Faculté des Sciences et Techniques

Université Abdelmalek Essaadi

Rapport du projet de fin de module

MASTER OF SCIENCE AND TECHNOLOGY IN ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

Food security in Morocco

Done by :

Mohamed Amine Fakhre-Eddine

Abdelali Ibn Tabet

Nouhaila Krimel

Soueilem Mohamed Abd Nasir

Mentored by :

Mr. Mostafa Ezziyyani

2023 / 2024


بسم الله الرحمن الرحيم
١٤٢٠

Dedication

"

To my dear parents

This work is dedicated to you.

My first teachers, my guides and

My biggest supporters, your unconditional love

And your sacrifices are always the reason.

to give the best of myself

"

-Mohamed Amine

-Abdeali

-Nouhaila

-Soueilem Mohamed

Abstract

Our research intends to solve crucial issues such as agricultural production, regional inequities, inefficiencies in food distribution, and resource utilization. Despite challenges such as a lack of comprehensive data and limited engagement with government agencies and local communities, we have achieved tremendous progress in our methods.

To date, we have successfully collected production statistics from each area as supplied by the High Commission for Planning. However, distribution and consumption statistics are currently lacking. To ensure accuracy and consistency across varied datasets, we conducted extensive analysis, standardization, and cleaning throughout the data preparation step.

In the Data Analysis and Modelling stage, we've started looking into machine learning methods like ARIMA and recurrent neural networks for forecasting agricultural output trends. While our geospatial analysis has yet to be completely deployed, we want to use GIS approaches to optimize transportation routes and logistics.

Moving ahead, our primary focus will be on creating optimization algorithms to improve resource allocation and eliminate surpluses and shortages. We want to build a prototype AI-powered food distribution system and use historical data to assess its effectiveness in optimizing food distribution.

Despite the obstacles, our process is iterative and adaptive. We are committed to developing our methodology based on stakeholder feedback and real-world experiences, with the goal of increasing food security and access to market information in the regions we serve.

Keywords: Agricultural Productivity, Regional Disparities, Inefficiencies in Food Distribution, Resource Utilization, Data Collection, Data Preprocessing, Data Analysis, Modeling, Machine Learning Algorithms, ARIMA, Recurrent Neural Networks, Geospatial Analysis, GIS Techniques, Optimization Algorithms, Prototype Implementation, Iterative Improvement, Food Security, Access to Market Information, Collaboration, Stakeholder Feedback

Table of contents

GENERAL INTRODUCTION 1

CHAPTER 1 : 2

PROJECT CONTEXT 2

1. HOST ORGANISATION : 3
2. PROBLEMATIC : 3
3. OBJECTIVES : 3
4. PROJECT MANAGEMENT AND COLLABORATION : 3

CHAPTER 2 : 5

STATE OF THE ART 5

1. RELATED WORK : 6
2. AI AND MACHINE LEARNING ALGORITHMS PRESENTATION : 6
    2.1. Definition : 6
    2.2. Machine learning : 6
        2.2.1. Definition : 7
        2.2.2. Process : 7
        2.2.3. ML model: 7
3. RELATED ALGORITHMS : 7
    3.1. Recurrent Neural Network (RNN) : 7
        3.1.1. Artificial neural networks (ANNs) : 7
        3.1.2. Neural network architecture : 8
        3.1.3. Neural network learning : 8
        3.1.4. RNNs : 9
    3.2. ARIMA : 9
        3.2.1. The meaning of p : 10
        3.2.2. The meaning of d : 10
        3.2.3. The meaning of q : 11
        3.2.4. AR model : 11
        3.2.5. MA model : 11
        3.2.5. ARIMA model : 11

CHAPTER 3 : 12

SUGGESTED SOLUTION : HARVEST GUARD 12

1. DATA PREPROCESSING AND METHODOLOGY : 13
    1.1. Data collection process : 13
    1.2. Data cleaning and preprocessing : 14
    1.3. Data visualization: 14

2. MODELING AND ANALYSIS : 15
2.1. ARIMA : 15
2.1.1. How to find the order of differencing (d) in ARIMA model : 15
2.1.2. How to find the order of the AR term (p) : 15
2.1.3. How to find the order of the MA term (q) : 16
2.1.4. Results : 17
3. APP DESIGN AND SOLUTION : 18
3.1. User experience : 18
3.2. User interface : 19
3.2.1. Landing page : 19
3.2.2. Statistics page : 19
3.2.3. Login page : 20
3.2.4. Dashboard : 21
3.2.5. Distribution : 22
GENERAL CONCLUSION 24
WEBOGRAPHIE 25

Liste des figures

Figure 1 : Github logo 4
Figure 2 : Linear logo 4
Figure 3 : Screenshot of working with linear 4
Figure 4 : Relation between AI, ML, NNs and DL 6
Figure 5 : Neural networks architecture 8
Figure 6 : A reccurent neural network 9
Figure 7 : Architecture of a neuron in a RNN 9
Figure 8 : ARIMA process workflow 10
Figure 9 : Example of data collected raw 13
Figure 10 : Data preprocessing preview 14
Figure 11 : Example of a product production visualization 15
Figure 12 : Partial Autocorrelation (PACF) plot 16
Figure 13 : Autocorrelation Function (ACF) 17
Figure 14 : Residuals plot 17
Figure 15 : Density plot 18
Figure 16 : UX workflow 18

General introduction

In Morocco, achieving food security is a complicated task that includes agricultural production, regional imbalances, and effective food delivery. Meeting the nutritional needs of all populations requires effective resource utilization and fair access to food. This study focuses on our current research to address food security concerns in Morocco using sophisticated technologies, including artificial intelligence (AI). Our project seeks to optimize resource allocation, eliminate surpluses and shortages, improve food security, increase access to market information, and ensure adaptation and scalability in the Moroccan setting.

Despite constraints such as data scarcity and restricted collaboration, our technique, which is based on rigorous data analysis, modelling, and iterative improvement, aims to contribute to the creation of a more robust and equitable food distribution system customized to Morocco's specific needs. Through joint efforts and stakeholder involvement, we want to pave the way for a future in which food security is more than just an ambition for Moroccan communities.

Chapter 1 :

Project Context

1. Host organisation :

Our project is hosted and mentored by Mr. Mostafa Ezziyyani, a distinguished figure within the Faculty of Sciences and Technology in Tangier. With a wealth of experience and expertise in the field of technology and its applications in various domains, Mr. Ezziyyani brings invaluable guidance and mentorship to our project. As a respected member of the academic community, his leadership ensures that our efforts are aligned with the goals and objectives of the faculty. Under his mentorship, we benefit from access to resources, expertise, and networking opportunities that are pivotal in driving the success of our project. Mr. Ezziyyani's commitment to fostering innovation and excellence serves as a cornerstone in our journey towards addressing food security challenges in Morocco.

2. Problematic :

Our research tackles various major food security issues in Morocco. These difficulties include agricultural productivity, geographical inequities, inefficiencies in food delivery, and poor resource utilization. Varying agricultural terrain throughout Morocco, along with discrepancies in infrastructure and market access, lead to uneven food distribution and availability. Furthermore, inefficiencies in resource allocation and logistics networks increase regional surpluses and shortages. To address these problems, a thorough grasp of local contexts, strong data-driven techniques, and new solutions suited to Moroccan communities' particular requirements are required. By identifying and tackling these main issue areas, our study hopes to establish the framework for Morocco's more egalitarian and resilient food distribution system.

3. Objectives :

Our initiative is motivated by a set of clear and practical objectives aimed at tackling Morocco's multiple food security concerns. These goals include optimising resource allocation to maximise efficiency and minimise waste, reducing surpluses and shortages through improved distribution strategies, enhancing overall food security by ensuring reliable access to nutritious food for all citizens, improving access to market information to facilitate informed decision-making among stakeholders, and ensuring adaptability and scalability of our solutions to accommodate changing agricultural and economic By carefully aligning our efforts with these goals, we want to achieve meaningful progress towards a more sustainable and inclusive food distribution system in Morocco, eventually benefiting the well-being and lives of its population.

4. Project management and collaboration :

To improve project management and communication, we've implemented two crucial tools: Linear and GitHub. Linear is our primary project management software, with straightforward task tracking, problem management, and progress monitoring features. Linear allows team members to effortlessly assign tasks, track progress, and prioritise activities, ensuring effective project execution.

Additionally, GitHub serves as our primary repository for code collaboration and version control. Using GitHub's powerful tools, such as branching, pull requests, and code reviews, we can preserve code integrity, encourage smooth communication among team members, and track changes throughout the development process. By leveraging the advantages of Linear and

GitHub, we improve openness, accountability, and efficiency within our project team, propelling us closer to achieving our goals in tackling food security concerns in Morocco.

FIGURE 1 : GITHUB LOGO

FIGURE 2 : LINEAR LOGO

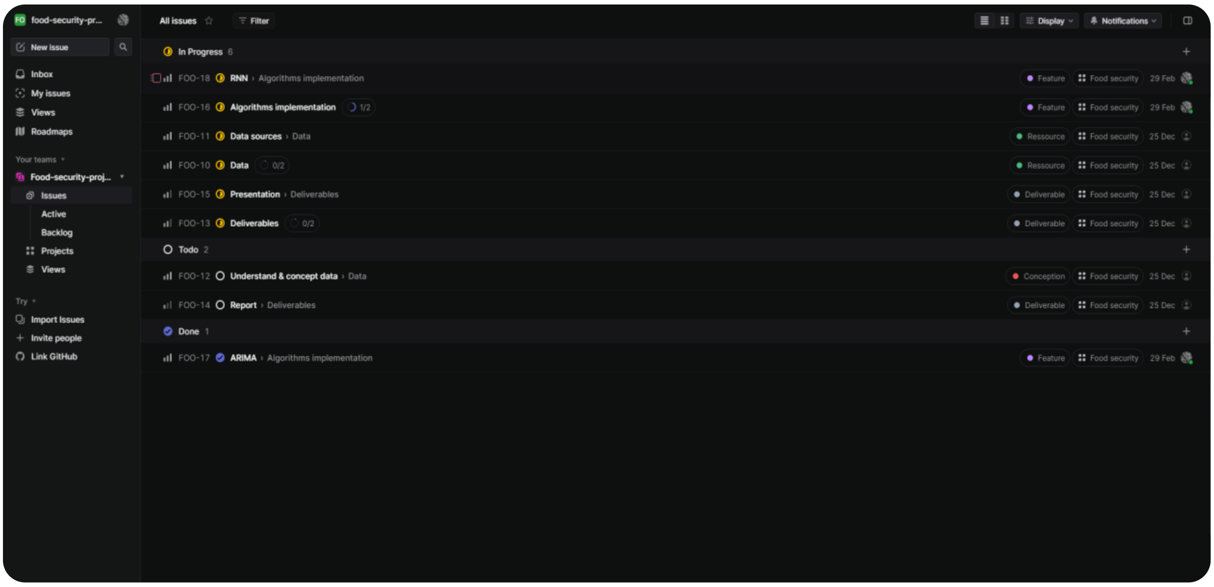


FIGURE 3 : SCREENSHOT OF WORKING WITH LINEAR

Chapter 2 :

State of the art

1. Related work :

The use of artificial intelligence (AI) models for food security has been studied to address a variety of issues, including climate change, population expansion, and economic inequality. Artificial intelligence models have been applied to improve food supply chain management, consumer health and nutrition, waste reduction, and food safety and quality.

Research was carried out at many geographical levels, including local, sub-regional, national, regional, and global levels. The models have been used in a variety of ways, including involving stakeholders in AI modelling for food security, using only AI models in food security scenarios, and incorporating various factors influencing food security, such as climate change, land use, crop yields, and socioeconomic factors. The study also determined the most critical elements impacting food security and their relative relevance.

Research on artificial intelligence (AI) for food security has limitations, including specific modeling frameworks, data availability, generalization vs. localization, data sources and time resolution, stakeholder involvement, and translating model outputs into policies. Some studies propose specific frameworks without comparisons, while others require up-to-date household-level data. Global models may limit the discovery of local patterns, while training separate models based on historical data from individual countries could enhance understanding. Incorporating secondary data with varying time resolutions can impact predictions' accuracy. Stakeholder involvement is crucial for comprehensive outcomes. Translation of model outputs into policies is challenging, and ethical considerations are essential when applying AI in food security contexts.

2. AI and Machine Learning Algorithms Presentation :

2.1. Definition :

Artificial intelligence is a set of techniques that seek to simulate human intelligence in a machine or system. By feeding these systems with algorithms, which use techniques such as machine learning and deep learning to demonstrate "intelligent" behavior comparable to that of a human being. By mining massive amounts of data from previous examples of similar behavior.

2.2. Machine learning :

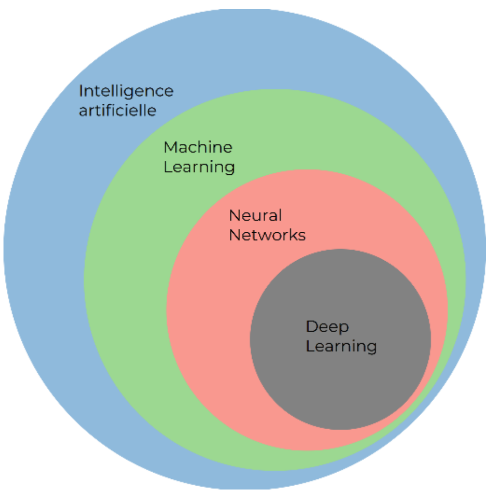


FIGURE 4 : RELATION BETWEEN AI, ML, NNS AND DL

2.2.1. Definition :
Machine Learning is a sub-branch of Artificial Intelligence (AI) focused on the creation of algorithms and models that enable machines (such as computers) to learn from the data they process and improve their performance on specific tasks without being highly programmed.

In simple terms, it's a set of applied statistical methods designed to teach machines to recognize patterns in data and then make predictions or decisions based on statistics, becoming more accurate over time.

2.2.2. Process :
The process of building a Machine Learning model can be divided into different phases :

1. Data collection : we gather the relevant data that will be used to train the Machine Learning model.

2. Data preparation : the collected data is cleaned, transformed and pre-processed to ensure that it is in a suitable format for analysis.

3. Model selection and construction : In this phase, an appropriate Machine Learning model is selected and constructed according to the problem to be solved and the characteristics of the data used.

4. Training : the Machine Learning model is trained using the prepared data, adjusting its internal parameters to minimize the error between the model's predictions and the actual values of the training data.

5. Test : To estimate the model's performance, it is evaluated using a separate dataset which it did not process during the training phase.

6. Using the model : Once the Machine Learning model has been trained and evaluated, it can be integrated into real-world applications to make predictions or decisions based on new data.

2.2.3. ML model:
A machine learning model is a kind of mathematical representation of a real process that is learned from data. The model is trained on a dataset, then used to make predictions or decisions based on new data, so once trained, it must be able to generate results from data it has never processed.

All ML models fall into three main categories, adapted to different tasks, types of input data and the problem to which they apply.

Those categories are supervised learning, non-supervised learning, and neural networks.

3. Related algorithms :

3.1. Recurrent Neural Network (RNN) :

Before introducing RNN, let’s have a brief introduction on Artificial Neural Nerworks (ANN).

3.1.1. Artificial neural networks (ANNs) :
Artificial neural networks (ANNs) or simulated neural networks (SNNs) are the most sophisticated and powerful artificial intelligence algorithms, enabling data to be classified and grouped very rapidly. They are designed to mimic the physiological model of electrical signal transmission in the real neurons of the human brain. These networks are made up of several

layers of artificial neurons (or nodes), each layer of which is linked to all the neurons in the preceding layer. The greater the number of layers, the "deeper" the network, and so, to function properly, a neural network needs to be trained on a very large quantity of data, sometimes in excess of millions or even tens of millions.

3.1.2. Neural network architecture :

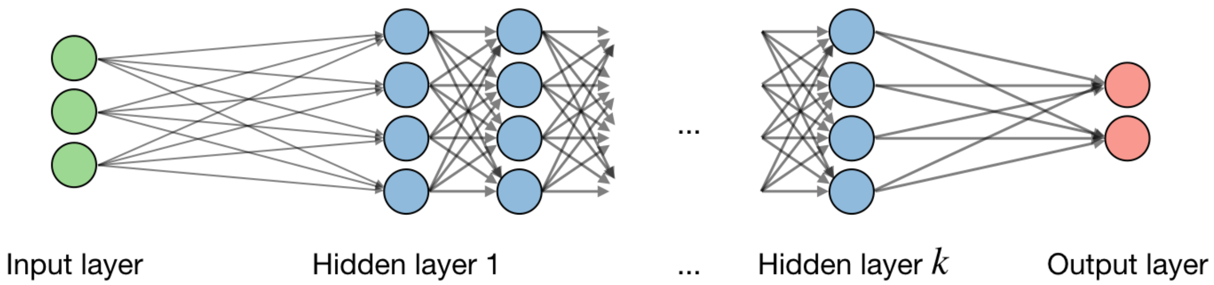


FIGURE 5 : NEURAL NETWORKS ARCHITECTURE

A neural network will contain three types of layers:

An input layer: Input nodes are the point of entry for data, processing and analyzing it for subsequent transmission to the next layer.

Hidden layers: Artificial neural networks can have a large number of "hidden layers" which rework and analyze the output of the input layer or other previous hidden layers.

Output layer: after all the layers have been processed, the output layer returns the final result. The number of layers depends on the type of problem, so if, for example, it's a binary classification problem, the output layer will have an output node. However, in the case of multi-class classification, the output layer may consist of more than one output node.

Note that :

Artificial neural networks can be either :

Single-layer: consisting solely of input and output layers.

Multilayer: with a number of hidden layers.

A "deep" neural network has several hidden layers with millions of neurons.

3.1.3. Neural network learning :

In general, to develop and train artificial neural networks, the following four steps are repeated in a loop:

1. First step : we start by circulating data from the first layer to the last, to generate an output y at the end.

2. Second step : we use what we call Cout (Cost function) to calculate the error between the output produced y and the reference output y_actual that we wish to have.

3. Third step : we measure how this cost function varies in relation to each layer of our model, starting with the last and working backwards to the very first.

4. Fourth step : this is where we use the gradient slope algorithm to correct each model parameter, then loop back to the first step to start a new training cycle.

3.1.4. RNNs :

In recurrent neural networks, information is passed through feedback loops, and thus back to a previous layer, enabling the system to build up a memory. In this way, RNNs use the context of inputs when calculating output, to produce an output that depends on previously calculated inputs and outputs.

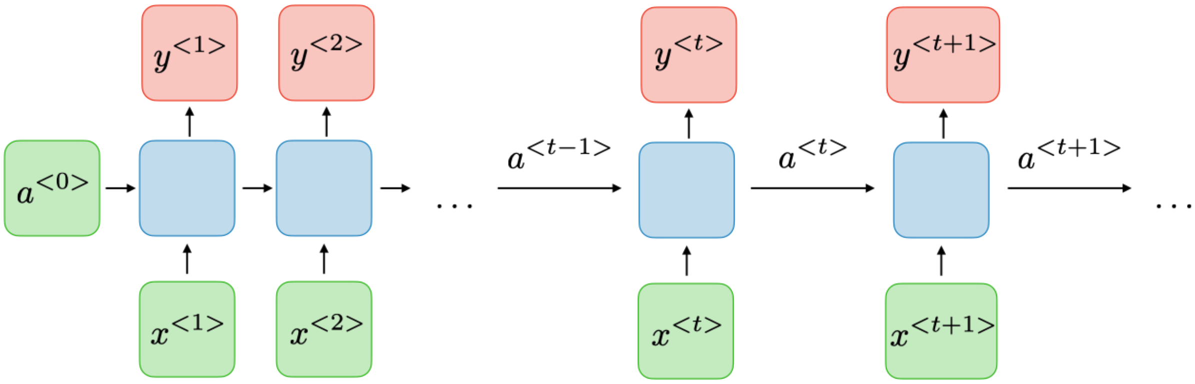


FIGURE 6 : A RECCURENT NEURAL NETWORK

For each timestep t, the activation $a^{<t>}$ and the output $y^{<t>}$ are expressed as follows:

$$a^{<t>} = g_1(W_{aaa}^{<t-1>} + W_{axx}^{<t>} + ba) \text{ and } y^{<t>} = g_2(W_{yaa}^{<t>} + by)$$

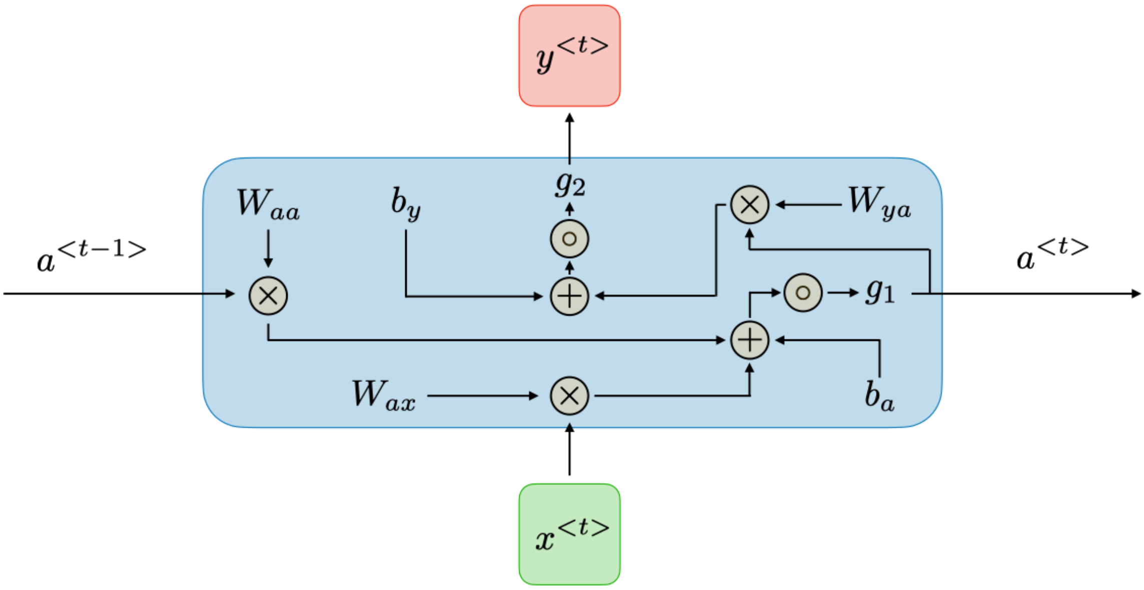


FIGURE 7 : ARCHITECTURE OF A NEURON IN A RNN

3.2. ARIMA :

ARIMA stands for Autoregressive Integrated Moving Average Model. It belongs to a class of models that explain a given time series using its own historical values, i.e., its own lags and lagged forecast error. The equation can be used to predict future values. ARIMA models may be used to represent any 'non-seasonal' time series that displays patterns rather than random white noise.

So, ARIMA, which stands for Auto Regressive Integrated Moving Average, is a forecasting technique based on the premise that information from previous values of a time series may be utilized to predict future values.

ARIMA models have three order parameters: (**p**, **d**, and **q**).
where,
p represents the order of the AR word.
q represents the order of the MA word.
d is the number of differences needed to make the time series stationary.

AR(p) Autoregression is a regression model that uses the dependent connection between a current observation and observations from a past period. An auto regressive (AR(p)) component is one that incorporates past values into the regression equation for a given time series.

I(d) Integration - makes the time series stable by differencing observations (subtracting an observation from the preceding time step). Differencing entails subtracting a series' current values from its prior values d times.

MA(q) Moving Average - a model that exploits the relationship between an observation and a residual error from a moving average model when applied to lagged data. A moving average component represents the model's error as the sum of past error terms. The order q specifies the number of terms to be included in the model.

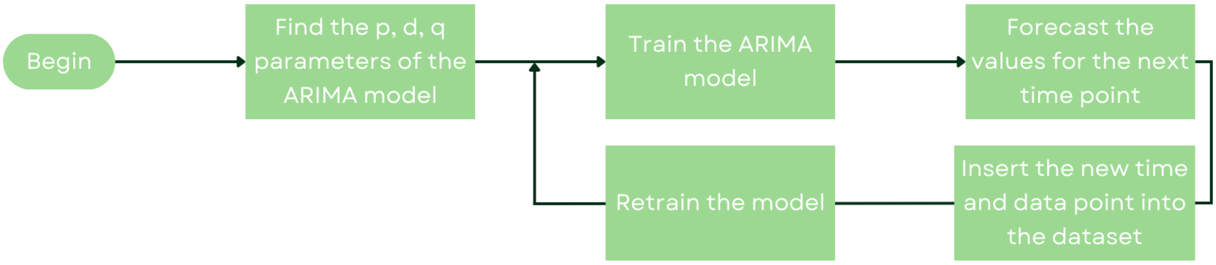


FIGURE 8 : ARIMA PROCESS WORKFLOW

ARIMA models include non-seasonal autoregressive integrated moving averages, the standard **ARIMA**.
SARIMA : Seasonal ARIMA.
SARIMAX : Seasonal ARIMA with Exogenous Variables
If a time series contains seasonal trends, we add seasonal words to make it SARIMA, which stands for Seasonal ARIMA.

3.2.1. The meaning of p :

p is the order of the **Auto Regressive (AR)** term. It refers to the number of lags of Y to be used as predictors.

3.2.2. The meaning of d :

The term "Auto Regressive" in ARIMA refers to a linear regression model that employs its own lags as predictors. Linear regression models, as we know, perform best when the predictors are uncorrelated and independent of one another. So we need to make the time series immobile.

The most popular method for making the series immobile is to differentiate it. That is, subtract the old number from the present one. Depending on the intricacy of the series, many differencings may be required.

The value of d therefore represents the smallest number of differencing required to make the series steady. If the time series is already stationary, d equals zero.

3.2.3. The meaning of q :

q denotes the order of the Moving Average (MA) phrase. It refers to the amount of lagged forecast mistakes that should be entered into the ARIMA model.

3.2.4. AR model :

An auto-regressive (AR) model is one in which Y_t is exclusively dependent on its own lags.

That is, Y_t is determined by Y_t 's lagged values. It is represented by the following equation:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \epsilon_1$$

Y_{t-1} is the lag1 of the series,

β_1 is the coefficient of lag1 that the model estimates, and

α is the intercept term, also estimated by the model.

3.2.5. MA model :

Similarly, in a Moving Average (MA) model, Yt is determined only by the delayed forecast errors. It is represented by the following equation:

$$Y_t = \alpha + \epsilon_1 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_q Y_{t-q}$$

where the error terms are the lag-specific autoregressive model errors.

The errors E_t and $E_{(t-1)}$ come from the following formulas:

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_0 Y_0 + \epsilon_t$$

$$Y_{t-1} = \beta_1 Y_{t-2} + \beta_2 Y_{t-3} + \ldots + \beta_0 Y_0 + \epsilon_{t-1}$$

3.2.5. ARIMA model :

An ARIMA model is one in which the time series is differenced at least once to make it stationary before combining the AR and MA components. Thus, the equation for an ARIMA model becomes

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} \epsilon_t + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_q Y_{t-q}$$

Chapter 3 :

Suggested solution : Harvest Guard

Introduction :

In this chapter, we begin the model development and implementation phase, which is the final phase of this report. Having introduced the concepts and algorithms of Machine Learning, we now focus on the practical aspects of our research.

1. Data Preprocessing and Methodology :

1.1. Data collection process :

The Harvest Guard project's data collecting procedure was mostly based on each region's Statistical Yearbooks, which were provided by the High Planning Commission. These comprehensive yearbooks include thorough information on area statistics, including agricultural output data. Given their distinctiveness and relevance to our project setting, these yearbooks serve as basic sources for collecting critical production data. It is crucial to highlight that these yearbooks cover a wide range of locations and are updated annually, offering a useful historical perspective on agricultural output patterns.

However, it is critical to recognize that the scope of the yearbooks is confined to production statistics, leaving out critical information about distribution and consumption. Despite this disadvantage, using the Statistical Yearbooks is an important first step towards gathering trustworthy and standardized data for our analytical and modelling efforts.

AGRICULTURE **الفلاحة**

Evolution de la superficie et de la production des céréales par région et province (ou préfecture) suite

تطور مساحة وإنتاج الحبوب حسب الجهة والإقليم (أو العمالة) تابع

Superficie en milliers d'hectares Production en milliers de quintaux	2016-2017*		2015-2016		2014-2015		المساحة بألف هكتار الإنتاج بألف قنطار
	الإنتاج Production	المساحة Superficie	الإنتاج Production	المساحة Superficie	الإنتاج Production	المساحة Superficie	
Casablanca- Settat	**23805.0**	**971.5**	**6045.3**	**758.2**	**26764.9**	**969.9**	**الدار البيضاء - سطات**
Benslimane	2226.5	85.7	700.5	83.7	2940.8	91.1	بنسليمان
Berrechid	3432.4	131.5	120.5	5.4	4083.4	128.8	برشيد
Casablanca	823.6	34.1	224.2	28.7	805.4	32.3	الدار البيضاء
El Jadida	4978.2	206.9	3793.9	357.1	5032.2	205.4	الجديدة
Mohammadia	19.2	0.8	…	…	…	…	المحمدية
Settat	6865.0	321.5	1201.1	282.6	7760.6	323.8	سطات
Sidi Bennour	5460.0	190.9	5.1	0.7	6142.5	188.5	سيدي بنور
Marrakech - Safi	**12569.9**	**1355.4**	**3982.9**	**560.9**	**21901.6**	**1294.0**	**مراكش - آسفي**
Al Haouz	1587.1	99.1	0.4	-	0.2	-	الحوز
Chichaoua	485.4	105.5	64.3	30.3	1014.0	154.1	شيشاوة
El Kelâa des Sraghna	2670.5	224.9	1969.1	204.6	7655.6	403.5	قلعة السراغنة
Essaouira	2148.7	203.9	986.1	99.2	2877.7	203.4	الصويرة
Marrakech	1388.6	100.6	82.2	16.2	4169.0	180.1	مراكش
Rehamna	1099.5	230.4	6.0	0.3	0.6	0.3	الرحامنة
Safi	2162.8	202.8	874.8	210.3	3558.2	170.6	آسفي
Youssoufia	1027.3	188.3	-	-	2626.3	182.0	اليوسفية
Drâa- Tafilalet	**1929.9**	**97.1**	**2006.2**	**103.5**	**2520.0**	**111.3**	**درعة - تافيلالت**
Errachidia	717.0	32.3	80.7	8.0	632.4	17.3	الرشيدية
Midelt	486.2	26.9	576.9	39.8	815.0	40.2	ميدلت
Ouarzazate	191.9	12.7	895.8	33.5	445.3	20.9	ورزازات
Tinghir	151.0	7.8	…	…	…	…	تنغير
Zagora	383.7	17.4	452.8	22.2	627.3	32.9	زاكورة

FIGURE 9 : EXAMPLE OF DATA COLLECTED RAW

1.2. Data cleaning and preprocessing :

First, we should extract relevant data to our research from the large amount of data existing in the yearbooks, our treatable data will look as the following:

	région de Tanger-Tétouan-Al hoceima																					
											Production (en 1000 Qx)											
	Cereales	Ble dur	Ble tendre	Mais	Orge	Legumineuses	Feves	Petits pois	Pois chiches	Lentilles	Tomates	Pomme de Terre	Carottes	Navet	Melon	Pastèque	Haricots verts	Fèves Vertes	Oignons	Courgettes	Concombre	Fraisier
2003-04	2603	1235.7	1043.1	15.3	598.1	67.5	131.6	35.8	91.6	9.5												
2004-05	3190.1	909.1	4124.8	1.7	649.7	47.9	134.1	19.8	80.3	35.1												
2005-06	2250.5	896.1	949.1	24.3	1224	124.6	20	7.8	26	35												
2006-07	2651.6	727.9	853.1	19.2	601.3	120.9	69.5	17.4	38.1	1.8												
2007-08	2254.9	605.8	1422.6	21.1	482.8	106.8	25.7	8.6	28	10.6												
2008-09	2222.5	605.8	898.2	26.9	874.3	100.5	20.3	14,4	39,9	6,6												
2009-10	3824.4	930.5	2042.8	1.6	849.3	119.6	31	26,5	38,9	13,5	113	247.4	27.5	12.8	559.8	96.4	1.4	18.3	81.2	13.8	0.8	93.7
2010-11	5282.9	1303.7	2082.4	34.1	1789.3	216.6	90	22,8	61,5	21,7												
2011-12	5117.3	1628.8	1901.6	0.5	1360.7	235.8	76	38,5	39,5	7,8												
2012-13	4577.6	1203.1	1422.4	22.5	1971.9	218.2	79.4	29,1	46,7	17,9	156.2	1977.7	53	19.8	780.2	112.4	6.2	39.4	177.1	20.2	4	150.4
2013-14	5250.3	1445.8	1880.7	32.4	1038.5	265.1	64.8	21,9	90,8	13,1	195.3	1891.9	62	28.3	723.9	316.4	6.6	51.4	200.1	53.8	6.5	173.8
2014-15	6808.8	1626.7	2701.6	20.1	2286.8	480.5	93.8	213,7	88,5	18,6	282.7	1904.5	50.1	37.4	810.2	300.1	2.4	146.1	166.3	22	12.4	281.6
2015-16	5005.9	1651.1	1975.1	24.5	1212	277.1	62.5	25,9	121,3	23,3	271.1	1840	34.9	28.7	827.9	217.1	2.7	53.7	142.8	15.7	7.9	281.6

Since our data have the same unit and order of magnitude, our data preprocessing will only consist of replacing missing values with 0.

	Cereales	Ble dur	Ble tendre	Mais	Orge	Legumineuses	Feves	Petits pois	Pois chiches	Lentilles	...	Carottes	Navet	Melon	Pastèque	Haricots verts	Fèves Vertes	Oignons	Courgettes	Concombre	Fraisier
2003-01-01	2603.0	1235.7	1043.1	15.3	598.1	67.5	131.6	35.8	91.6	9.5	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2004-01-01	3190.1	909.1	4124.8	1.7	649.7	47.9	134.1	19.8	80.3	35.1	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2005-01-01	2250.5	896.1	949.1	24.3	1224.0	124.6	20.0	7.8	26	35	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2006-01-01	2651.6	727.9	853.1	19.2	601.3	120.9	69.5	17.4	38.1	1.8	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2007-01-01	2254.9	605.8	1422.6	21.1	482.8	106.8	25.7	8.6	28	10.6	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2008-01-01	2222.5	605.8	898.2	26.9	874.3	100.5	20.3	14,4	39,9	6,6	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2009-01-01	3824.4	930.5	2042.8	1.6	849.3	119.6	31.0	26,5	38,9	13,5	...	27.5	12.8	559.8	96.4	1.4	18.3	81.2	13.8	0.8	93.7
2010-01-01	5282.9	1303.7	2082.4	34.1	1789.3	216.6	90.0	22,8	61,5	21,7	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2011-01-01	5117.3	1628.8	1901.6	0.5	1360.7	235.8	76.0	38,5	39,5	7,8	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2012-01-01	4577.6	1203.1	1422.4	22.5	1971.9	218.2	79.4	29,1	46,7	17,9	...	53.0	19.8	780.2	112.4	6.2	39.4	177.1	20.2	4.0	150.4
2013-01-01	5250.3	1445.8	1880.7	32.4	1038.5	265.1	64.8	21,9	90,8	13,1	...	62.0	28.3	723.9	316.4	6.6	51.4	200.1	53.8	6.5	173.8
2014-01-01	6808.8	1626.7	2701.6	20.1	2286.8	480.5	93.8	213,7	88,5	18,6	...	50.1	37.4	810.2	300.1	2.4	146.1	166.3	22.0	12.4	281.6
2015-01-01	5005.9	1651.1	1975.1	24.5	1212.0	277.1	62.5	25,9	121,3	23,3	...	34.9	28.7	827.9	217.1	2.7	53.7	142.8	15.7	7.9	281.6

	Cereales	Ble dur	Ble tendre	Mais	Orge	Legumineuses	Feves	Petits pois	Pois chiches	Lentilles	...	Carottes	Navet	Melon	Pastèque	Haricots verts	Fèves Vertes	Oignons	Courgettes	Concombre	Fraisier
2003-01-01	2603.0	1235.7	1043.1	15.3	598.1	67.5	131.6	35.8	91.6	9.5	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2004-01-01	3190.1	909.1	4124.8	1.7	649.7	47.9	134.1	19.8	80.3	35.1	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2005-01-01	2250.5	896.1	949.1	24.3	1224.0	124.6	20.0	7.8	26	35	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2006-01-01	2651.6	727.9	853.1	19.2	601.3	120.9	69.5	17.4	38.1	1.8	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2007-01-01	2254.9	605.8	1422.6	21.1	482.8	106.8	25.7	8.6	28	10.6	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2008-01-01	2222.5	605.8	898.2	26.9	874.3	100.5	20.3	14,4	39,9	6,6	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2009-01-01	3824.4	930.5	2042.8	1.6	849.3	119.6	31.0	26,5	38,9	13,5	...	27.5	12.8	559.8	96.4	1.4	18.3	81.2	13.8	0.8	93.7
2010-01-01	5282.9	1303.7	2082.4	34.1	1789.3	216.6	90.0	22,8	61,5	21,7	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2011-01-01	5117.3	1628.8	1901.6	0.5	1360.7	235.8	76.0	38,5	39,5	7,8	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2012-01-01	4577.6	1203.1	1422.4	22.5	1971.9	218.2	79.4	29,1	46,7	17,9	...	53.0	19.8	780.2	112.4	6.2	39.4	177.1	20.2	4.0	150.4
2013-01-01	5250.3	1445.8	1880.7	32.4	1038.5	265.1	64.8	21,9	90,8	13,1	...	62.0	28.3	723.9	316.4	6.6	51.4	200.1	53.8	6.5	173.8
2014-01-01	6808.8	1626.7	2701.6	20.1	2286.8	480.5	93.8	213,7	88,5	18,6	...	50.1	37.4	810.2	300.1	2.4	146.1	166.3	22.0	12.4	281.6
2015-01-01	5005.9	1651.1	1975.1	24.5	1212.0	277.1	62.5	25,9	121,3	23,3	...	34.9	28.7	827.9	217.1	2.7	53.7	142.8	15.7	7.9	281.6

FIGURE 10 : DATA PREPROCESSING PREVIEW

1.3. Data visualization:

Our data is numerical time series across the years, so it can be simply visualized with a line plot.

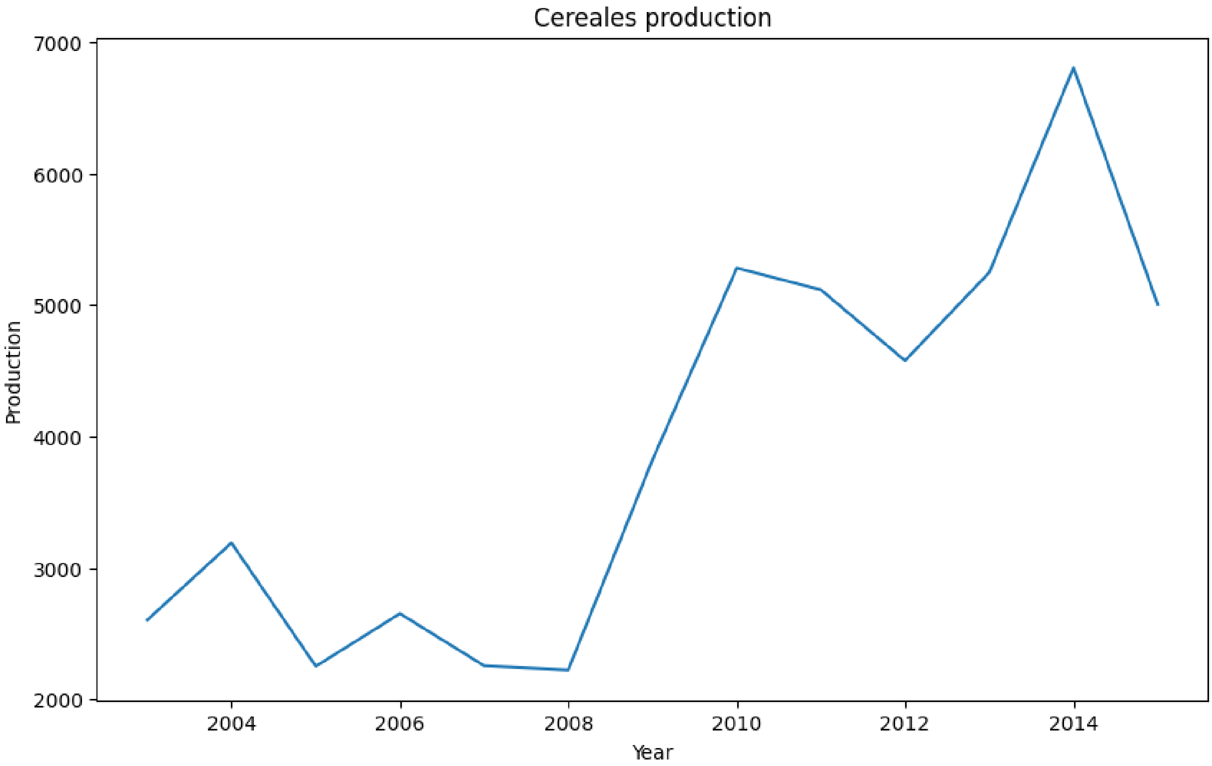


FIGURE 11 : EXAMPLE OF A PRODUCT PRODUCTION VISUALIZATION

2. Modeling and analysis :

As we stated earlier, the models we are using at the present time are ARIMA and RNNs.

2.1. ARIMA :

Starting with ARIMA, we first need to find it's parameters (**p**, **d**, and **q**).

2.1.1. How to find the order of differencing (d) in ARIMA model :

First, I'll use the Statsmodels package's Augmented Dickey Fuller test (ADF Test) to determine whether the series is stationary. The rationale for this is because we only require differencing if the series is not stationary. Otherwise, no differencing is required; that is, d=0.

The null hypothesis (Ho) for the ADF test is that the time series is not stationary. So, if the test's p-value is smaller than the significance level (0.05), we reject the null hypothesis and conclude that the time series is really stationary.

So, in our situation, if the P value is more than 0.05, we proceed with determining the order of difference.

```
ADF Statistic: -2.115530
p-value: 0.238316
```

Since p is smaller than 0.5, there is no need to difference the series, we can fix d on 0.

2.1.2. How to find the order of the AR term (p) :

The next step is to determine whether the model requires any AR terms. We will determine the needed number of AR terms by examining the **Partial Autocorrelation (PACF)** plot. **Partial autocorrelation** may be defined as the correlation between a series and its lag after eliminating the effects of intermediate delays. So, PACF represents the pure correlation between a lag and the series. This will allow us to determine whether or not the lag is required in the AR term.

The partial autocorrelation of lag (k) of a series is the lag's coefficient in the autoregression equation of Y.

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_3 Y_t - 3$$

Assuming Y_t is the current series and Y_{t-1} is the lag 1 of$Y$, the partial autocorrelation of lag 3 Y_{t-3} is the coefficient α_3 of Y_{t-3} in the equation above.

Now we need to determine the number of AR words. Any autocorrelation in a stationarized series can be eliminated by including enough AR terms. So, we set the order of the AR term to be equal to the number of delays that pass the significance limit in the PACF plot.

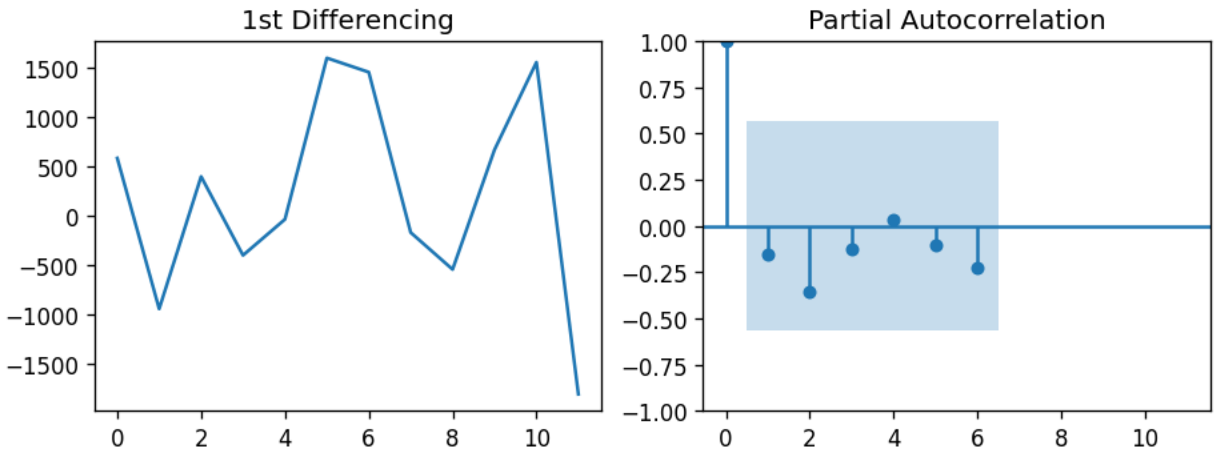


FIGURE 12 : PARTIAL AUTOCORRELATION (PACF) PLOT

We can see that the PACF lag 1 is quite significant since it is well above the significance line. So, we will fix the value of p as 1.

2.1.3. How to find the order of the MA term (q) :

We will examine the ACF plot to determine the number of MA terms in the same way that we examined the PACF plot for the number of AR terms. An MA term is officially defined as the delayed forecast's inaccuracy.

The ACF indicates how many MA terms are necessary to eliminate autocorrelation from the stationarized series.

Let us look at the autocorrelation plot of the differenced series:

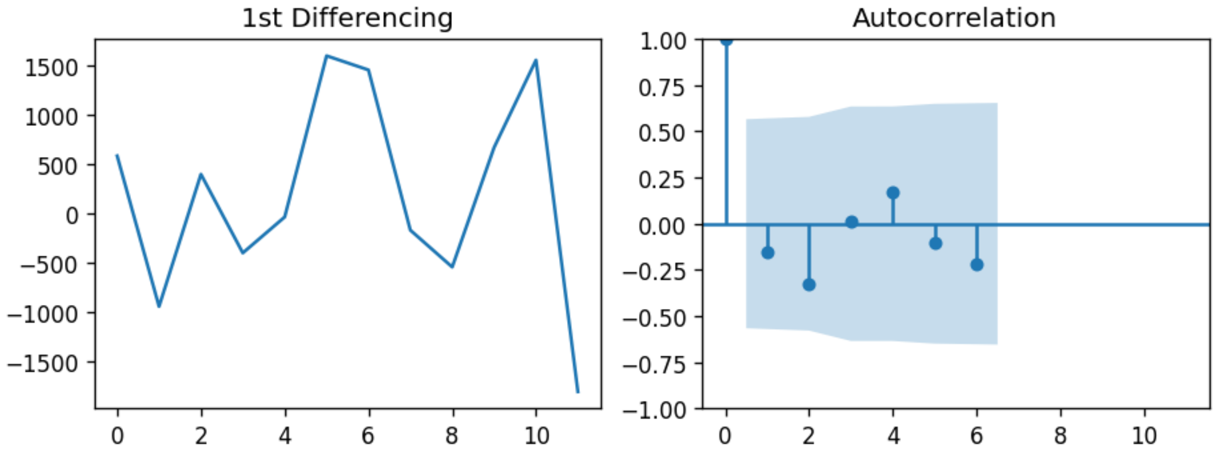


FIGURE 13 : AUTOCORRELATION FUNCTION (ACF)

We can see that one lag is well above the significance line. So, we will fix q as 1. If there is any doubt, we will go with the simpler model that sufficiently explains the Y.

2.1.4. Results :

The resulting model is the following :

```
                               SARIMAX Results
==============================================================================
Dep. Variable:               Cereales   No. Observations:                   13
Model:                 ARIMA(0, 1, 1)   Log Likelihood                -100.244
Date:                Thu, 29 Feb 2024   AIC                            204.487
Time:                        22:14:48   BIC                            205.457
Sample:                    01-01-2003   HQIC                           204.128
                         - 01-01-2015
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.0846      0.258     -0.328      0.743      -0.590       0.421
sigma2      1.049e+06   5.64e+05      1.859      0.063   -5.68e+04    2.16e+06
===================================================================================
Ljung-Box (L1) (Q):                   0.27   Jarque-Bera (JB):                 0.39
Prob(Q):                              0.60   Prob(JB):                         0.82
Heteroskedasticity (H):               4.03   Skew:                            -0.12
Prob(H) (two-sided):                  0.21   Kurtosis:                         2.15
===================================================================================
```

Let's plot the residuals to ensure there are no patterns (that is, look for constant mean and variance).

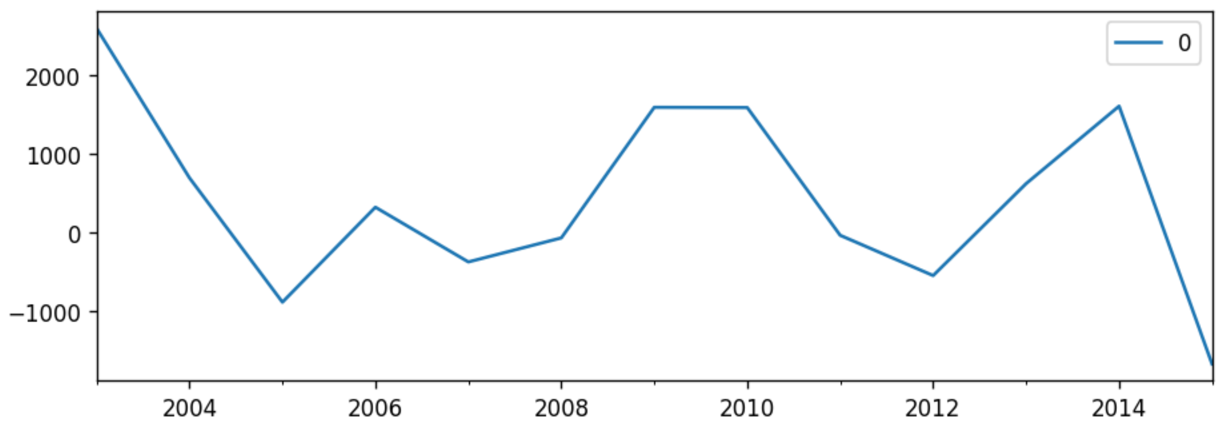


FIGURE 14 : RESIDUALS PLOT

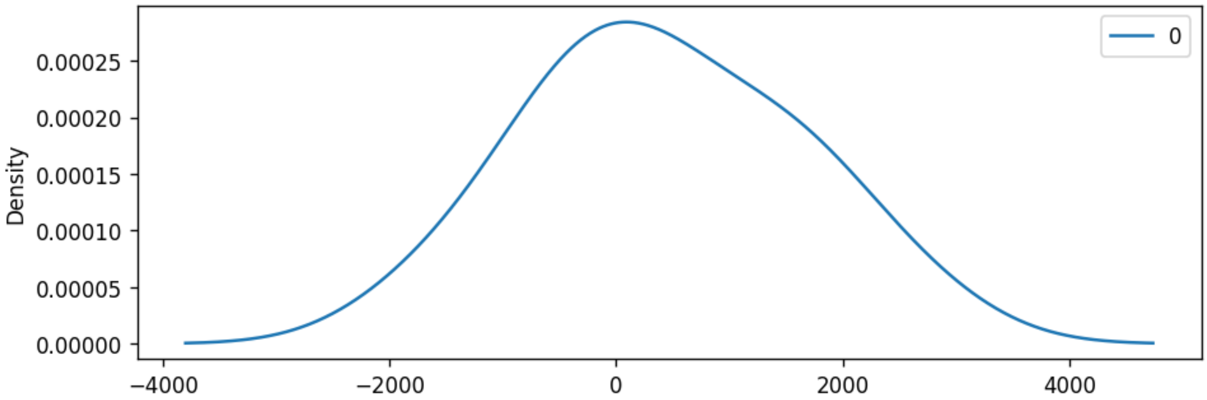


FIGURE 15 : DENSITY PLOT

The residual errors seem fine with near zero mean and uniform variance.

The prediction for the 2016 year seems to be logical :

```
2016-01-01    5146.949716
```

3. App design and solution :

Harvest Guard is a unique programme that provides the public with access to extensive agricultural statistics and data, as well as a secure login section for authorised members. The Harvest Guard app's purpose is to democratise access to agricultural information by providing users with important insights into production patterns, regional data, and other topics. Harvest Guard provides a user-friendly interface and easy navigation, allowing users to easily study and interact with agricultural data.

Furthermore, Harvest Guard provides a special dashboard for authorised members that includes dynamic data and distribution insights, ensuring that stakeholders have real-time knowledge to make educated decisions. Harvest Guard aims to bridge the gap between data and action, resulting in a more transparent and efficient agricultural ecosystem for all stakeholders.

3.1. User experience :

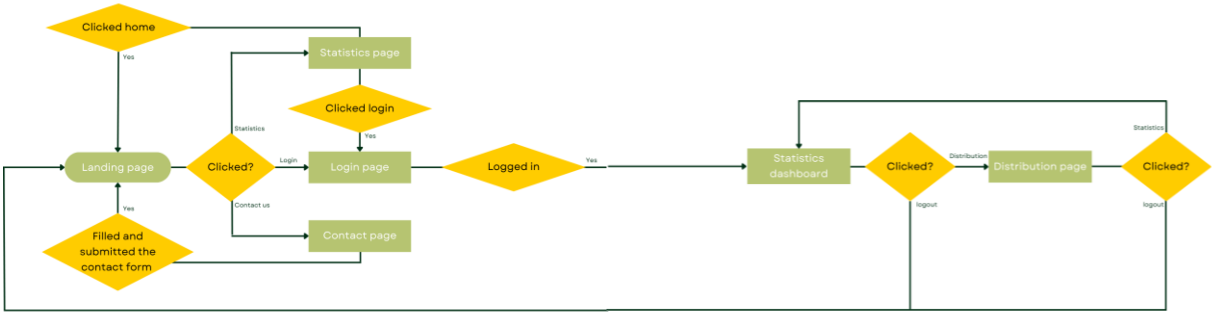


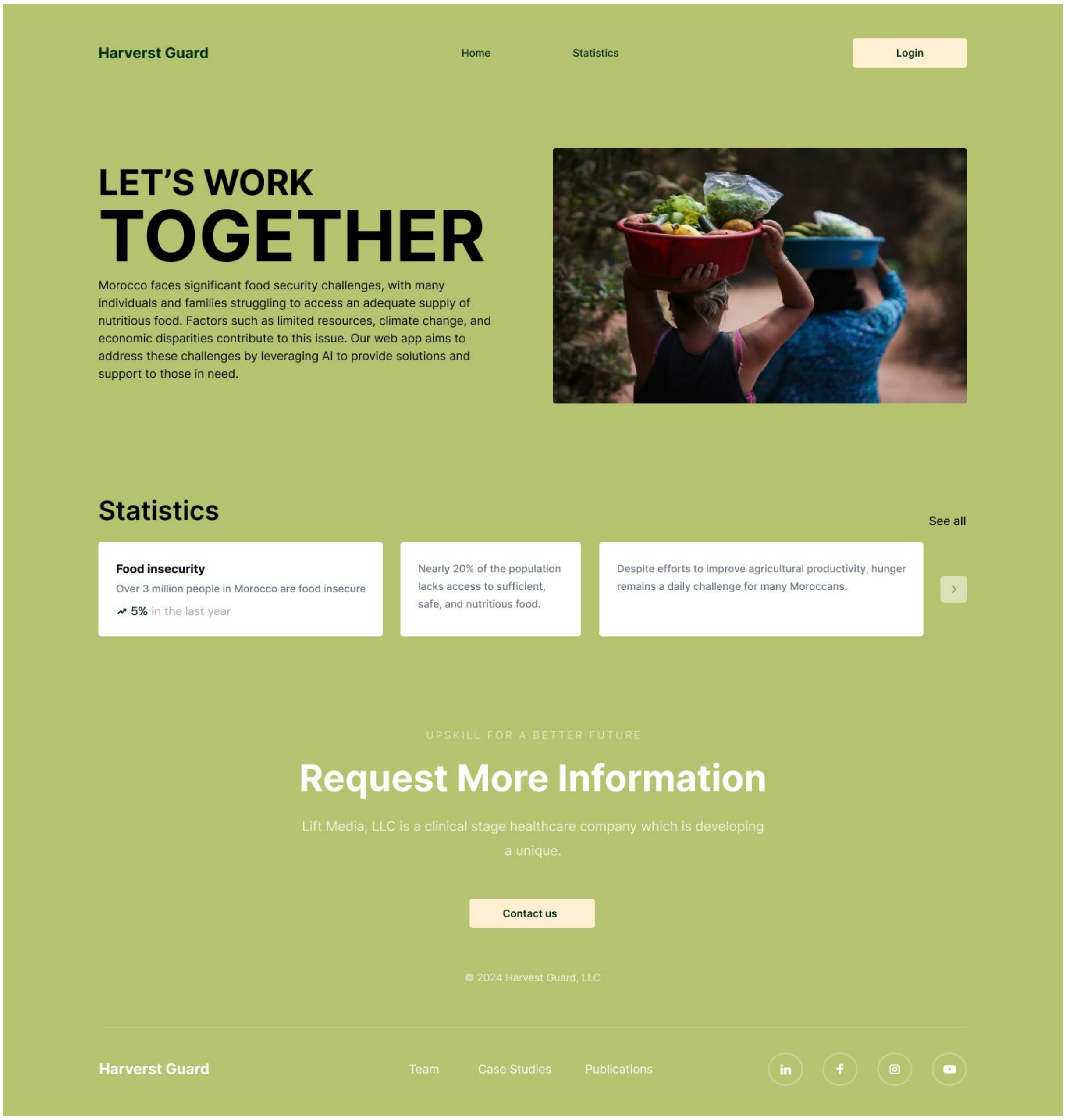
FIGURE 16 : UX WORKFLOW

3.2. User interface :

3.2.1. Landing page :

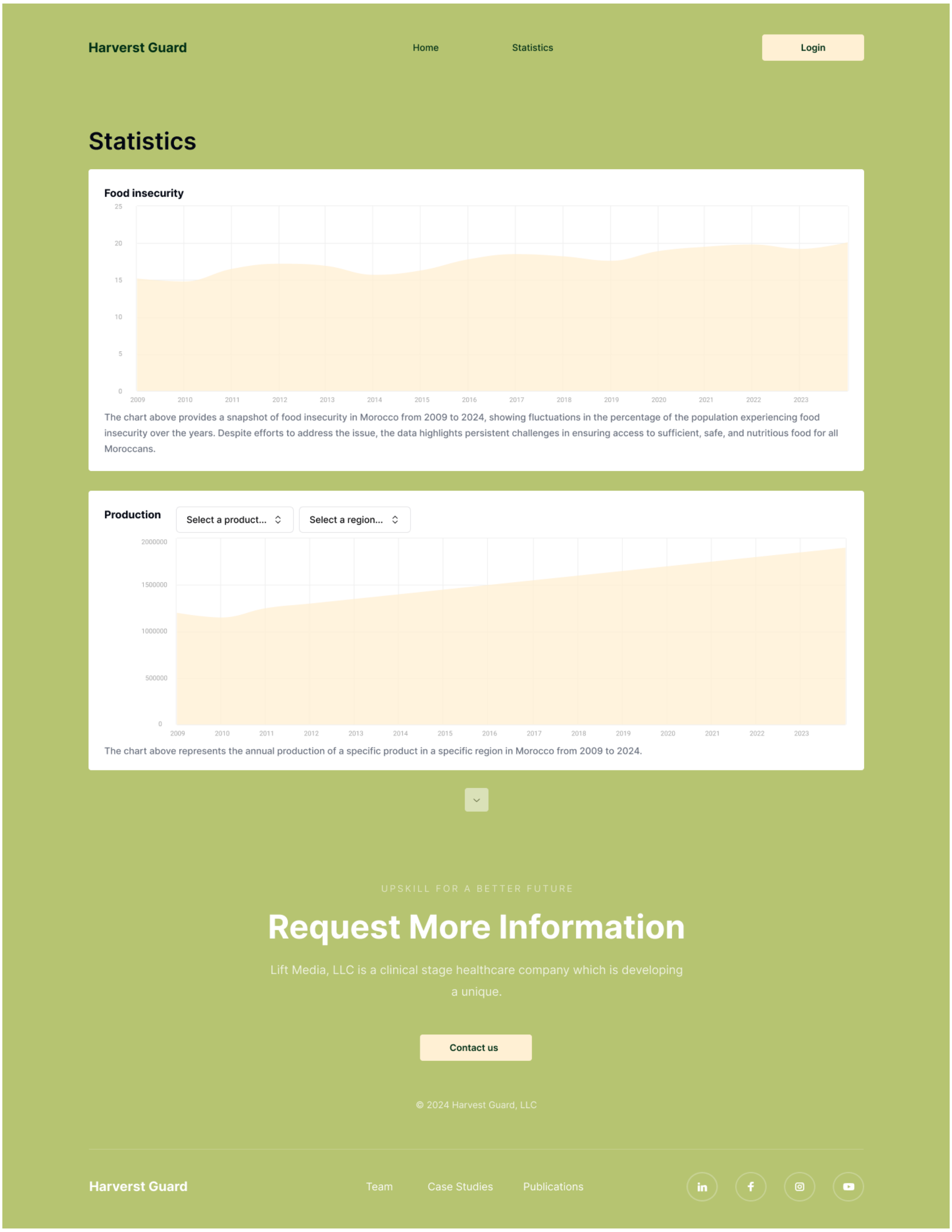
The landing is the page where the user lands when entering the app for the first time.

It consists of a navbar with a login button, a small introduction and small statistics section.

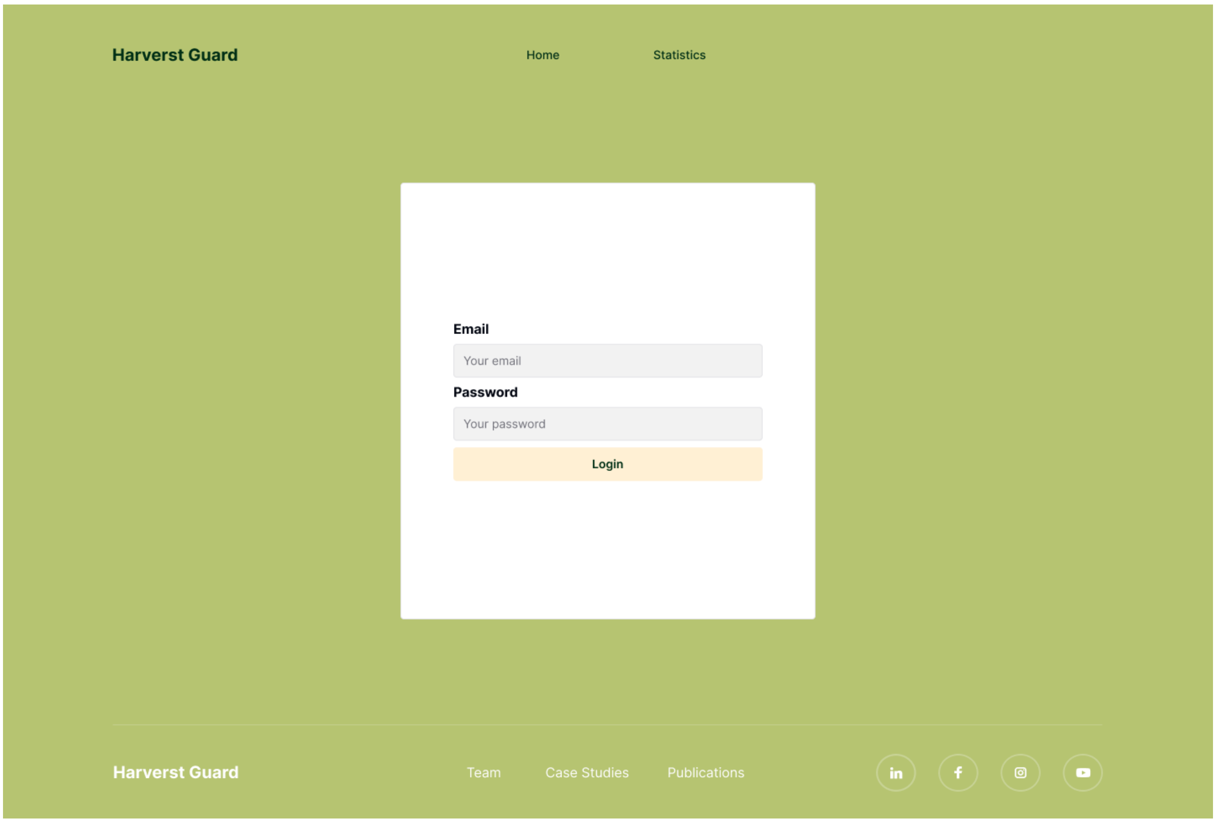


3.2.2. Statistics page :

It's a simple page with different statistics and graphs about the food security problem.

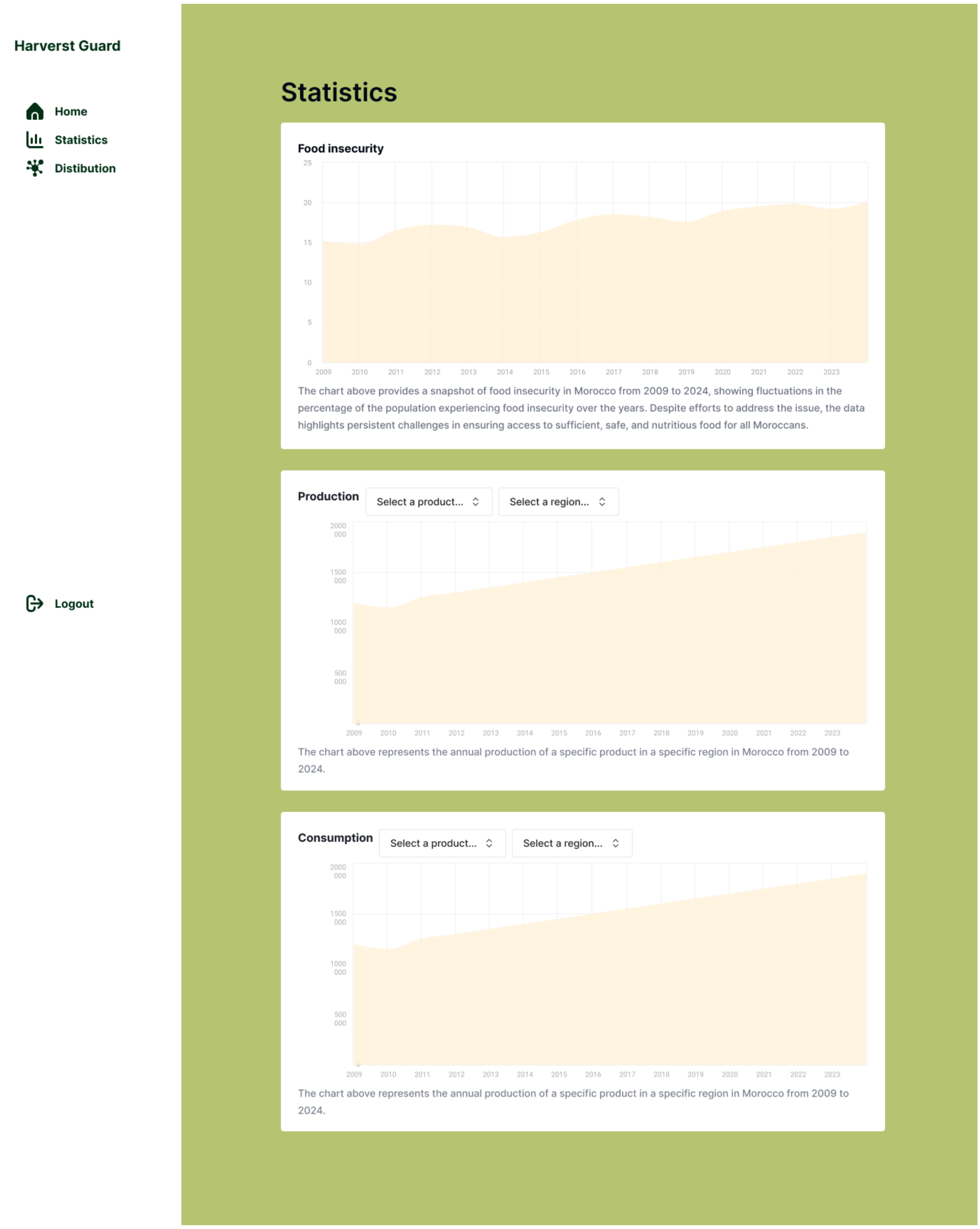


3.2.3. Login page :
The page where authorized user can login to the dashboard.



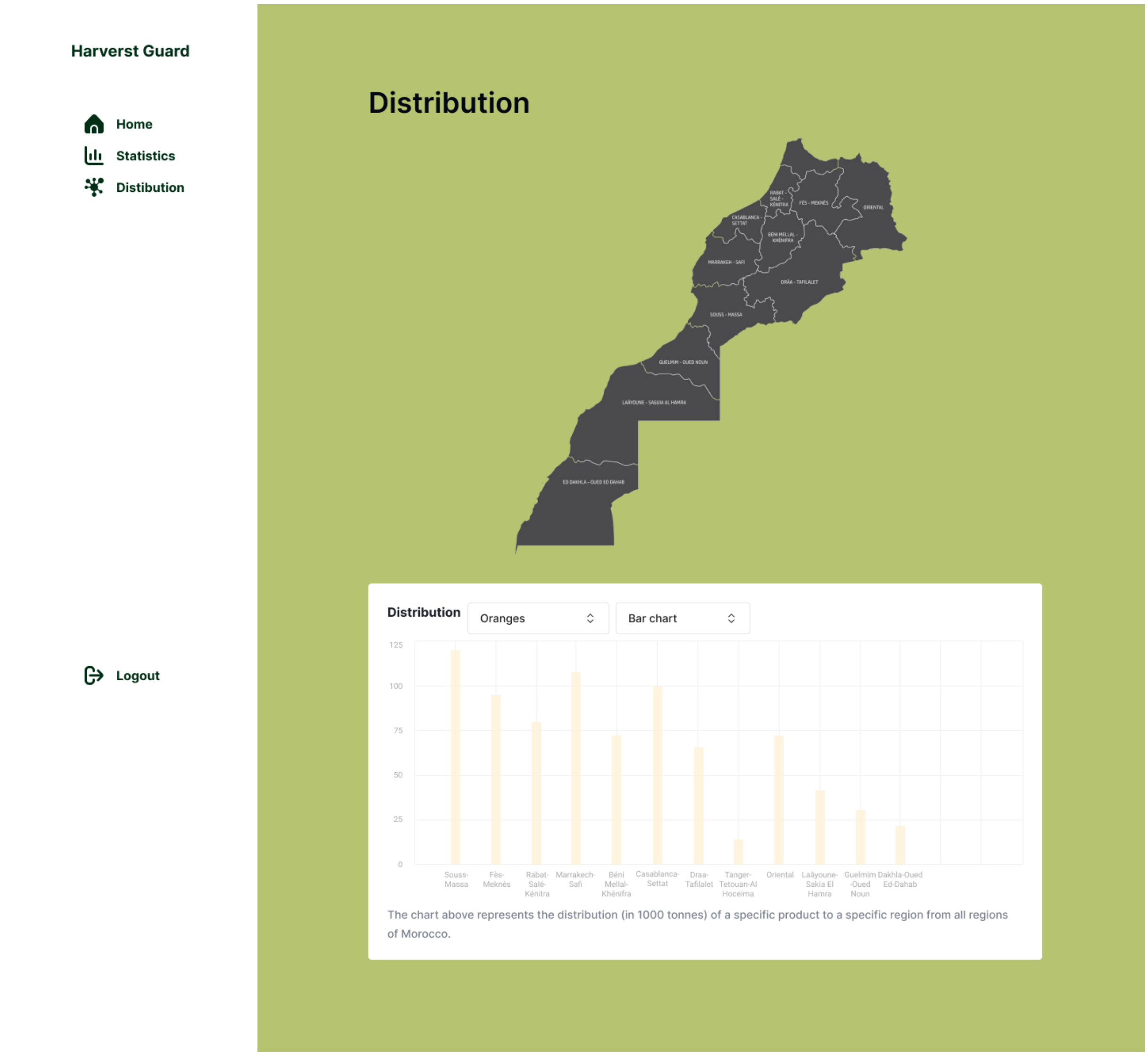
3.2.4. Dashboard :

Once the authorized user logs in, he/she will land on a dashboard with extensive and dynamic statistics for a more in depth view about the situation.



3.2.5. Distribution :

The distribution page is where the authorized can choose a region and the optimized distribution to that region with the shipment size of each product from each region.


Harverst Guard
Home
Statistics
Distibution
Logout
Distribution
Distribution
Oranges
Bar chart
125
100
75
50
25
0
Souss-Massa
Fès-Meknès
Rabat-Salé-Kénitra
Marrakech-Safi
Béni Mellal-Khénifra
Casablanca-Settat
Draa-Tafilalet
Tanger-Tetouan-Al Hoceima
Oriental
Laâyoune-Sakia El Hamra
Guelmim-Oued Noun
Dakhla-Oued Ed-Dahab
The chart above represents the distribution (in 1000 tonnes) of a specific product to a specific region from all regions of Morocco.

General conclusion

Finally, the Harvest Guard initiative is an important step towards tackling Morocco's complicated difficulties with agricultural data accessibility and openness. We want to democratise access to agricultural information by creating the Harvest Guard application, which provides stakeholders with useful insights into production patterns and regional dynamics.

However, it is critical to recognise the ongoing challenge provided by a lack of thorough data, particularly on distribution and consumption trends. Despite our best efforts, this constraint impedes the full realisation of the project's aims and its future development. Moving forward, resolving the issue of data scarcity is a vital goal, necessitating collaboration among government agencies, local communities, and other stakeholders to ensure the availability of trustworthy and standardised data. Despite this hurdle, we are dedicated to moving on with the Harvest Guard project and using technology to create good change in the agricultural sector, ultimately contributing to greater food security and sustainability in Morocco.

Webographie

Tags | Téléchargements | Site institutionnel du Haut-Commissariat au Plan du Royaume du Maroc. (n.d.). https://www.hcp.ma/downloads/?tag=Annuaires+statistiques+des+r%C3%A9gions

Bouzembrak, Y., & Marvin, H. (2019). Impact of drivers of change, including climatic factors, on the occurrence of chemical food safety hazards in fruits and vegetables: A Bayesian Network approach. Food Control, 97, 67–76. https://doi.org/10.1016/j.foodcont.2018.10.021

Marvin, H., & Bouzembrak, Y. (2020). A system approach towards prediction of food safety hazards: Impact of climate and agrichemical use on the occurrence of food safety hazards. Agricultural Systems, 178, 102760. https://doi.org/10.1016/j.agsy.2019.102760

Sarku, R., Lenfers, U. A., & Clemen, T. (2023). The Application of Artificial Intelligence Models for Food Security: a review. Agriculture, 13(10), 2037. https://doi.org/10.3390/agriculture13102037