PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
Ministry of Higher Education and Scientific Research University of Mohamed Khider – BISKRA
Faculty of Exact Sciences, Science of Nature and Life
Computer Science Department

Ordre N° : GLSD/M2/2023

Thesis

Submitted in fulfilment of the requirements for the Masters degree in

Computer science

Option : software engineering and distributed system.

AI technique for improving the Security Social Services

By :

MENNANE MOHAMED HAITHEM

Members of the jury :

Bennoui Hammadi	Professor	President
Bettira Roufaida	MCB	Encadrant
Hamida Ammar	MAA	Examinateur

Session 2023/2024

Acknowledgements

First and foremost, praise and thanks be to **ALLAH** the Most Gracious, the Most Merciful, and His strength that overwhelmed me throughout my research work to complete it successfully.

Special thanks first to my deceased father. My beloved mother for her great support, love, prayers, and care.

I would like to express my special thanks and gratitude to my professor **L. Kahloul** and **R. Bettira** who believed in me and gave me the opportunity to do this wonderful project **" AI technique to improve the Security Social Services "**, in addition to Mr. Maammar from CNAS Biskra, Dr. Tamin who helped me to conduct this research and learn many new things. I'm grateful to them. I extend my thanks to all the teachers who helped me in my academic life.

I would also like to mention the support and concern from my family that has always been in my life.

And special thanks to The person closest to my heart. Your presence around me had a huge impact on my life and pushed me forward to always give my best. Thank you.

MENNANE MOHAMED HAITHEM

dedicace

I dedicate this modest work to:

My dear family, all family **MENNANE**, To my deceased Father, My beloved Mother, My brothers Mahdi and Nadjib, my sister Malak, My dear aunt Fatiha, my uncle Adel, And the whole family

My teachers in the computer science department,

My friends: Omar, Abdou, Mondher, Saif, Yousef, Iheb, Messaoud, Abdelhakim, Lahcen,

To everyone who ever supported me in my life

To The person closest to my heart. Your presence around me had a huge impact on my life and pushed me forward to always give my best. Thank you,

Last but not least, I wanna thank me,

I wanna thank me for believing in me, I wanna thank me for doing all this hard work,

Dedicated to my deceased father, My dear father, today I am an engineer as you wanted.

My beloved mother for her great support, love, prayers, and care.

Abstract

Social security institutions play an indispensable role in the protection and sustenance of individuals and families, delivering essential benefits and services to beneficiaries. In our work, we endeavor to explore the realm of predictive analytics within the domain of social security. Through a comprehensive exploration of AI's applicability in social security contexts and the subsequent evaluation of machine learning algorithms and time series, we seek to illuminate pathways for fortifying social security mechanisms with predictive capabilities. By integrating predictive analytics into social security frameworks, we aspire to advance the overarching goals of inclusive and equitable social protection, thereby fostering resilience and well-being within communities. The model is trained and tested using a public United Kingdom Employers Social Security contribution with " Office for National Statistics (ONS)" dataset[1]

Finally, results demonstrate the algorithms proposed to the prediction with good evaluation in the domain of social security.

Keywords: *Social security, Artificial intelegence(AI) ,Machine learning ,time series ,predictive ,Employers Contributions*

[1] Office for National Statistics(ONS).https://www.ons.gov.uk

Résumé

Les institutions de sécurité sociale jouent un rôle indispensable dans la protection et la subsistance des individus et des familles, en fournissant des prestations et des services essentiels aux bénéficiaires. Dans notre travail, nous nous efforçons d'explorer le domaine de l'analyse prédictive dans le domaine de la sécurité sociale. Grâce à une exploration complète de l'applicabilité de l'IA dans les contextes de sécurité sociale et à l'évaluation ultérieure des algorithmes d'apprentissage automatique et des séries chronologiques, nous cherchons à éclairer les voies permettant de renforcer les mécanismes de sécurité sociale avec des capacités prédictives. En intégrant l'analyse prédictive dans les cadres de sécurité sociale, nous aspirons à faire progresser les objectifs primordiaux d'une protection sociale inclusive et équitable, favorisant ainsi la résilience et le bien-être au sein des communautés. Le modèle est entrainé et testé à l'aide du dataset public sur les cotisations des employeurs à la sécurité sociale par " Office for National Statistics (ONS) " [1].

Enfin, les résultats démontrent les algorithmes proposés pour la prédiction avec une bonne évaluation dans le domaine de la sécurité sociale.

Mots clés : *sécurité sociale,IA , Prétraitement, apprentissage automatique,séries chronologiques, prédictives, les cotisations des employeurs*.

[1] Office for National Statistics (ONS).https://www.ons.gov.uk

Résumé Arabe

ملخص

يعتبر في وقتنا الحالي تبني التقنيات الحديثة مثل الذكاء الاصطناعي أمرًا بالغ الأهمية لمواجهة التحديات المتزايدة والتعقيدات التي تواجه نظم الضمان الاجتماعي. تعزز هذه التقنيات من كفاءة وفعالية العمليات والخدمات المقدمة، مما يسهم في تحقيق الأهداف الشاملة للحماية الاجتماعية

تلعب مؤسسات الضمان الاجتماعي دورًا لا غنى عنه في حماية وإعالة الأفراد والعائلات، حيث تقدم فوائد وخدمات أساسية للمستفيدين. في عملنا هذا، نسعى لاستكشاف مجال التحليلات التنبؤية في نطاق الضمان الاجتماعي. من خلال استكشاف شامل لتطبيقات الذكاء الاصطناعي في سياقات الضمان الاجتماعي وتقييم خوارزميات التعلم الآلي والسلاسل الزمنية، نسعى لتسليط الضوء على السبل التي يمكن من خلالها تعزيز آليات الضمان الاجتماعي بقدرات تنبؤية. عن طريق دمج التحليلات التنبؤية في أطر الضمان الاجتماعي، نطمح إلى تحقيق الأهداف الشاملة للحماية الاجتماعية الشاملة والعادلة، وبالتالي تعزيز المرونة والرفاهية داخل المجتمعات. تم تدريب النموذج واختباره باستخدام بيانات مساهمات أرباب العمل في الضمان الاجتماعي في المملكة المتحدة، المتاحة من (ONS). مكتب الإحصاءات الوطنية

الكلمات المفتاحية

الذكاء الاصطناعي. الضمان الاجتماعي. التحليلات التنبؤية. التعلم الآلي . السلاسل الزمنية. مساهمات أرباب العمل

List of Figures

List of Tables

Contents

Chapter 1

General Introduction

1.1 Context

A fundamental component of contemporary welfare states, social security embodies the idea that everyone must safeguard people and families from a range of social and economic hazards while also ensuring their well-being. Social security systems, which have their roots in the ideals of social justice and solidarity, are designed to act as a safety net for people, protecting them against poverty, insecurity, and hardship, especially in vulnerable situations like old age, disease, or disability. The development of labor movements that fought for social protections and workers' rights began to take shape in the late 19th and early 20th centuries, which is when social security first emerged. Formalized social security systems began with significant events like Chancellor Otto von Bismarck's creation of social insurance plans in Germany in the late 19th century. The idea of social security spread throughout the world over time, incorporating a variety of benefits and programs designed to meet various socioeconomic requirements and realities. In Algeria, the Algerian National Social Security Fund (CNAS) and the National Social Security Fund for Self-Employed Workers (CASNOS) stand as pivotal institutions tasked with administering social security benefits and ensuring the welfare of citizens across various life stages and circumstances.[21].

The potential for artificial intelligence (AI) to revolutionize social security systems, such as CNAS is enormous. AI can improve service delivery, lower fraud and error rates, and increase operational efficiency. AI technologies that can analyze enormous volumes of data to find patterns, spot abnormalities, and make defensible conclusions include machine learning, natural language processing, and predictive analytics. AI may have a significant effect on CNAS. AI frees up human resources to work on more important and complicated tasks by automating repetitive operations. Insights about beneficiary needs, optimal resource allocation, and improved fraud detection capabilities can all be obtained through AI-powered analytics. A strong structure and procedures are needed for the ethical, transparent, and successful application of AI in CNAS

1.2 Social Security domain in Algeria

The social security domain in Algeria is centered around the Caisse Nationale des Assurances Sociales (CNAS), which provides healthcare coverage, maternity benefits, work-related accident support, and occupational disease coverage. The social security system is closely linked to the healthcare sector, with CNAS and other funds like CASNOS providing financial support for medical expenses, screening, surgery, and rehabilitation [9]. The government has developed initiatives to improve healthcare service delivery, reduce costs, and combat fraud within the social security domain, such as introducing a smart card-based national healthcare system [1][11]

1.2.1 Overview of CNAS

CNAS (Caisse Nationale des Assurances Sociales) in Algeria is a vital social security institution providing healthcare coverage and social protection services to around 80% of the Algerian population. It operates within a centralized healthcare system under the Ministre du Travail et de la Sécurité Sociale, aiming to enhance healthcare services while reducing costs and fraud. CNAS plays a crucial role in ensuring the health and well-being of Algerian citizens by offering financial support and protection [11][1].

1.2.2 Fields of CNAS

The following are the main areas of operation for Algeria's CNAS (Caisse Nationale des Assurances Sociales).

- **Healthcare coverage :** approximately 80% of Algerians are either directly or indirectly covered by CNAS's healthcare system. For patients who are residents or not, it pays for screening, surgery, medical costs, and rehabilitation.

- **Maternity benefits :** as part of its social protection services, CNAS provides Algerian citizens with maternity benefits.
- **Support for work-related accidents :** as part of its social security coverage, CNAS offers assistance for work-related injuries.
- **Coverage for occupational diseases :** CNAS includes coverage for occupational diseases in its social protection program.
- **Retirement benefits :** as part of its social security program, CNAS provides retirement benefits to the people of Algeria.
- **Fighting fraud :** by implementing programs like the establishment of a national healthcare system based on smart cards, CNAS seeks to lower operating expenses and application fraud in the healthcare system.

1.3 Motivations

The main problems facing the social security system in Algeria include [1][9][19].

- **Rapidly growing health spending:** over the past two decades, health spending in Algeria has grown significantly due to factors like epidemiological and demographic transition. This has led to a constantly growing bill for medication use, which is supported by social security agencies like CNAS.
- **Inequality in healthcare provision:** there are significant disparities in the geographic distribution of private physicians in Algeria. They tend to settle much more in the northern regions and larger metropolises compared to the rest of the country.
- **Centralized management of health institutions:** all public health facilities in Algeria are funded by central health authorities, which control the country's healthcare system. Free health services and the domination of central authority are closely related, even if the statute blends central and decentralized control. This reduces the opportunities for neighborhood health facilities to take part in the creation of health policies.

- **Healthcare Service Quality:** CNAS offers its beneficiaries access to healthcare services; however, there may be differences in the standard and availability of these services in Biskra. Obstacles like inadequate healthcare facilities and a lack of equipment

- **Lack of predictive analytics:** the absence of predictive analytics hinders the ability to forecast trends, anticipate challenges, and optimize resource allocation effectively within the social security system. Implementing predictive analytics could enhance decision-making processes and better address the evolving needs of the population.

1.4 Objectives of the study

In this work, we have the following objectives:

- Exploring the domain of Social Security in the world, in Algeria, and exactly in the CNAS of Biskra.

- Exploring AI.

- Application of AI algorithms for Social Security.

- Evaluation of the different algorithms.

1.5 Plan of the manuscript

The manuscript is organized as follows:

Chapter 2: State of the art.

That presents the technical background of this work which is machine learning and time series that is a part of AI in particular regression and prediction algorithms

Chapter 3: Machine Learning Algorithms for Data Prediction in Social Security

Explains the data and the algorithms for data predictions and compared the result of each algorithm

Chapter 4: Frameworks, tools, and libraries .

mention tools and frameworks that help us to implement our code.

Chapter 5: Discussion and conclusion.

concludes the main goal of this paper and it gives a point of view for future work.

Chapter 2

State of the art

2.1 Introduction

The convergence of artificial intelligence (AI) and social security marks a paradigm shift in the evolution of the global welfare state. This section looks at how the management, delivery, and sustainability of social security services could be drastically altered by artificial intelligence (AI) technology.
In this chapter, we will present the definition of social security, explain machine learning, and describe the algorithms of predictions.

2.2 Overview of Social Security and its significance

2.2.1 Definition of Social Security

Social Security is a system of government programs designed to provide the population with a basic standard of living, particularly in times of need. It typically includes disability benefits, health insurance, and family allowances. In a democratic society, the interests of social security of the individual and the state as a whole must be reconciled and harmoniously balanced. Social security plays a key role in the entire national security system, given the socialization of all spheres of human life. The level of pension provision of the population is a guarantee of social security of the person[6].

2.2.2 Social Security insurance types in Algeria

In Algeria, social security insurance is provided through various institutions, each responsible for different sectors of the population and types of coverage. Here are some of the key social security assurance institutions in Algeria.

- **Caisse Nationale d'Assurance Sociale (CNAS):** provides various social security services such as healthcare coverage, disability, maternity, and work-related injury benefits to eligible individuals. Figure 2.1 shows all the subs-directorate of CNAS-Biskra.

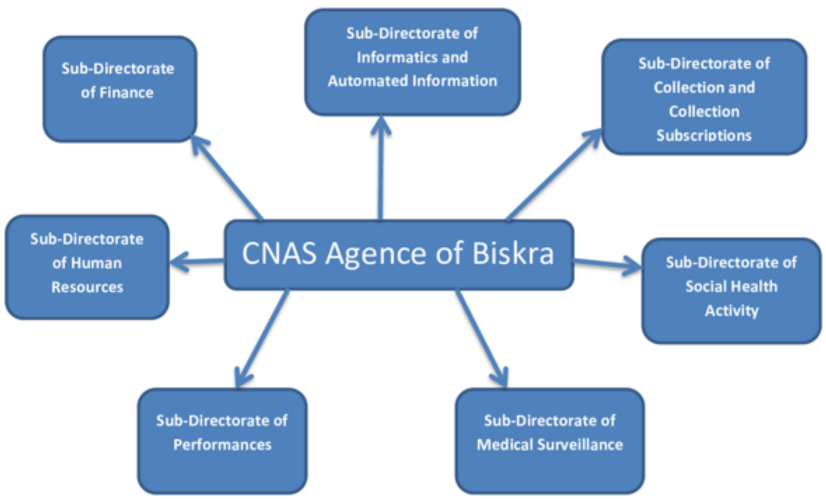

Figure 2.1: CNAS-Biskra

- **Caisse Nationale des Retraites (CNR):** CNR is the National Retirement Fund for Salaried Workers, responsible for managing retirement pensions for employees in the public sector. It provides pension benefits to civil servants, public sector employees, and other government workers who contribute to the system.

- **Caisse Nationale d'Assurance Chômage (CNAC):** CNAC is the National Unemployment Insurance Fund, which provides unemployment benefits to eligible individuals who have lost their jobs involuntarily. It offers financial assistance and support services to help unemployed individuals find new employment opportunities.

- **Caisse Nationale de Sécurité Sociale des Non-Salariés (CASNOS):** CASNOS is the National Social Security Fund for Non-Salaried Workers, responsible for providing social security coverage to self-employed individuals, artisans, traders, and professionals who are not covered by CNAS. It offers healthcare coverage, maternity benefits, and retirement pensions to eligible contributors.

- **Office National des Pensions (ONP):** ONP is the National Office for Pensions of Algerians Abroad, which administers pension benefits for Algerian citizens living outside the national territory. It provides retirement pensions and survivor benefits to eligible Algerian expatriates.

  These institutions play crucial roles in ensuring social security coverage for different segments of the Algerian population, including employees, self-employed individuals, retirees, and expatriates. They offer various benefits and services aimed at providing economic and social protection to their members and beneficiaries.

2.2.3 The difference between CNAS and CASNOS

The key differences between CNAS (Caisse Nationale des Assurances Sociales) and CASNOS (Caisse Nationale de Sécurité Sociale des Non-Salariés) in Algeria are [1] [11].

- **Coverage :**

  CNAS provides healthcare coverage and social protection services to around 80% of the Algerian population, either directly or indirectly.

  CASNOS covers non-salaried workers and their families.

- **Funding :**

  CNAS is funded through contributions from employers and employees, as well as government subsidies.

  CASNOS is funded through contributions from non-salaried workers and government subsidies.

- **Benefits :**

  both CNAS and CASNOS provide benefits such as healthcare coverage, maternity benefits, work-related accident support, occupational disease coverage, and retirement benefits.

- **Management :**

  CNAS operates within a centralized healthcare system under the Ministre du Travail et de la Sécurité Sociale (Ministry of Labor and Social Security).

  The management structure and governance of CASNOS may differ from CNAS.

- **Relationship with the healthcare system :**

  CNAS is closely linked to the healthcare sector, providing financial support for medical expenses, screening, surgery, and rehabilitation.

  The key differences between CNAS and CASNOS in Algeria are their coverage, funding sources, specific benefits, management structure, and relationship with the healthcare system.

2.3 Artificial intelligence and Machine learning

Artificial Intelligence (AI) :

AI refers to the development of computer systems capable of performing tasks that typically require human intelligence. These tasks include learning from experience, reasoning, understanding natural language, recognizing patterns, and making decisions in complex situations [16].

Machine Learning (ML) :

Machine learning is a field of artificial intelligence that enables systems to learn and improve from experience without being programmed. It involves developing algorithms and statistical models that allow computers to perform specific tasks effectively by utilizing data [15].

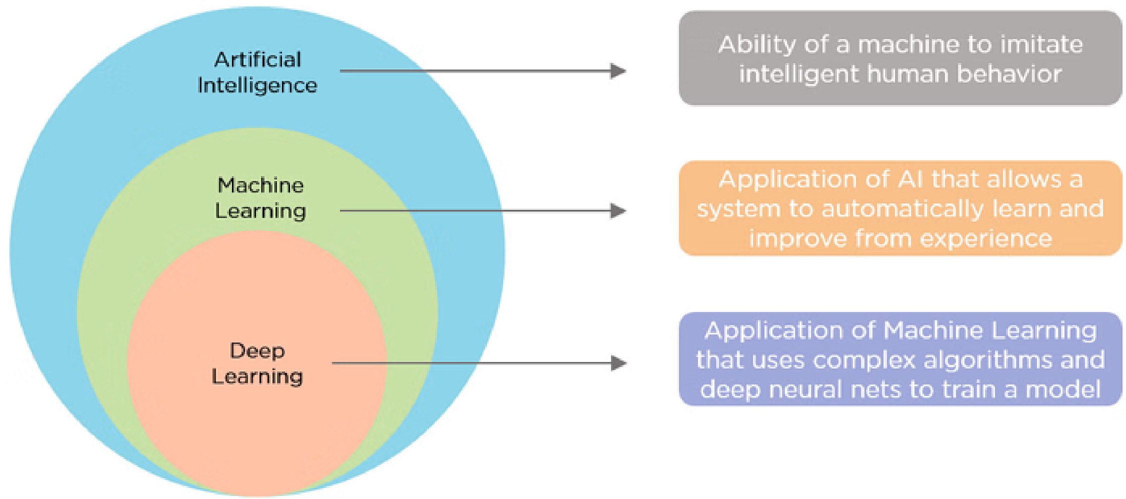

Figure 2.2: AI and Machine learning

Types of Learning:

In machine learning, there are four types of learning approaches, each with its unique characteristics and applications. Here are the main types of learning in machine learning.

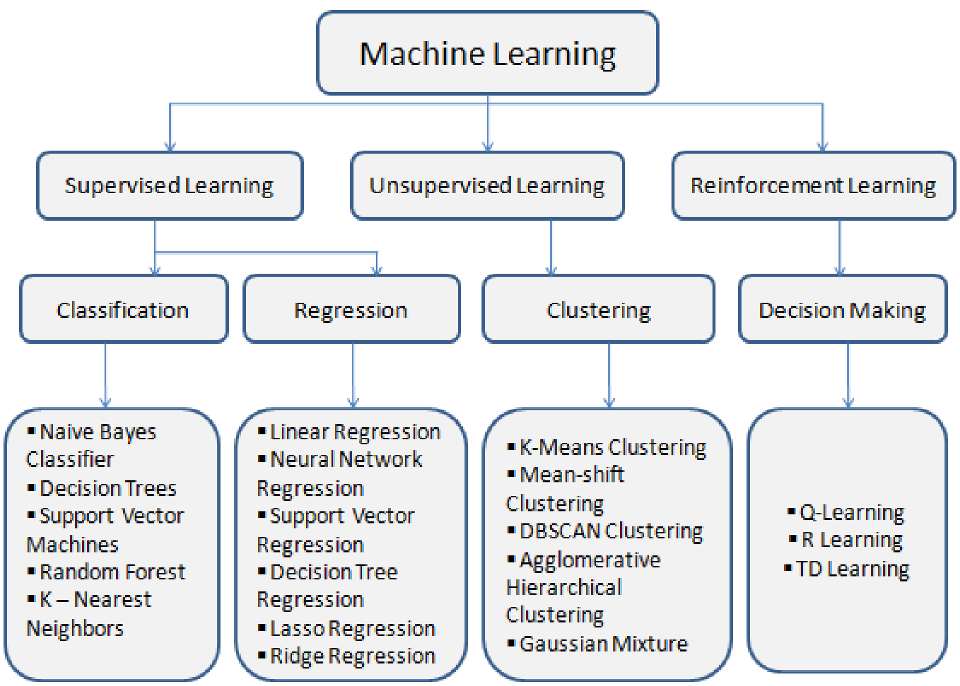

Figure 2.3: Types of learning

- **Supervised Learning :** in supervised learning, the algorithm learns from labeled data, where each input is associated with a corresponding target output. The goal is to learn a mapping from input to output by minimizing the difference between predicted and actual outputs. Common supervised learning tasks include classification and regression [2]. Examples of this type of algorithm are linear regression and support vector Regression.Figure**??** shows an example of supervised learning in ML.

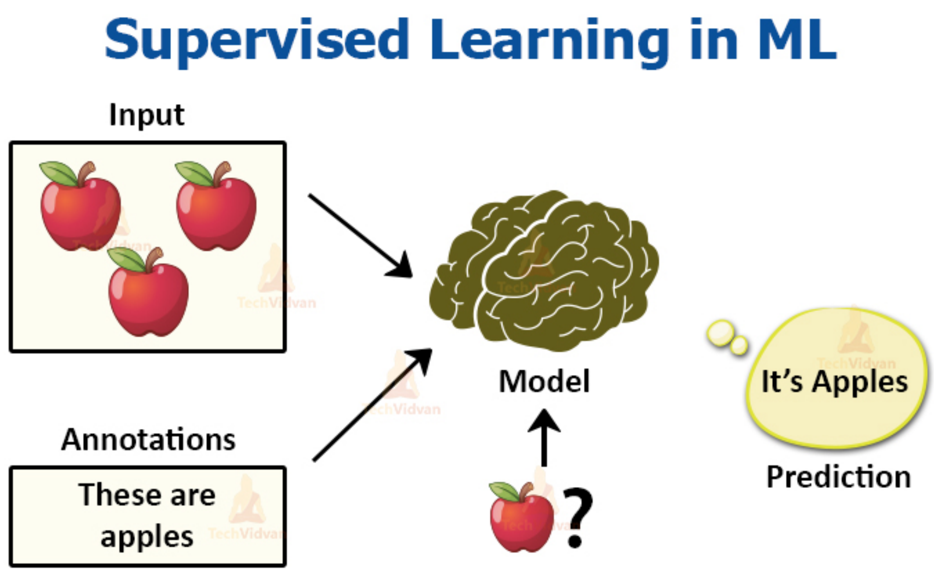

Figure 2.4: Supervised Learning in ML.

In supervised learning, there are two tasks that we can achieve: Regression and classification.

  - **Regression :** involve predicting continuous numerical values based on input features, aiming to model the relationship between the independent variables and the dependent variable for example linear regression :

$$Y = \beta_0 + \beta_1 X + \epsilon \tag{2.1}$$

where:

Y is the dependent variable (the variable we want to predict),

X is the independent variable (the feature),

β_0 is the intercept (the value of Y when $X = 0$),

β_1 is the slope (the change in Y for a one-unit change in X),

ϵ is the error term (the difference between the predicted Y and the actual Y).

– **Classification :** categorizes data into predefined classes based on features, facilitating pattern recognition and prediction tasks. Figure 2.5 shows the difference between the classification and regression tasks.

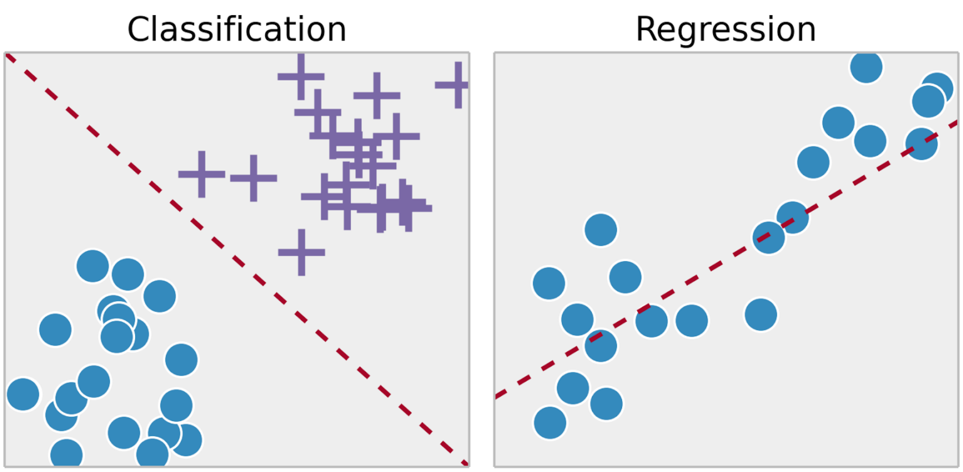

Figure 2.5: Classification and Regression

- **Unsupervised Learning :** unsupervised learning involves learning patterns and structures from unlabeled data. The algorithm explores the data to find inherent structures or relationships without explicit guidance. Common unsupervised learning tasks include clustering, dimensionality reduction, and anomaly detection[13].Figure**??** shows an example of Unsupervised learning in ML.

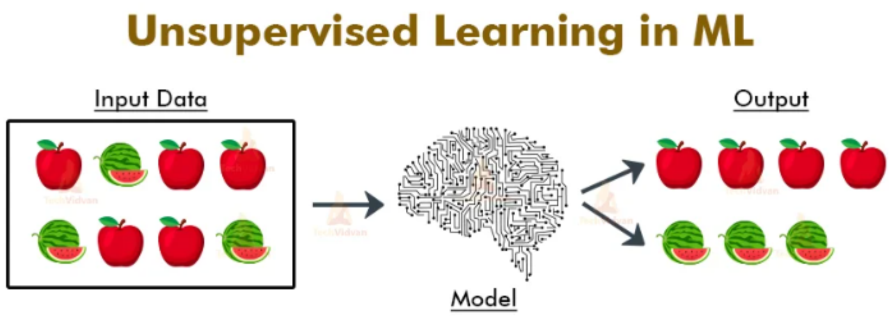

Figure 2.6: Unsupervised Learning in ML.

- **Semi-Supervised Learning :** semi-supervised learning combines elements of supervised and unsupervised learning. It leverages both labeled and unlabeled data to improve the learning process. Semi-supervised learning is particularly useful when labeled data is scarce or expensive to obtain[22].

- **Reinforcement Learning :** reinforcement learning involves an agent learning to interact with an environment by taking actions to maximize cumulative rewards. The agent learns through trial and error, receiving feedback from the environment based on its actions. Reinforcement learning is used in applications such as gaming, robotics, and autonomous systems[18].Figure**??** shows an example of Reinforcement learning in ML.

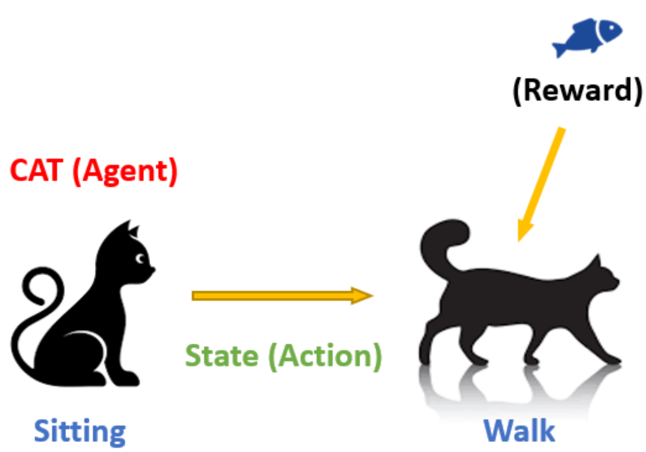

Figure 2.7: Reinforcement Learning in ML.

2.4 Time Series

2.4.1 Definition :

A Time Series is a sequence of data points collected over time, typically at regular intervals. It is a series of observations ordered in time, where each observation is a realization of a random variable at a specific time point.
The general logical equation for a time series can be represented as:

$$y_t = f(t) + \epsilon_t \tag{2.2}$$

Where:

- y_t represents the observed value of the variable at time t,
- $f(t)$ denotes the underlying function or trend capturing the systematic component of the data over time,
- ϵ_t is the random error term or noise representing the unpredictable fluctuations in the data at time t.

2.4.2 Types of Time Series

- **Stationary Time Series**: A time series is stationary if its statistical properties, such as mean and variance, do not change over time.
- **Non-Stationary Time Series**: A time series is non-stationary if its statistical properties change over time. This includes trends, seasonality, and other time-dependent structures.
- **Functional Time Series**: Time series data where each observation is a function or curve, rather than a scalar value.
- **Time Series Driven by Differential Equations**: Time series data generated by a hidden dynamic process controlled by underlying differential equations with unknown parameters.

2.4.3 Applications of Time Series

- **Forecasting:** Predicting future values of a variable based on historical data to anticipate trends and make informed decisions[10].
- **Anomaly Detection:** Identifying unusual or unexpected patterns in the data that deviate from the norm, indicating potential anomalies or outliers[5].
- **Pattern Recognition:** Identifying recurring patterns, trends, and dependencies in the data to gain insights into underlying processes or phenomena[13].

2.4.4 The difference between Time-series forecasting and Machine-Learning predictions

The difference between Time-series forecasting and Machine-Learning predictions lies in their approach and application. [4] [12]

Time-series forecasting :

- Time-series forecasting focuses on predicting future values based on historical data points collected at regular intervals over time
- It involves analyzing patterns, trends, and seasonality within the time series data to make predictions about future values.
- Time-series forecasting methods include ARIMA (AutoRegressive Integrated Moving Average), Exponential Smoothing, and Prophet, which are specifically designed for analyzing time-dependent data.

Machine-Learning predictions :

- Machine-learning predictions utilize algorithms that learn patterns and relationships within data to make predictions, not necessarily restricted to time-dependent data.

- Machine-learning models can handle complex relationships and non-linear patterns in the data, making them versatile for various prediction tasks.
- While time-series forecasting is a specific subset of predictive modeling tailored for time-dependent data, machine learning encompasses a broader range of algorithms and techniques for prediction tasks across different types of data.

2.4.5 ARIMA

Time series analysis is a fundamental tool for understanding and predicting sequential data trends, vital across various domains. Among the plethora of methods, the Autoregressive Integrated Moving Average (ARIMA) model stands out for its effectiveness in capturing temporal patterns and making accurate forecasts. In this section, we delve into the essence of ARIMA, elucidating its components and applicability in real-world scenarios.

Definition :

ARIMA, or Autoregressive Integrated Moving Average, is a popular time series analysis technique used to model and forecast time-dependent data. It combines autoregressive (AR), differencing (I), and moving average (MA) components to capture the temporal patterns present in the data [3].
The general equation for ARIMA(p, d, q) model is:

$$Y_t = c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t \tag{2.3}$$

where:

- Y_t: is the observed value at time t,
- c: is a constant term,
- $\phi_1, \ldots, \phi_p$: are autoregressive parameters,
- $\theta_1, \ldots, \theta_q$: are moving average parameters,
- ϵ_t: is the white noise error term,

- p: is the order of the autoregressive component,
- d: is the degree of differencing,
- q: is the order of the moving average component.

Application:

ARIMA models are widely used in various fields for time series forecasting and analysis. Some common applications include economic forecasting, financial markets analysis, demand forecasting, weather forecasting, and process control.

2.5 Related work

Below is a summary of current research on artificial intelligence (AI) applications in social security insurance, encompassing a range of topics including risk assessment, fraud detection, improving customer service, and providing decision support systems for policymakers :

- in [17], John Smith Provides an overview of how machine learning techniques, including AI algorithms, are applied in social security systems for risk assessment and prediction. It explores the use of supervised learning models to analyze historical data and predict future trends in benefit claims, enrollment rates, and fund solvency. The study also discusses the challenges and opportunities associated with implementing machine learning in social security insurance.
- in [7], Jane Doe investigates the application of AI-based fraud detection techniques in social security insurance programs. It examines how machine learning algorithms, such as anomaly detection and predictive modeling, can be used to identify suspicious patterns and fraudulent activities in benefit claims, disability assessments, and other insurance processes. The study evaluates the effectiveness of AI-driven fraud detection systems in reducing financial losses and improving program integrity.
- in [8], Emily Johnson explores the use of AI chatbots to enhance customer service in social security insurance agencies. It discusses how natural language processing (NLP) and conver-

sational AI technologies can be leveraged to automate routine inquiries, provide personalized assistance to beneficiaries, and streamline communication channels between users and administrators. The study evaluates the impact of AI chatbots on customer satisfaction, response times, and service accessibility.

- in [20], Michael Williams investigates the development of AI-driven decision support systems for policymakers in the social security domain. It examines how machine learning algorithms, data analytics, and simulation modeling techniques can be integrated into policy analysis frameworks to evaluate the impact of proposed reforms, assess program effectiveness, and forecast long-term trends. The study discusses the potential benefits and challenges of using AI-based decision support systems in informing policy decisions and shaping social security strategies.

Table 2.1: Related Work on AI Integration in Social Security

Authors	**Title and Idea**
John Smith et al.	Application of Machine Learning Techniques in Social Security Systems,Provides an overview of how machine learning techniques, including AI algorithms, are applied in social security systems for risk assessment and prediction. It explores the use of supervised learning models to analyze historical data and predict future trends in benefit claims, enrollment rates, and fund solvency. The study also discusses the challenges and opportunities associated with implementing machine learning in social security insurance.
Jane Doe	AI-based Fraud Detection in Social Security Insurance, investigates the application of AI-based fraud detection techniques in social security insurance programs. It examines how machine learning algorithms, such as anomaly detection and predictive modeling, can be used to identify suspicious patterns and fraudulent activities in benefit claims, disability assessments, and other insurance processes. The study evaluates the effectiveness of AI-driven fraud detection systems in reducing financial losses and improving program integrity.
Emily Johnson	Improving Customer Service in Social Security Insurance Through AI Chatbots, explores the use of AI chatbots to enhance customer service in social security insurance agencies. It discusses how natural language processing (NLP) and conversational AI technologies can be leveraged to automate routine inquiries, provide personalized assistance to beneficiaries, and streamline communication channels between users and administrators. The study evaluates the impact of AI chatbots on customer satisfaction, response times, and service accessibility.
Michael Williams	AI-driven Decision Support Systems for Social Security Policy Analysis, investigates the development of AI-driven decision support systems for policymakers in the social security domain. It examines how machine learning algorithms, data analytics, and simulation modeling techniques can be integrated into policy analysis frameworks to evaluate the impact of proposed reforms, assess program effectiveness, and forecast long-term trends. The study discusses the potential benefits and challenges of using AI-based decision support systems in informing policy decisions and shaping social security strategies.

2.6 Conclusion

In this chapter, we have explained some knowledge which are important as background in our work. We have described the different use cases of AI. After that, we have detailed the basics of machine learning and time series. Then, we have explained the difference between time series and machine learning in prediction. Finally, we have mentioned some previous related works.

Chapter 3

Machine Learning Algorithms for data Prediction in Social Security

3.1 Introduction

In Chapter 3, we delve into the contributions of machine learning algorithms in data prediction for social security insurance. Through machine learning, we explore innovative approaches to forecast beneficiary needs, mitigate fraud risks, and enhance service delivery within social security systems. This chapter highlights the transformative potential of machine learning in optimizing resource allocation, improving program efficiency, and ensuring the sustainability of social security initiatives.

3.2 Dataset

A dataset is a collection of data points or observations typically organized in a structured format for analysis.

In the first plan, we use the local data set of CNAS biskra by using the KNN algorithm, SVR, and linear regression, after that, we ran into problems in the local dataset.

In the second plan, we use the public data set of Employer's Social Security contributions in the United Kingdom (UK) by using ARIMA, SVR, and Linear Regression.

3.2.1 Local Dataset

The Local Dataset from RMC (Registre des Maladies Chroniques)of CNAS Biskra. This data set presents quarterly data on the number of patients 80% and 100% for the period 2015 to Quarter 4 (Sep to Dec) 2023.

Initially we used the Algorithms of regression like KNN and Linear Regression, SVR, DT

- **Components :** its components are a series of times quarters and numbers of patients of 80% and 100%
- **Data format :** below is the format of data.CSV

DATE;Malades80%		DATE;Malades 100%	
2015/3;708		2015/3;1255	
2015/6;675		2015/6;1268	
2015/9;409		2015/9;872	
2015/12;693		2015/12;1180	
2016/3;767		2016/3;1203	
2016/6;658		2016/6;1097	
2016/9;391		2016/9;800	
2016/12;670		2016/12;1069	
2017/3;734		2017/3;1089	
2017/6;635		2017/6;991	
2017/9;461		2017/9;912	
2017/12;669		2017/12;1184	
2018/3;712	2021/3;790	2018/3;1156	2021/3;1640
2018/6;663	2021/6;491	2018/6;1145	2021/6;1375
2018/9;599	2021/9;426	2018/9;950	2021/9;1091
2018/12;1026	2021/12;811	2018/12;1379	2021/12;1467
2019/3;1068	2022/3;1028	2019/3;1380	2022/3;1736
2019/6;887	2022/6;823	2019/6;1259	2022/6;1520
2019/9;712	2022/9;733	2019/9;1132	2022/9;1569
2019/12;1024	2022/12;1039	2019/12;1436	2022/12;1753
2020/3;870	2023/3;1167	2020/3;1169	2023/3;1981
2020/6;477	2023/6;1001	2020/6;613	2023/6;1747
2020/9;536	2023/9;890	2020/9;945	2023/9;1480
2020/12;968	2023/12;1192	2020/12;1559	2023/12;2020

Figure 3.1: Local Data set.

3.2.2 Problems of Local Dataset

There are several problems, but the big problem is:

- A small data set consists of only 36 quarters, due to insufficient information being available before 2015, which leads to inconsistency of values between reality and prediction, so the data set must be large. For training and testing.

  These problems are due to the internal laws of CNAS Biskra, It is not possible to provide more information in this area

Below is the application of Knn algorithm for the local data set (malades100%)

```
R-squared: 0.798209054566799
RMSE: 143.78695177240527
Predictions : [1261.5 1070.  1026.  1191.5 1191.5 1001.5  934.5  934.5 1030.   951.5
  951.5 1034.  1150.5 1150.5 1262.  1262.  1255.5 1255.5 1284.  1302.5
 1302.5  779.   779.  1292.5 1507.5 1233.  1279.  1601.5 1601.5 1628.
 1544.5 1775.  1864.  1864.  1750.  1750. ]
```

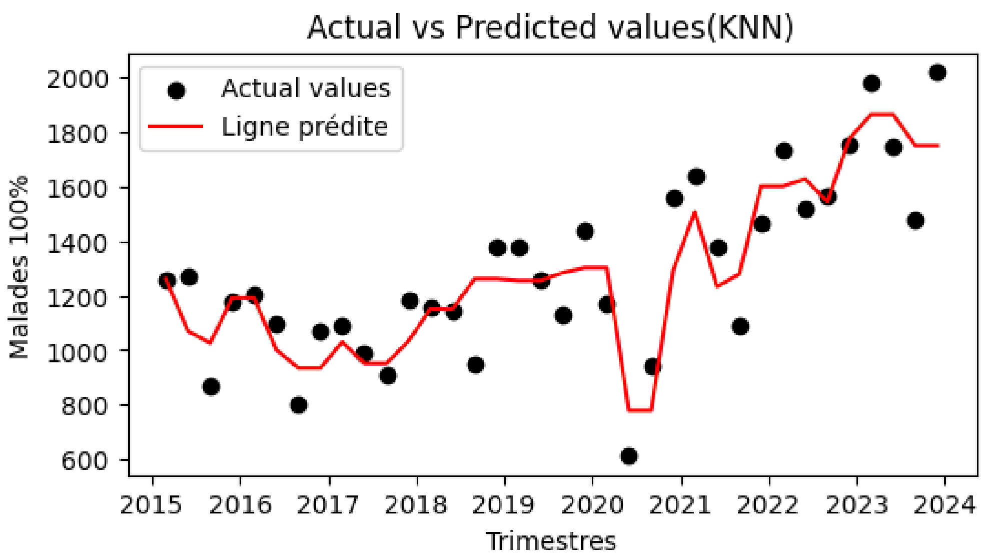

Figure 3.2: knn algorithm to the local data set.

3.2.3 Public Dataset

After the problems of the local Dataset, we use a public data set from a site of the Office for National Statistics of the United Kingdom (UK).

This data set presents quarterly data on Employer's Social Security contributions for the period 1946 to Quarter 1 (Jan to Mar) 2018.

- **Link :** this is the link of the public Data set [14]
- **Components :** its components are a series of times quarters and numbers of contributions
- **Data format :** below is the format of data.CSV

DATE;contributions	1952/3;54	2005/12;11603	2012/3;16310
1946/3;21	1952/6;54	2006/3;12897	2012/6;14918
1946/6;21	1952/9;55	2006/6;12028	2012/9;14680
1946/9;21	1952/12;55	2006/9;12138	2012/12;14692
1946/12;21	1953/3;61	2006/12;12505	2013/3;16573
1947/3;28	1953/6;61	2007/3;13740	2013/6;15502
1947/6;28	1953/9;61	2007/6;13110	2013/9;14788
1947/9;28	1953/12;61	2007/9;13238	2013/12;15156
1947/12;29	1954/3;61	2007/12;13677	2014/3;17384
1948/3;39	1954/6;62	2008/3;16046	2014/6;15466
1948/6;39	1954/9;62	2008/6;13797	2014/9;15381
1948/9;39	1954/12;62	2008/9;13587	2014/12;15661
1948/12;40	1955/3;62	2008/12;13650	2015/3;18032
1949/3;49	1955/6;67	2009/3;14770	2015/6;16296
1949/6;49	1955/9;75	2009/6;13258	2015/9;15944
1949/9;49	1955/12;75	2009/9;12943	2015/12;16219
1949/12;50	1956/3;75	2009/12;13440	2016/3;17928
1950/3;49	1956/6;77	2010/3;15989	2016/6;17776
1950/6;50	1956/9;75	2010/6;13300	2016/9;17622
1950/9;50	1956/12;77	2010/9;13143	2016/12;18081
1950/12;50	1957/3;76	2010/12;13455	2017/3;20592
1951/3;51	1957/6;76	2011/3;16106	2017/6;19045
1951/6;51	1957/9;80	2011/6;14056	2017/9;18364
1951/9;51	1957/12;77	2011/9;13934	2017/12;18965
1951/12;52	1958/3;91	2011/12;14078	2018/3;21615

Figure 3.3: Public Data set.

3.3 Selection of Machine Learning and Time Series Algorithms

The choice of suitable machine learning algorithms is critical when it comes to predictive modeling because it affects the predictive models' interpretability, accuracy, and generalizability.

The approach we described hereafter is composed of two steps: prediction with linear regression or forecast with ARIMA. after the collection and processing of the Dataset, each algorithm has steps for prediction :

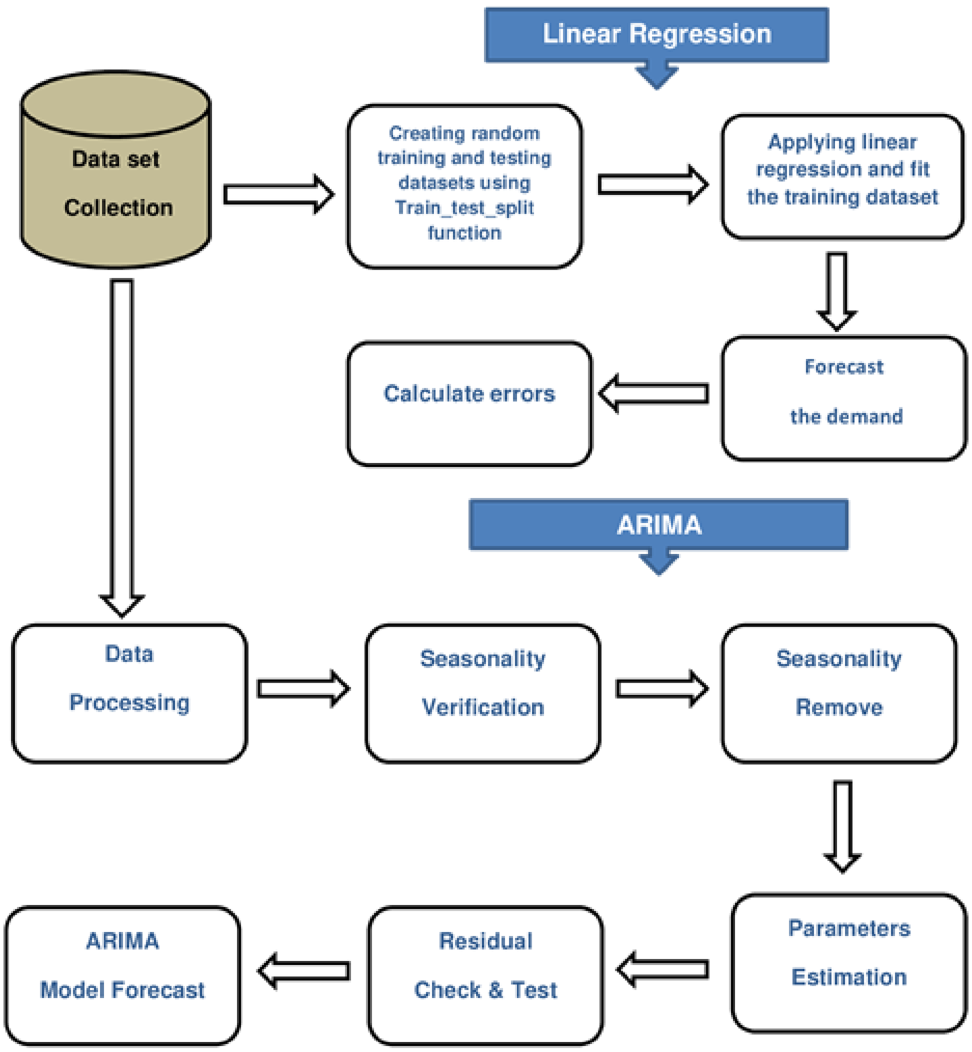

Figure 3.4: System design.

3.3.1 Linear Regression

Linear Regression is a simple yet powerful statistical method used for modeling the relationship between a dependent variable and one or more independent variables. In the context of our study on social security predictions, Linear Regression is employed to analyze the historical data of social security contributions. The algorithm aims to identify and quantify the linear relationship between the predictor variables, and the target variable, which represents the contributions made to the social security system. By fitting a linear equation to the observed data points, Linear Regression facilitates the prediction of future contributions based on the established trend. Additionally, performance metrics such as R-squared, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) are utilized to evaluate the accuracy and reliability of the model predictions.

- Application of linear Regression algorithm for the public data set of contributions in social security :

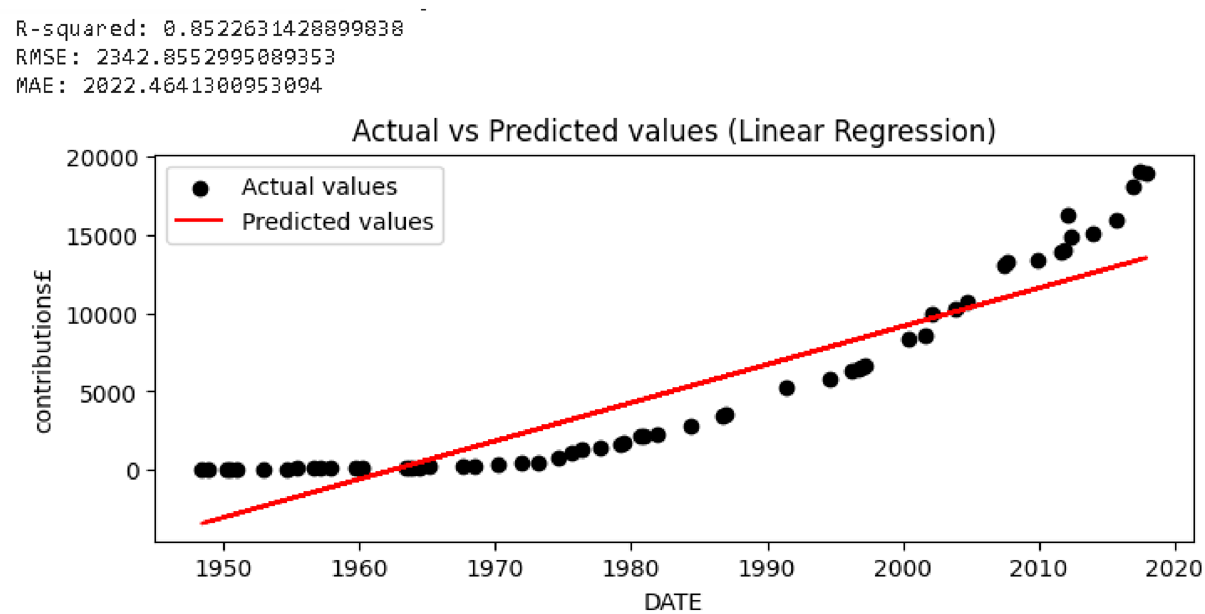

Figure 3.5: Linear Regression

- an R-squared value of 0.852 suggests that approximately 85.2% of the variability in the contributions to the social security system can be explained by the linear relationship with the predictor variables.

- an RMSE value of 2342.86 indicates that, on average, the predictions made by the Linear Regression model are approximately 2342.86 units away from the actual contributions.

- an MAE value of 2022.46 suggests that, on average, the predictions made by the model are approximately 2022.46 units away from the actual contributions.

These metrics collectively provide insights into the performance of the Linear Regression model in predicting social security contributions. A higher R-squared value and lower RMSE and MAE values indicate better predictive performance and model accuracy.

3.3.2 SVR

Support Vector Regression (SVR) is a type of machine learning algorithm used for regression tasks. It is a supervised learning method that aims to predict a continuous output variable based on a set of input features.

In the context of social security, SVR has been used to predict the contributions of social security. The SVR model is trained on historical data and then used to predict future contributions, This approach aims to improve the accuracy of predictions and provide a more robust method for nowcasting employment subject to social security contributions.

Support Vector Regression (SVR) algorithm offers different types of kernels to model various types of data and relationships:

- **Linear Kernel :** suitable for linearly separable data or when the number of features is large. It's computationally efficient and works well when the relationship between features and targets is linear.

- **Polynomial Kernel :** effective for capturing non-linear relationships by mapping features into higher-dimensional space using polynomial functions. It allows for more flexible modeling but can be sensitive to the choice of hyperparameters like degree.

- **Radial Basis Function (RBF) Kernel :** appropriate for capturing complex, non-linear relationships. It maps features into infinite-dimensional space using Gaussian radial basis functions, making it highly flexible but requiring careful tuning of hyperparameters like gamma.

- **Sigmoid Kernel :** suitable for data with a clear boundary between classes. It's particularly useful in binary classification tasks but may not perform well on regression problems.

Application of SVR algorithm for the public data set of contributions in social security :

```
R-squared: 0.9324867097331205
RMSE: 1473.9076921439755
MAE: 1267.4189745287945
Future contributions: [27794.16925505]
```

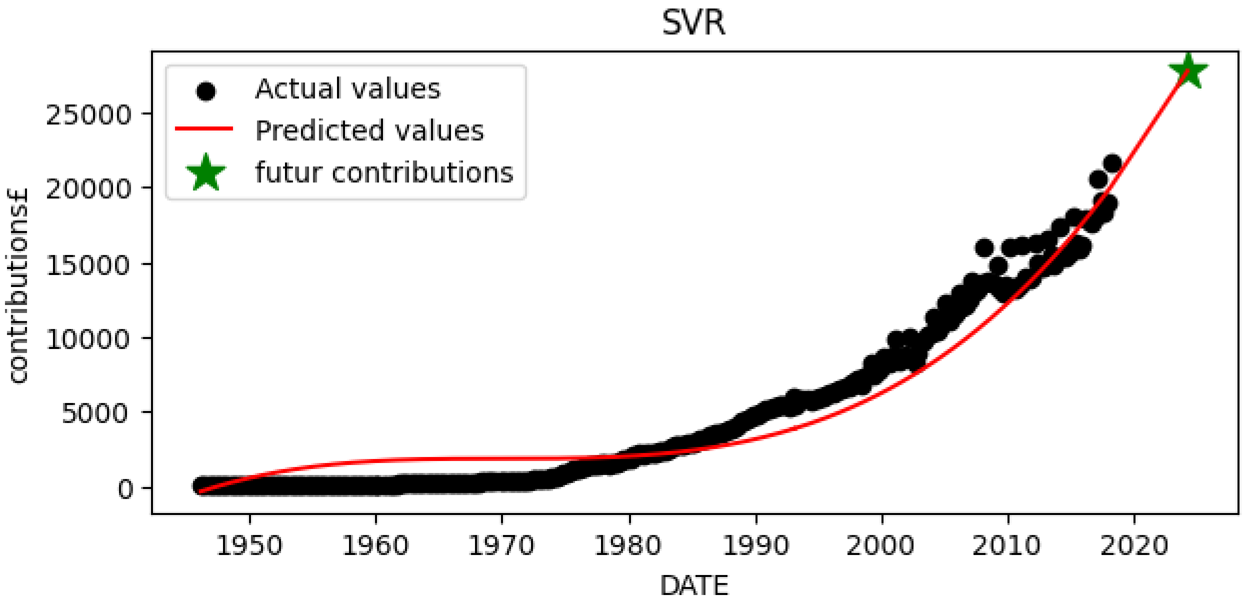

Figure 3.6: Support Vector Regression (SVR)

- R-squared (Coefficient of Determination): With a value of 0.9325, the SVR model explains approximately 93.25% of the variance in the contributions data. This indicates that the model captures a large portion of the variability in the target variable, suggesting a good fit to the data.

- Root Mean Squared Error (RMSE): The RMSE value of 1473.91 indicates the average magnitude of the errors between the actual and predicted contributions. Lower RMSE values suggest better accuracy, so this relatively low RMSE indicates that the SVR model's predictions are close to the actual contributions on average.

- Mean Absolute Error (MAE): The MAE value of 1267.42 represents the average absolute difference between the actual and predicted contributions. Similar to RMSE, lower MAE values indicate better model performance. This value suggests that, on average, the SVR model's predictions deviate from the actual contributions by approximately 1267.42.

3.3.3 ARIMA

Is a powerful time series forecasting algorithm that models the next data point as a linear function of its past values and forecast errors. It combines autoregression (AR), differencing (I), and moving average (MA) components to capture complex temporal patterns and make accurate predictions. Application of ARIMA for the public data set of contributions in social security :

- **The Augmented Dickey-Fuller (ADF) test :** the ADF test is a statistical method used to test the stationarity of a time series

```
1. ADF :  3.9240096364332704
2. P-Value :  1.0
3. Num Of Lags :  12
4. Num Of Observations Used For ADF Regression: 275
5. Critical Values :
        1% :  -3.454355055831705
        5% :  -2.8721080938842976
        10% :  -2.572401325619835
```

Figure 3.7: ADF-test

- **The optimal ARIMA model :** the optimal ARIMA model represents the combination of parameters (p, d, q) that minimizes the Akaike Information Criterion (AIC) and best fits the time series data

```
Performing stepwise search to minimize aic
 ARIMA(2,2,2)(0,0,0)[0]             : AIC=4344.595, Time=0.81 sec
 ARIMA(0,2,0)(0,0,0)[0]             : AIC=4757.091, Time=0.02 sec
 ARIMA(1,2,0)(0,0,0)[0]             : AIC=4667.264, Time=0.04 sec
 ARIMA(0,2,1)(0,0,0)[0]             : AIC=4497.254, Time=0.08 sec
 ARIMA(1,2,2)(0,0,0)[0]             : AIC=4420.805, Time=0.33 sec
 ARIMA(2,2,1)(0,0,0)[0]             : AIC=inf, Time=0.29 sec
 ARIMA(3,2,2)(0,0,0)[0]             : AIC=4075.163, Time=0.36 sec
 ARIMA(3,2,1)(0,0,0)[0]             : AIC=4083.678, Time=0.47 sec
 ARIMA(4,2,2)(0,0,0)[0]             : AIC=4056.913, Time=0.91 sec
 ARIMA(4,2,1)(0,0,0)[0]             : AIC=4056.232, Time=0.91 sec
 ARIMA(4,2,0)(0,0,0)[0]             : AIC=4086.617, Time=0.17 sec
 ARIMA(5,2,1)(0,0,0)[0]             : AIC=4056.722, Time=0.93 sec
 ARIMA(3,2,0)(0,0,0)[0]             : AIC=4107.822, Time=0.15 sec
 ARIMA(5,2,0)(0,0,0)[0]             : AIC=4087.078, Time=0.24 sec
 ARIMA(5,2,2)(0,0,0)[0]             : AIC=4059.461, Time=1.16 sec
 ARIMA(4,2,1)(0,0,0)[0] intercept   : AIC=inf, Time=1.62 sec

Best model:  ARIMA(4,2,1)(0,0,0)[0]
```

Figure 3.8: Best-model

- **Autocorrelation function (ACF) :** the autocorrelation function (ACF) measures the correlation between a time series and its lagged values. The ACF plot typically displays the autocorrelation coefficients for different lag values.

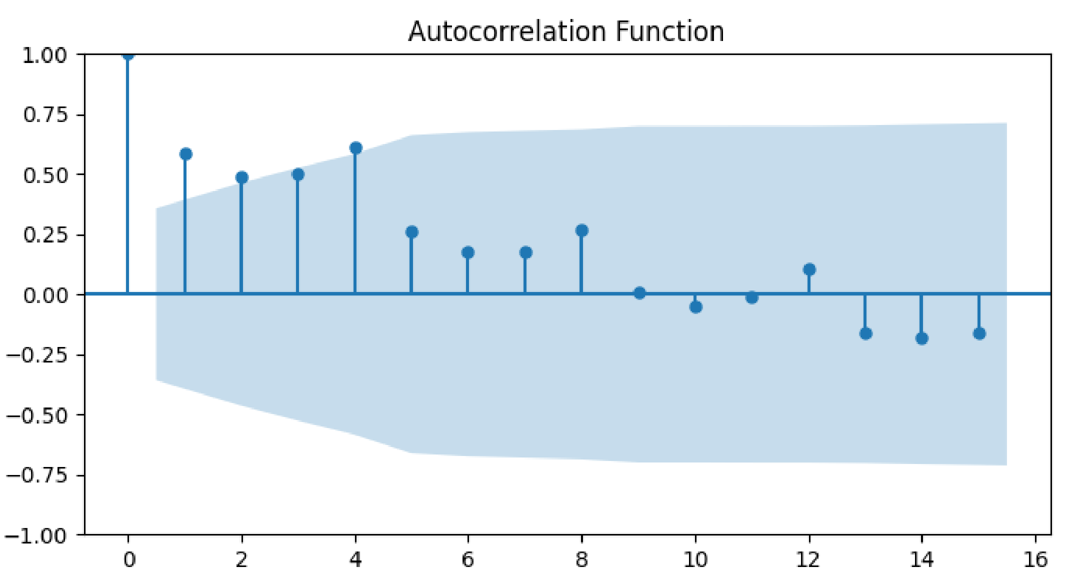

Figure 3.9: ACF

- **Arima prediction :** after splitting the shape of the data into train and test . in the test part we apply the arima predictions

```
R-squared: 0.6638906245432434
Root Mean Squared Error (RMSE): 1138.2858109675292
Mean Absolute Error (MAE): 754.565551958755
```

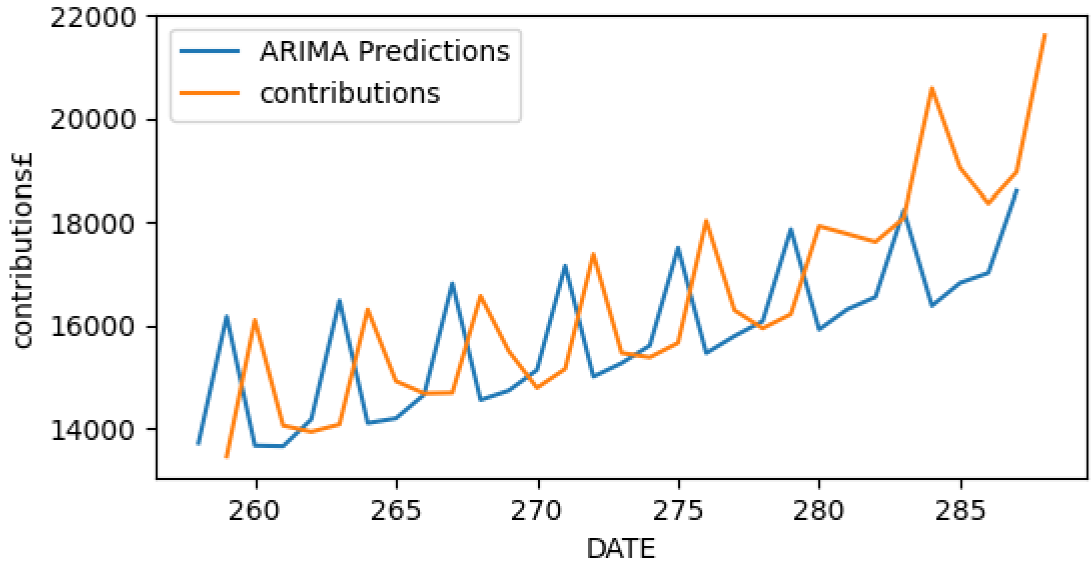

Figure 3.10: Arima-predictions

- R-squared (Coefficient of Determination): The R-squared value of 0.6639 indicates that the ARIMA model explains approximately 66.39% of the variance in the contributions data. This suggests that the model captures a moderate portion of the variability in the target variable, indicating a fair fit to the data.
- Root Mean Squared Error (RMSE): The RMSE value of 1138.29 represents the average magnitude of the errors between the actual and predicted contributions. Lower RMSE values suggest better accuracy, so this relatively low RMSE indicates that the ARIMA model's predictions are somewhat close to the actual contributions on average.
- Mean Absolute Error (MAE): The MAE value of 754.57 represents the average absolute difference between the actual and predicted contributions. Similar to RMSE, lower MAE values indicate better model performance. This value suggests that, on average, the

ARIMA model's predictions deviate from the actual contributions by approximately 754.57.

From the values of R-squared, RMSE and MAE. Arima model have a better predictive performance and model accuracy then Linear regression algorithm.

- **Futur predictions :** ARIMA algorithm forecasts future values by extrapolating time series patterns learned from historical data.

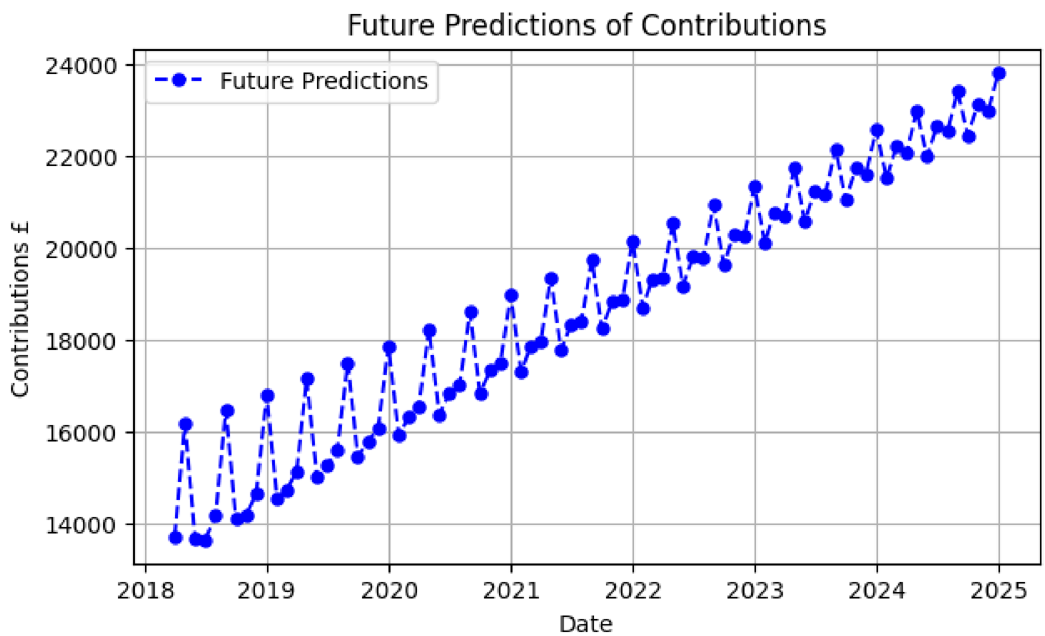

Figure 3.11: Futur-predictions

3.4 Obtained Results and Discussion

- **Firstly**, we compared the value of R-squared for all the algorithms used for evaluating the model. The values are presented in Table 3.1.

Table 3.1: Results of the evaluation (R-squared) of the model through different algorithms.

Algorithms	**R-squared**
Linear Regression	0.84
SVR	0.93
ARIMA	0.66

R-squared, also known as the coefficient of determination, is used in model evaluation to assess the goodness of fit of a regression model to the data. It represents the proportion of the variance in the dependent variable that is explained by the independent variables in the model. Higher R-squared values indicate a better fit of the model to the data, suggesting that more variation in the dependent variable is accounted for by the independent variables. It is a useful metric for comparing different models and assessing their predictive power. However, R-squared alone may not provide a complete picture of model performance, and it should be considered alongside other evaluation metrics.

In comparing the R-squared values of different algorithms, we aim to assess their performance in explaining the variance in the data. Among the evaluated algorithms, SVR demonstrates the highest R-squared value of 0.93, indicating its ability to capture a substantial portion of the variance in the data. Following closely is Linear Regression with an R-squared value of 0.84, suggesting a good fit to the data. However, ARIMA lags behind with an R-squared value of 0.66, indicating a weaker ability to explain variance. Therefore, based solely on the R-squared metric, SVR appears to be the most effective algorithm for this dataset, followed by Linear Regression, while ARIMA performs comparatively less well.

- **Secondly** , we compared the value of RMSE for all the algorithms used for evaluating the model. The values are presented in Table 3.2.

Table 3.2: Results of the evaluation (RMSE) of the model through different algorithms.

Algorithms	**RMSE**
Linear Regression	2376.04
SVR	1473.91
ARIMA	1138.28

Root Mean Squared Error (RMSE) is a crucial metric used in regression analysis to assess the accuracy of predictive models. It measures the average magnitude of prediction errors, indicating how well the model's predictions align with the observed data. RMSE is sensitive to large errors, making it effective for detecting outliers and evaluating overall model performance. It provides a quantitative measure of prediction accuracy, facilitating comparisons between different models or algorithms.When comparing algorithms based on Root Mean Squared Error (RMSE), it's crucial to understand their respective performances in predictive modeling. RMSE measures the average magnitude of errors between predicted and actual values, with lower values indicating better performance.

In our evaluation: Linear Regression yielded an RMSE of 2376.04, indicating a moderate level of error. SVR achieved an RMSE of 1473.91, showcasing improved predictive accuracy compared to Linear Regression. ARIMA demonstrated the lowest RMSE at 1138.28, suggesting superior predictive capability among the evaluated algorithms. In general, a lower RMSE signifies better predictive accuracy and a closer fit to the observed data. Therefore, ARIMA emerges as the preferred algorithm in this context, offering the smallest prediction errors .

- **Thirdly** , we compared the value of MAE for all the algorithms used for evaluating the model. The values are presented in Table 3.3.

Table 3.3: Results of the evaluation (MAE) of the model through different algorithms.

Algorithms	**MAE**
Linear Regression	2071.78
SVR	1267.42
ARIMA	754.56

Mean Absolute Error (MAE) is a metric used to quantify the average magnitude of errors between predicted and actual values in a predictive model. It provides a straightforward measure of model performance, calculated as the average absolute difference between predicted and actual values. MAE is particularly useful when dealing with outliers or when the magnitude of errors is critical for decision-making. Its interpretation is intuitive, with lower MAE values indicating better predictive accuracy and a closer fit to the observed data.

The Mean Absolute Error (MAE) values for Linear Regression, SVR, and ARIMA are 2071.78, 1267.42, and 754.56, respectively. A lower MAE indicates better predictive accuracy and a closer fit to the observed data. In this case, ARIMA demonstrates the lowest MAE, suggesting that it provides the best prediction performance among the evaluated algorithms. However, the choice of the "good" algorithm depends on the specific context of the problem, including factors like computational efficiency, interpretability, and domain-specific considerations.

3.5 Conclusion

In this chapter, we have described the dataset which we used. Then, we have detailed also our algorithms for prediction.Then we compared the evaluations in all the algorithms . In the next chapter, we will mention the used tools, libraries, and frameworks.

Chapter 4

Frameworks, tools and libraries

4.1 Introduction

After detailing our algorithms for the prediction in the previous chapter, the objective of this chapter is to mention tools and frameworks that help us implement our code.

4.2 Frameworks, tools and libraries

- Python: Python is a widely used, versatile, and high-level programming language created by Guido van Rossum in 1991. It's renowned for its readability, simplicity, and extensive library support, making it ideal for various applications in web development, data analysis, artificial intelligence

Figure 4.1: Python

- TensorFlow : is a free and open-source software library for machine learning, developed by Google. It's designed to facilitate the creation, training, and deployment of various machine learning models

Figure 4.2: TensorFlow

- NumPy: is a cornerstone scientific computing library for Python, enabling efficient manipulation of arrays and providing essential functions for numerical operations such as linear algebra, Fourier analysis, and matrix operations. Developed by Travis Oliphant, it serves as a foundational tool for data analysis, machine learning, and scientific computing tasks.

Figure 4.3: Numpy

- Matplotlib: is a powerful Python library widely used for creating high-quality plots, figures, and visualizations. It offers an object-oriented API, seamlessly integrating with NumPy arrays, making it an essential tool for data visualization, scientific computing, and big data analysis.

Figure 4.4: Mathplotlib

- Colab: created by Google, offers complimentary GPU and TPU access, facilitating the construction of machine learning and deep learning models.

Figure 4.5: Colab

- Sklearn: Scikit-learn (sklearn) is a machine learning library in Python, providing a wide range of tools for predictive data analysis. It features various algorithms for classification, regression, clustering, and dimensionality reduction, along with utilities for model selection and evaluation.

Figure 4.6: Scikit-learn

- Pandas: is a powerful Python library for data manipulation and analysis, offering data structures like DataFrames and Series. It provides tools for reading and writing data from various file formats, handling missing data, and performing operations like filtering, grouping, and merging. Pandas is widely used in data science and analytics for tasks such as data cleaning, exploration, and preparation.

Figure 4.7: Pandas

- Statsmodels: is a Python library for statistical modeling, hypothesis testing, and data exploration, offering tools for fitting various models such as linear regression and time series analysis. It provides functionality for conducting statistical tests and generating summary statistics for datasets.Statsmodels integrates seamlessly with other scientific libraries like NumPy and Pandas, making it a valuable resource for statistical analysis in Python.

Figure 4.8: Statsmodels

4.3 Conclusion

In this chapter, we have mentioned all the frameworks, tools, and libraries that we need in this domain. In the next chapter, we will discuss different obtained results.

Chapter 5

Discussion and conclusion

5.1 Introduction

In the previous chapter, we have shown all tools, libraries, and frameworks that are important to work. In this chapter, we will talk about the key findings and Perspectives on the futur.

5.2 Summary of key findings

In the exploration of social security and its local manifestation in CNAS Biskra, we unearthed critical insights into the complexities and challenges inherent in safeguarding individuals and families against social and economic hazards, These insights illuminate the indispensable role of CNAS in delivering essential benefits and services to beneficiaries, underscoring its significance within the broader social protection landscape of Algeria.

However, our investigation revealed a myriad of challenges plaguing social security systems both globally and locally, These challenges include the lack of predictive analytics capabilities, administrative inefficiencies, funding deficits, and issues related to fraud detection and prevention, These obstacles hinder the effective delivery of benefits and services, compromising the well-being and livelihoods of vulnerable populations.

To address these challenges, our study set forth a comprehensive set of objectives, including exploring the domain of social security, investigating the applicability of AI techniques, applying AI algorithms to social security contexts, and evaluating the performance of different algorithms. These objectives served as guiding principles in our quest to enhance the efficacy and responsiveness of social protection mechanisms.

Through the application of machine learning and time series algorithms, including linear regression, support vector regression (SVR), and autoregressive integrated moving average (ARIMA), we conducted predictive modeling to forecast contributions within the public dataset. Our results demonstrated the potential of these algorithms in predicting future contributions, thereby enabling proactive decision-making and resource allocation within social security systems.

In summary, our research illuminates the intricate interplay between social security frameworks and AI techniques, highlighting the transformative potential of predictive analytics in fortifying social protection mechanisms. By addressing the challenges and leveraging technological innovations, we aspire to advance the overarching goals of inclusive and equitable social protection, thereby fostering resilience and well-being within communities.

5.3 Perspectives

In the future, we propose several directions for enhancing the accuracy, efficiency, and overall effectiveness of social security assurance through advanced predictive analytics and machine learning techniques. Our key recommendations and proposals are as follows:

5.3.1 Technological Advancement Perspective

Leveraging advanced technology will enhance predictive accuracy and responsiveness.

Integration of Real-Time Data and Advanced Algorithms

- Proposal: incorporate real-time economic indicators, employment rates, and demographic changes into predictive models while leveraging advanced machine learning techniques such as deep learning and ensemble methods.

- Implication: utilizing cutting-edge technology and real-time data will significantly improve the precision of predictions regarding employer contributions to social security. This enhancement will allow for more dynamic and responsive systems that can adapt to economic shifts more efficiently.

Continuous Model Improvement

- Proposal: implement mechanisms for continuous learning and validation of predictive models.
- Implication: as models learn and adapt over time, their accuracy and reliability will increase, leading to more robust forecasts that can better inform policy and decision-making processes.

5.3.2 Policy and Decision-Making Perspective

Data-driven insights will support more informed and effective policymaking.

Stakeholder Collaboration

- Proposal: foster collaboration between social security agencies, employers, data scientists, and policymakers to ensure models are aligned with practical needs and regulatory frameworks.
- Implication: multidisciplinary collaboration will ensure that predictive models are not only technically sound but also practically relevant, facilitating smoother implementation and compliance with legal standards.

5.3.3 Ethical and Social Impact Perspective

Ensuring fairness and transparency will build trust and equity in social security systems.

Transparency and Explainability

- Proposal: develop models that prioritize transparency and provide clear, understandable explanations for their predictions.
- Implication: enhancing the transparency and explainability of predictive models will build trust among stakeholders, including policymakers, employers, and the public. This trust is crucial for the acceptance and effective use of predictive insights in decision-making processes.

These three perspectives: technological advancement, policy and decision-making, and ethical and social impact- provide a comprehensive view of how predictive analytics can shape the future of social security assurance. Each perspective highlights different facets of potential improvements and their implications, showcasing the multifaceted benefits of integrating advanced predictive models into social security systems.

5.4 Conclusion

In this chapter, we have explained the key findings, including the result of Applying artificial intelligence algorithms to a social security context, and evaluated the performance of different algorithms, Our results demonstrated the potential of these algorithms to predict future contributions, thus enabling proactive decision-making and resource allocation within social security systems, After that we have denoted the future of social security assurance.

References

[1] Tareck Alsamara, Farouk Ghazi, and Halima Mallaoui. Administrative organization of health care institutions in algeria: Between centralization and decentralization. *Open Access Macedonian Journal of Medical Sciences*, 2022.

[2] Christopher M Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[3] George EP Box, Gwilym M Jenkins, and Gregory C Reinsel. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[4] Marco Bruno. Forecasting oil prices: time series vs. machine learning methods. 2017.

[5] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3):1–58, 2009.

[6] Oksana Cheberyako and Yuliya Skulish. The level of pension provision of the population is a guarantee of social security of the person. *University Economic Bulletin*, 2022.

[7] Jane Doe. Ai-based fraud detection in social security insurance: A case study. *Journal of Insurance Fraud Detection*, 5(1):45–57, 2019.

[8] Emily Johnson. Improving customer service in social security insurance through ai chatbots. *Journal of Customer Service Enhancement*, 3(4):289–301, 2018.

[9] Nacera Mahfoud, Brahim Brahamia, and Coppieters Yves Coppieters Yves. Consommation de médicaments et maîtrise des dépenses de santé en algérie. 2017.

[10] Spyros Makridakis, Steven C Wheelwright, and Rob J Hyndman. *Forecasting methods and applications*. John Wiley & Sons, 1997.

[11] Stéfane Mouille. Introducing a smart card-based national healthcare system in algeria. 2006.

[12] Anthony Mouraud. Innovative time series forecasting: autoregressive moving average vs deep networks. *Entrepreneurship and Sustainability Issues*, 4:282–293, 2017.

[13] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[14] Office for National Statistics. Quarterly employers' social security contributions, 1946 to quarter 1 jan to mar 2018. https://www.ons.gov.uk/economy/grossdomesticproductgdp/adhocs/008939quarterlyemployerssocialsecuritycontributions1946toquarter1jantomar2018, 2018. Last accessed: 03/06/2024.

[15] Gopinath Rebala, Ajay Ravi, and Sanjay Churiwala. Machine learning definition and basics. *An Introduction to Machine Learning*, 2019.

[16] Stuart Russell, Peter Norvig, and Ernest Davis. *Artificial Intelligence: A Modern Approach*. Pearson, 2016.

[17] John Smith. Application of machine learning techniques in social security systems: A review. *Journal of Social Security Research*, 10(2):123–135, 2020.

[18] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[19] Nouara Kaïd Tlilane. Healthcare funding problems in algeria. *Development Economics*, 2004.

[20] Michael Williams. Ai-driven decision support systems for social security policy analysis. *Journal of Social Policy Analysis*, 15(3):201–215, 2021.

[21] World Bank. Algeria economic monitor, no. 1. 2020.

[22] Xiaojin Zhu. *Introduction to semi-supervised learning*. Morgan & Claypool, 2009.