all the data according to their available attributes, in order to classify them into homogeneous groups of examples based on the similarity.

The similarity is generally calculated according to a distance function between pairs of examples. It is then up to the operator to associate or deduce meaning for each group and for the patterns of groups appearance, or groups of groups, in their space. Various mathematical tools and software can help for this. We also talk about regression data analysis [44] (adjustment of a model by a least squares procedure or other optimization of a cost function).

Unsupervised learning models are utilized for three main tasks—clustering, association, and dimensionality reduction. Below we'll define each learning method and highlight common algorithms and approaches to conduct them effectively [41].

a. **Clustering:** Clustering is a statistical analysis method used to organize raw data into homogeneous silos. Within each cluster, the data are grouped according to a common characteristic. Clustering aims to partition a set of data into several groups so that "similar" data are found in the same group according to a similarity function, identifying similar subpopulations in the data. For example, a cluster can be a group of customers with purchase histories, interactions, and other factors.
b. **Association Rules:** An association rule identifies relationships between variables in a dataset by using a rule. Using these methods, companies are able to better understand the relationships between different products by analyzing the consumer basket. Companies can develop better cross-selling strategies and recommendation engines by understanding customers' consumption habits. The playlist "Customers who bought this item also bought" on Amazon or Spotify's "Discover the Week" are examples of this. Several algorithms are used to generate association rules, including Apriori, Eclat, and FP-Growth, but the Apriori algorithm is most commonly used.

c. **Dimensionality reduction:** Although more data leads to more accurate results, it can also negatively impact the performance of machine learning algorithms (e.g. overfitting) and make it difficult to visualize datasets. A dimensionality reduction technique is used when the number of features in a dataset is too high. Data inputs are reduced to a manageable size while also remaining as accurate as possible. Dimensionality reduction is commonly used in the pre-processing stage of data analysis, and there are a few methods available, such as principal component analysis (PCA) [45] and singular value decomposition (SVD) [46].

#### 3.5.2.1.Unsupervised learning algorithms

Among the algorithm used for the unsupervised learning, there are:

- K-mean (K-means clustering);
- K-NN (k nearest neighbors);
- Principal Component Analysis;
- Singular Value Decomposition;
- Independent Component Analysis;
- Distribution models;
- Hierarchical clustering.

#### 3.5.2.2.Unsupervised learning applications

Among the most common applications of unsupervised learning in the real world are [41]:

a. **News Sections:** Google News uses unsupervised learning to categorize articles about the same story from various online media.

b. **Computer vision:** unsupervised learning algorithms are used for visual perception tasks, such as object recognition.

c. **Medical imaging:** unsupervised machine learning provides essential features for medical imaging devices, such as image detection, classification and segmentation, used in radiology and pathology to diagnose patients quickly and accurately.

d. **Anomaly detection:** unsupervised learning models can analyze large amounts of data and discover atypical data points in a data set. These anomalies can lead to faulty equipment, human error or security breaches.

e. **Customer personalities:** The definition of customer personalities makes it possible to better understand the common traits and shopping habits of commercial customers. Unsupervised learning allows companies to create better buyer personality profiles, which allows organizations to align their messages with products in a more appropriate way.

f. **Recommendation engines:** Using past purchase behaviour data, unsupervised learning can help uncover data trends that can be used to develop more effective cross-selling strategies. This is used to make relevant additional recommendations to customers during the payment process for online retailers.

### 3.5.3. Reinforcement learning

Reinforcement learning is a behavioural learning model which consists of receiving feedback from data analysis in order to guide users towards the best outcomes.

Reinforcement learning differs from other types of supervised learning because the system is not trained with the sample data set. On the contrary, the system learns by trial and error. Therefore, a sequence of successful decisions will lead to the "strengthening" of the process because it best solves the problem in question.

A reinforcement learning agent (robot, etc.) learns the actions to take, based on experiments, so as to optimize a quantitative reward over time. Based on his current state, the agent makes decisions based on the environment he is immersed in. As a result, the environment rewards the agent, whether it be positive or negative. Towards the goal of maximizing the sum of rewards over time, the agent seeks, through iterative experiments, a strategy or policy that aligns the action to be taken with the current state [47].

#### 3.5.3.1.Reinforcement learning algorithms

Among the most common frameworks used in Reinforcement Learning are [48]:

- Temporal difference learning (TD-learning);
- Q-learning;
- Monte Carlo algorithm;

- Bootstrap;
- Dynamic programming.
- Markov Decision Process (MDP)

#### 3.5.3.2.Reinforcement learning applications

Reinforcement learning (RL) has found numerous applications across various domains. Here are some notable applications of reinforcement learning [49]:

- Robotic control;
- Text mining or Text mining;
- Finance;
- Health;
- Reinforcement learning is used to solve optimization problems.

### 3.5.4. Deep learning

Deep learning (known also: deep structured learning, or hierarchical learning) is a machine learning technique (often called a sub-discipline of machine learning) that uses hierarchical neural networks (NN) in successive layers in order to learn data iteratively to learn from a combination of unsupervised and supervised algorithms. Deep learning is especially useful when you are trying to learn models from unstructured data. Although deep learning is very similar to a traditional neural network, it will have many more hidden layers. The more complex the problem, the more hidden layers there will be in the model [50].

Deep learning as a complex neural networks are designed to mimic the functioning of the human brain so that computers can be trained to deal with abstractions and ill-defined problems.

Neural networks and deep learning are often used in image recognition, speech and computer vision applications.

#### 3.5.4.1.Deep learning algorithms

The artificial neural network: Deep learning is essentially based on an artificial neural network (ANN) which inspired by the human brain.

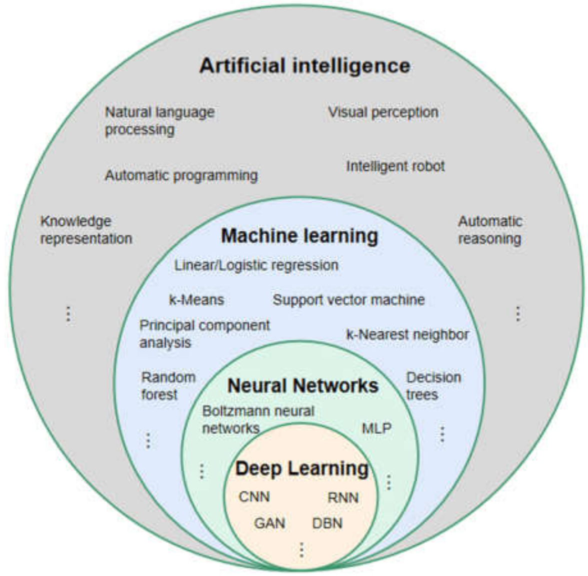


Fig.3. 2 the training a machine learning model.

### 3.5.4.2.Deep learning applications

The area of deep learning applies to a variety of sectors of information and communication technologies, including [51]:

- Visual recognition, for example, of a traffic sign by a robot or an autonomous and vocal car (Analyze the emotions revealed by a photographed or filmed face;
- Robotics ;
- Bioinformatics, for example, for the study of DNA and non-coding segments of the genome, or Cytometry;
- Shape recognition or comparison: Better recognize highly deformable objects;
- Security ;
- Health: To make, in certain cases, a medical diagnosis (e.g. automatic recognition of a cancer in medical imaging, or automatic detection of Parkinson's disease by voice), or prospective or prediction (e.g. prediction of the properties of a soil filmed by a robot) ;
- Computer-assisted pedagogy ;
- Art: reproducing an artistic work from a photo on the computer;
- Artificial intelligence in general ;
- Translation: The deep learning method is now used for the development of machine translation engines.

### 3.6.Conclusion

Today, the use of machine learning techniques has become very common in a wide range of applications. However, machine learning is a very powerful tool that allows us to perform several actions such as classifying data, making a program learn from experiments or even creating an evolutionary program that is constantly improving.

In this chapter, we have provided a general view of machine learning. We have seen the different methods and different approaches that exist for machine learning. Also, we have seen the main algorithms used for learning, as well as some applications for each.

One of the key problems associated with the RESs (Renewable energy sources) is the problem of intermittence of energy production that the time period in which the RESs produce energy often does not coincide with the period when the energy is demanded (), for example the PV production is intermittent and random, and could be strongly fluctuating according to weather conditions (i.e. wind velocity, radiation intensity…), which can occur as a result of variations in the availability of renewable energy due to the intermittent nature of these sources, such as wind and solar radiation There are many concerns about the flexibility, variability, non-controllability of these sources, and they have an impact on the ability maintain the power balance between supply and demand.

In this chapter, we will focus on presenting the proposed approach and methodology of our study. We will delve into the details of the algorithm adopted in our research, providing a comprehensive explanation of its workings.

## 4.1.The architecture of HEMS with IS

The approach for HEMS (Home Energy Management System) proposed by A. Bouakkaz [52], include the IS which aims to balance the fluctuations found in electricity generated in power systems from renewables energies (photovoltaic and wind power …) also, minimizing the dependency on the utility grid and so, reducing the cost of energy. The system is designed in multi-layer system structure, each layer with a different function, which based on multiple criteria decision:

### 4.1.1. A controller layer

Presents the monitoring and controlling system (pilotage system) which considered as the brain of the energy management system in the home that is responsible for planning, scheduling and issuing decisions and orders. The solutions to the optimization problems are based on meta-heuristic optimization algorithms and machine learning algorithms. The controller layer contains two control unites which performed two different function of control:

a. **An anticipative control unit (ACU):** has been used to control energy (production and consumption) in the system for long-term scheduling. The main objective of the anticipative control unit (ACU) is the anticipation of probable scenarios of energy over a long term horizon through:
   - The prediction of the daily energy consumption based on the time of use;
   - The prediction of the energy production from renewable sources based on meteorological data (electricity supply from renewable sources is calculated from a time series of hourly values of wind and solar irradiation in the system as well as weather data);
   - The prediction of energy consumption cost based on the predicted energy demand , energy production and the price of electricity,
   - Determination of the appropriate time to shift household appliances (Scheduling) using meta-heuristics algorithms.

b. **A reactive control unit (RCU):** based on real-time control, which used for a real-time intervention to manage the unexpected variations that may occur in the system.

### 4.1.2. Information layer

Which contain an acquisition data system (ADS) that collect the information and the states of the system. The information layer collects the information from both inside and outside of the building, which includes the recent status of all components of the building (from inside and outside of the building) climate status, occupant status, appliance status, system storage status etc.

### 4.1.3. Operation layer

Presents the active part in the control and supervision mechanism in the building which is responsible for the execution of the orders established by the control layer, it allows the interaction between the different system components (sources, electrical charges, etc.). Its role is to set commend on/off for the devices also is responsible for collecting the measurements from all devices (sources, batteries and appliances) and send it to information layer.

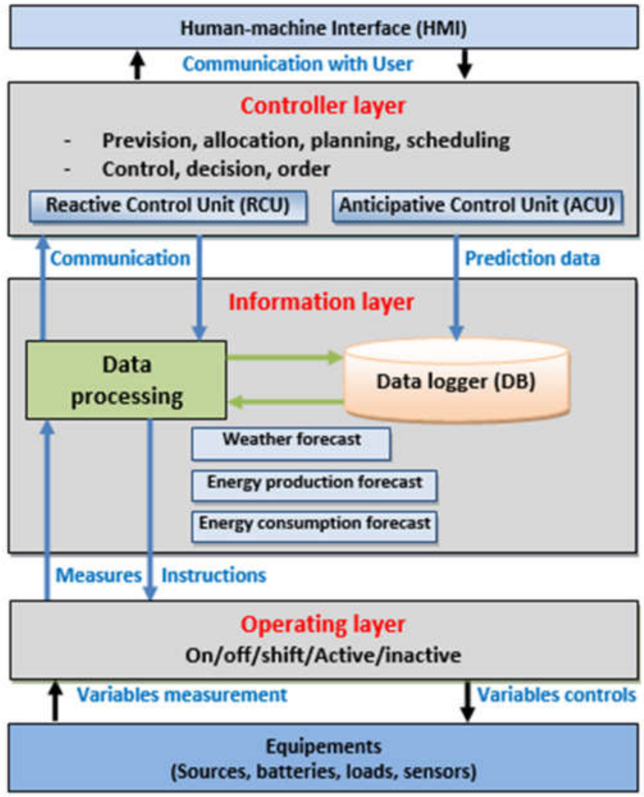


Fig.4. 1 the multi-layer control system.

## 4.2.Machine Learning techniques used for prediction

In order to control the energy in home energy management system (HEMS), we need to anticipate some variable such as the energy consumption and the cost of energy regularly. In our study, we used a machine learning techniques and we selected a four (4) different algorithms to perform the prediction of energy consumption and its cost for a different seasons. And in this section we present those algorithms:

### 4.2.1. Linear regression

The linear regression technique is one of the fundamentals of statistics and machine learning. There's a good chance that we'll need it whether you do statistics, machine learning, or scientific computing. It's best to build a solid foundation first and then proceed toward more complex methods [2].

The concept of regression is used in many fields, including economics, computer science, and social sciences. Increasing data availability and greater awareness of data's practical value make it more important than ever.

The purpose of regression is to forecast a response using a new set of predictors. An example would be to predict a household's electricity consumption based on its outdoor temperature, time of day, and number of residents.

#### 4.2.1.1.Linear regression formulation

In linear regression, we assume that the dependent variable $y$ and the independent variable $x$, are linearly related [53]:

$$x = (x_1, \ldots, x_r) \tag{1}$$

Where $r$ is the number of predictors

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_r x_r + \varepsilon \tag{2}$$

This equation is the regression equation. $\beta_0$, $\beta_1$, …, $\beta_r$ are the regression coefficients, and $\varepsilon$ is the random error.

The linear regression procedure calculates the estimators of the regression coefficients (the predicted weights), denoted with ($b_0$, $b_1$, …, $b_r$). These estimators define the estimated regression function:

$$(x) = b_0 + b_1 x_1 + \cdots + b_r x_r \tag{3}$$

It is essential that the function $f(x)$ should captures as much as possible the dependencies between the inputs and outputs.

For each observation $i = 1, \ldots, n$, the estimated or predicted response $f(x_i)$ should be as close as possible to the corresponding actual response $yi$. The differences actual response $(yi)$ and the predicted response $(x_i)$ for all observations $i = 1, \ldots, n$, are called the residuals.

Regression is about determining the best-predicted weights that is, the weights corresponding to the smallest residuals.

To get the best weights, it usually minimizes the sum of squared residuals (SSR) for all observations:

$$SSR = \Sigma_i (y_i - f(x_i))^2 \tag{4}$$

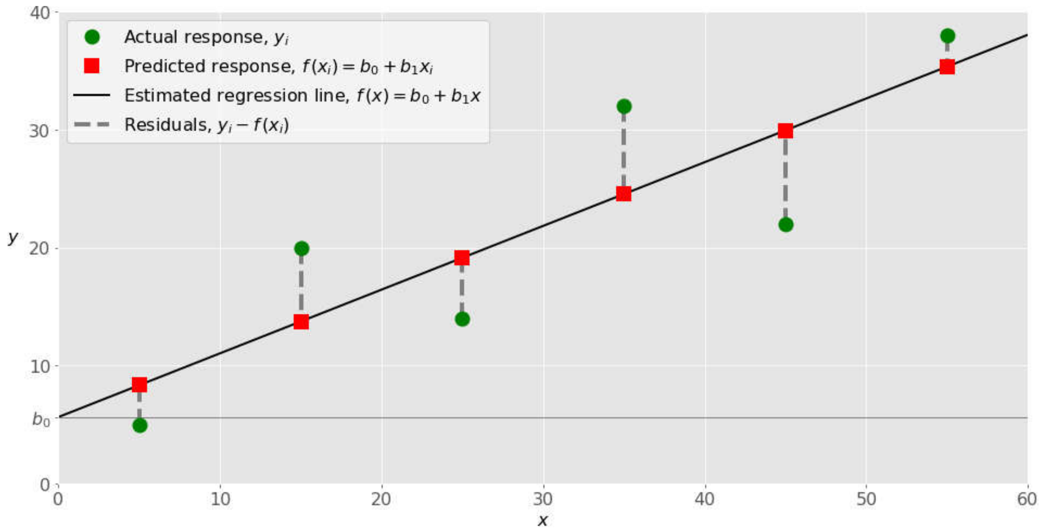


Fig.4. 2 Presentation of a simple linear regression [53].

#### 4.2.1.2.Linear regression Performance

There is a dependence between actual responses $y_i$, $i = 1, ..., n$, and predictors xi. However, there's also an additional inherent variance of the output.

The coefficient of determination denoted as $R^2$, tells us which amount of variation in $y$ can be explained by the dependence on x, using the particular regression model. The value $R^2 = 1$ corresponds to SSR = 0 which presents the perfect fit.

So, a larger $R^2$ indicates a better fit and means that the model can better explain the variation of the output with different inputs.

### 4.2.2. Random forest

A random forest also known as a random decision forest, is an ensemble learning method that operates by constructing a multitude of decision trees at training time to handle classification, regression, and other tasks. In classification tasks, the outcome of the random forest is the class that is selected by the majority of trees. In regression tasks, the average prediction is returned for each tree [54].

Random forests can be used to rank the importance of variables in a regression or classification problem in a natural way.

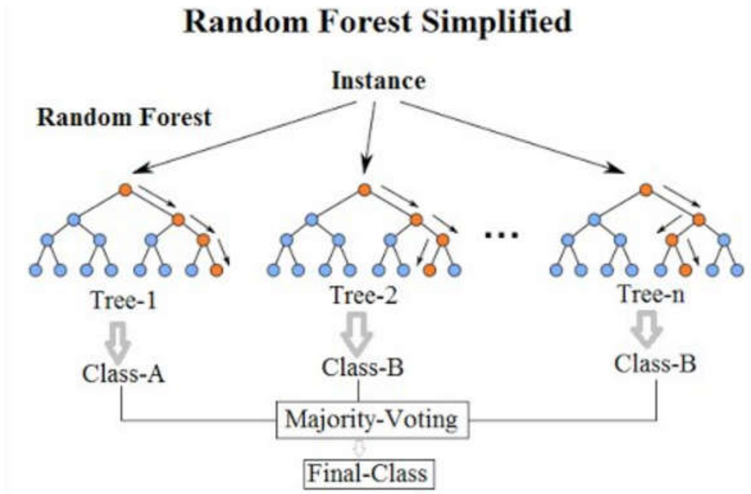


Fig.4. 3 Presentation of a random decision forest (*source: Wikipedia*).

The first step in measuring the variable importance in a data set $D_n$ is to fit a random forest to the data.

$$D_n = \{(X_i; Y_i)\}_{i=1}^{n} \tag{5}$$

Data point out-of-bag errors are recorded and averaged during the fitting process (errors on an independent test set can be substituted if bagging is not used during training).

In order to determine the importance of the $j^{th}$ feature after training, the values of the $j^{th}$ feature are permuted among the training data, and the out-of-bag error is calculated again on this perturbed set. We calculate the importance score for the $j^{th}$ feature by averaging the difference between out-of-bag error before and after permutation over all trees. Standard deviation is used to normalize the score [54].

### 4.2.3. Support Vector Machines

A support vector machine (SVM), also known as a support vector network, is a supervised learning model used for classification and regression analysis.

SVMs are one of the most robust prediction methods, being based on statistical learning frameworks.

Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting) [55].

#### 4.2.3.1.Linear SVM

A form of a given training dataset $D_n$ of n points (Equation (5)) where the $Y_i$ are either 1 or $-1$, each indicating the class to which the point $X_i$ belongs. Each $X_i$ is a p-dimensional real vector. We want to find the "maximum-margin hyperplane" that divides the group of points $X_i$ for which $Y_i$=1 from the group of points for which $Y_i$ =-1, which is defined so that the distance between the hyperplane and the nearest point $X_i$ from either group is maximized.

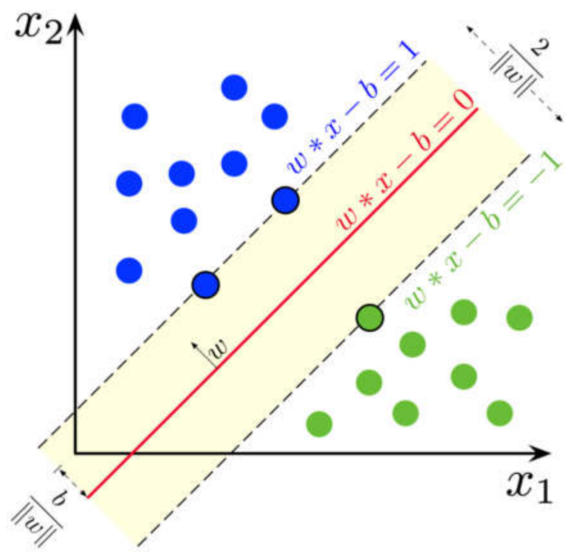


Fig.4. 4 Samples on the margin are called the support vectors. *(Source: Wikipedia)*

### 4.2.4. Artificial Neural Networks

Artificial neural networks, also known as neural networks, are massively parallel distributed processors consisting of simple processing units that store experiential knowledge and make it available. There are two ways in which it resembles the brain [5]:

- Through a learning process, the network acquires knowledge from its environment.
- During learning, synaptic weights are used to store the connection strengths among neurons.

A neural network learns to make predictions by following these steps [57]:

- Taking the input data;
- Making a prediction;
- Comparing the prediction to the desired output;
- Adjusting its internal state to predict correctly the next time

A neuron, as shown in Fig.4.5, is the basic information processing unit of a neural network. A neuron is the fundamental building block for a wide variety of neural networks. The neural model consists of three basic elements [58]:

- A set of synapses, or connecting links: each synapses or connecting link is characterized by its own weight ($w_{kj}$ ) or strength.
- A linear combiner composed of a sum of synaptic strength-weighted input signals.
- A neuron's activation function is a squashing function that limits its output amplitude.

The neural model presented in Fig.4.5 also includes an externally applied bias *($b_k$)*. This bias affects the net input of the activation function in either a positive (increasing) or negative (decreasing) way, depending on its sign.

In mathematical terms, the neuron *k* depicted in Fig. 4.5 can be described by the following pair of equations:

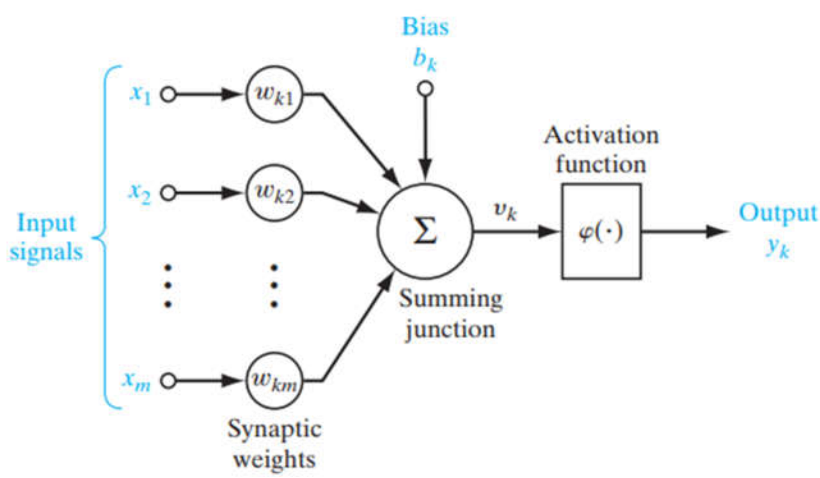


Fig.4. 5 Samples on the margin are called the support vectors. [58]

$$u_k = \sum_{j=1}^{m} w_{kj}\, x_j \tag{6}$$

$$y_i = \varphi(u_k + b_k) \tag{7}$$

where:

- $x_1, x_2, \ldots, x_m$ are the input signals;
- $w_{k1}, w_{k2}, \ldots, w_{km}$ are the respective synaptic weights of neuron $k$;
- $u_k$ is the linear combiner output due to the input signals;
- $b_k$ is the bias;
- $\varphi$ is the activation function;
- $y_k$ is the output signal of the neuron.

### 4.3.Cross-validation technique

Cross-validation is a valuable technique in machine learning and statistical analysis for assessing the generalization performance of models and selecting the best model based on performance metrics. In cross-validation, the data is instead split repeatedly and multiple models are trained. The most commonly used version of cross-validation is k-fold cross-validation, where k is a user-specified number, usually 5 or 10.

When performing five-fold cross-validation, the data is first partitioned into five parts of (approximately) equal size, called folds. Next, a sequence of models is trained. Each model is trained using four folds as the training set and the remaining fold as the validation set. This process is repeated five times, with each fold serving as the validation set once. The performance of each model is then evaluated by calculating a performance metric, such as accuracy or mean squared error, on the validation set. The final performance of the model is usually reported as the average of the performance metrics obtained from the five folds [57].

The advantage of using k-fold cross-validation is that it provides a more robust estimate of a model's performance by using multiple training and validation sets. It helps to reduce the bias and variance in the performance estimate compared to a single train-test split. Additionally, cross-validation allows for a more thorough evaluation of the model's ability to generalize to unseen data, as it tests the model on different subsets of the data.

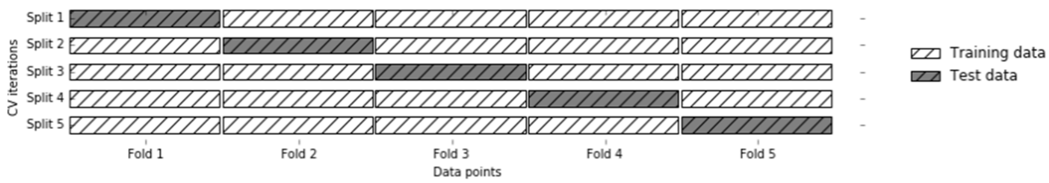


Fig.4. 6 Data splitting in five-fold cross-validation. [57]

## 4.4.Mean Squared Error

Mean Squared Error (MSE) is a commonly used metric to measure the average squared difference between the predicted and actual values in a regression problem. It quantifies the average magnitude of the error or the discrepancy between the predicted values and the ground truth values [59].

To calculate the MSE, you need a set of predicted values and their corresponding actual values. The formula for MSE is:

$$MSE = \frac{1}{n} \times \sum (y_i - \hat{y}_i)^2 \quad (8)$$

where:

- n is the total number of samples or data points;
- $y_i$ represents the actual value of the $i^{th}$ sample
- $\hat{y}_i$ represents the predicted value of the $i^{th}$ sample

The MSE is computed by taking the average of the squared differences between the predicted and actual values. It is called "mean" squared error because it involves calculating the mean (average) of the squared errors.

A lower MSE value indicates better model performance, as it signifies that the predicted values are closer to the actual values on average. On the other hand, a higher MSE suggests greater errors and a poorer fit of the model to the data.

## 4.5.Energy demand and cost prediction

For this study, the prediction of energy consumption and cost of energy consumption will be performed using several machine learning algorithms. The following steps will be followed:

- **Step 1: Selecting training data:** The first step involves selecting a representative set of data that will be used to train the machine learning models. This training data should encompass a wide range of scenarios and cover different variables that impact energy consumption and cost.

  In this study, we used three main parameters as features (global horizontal irradiation, temperature and the power demand) which are directed in the figures 7, 8 and 9 respectively.

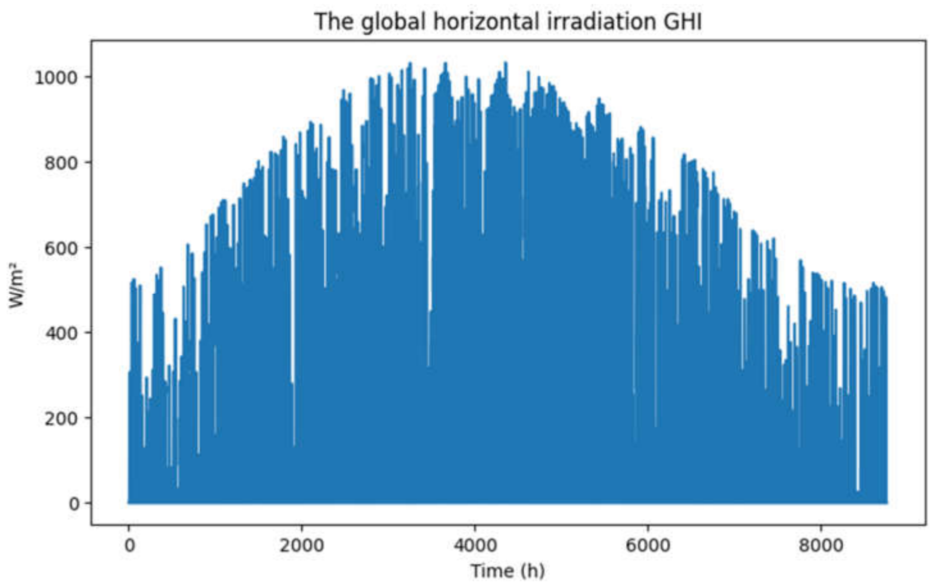


Fig.4. 7 Data of global horizontal irradiation (GHI).

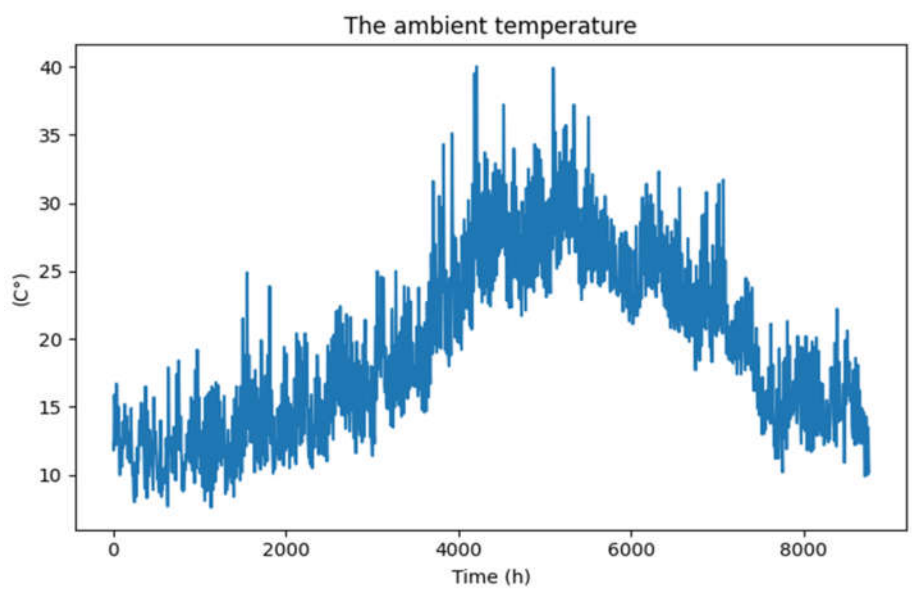


Fig.4. 8 Data of ambient temperature (C°).

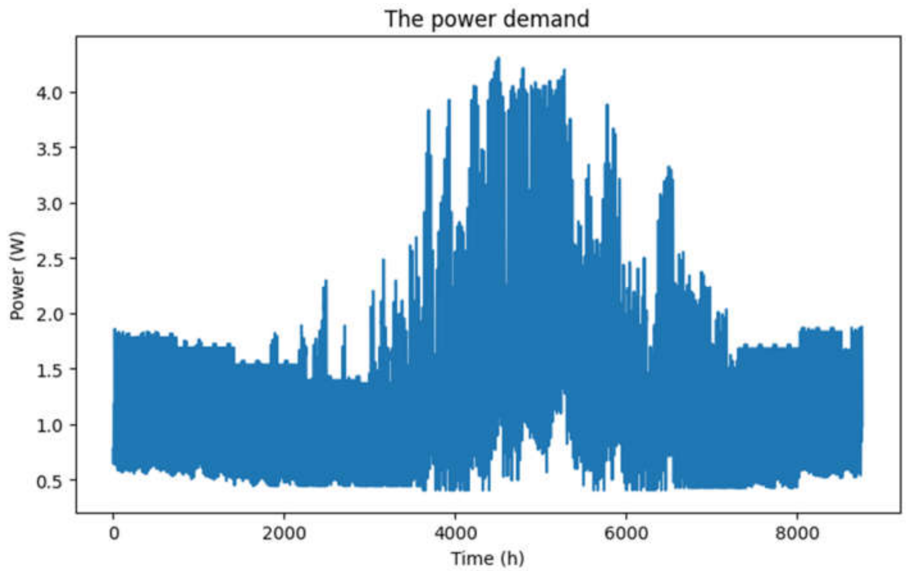


Fig.4. 9 Data of power demand (W).

- **Step 2: Selecting a machine learning algorithm:** Based on the specific requirements and characteristics of the data, several regression machine learning algorithms will be chosen for the prediction task. Various algorithms, such as linear regression, random forests, support vector machines, or artificial neural networks, may be considered.
- **Step 3: Applying cross-validation technique:** To evaluate the performance and generalization ability of the chosen machine learning algorithm, cross-validation will be applied. This involves splitting the training data into multiple subsets or folds (in our case, we chose a k = 10) and iteratively training and testing the model on different combinations of these subsets. The results obtained from cross-validation provide a more robust assessment of the model's predictive capability.
- **Step 4: Fit the training data and calculate the Mean Squared Error (MSE):** The selected machine learning algorithm will be trained using the training data. During the training process, the algorithm will learn patterns and relationships within the data to make accurate predictions. The MSE will be calculated by comparing the predicted values with the actual values from the training data. The MSE provides a quantitative measure of the prediction error.
- **Step 5: Making predictions for new data:** Once the machine learning model is

trained and evaluated, it can be used to make predictions for new, unseen data. This involves providing the model with input data related to energy consumption and cost and obtaining the corresponding predicted values. These predictions can then be used to gain insights, make informed decisions, and optimize energy management in residential homes.

By following these steps, the study aims to leverage machine learning algorithms to accurately predict energy consumption and cost, enabling effective energy management strategies in residential.

### 4.6.Conclusion

In this chapter, our focus was on proposing a decision support system that incorporates an information system for efficient energy management in residential homes. We provided a comprehensive overview of the system and its components, emphasizing its role in supporting decision-making processes related to energy consumption and cost.

To enable accurate predictions of energy consumption and energy cost, we explored various machine learning algorithms. These algorithms, such as linear regression, random forests, support vector machines, and artificial neural networks, are utilized to analyze historical data and forecast future energy patterns. By leveraging these algorithms, our system can provide valuable insights into energy usage and cost, empowering homeowners to make informed decisions and optimize their energy management strategies.

Furthermore, we outlined the main steps involved in making predictions using machine learning algorithms. These steps include selecting training data, choosing an appropriate algorithm, applying cross-validation techniques to evaluate performance, fitting the data to the algorithm, and generating predictions.

This chapter, is devoted to present the obtained results for both energy and cost predictions for different seasons using different algorithms such as: linear regression (LR), random forest (RF), support vector machine (SVM) and artificial neural networks (ANN) in order to select the performance of each algorithm. The selected machine learning algorithms were performed by Python version 3.11.

For energy consumption prediction, we used the solar irradiation and temperature for three months with time step of one hour.

For energy cost prediction, we used both solar irradiation and temperature in addiction the energy demand.

## 5.1.First quarter of the year

### 5.1.1. Energy consumption prediction

The results of energy consumption prediction for the first quarter (January, February, and March) is summarized in Tab.5.1 while the regression curves for each algorithm is shown in Fig.5.1.

For the first quarter, the total real energy consumption equal 2042.322 kWh, while the predicted energy consumption for each algorithm is:

- ✓ Linear regression (LR) algorithm; the total predicted energy was 2044.005 kWh with error equal 1.683 kWh (0.082%).
- ✓ Random forest algorithm (RF), the total predicted energy was 2040.781 kWh with error equal - 1.542 kWh (0.075 %).
- ✓ Support vector machine (SVM), the total predicted energy was 1950.026 kWh with error equal -92.296 kWh (4.519%).
- ✓ Artificial neural networks (ANN), the total predicted energy was 2043.567 kWh with error equal 1.245 kWh (0.061%).

Tab.5. 1 presentation the results of the energy consumption prediction (first quarter)

| Predictor algorithm | Predicted energy consumption (kWh) | Real energy consumption (kWh) | Error (kWh) | Error (%) | MSE |
|---|---|---|---|---|---|
| **LR** | 2044.005 | 2042.322 | 1.683 | 0.082 | 0.121 |
| **RF** | 2040.781 | 2042.322 | - 1.542 | 0.075 | 0.099 |
| **SVM** | 1950.026 | 2042.322 | -92.296 | 4.519 | 0.122 |
| **ANN** | 2043.567 | 2042.322 | 1.245 | 0.061 | 0.117 |

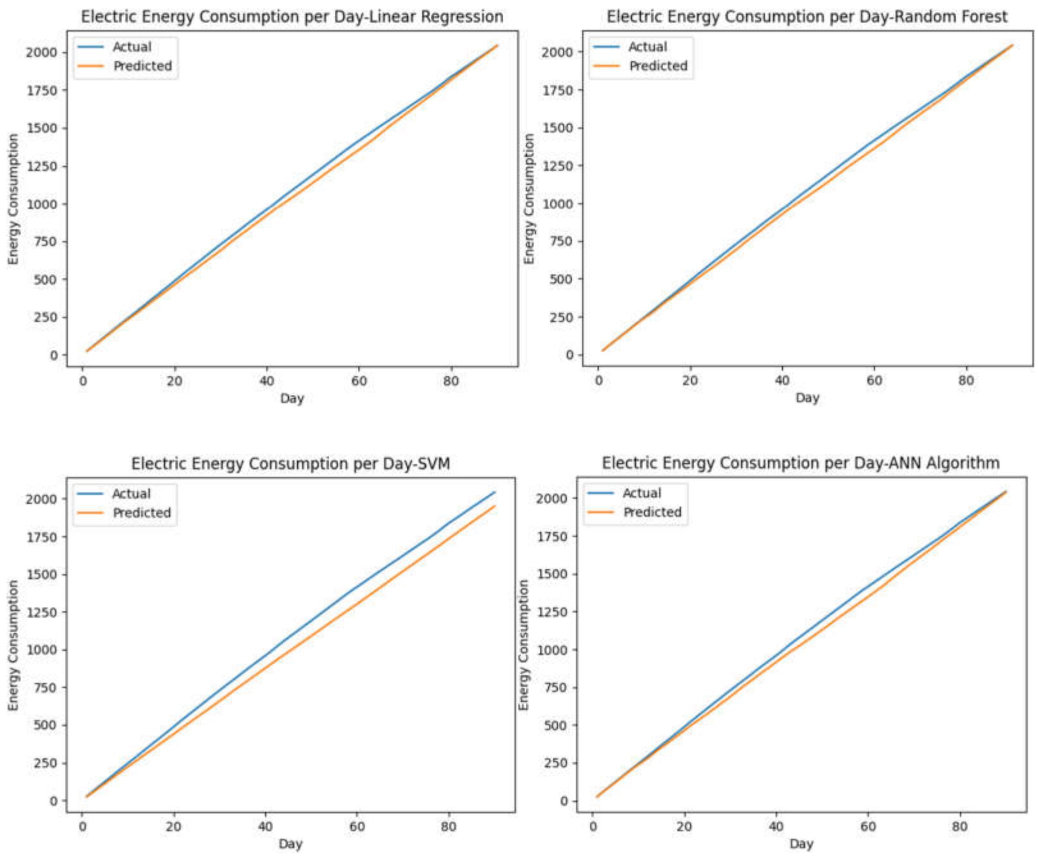


Fig.5. 1 The energy consumption for first quarter.

### 5.1.2. Energy cost prediction

The results of energy cost prediction for the first quarter (January, February, and March) is summarized in Tab.5.2 while the regression curves for each algorithm is shown in Fig.5.2.

For the first quarter, the total real energy cost equal 14313.814 DZD, while the predicted energy consumption for each algorithm is:

- ✓ Linear regression (LR) algorithm; the total predicted energy was 14275.374 DZD with error equal -38.44 DZD (0.269%).

- ✓ Random forest algorithm (RF), the total predicted energy was 14258.875 DZD with error equal -26.94 DZD (0.188%).
- ✓ Support vector machine (SVM), the total predicted energy was 5591.887 DZD with error equal -8721.928 DZD (60.934%).
- ✓ Artificial neural networks (ANN), the total predicted energy was 14432.677 DZD with error equal 118.863 DZD (0.830%).

Tab.5. 2 presentation the results of the energy cost prediction (first quarter)

| Predictor algorithm | Predicted energy cost (DZD) | Real energy cost (DZD) | Error (DZD) | Error (%) | MSE |
|---|---|---|---|---|---|
| **LR** | 14275.374 | 14313.814 | -38.44 | 0.269 | 48.968 |
| **RF** | 14258.875 | 14313.814 | -26.94 | 0.188 | 3.851 |
| **SVM** | 5591.887 | 14313.814 | -8721.928 | 60.934 | 137.604 |
| **ANN** | 14432.677 | 14313.814 | 118.863 | 0.830 | 28.479 |

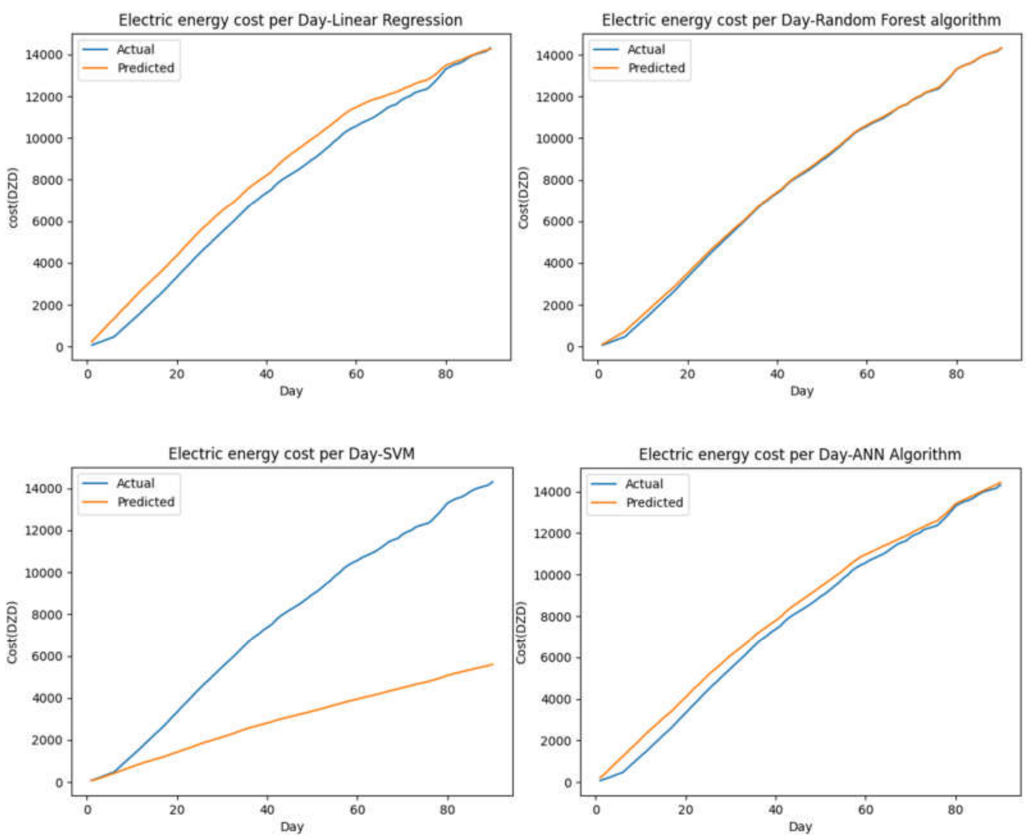


Fig.5. 2 The energy cost for first quarter.

## 5.2. Second quarter of the year

### 5.2.1. Energy consumption prediction

The results of energy consumption prediction for the second quarter (April, May and June) is summarized in Tab.5.3 while the regression curves for each algorithm is shown in Fig.5.3.

Tab.5. 3 presentation the results of the energy consumption prediction (second quarter)

| Predictor | Predicted energy consumption (kWh) | Real energy consumption (kWh) | Error (kWh) | Error (%) | MSE |
|---|---|---|---|---|---|
| **LR** | 2564.643 | 2572.678 | - 8.035 | 0.312 | 0.384 |
| **RF** | 2574.152 | 2572.678 | 1.475 | 0.057 | 0.142 |
| **SVM** | 1981.517 | 2572.678 | -591.16 | 22.978 | 0.519 |
| **ANN** | 2646.02 | 2572.678 | 73.342 | 2.851 | 0.418 |

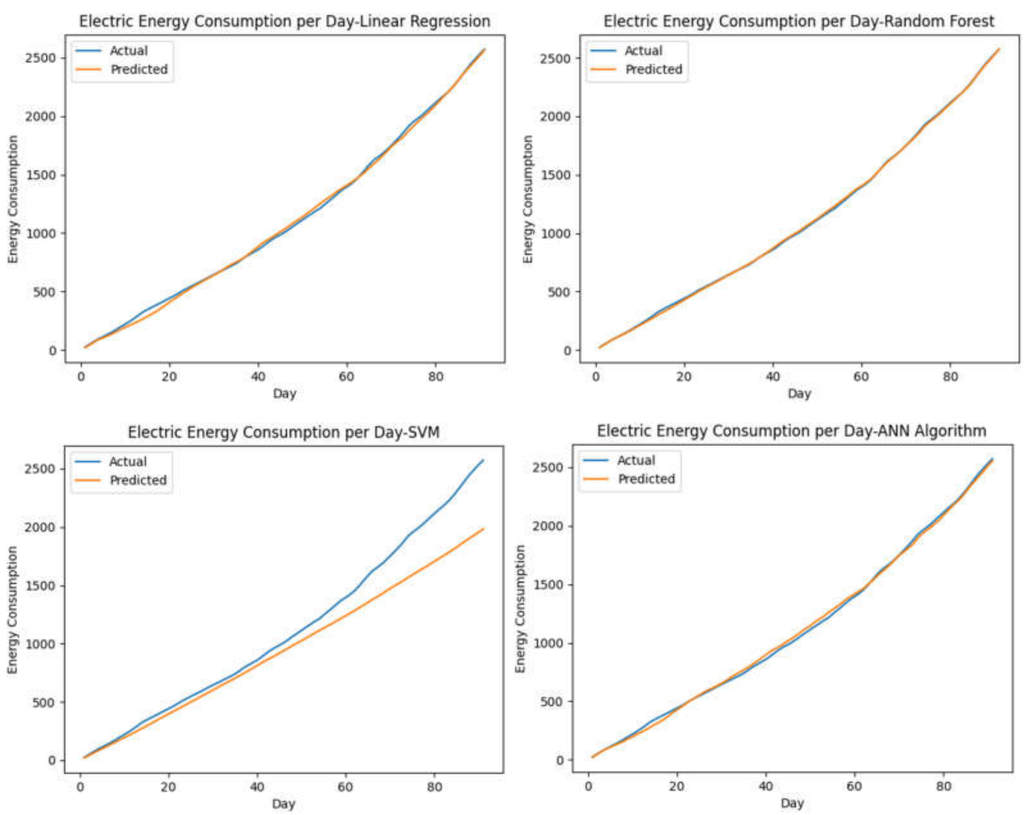


Fig.5. 3 The energy consumption for second quarter: (a: LR; b: RF; c: SVM; d: ANN)

For the second quarter, the total real energy consumption equal 2572.678 kWh, while the predicted energy consumption for each algorithm is:

- ✓ Linear regression (LR) algorithm; the total predicted energy was 2564.643 kWh with error equal - 8.035 kWh (0.312%).

- ✓ Random forest algorithm (RF), the total predicted energy was 2574.152 kWh with error equal 1.475 kWh (0.057%).
- ✓ Support vector machine (SVM), the total predicted energy was 1981.517 kWh with error equal -591.16 kWh (22.978%).
- ✓ Artificial neural networks (ANN), the total predicted energy was 2646.02 kWh with error equal 73.342 kWh (2.851%).

### 5.2.2. Energy cost prediction

The results of energy cost prediction for the second quarter (April, May and June) is summarized in Tab.5.4 while the regression curves for each algorithm is shown in Fig.5.4.

For the second quarter, the total real energy cost equal 15592.805 DZD, while the predicted energy consumption for each algorithm is:

- ✓ Linear regression (LR) algorithm; the total predicted energy was 15564.629 DZD with error equal -28.176 DZD (0.181%).
- ✓ Random forest algorithm (RF), the total predicted energy was 15645.207 DZD with error equal 52.403 DZD (0.336%).
- ✓ Support vector machine (SVM), the total predicted energy was 2792.304 DZD with error equal -12800.501 DZD (82.092%).
- ✓ Artificial neural networks (ANN), the total predicted energy was 15335.168 DZD with error equal -257.637 DZD (1.652%).

Tab.5. 4 presentation the results of the energy cost prediction (second quarter)

| **Predictor algorithm** | **Predicted energy cost (DZD)** | **Real energy cost (DZD)** | **Error (DZD)** | **Error (%)** | **MSE** |
|---|---|---|---|---|---|
| **LR** | 15564.629 | 15592.805 | -28.176 | 0.181 | 109.218 |
| **RF** | 15645.207 | 15592.805 | 52.403 | 0.336 | 4.974 |
| **SVM** | 2792.304 | 15592.805 | -12800.501 | 82.092 | 255.306 |
| **ANN** | 15335.168 | 15592.805 | -257.637 | 1.652 | 33.248 |

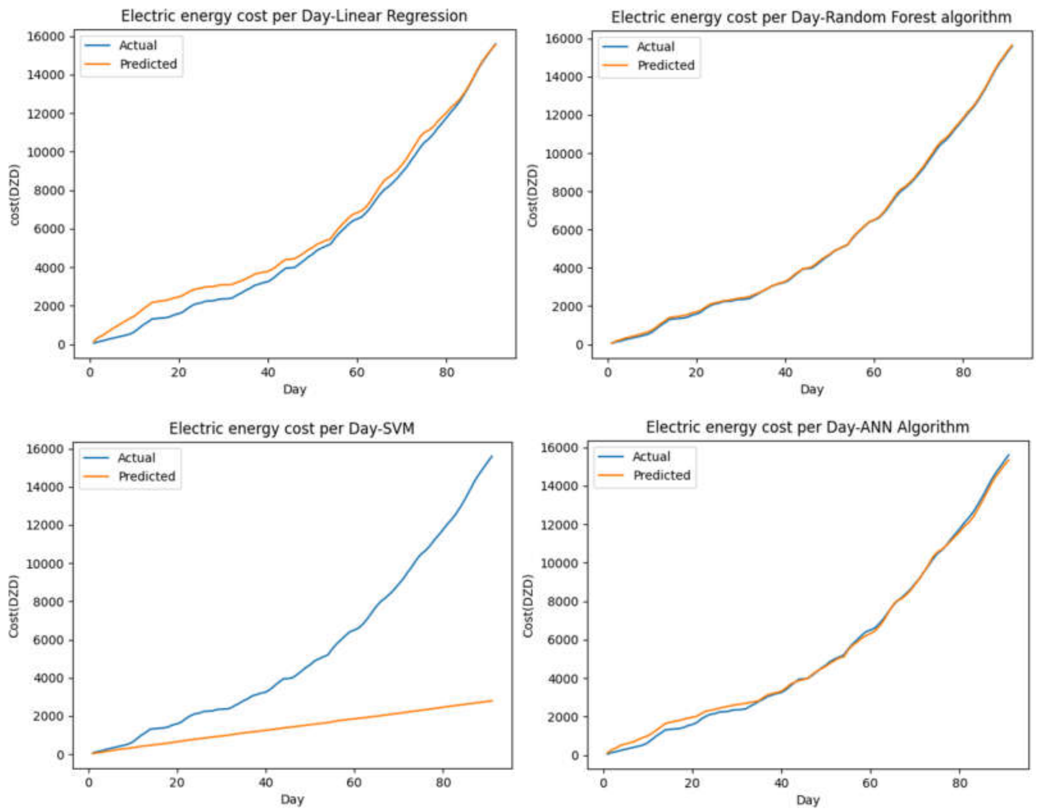


Fig.5. 4 The energy cost for second quarter.

## 5.3.Third quarter of the year

### 5.3.1. Energy consumption prediction

The results of energy consumption prediction for the third quarter (July, August and September) is summarized in Tab.5.5 while the regression curves for each algorithm is shown in Fig.5.5.

For the third quarter, the total real energy consumption equal 4004.297 kWh, while the predicted energy consumption for each algorithm is:

- ✓ Linear regression (LR) algorithm; the total predicted energy was 3987.462 kWh with error equal -16.835 kWh (0.420%).
- ✓ Random forest algorithm (RF), the total predicted energy was 4005.775 kWh with error equal 1.479 kWh (0.037%).
- ✓ Support vector machine (SVM), the total predicted energy was 3292.043 kWh with error equal -712.254 kWh (17.787%).
- ✓ Artificial neural networks (ANN), the total predicted energy was 4067.916 kWh with error equal 63.619 kWh (1.589%).

Tab.5. 5 presentation the results of the energy consumption prediction (third quarter)

| Predictor | Predicted energy consumption (kWh) | Real energy consumption (kWh) | Error (kWh) | Error (%) | MSE |
|---|---|---|---|---|---|
| **LR** | 3987.462 | 4004.297 | -16.835 | 0.420 | 0.832 |
| **RF** | 4005.775 | 4004.297 | 1.479 | 0.037 | 0.347 |
| **SVM** | 3292.043 | 4004.297 | -712.254 | 17.787 | 1.036 |
| **ANN** | 4067.916 | 4004.297 | 63.619 | 1.589 | 0.878 |

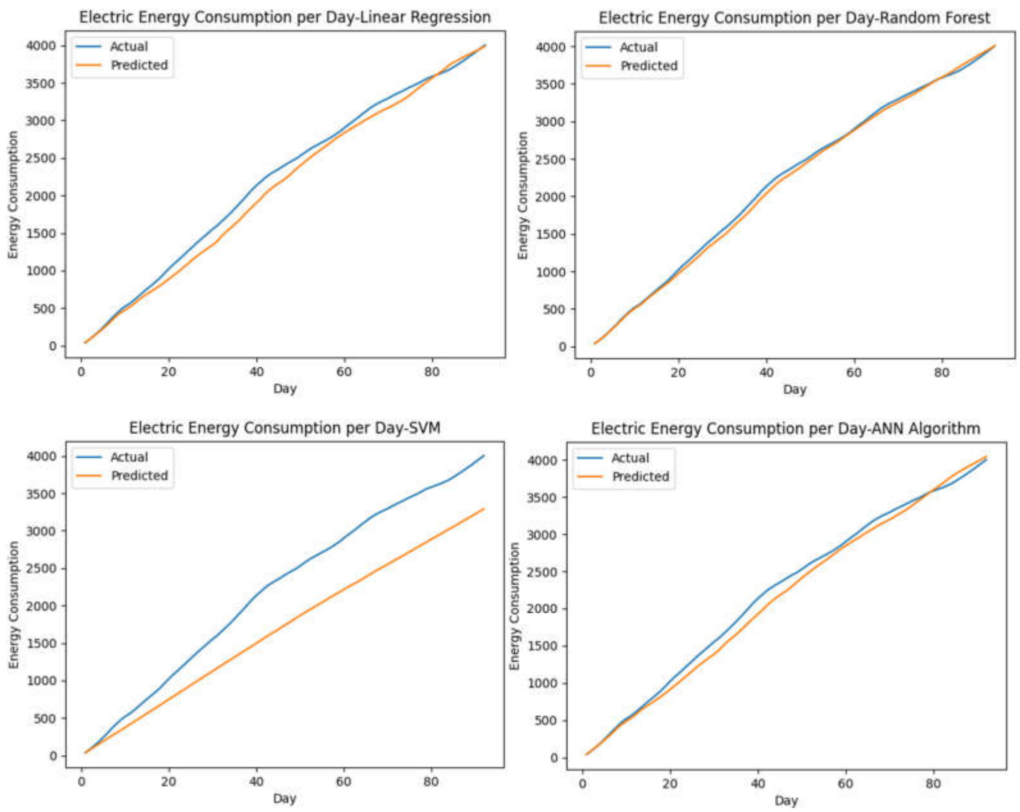


Fig.5. 5 The energy consumption for third quarter.

### 5.3.2. Energy cost prediction

The results of energy cost prediction for the third quarter (July, August and September) is summarized in Tab.5.6 while the regression curves for each algorithm is shown in Fig.5.6.

Tab.5. 6 presentation the results of the energy cost prediction (third quarter)

| Predictor algorithm | Predicted energy cost (DZD) | Real energy cost (DZD) | Error (DZD) | Error (%) | MSE |
|---|---|---|---|---|---|
| **LR** | 32836.864 | 32884.313 | -47.449 | 0.144 | 218.758 |
| **RF** | 32954.907 | 32884.313 | 70.594 | 0.215 | 16.513 |
| **SVM** | 9006.306 | 32884.313 | -23878.008 | 72.612 | 652.963 |
| **ANN** | 33420.703 | 32884.313 | 536.39 | 1.631 | 110.376 |

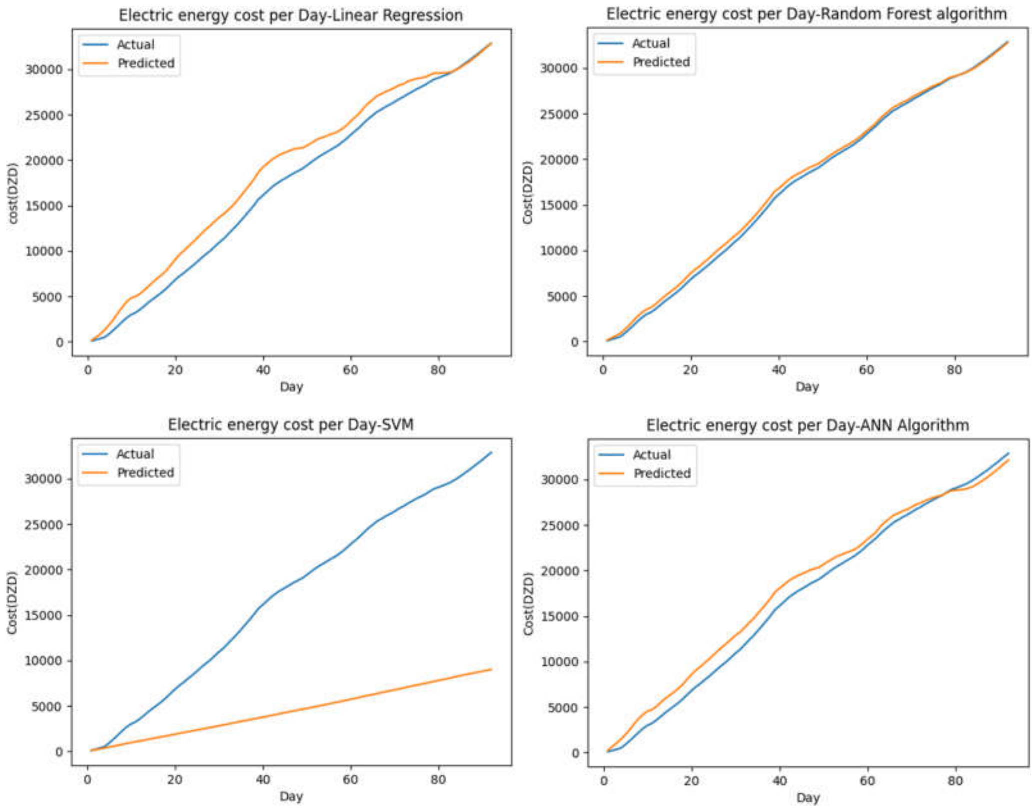


Fig.5. 6 The energy cost for first quarter.

For the third quarter, the total real energy cost equal 32884.313 DZD, while the predicted energy consumption for each algorithm is:

- ✓ Linear regression (LR) algorithm; the total predicted energy was 32836.864 DZD with error equal -47.449 DZD (0.144%).
- ✓ Random forest algorithm (RF), the total predicted energy was 32954.907 DZD with error equal 70.594 DZD (0.215%).
- ✓ Support vector machine (SVM), the total predicted energy was 9006.306 DZD with error equal -23878.008 DZD (72.612%).
- ✓ Artificial neural networks (ANN), the total predicted energy was 33420.703 DZD with error equal 536.39 DZD (1.631%).

## 5.4.Fourth quarter of the year

### 5.4.1. Energy consumption prediction

The results of energy consumption prediction for the fourth quarter (October, November and December) is summarized in Tab.5.7 while the regression curves for each algorithm is shown in Fig.5.7.

Tab.5. 7 presentation the results of the energy consumption prediction (fourth quarter)

| Predictor | Predicted energy consumption (kWh) | Real energy consumption (kWh) | Error (kWh) | Error (%) | MSE |
|---|---|---|---|---|---|
| **LR** | 2194.975 | 2210.039 | -15.064 | 0.682 | 0.209 |
| **RF** | 2211.515 | 2210.039 | 1.477 | 0.067 | 0.137 |
| **SVM** | 1975.587 | 2210.039 | -234.452 | 10.608 | 0.217 |
| **ANN** | 2200.607 | 2210.039 | -9.432 | 0.427 | 0.201 |

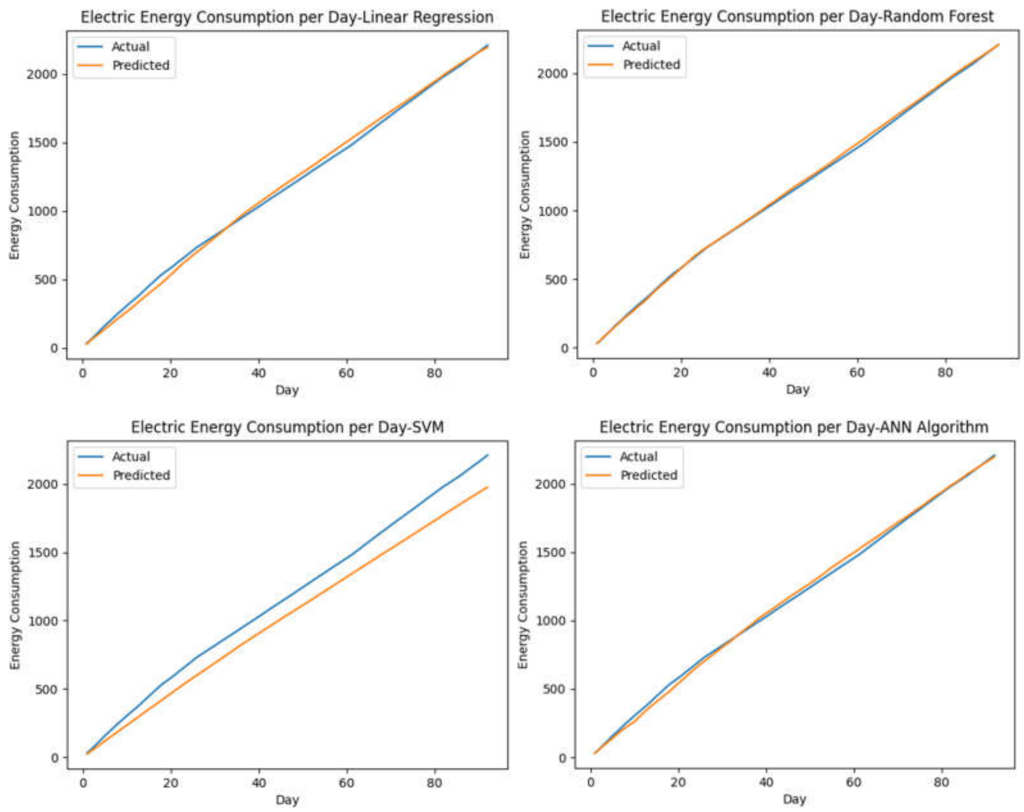


Fig.5. 7 The power consumption for first quarter.

For the fourth quarter, the total real energy consumption equal 2210.039 kWh, while the predicted energy consumption for each algorithm is:

- ✓ Linear regression (LR) algorithm; the total predicted energy was 2194.975 kWh with error equal 1.477 kWh (0.682%).
- ✓ Random forest algorithm (RF), the total predicted energy was 2211.515 kWh with error equal 1.479 kWh (0.067%).
- ✓ Support vector machine (SVM), the total predicted energy was 1975.587 kWh with error equal -234.452 kWh (10.608%).
- ✓ Artificial neural networks (ANN), the total predicted energy was 2200.607 kWh with error equal -9.432 kWh (0.427%).

### 5.4.2. Energy cost prediction

The results of energy cost prediction for the fourth quarter (October, November and December) is summarized in Tab.5.8 while the regression curves for each algorithm is shown in Fig.5.8.

Tab.5. 8 presentation the results of the energy cost prediction (fourth quarter)

| Predictor algorithm | Predicted energy cost (DZD) | Real energy cost (DZD) | Error (DZD) | Error (%) | MSE |
|---|---|---|---|---|---|
| **LR** | 19127.688 | 18770.027 | 357.661 | 1.905 | 70.679 |
| **RF** | 18761.427 | 18770.027 | -8.599 | 0.046 | 1.709 |
| **SVM** | 6597.788 | 18770.027 | -12172.238 | 64.849 | 200.913 |
| **ANN** | 19194.615 | 18770.027 | 424.589 | 2.262 | 21.997 |

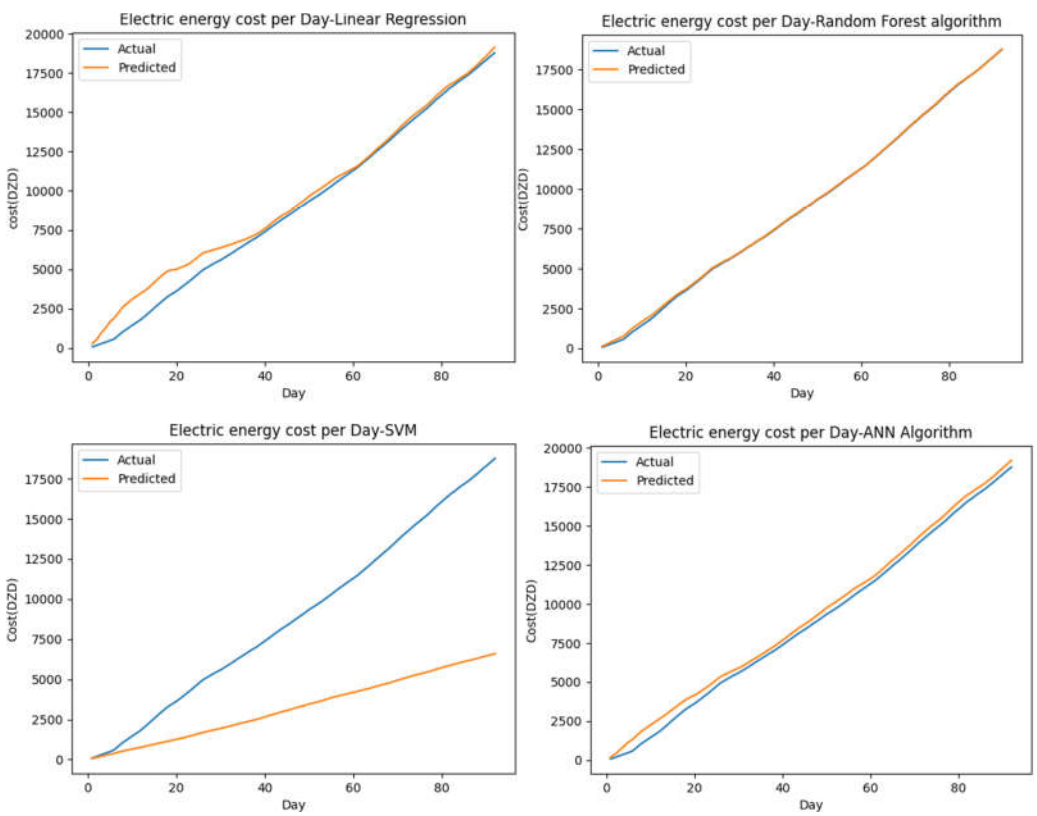


Fig.5. 8 The energy cost for first quarter.

For the third quarter, the total real energy cost equal 18770.027 DZD, while the predicted energy consumption for each algorithm is:

- ✓ Linear regression (LR) algorithm; the total predicted energy was 19127.688 DZD with error equal 357.661 DZD (1.905%).
- ✓ Random forest algorithm (RF), the total predicted energy was 18761.427 DZD

with error equal -8.599 DZD (0.046%).

- ✓ Support vector machine (SVM), the total predicted energy was 6597.788 DZD with error equal 12172.238 DZD (64.849%).
- ✓ Artificial neural networks (ANN), the total predicted energy was 19194.615 DZD with error equal 424.589 DZD (2.262%).

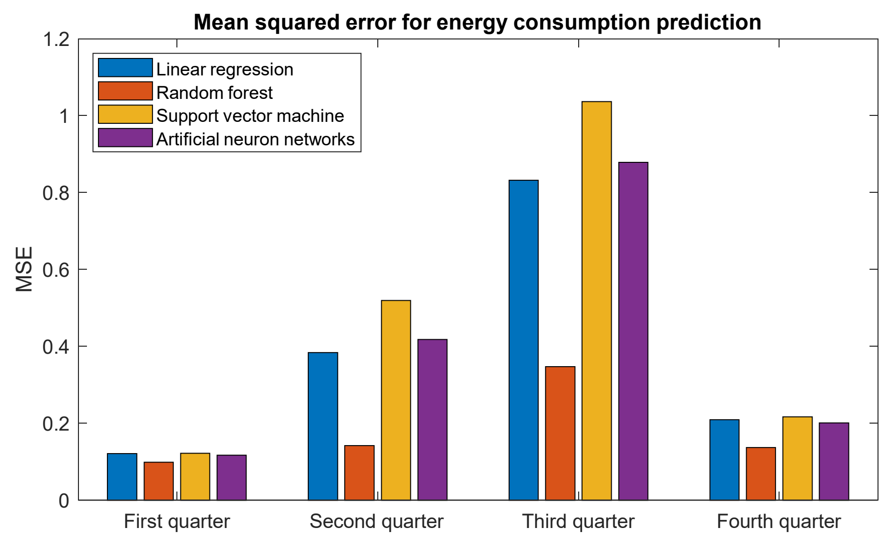


Fig.5. 9 The mean squared error for energy consumption prediction

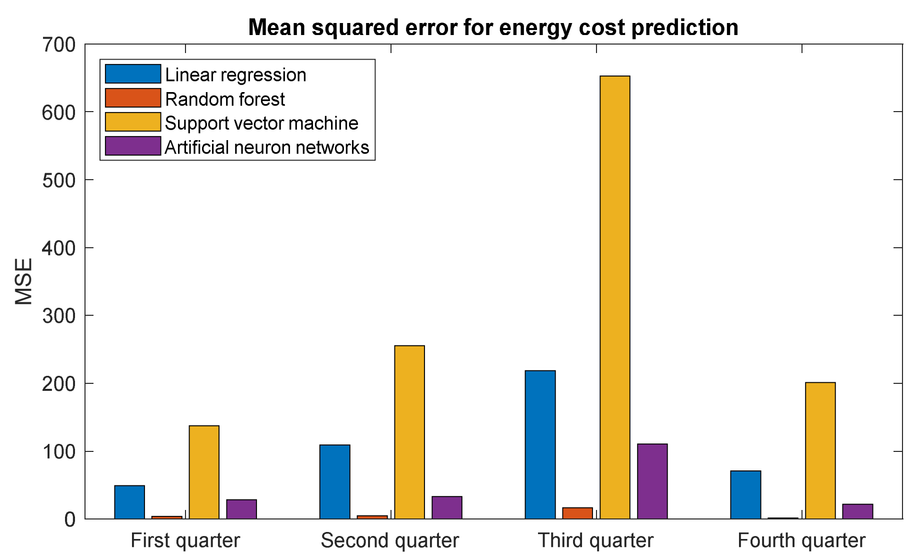


Fig.5. 10 The mean squared error for energy cost prediction

The Fig.5.9 shows the mean squared error (MSE) for the energy prediction obtained by each algorithm. The MSE recorded for each algorithm is:

- ✓ The first quarter: 0.121, 0.099, 0.122, and 0.117 for LR, RF, SVM and ANN respectively.

- ✓ The second quarter: 0.384, 0.142, 0.519 and 0.418 for LR, RF, SVM and ANN respectively.
- ✓ The third quarter: 0.832, 0.347, 1.036 and 0.878 for LR, RF, SVM and ANN respectively.
- ✓ The fourth quarter: 0.209, 0.137, 0.217 and 0.201 for LR, RF, SVM and ANN respectively.

The MSE for the cost prediction obtained by each algorithm is presented by Fig.5.10, and its value is as follow:

- ✓ The first quarter: 48.968, 3.851, 137.604 and 28.479 for LR, RF, SVM and ANN respectively.
- ✓ The second quarter: 109.218, 4.974, 255.306 and 33.248 for LR, RF, SVM and ANN respectively.
- ✓ The third quarter: 218.758, 16.513, 652.963 and 110.376 for LR, RF, SVM and ANN respectively.
- ✓ The fourth quarter: 70.679, 1.709, 200.913 and 21.997 for LR, RF, SVM and ANN respectively.

Based on the previous results, for the energy consumption and cost prediction, the random forest algorithm (RF) is the best with minimum errors regarding to the total energy consumption and cost errors with followed by the linear regression (LR) and artificial neural networks (ANN), while the support vector machine (SVM) is the worst regarding to these errors.

For the mean squared error (MSE), also, the random forest (RF) has the best performance (the best fit) with minimum MSE followed in this time by the ANN algorithm and LR, while the SVM has the worst regarding the MSE.

## 5.5.Conclusion

In this chapter, we used a different machine learning algorithms to predict energy consumption and energy cost. The results obtained from the prediction show us that the random forest algorithm has the best performance for both energy consumption and cost prediction. With the help of these forecasts, the system makes it possible to control energy consumption and also the cost of energy in the home.

# Conclusions

In this study, we have proposed a decision support system that incorporates an information system for efficient energy management in residential homes. To predict energy consumption and cost, we employed various machine learning algorithms, including linear regression, random forests, support vector machine, and artificial neural networks. Through our analysis, we found that the random forest algorithm exhibited the best performance in predicting both energy consumption and cost. Leveraging these predictions, the system enables the control of energy consumption and cost within the home.

As a future perspective, we aim to explore the application of a metaheuristics approach based on these findings to further optimize energy consumption and reduce energy costs. This approach would involve employing optimization techniques to dynamically control energy usage based on real-time data and varying factors. By integrating metaheuristics into our system, we anticipate achieving even greater efficiency and cost savings in residential energy management.

# Bibliography


[1] Men, R.L., 1979. SOLAR ENERGY M A G D E N T SYSTEM. PhD Propos. 1. https://doi.org/10.1017/CBO9781107415324.004

[2] Capehart, B.L., Muth, E.J., Stop, M.O., 1982. Minimizing residential electrical energy costs using microcomputer energy management systems. 6, 261–269.

[3] Rahman, S., Bhatnagar, R., 1986. Computerized energy management systems-why and how. J. Microcomput. Appl. 9, 261–270. https://doi.org/10.1016/0745-7138(86)90022-9.

[4] Wacks, K.P., 1991. Utility Load Management Using Home Automation. IEEE Trans. Consum. Electron. 37, 168–174. https://doi.org/10.1109/30.79325.

[5] Stum. K, Mosier. R, E.H.T., 1997. Energy management systems.

[6] Obayashi, F., Tokunaga, Y., 2003. A study of building energy management system based of multi agents. SICE 2003 Annu. Conf. (IEEE Cat. No.03TH8734) 2, 1526–1530.

[7] Fardadi, M., Ghafari, A.S., Hannani, S.K., 2005. PID Neural Network control of SUT building energy management system. IEEE/ASME Int. Conf. Adv. Intell. Mechatronics, AIM 1, 682–686. https://doi.org/10.1109/aim.2005.1511061.

[8] Rim MISSAOUI, B., 2012. Gestion Énergétique optimisée pour un bâtiment intelligent validations Résumé. Université de Grenoble.

[9] H. Zhang, A. Davigny, F. Colas, Y. Poste, and B. Robyns, “Fuzzy logic based energy management strategy for commercial buildings integrating photovoltaic and storage systems,” Energy Build., vol. 54, pp. 196–206, 2012.

[10] E. Kuznetsova, Y. F. Li, C. Ruiz, E. Zio, G. Ault, and K. Bell, “Reinforcement learning for microgrid energy management,” Energy, vol. 59, pp. 133–146, Sep. 2013..

[11] Nakabi, T. A., & Toivanen, P. (2021). Deep reinforcement learning for energy management in a microgrid with flexible demand. Sustainable Energy, Grids and Networks, 25, 100413. https://doi.org/10.1016/j.segan.2020.100413.

[12] Han, G., Lee, S., Lee, J., Lee, K., & Bae, J. (2021). Deep-learning- and reinforcement-learning-based profitable strategy of a grid-level energy storage system for the smart grid. Journal of Energy Storage, 41, 102868. https://doi.org/10.1016/j.est.2021.102868.

[13] A. Forootani, M. Rastegar, and M. Jooshaki, “An Advanced Satisfaction-Based Home Energy Management System Using Deep Reinforcement Learning,” *IEEE Access*, vol. 10, pp. 47896–47905, 2022.

[14] Dinh, H. T., Lee, K., & Kim, D. (2022). Supervised-learning-based hour-ahead demand response for a behavior-based home energy management system approximating MILP optimization. Applied Energy, 321, 119382. https://doi.org/10.1016/j.apenergy.2022.119382.

[15] Cai, W., Kordabad, A. B., & Gros, S. (2023). Energy management in residential microgrid using model predictive control-based reinforcement learning and Shapley value. Engineering Applications of Artificial Intelligence, 119, 105793. https://doi.org/10.1016/j.engappai.2022.105793.

[16] Donnée. (2022, November 20). In Wikipedia. https://fr.wikipedia.org/wiki/Donn%C3%A9e.

[17] Donnée (informatique). (2021, December 24). In Wikipedia. https://fr.wikipedia.org/wiki/Donn%C3%A9e(informatique).

[18] Information. (2023, February 23). In Wikipedia. https://fr.wikipedia.org/wiki/Information.

[19] Processus. (2022, March 30). In Wikipedia. https://fr.wikipedia.org/wiki/Processus.

[20] Processus (informatique). (2022, March 17). In Wikipedia. https://fr.wikipedia.org/wiki/Processus_(informatique).

[21] Poussin, Jean-Christophe. (1987). Notions de système et de modèle : note. Agronome ORSTOM, Laboratoire d'Économie rurale, INA-PG, 78850 Thiverval Grignon.

[22] Oliveira, Káthia & Thion, Virginie & Dupuy-Chessa, Sophie & Gervais, Marie-Pierre & Cherfi, Samira & Kolski, Christophe. (2012). Limites de l'évaluation d'un Système d'Information : une analyse fondée sur l'expérience pratique.

[23] Pierre-Emmanuel Arduin, Michel Grundstein, Camille Rosenthal-Sabroux. Système d'information et de connaissance. ISTE Éditions, 2015, 978-1-78405-081-8. ⟨hal-01194848⟩.

[24] Reix, R., Fallery, B., Kalika, M. & Rowe, F. (2016). Chapitre 2. Système d'information et décision. Dans : , R. Reix, B. Fallery, M. Kalika & F. Rowe (Dir), Systèmes d'information et management (pp. 71-125). Paris: Vuibert.

[25] Stair, R., & Reynolds, G. (2020). Principles of information systems. Cengage Learning.

[26] Laurini, R., & Thompson, D. (1992). Fundamentals of spatial information systems (Vol. 37). Academic press.

[27] Dakhli, Zakaria & Danel, Thomas & Lafhaj, Zoubeir. (2019). Smart Construction Site: Ontology of Information System Architecture. Modular and Offsite Construction (MOC) Summit Proceedings. 41-50. 10.29173/mocs75.

[28] Automated information system. (2022, June 29). In Wikipedia. https://en.wikipedia.org/wiki/Automated_information_system.

[29] B. Mahesh, Machine Learning Algorithms - A Review, 2020. https://doi.org/10.21275/ART20203995.

[30] N.J. Nilsson, Introduction to machine learning (an early draft of a proposed textbook), Laboratory, Robotics Science, Department of Computer University, Stanford, Stanford, 2005.

[31] A.M. Tuking, In a recent paper Alonzo Church f has introduced an idea of " effective calculability ", which is equivalent to my " computability ", but is very, 38 (1936) 173–198.

[32] J. Hurwitz, K. Daniel, Machine Learning For Dummies, John Wiley & Sons, Inc. 111 River St., Hoboken, 2018. www.wiley.com.

[33] Campbell, M., Hoane Jr, A. J., & Hsu, F. H. (2002). Deep blue. Artificial intelligence, 134(1-2), 57-83.

[34] Chen, Y., Argentinis, J. E., & Weber, G. (2016). IBM Watson: how cognitive computing can be applied to big data challenges in life sciences research. Clinical therapeutics, 38(4), 688-701.

[35] Neufeld, E., & Finnestad, S. (2020). In defense of the Turing test. AI & SOCIETY, 35, 819-827.

[36] Chen, J. X. (2016). The evolution of computing: AlphaGo. Computing in Science & Engineering, 18(4), 4-7.

[37] Roose, K. (2022). The brilliance and weirdness of ChatGPT. The New York Times.

[38] Wang, Q., Ma, Y., Zhao, K., & Tian, Y. (2020). A comprehensive survey of loss functions in machine learning. Annals of Data Science, 1-26.

[39] Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. Neurocomputing, 300, 70-79.

[40] Rençberoğlu, E. (2019). Fundamental techniques of feature engineering for machine learning. Towards Data Sci.

[41] E. Birmelé, Apprentissage supervisé, 2020.

[42] Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Tibshirani, R., & Friedman, J. (2009). Overview of supervised learning. The elements of statistical learning: Data mining, inference, and prediction, 9-41.

[43] Kassambara, A. (2017). Practical guide to cluster analysis in R: Unsupervised machine learning (Vol. 1). Sthda.

[44] D.E.S.D. Ees, V.L.E. Hay, W. Ackermann, B.L.E. Roux, mots cl e, Math. Sci. Hum. (2002) 13–45.

[45] Salem, B. (2021). Principal Component Analysis (PCA). La Tunisie Medicale, 99(4), 383-389.

[46] Lange, K., & Lange, K. (2010). Singular value decomposition. Numerical analysis for statisticians, 129-142.

[47] Barto, A. G. (1997). Reinforcement Learning. Neural Systems for Control, 7-30. https://doi.org/10.1016/B978-012526430-3/50003-9.

[48] Padakandla, S. (2021). A survey of reinforcement learning algorithms for dynamically varying environments. ACM Computing Surveys (CSUR), 54(6), 1-25.

[49] Li, Y. (2019). Reinforcement Learning Applications. ArXiv. /abs/1908.06973.

[50] Li, Song & Deng, Yu-Qin & Zhu, Zhi-Ling & Hua, Hong-Li & Tao, Ze-Zhang. (2021). A Comprehensive Review on Radiomics and Deep Learning for Nasopharyngeal Carcinoma Imaging. Diagnostics. 11. 1523. 10.3390/diagnostics11091523.

[51] Shinde, P. P., & Shah, S. (2018, August). A review of machine learning and deep learning applications. In 2018 Fourth international conference on computing communication control and automation (ICCUBEA) (pp. 1-6). IEEE.

[52] A. Bouakkaz, «Energy pilotage of habitat with multi-sources of energy,» University of Aout 20, 1955 Skikda, Skikda, 2021.

[53] Stojiljkovic, M. (2021). Linear regression in Python. Real Python. https://realpython.com/linear-regression-in-python/. Accessed, 8.

[54] Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. Ensemble machine learning: Methods and applications, 157-175.

[55] Steinwart, I., & Christmann, A. (2008). Support vector machines. Springer Science & Business Media.

[56] S. Guido et M. Andreas C, A GUIDE FOR DATA SCIENTISTS: Introduction to Machine Learning with Python, Sebastopol: O'Reilly Media, 2016.

[57] Steinwart, I., & Christmann, A. (2008). Support vector machines. Springer Science & Business Media.

[58] H. Simon, Neural Networks and Learning Machine, Canada: Pearson, 2009.

[59] Das, K., Jiang, J., & Rao, J. N. K. (2004). Mean squared error of empirical predictor.