

University Mohamed Boudiaf - M'sila
Faculty of Mathematics and Informatics
Computer Science Departement



Dissertation submitted to obtain Master degree

By:

Bentameur Achwak

SUBJECT

**Toward an agricultural investor space:
A plant prediction framework using KNN**

e

defended in front of the jury composed of

Mr. BAHACHE MOHAMED

University of M'sila

Chair

Mr. MEHENNI TAHAR

University of M'sila

Supervisor

Mr. AMROUNE NASEREDDINE

University of M'sila

Examiner

Academic year 2020/2021

University Mohamed Boudiaf - M'sila
Faculty of Mathematics and Informatics
Computer Science Departement



Dissertation submitted to obtain Master degree

By:

Bentameur Achwak

SUBJECT

**Toward an agricultural investor space:
A plant prediction framework using KNN**

dedefended in front of the jury composed of

Mr. BAHACHE MOHAMED

University of M'sila

Chair

Mr. MEHENNI TAHAR

University of M'sila

Supervisor

Mr. AMROUNE NASEREDDINE

University of M'sila

Examiner

Academic year 2020/2021

Acknowledgment

FIRST OF ALL, I'D LOVE TO SHOW MY APPRECIATION AND GRATITUDE TO ALLAH THE ALMIGHTY WHO MADE THE PERFECT CONDITION FOR THIS WORK TO SEE THE LIGHT.

ANXIETY AND DEPRESSION WERE OVERWHELMING WHILE TRYING TO FINISH THIS PROJECT AND IT WAS ONLY POSSIBLE TO FIGHT THROUGH IT WITH THE SUPPORT OF SOME PEOPLE THAT I WANT TO THANK INDIVIDUALLY.

I SHALL GIVE THE CREDIT OF LEADING THIS PROJECT TO MY SUPERVISOR P. MHENNI TAHAR , WHO OFFERED ME THE OPPORTUNITY TO WORK ON THIS PROJECT . LET ALONE SHARING HIS WISE GUIDANCE AND UNCLOUDED THOUGHTS TO MAKE THIS PROJECT GO ON THE RIGHT PATH. SO THANK YOU

THANKS TO MY FAMILY MEMBERS WHO FOR THE SUPPORT DURING MY EDUCATIONAL CAREER OVER THE YEARS .

THANKS ENGINEER JDILAT LAKHDAR ,SABER AND MY FRIEND YUCEF, THANKS TO MY COLLEAGUES WHO MADE THE PERFECT ATMOSPHERE TO MAKE THE LAST TWO YEARS BEARABLE.

THANK YOU ALL.

List of Tables

Table 1.1: Coordinates of M'sila	3
Table 1.2 : Monthly rainfall distribution. Source : ONM M'sila (1990-2019).....	4
Table 1.3: Seasonal distribution of rainfall. Source : ONM M'sila (1990-2019).....	5
Table 1.4 :Breakdown of minimum (m), maximum (M) and average temperatures per month.....	7
Table 1.5: Monthly average wind speed (m/s) Source : ONM (1990-2019).....	8
Table 1.6: Monthly average relative air humidity Source : ONM M'sila (1990-2019)	8
Tableau 1.7: Calculation of evapotranspiration Source : ONM M'SILA (1990-2019)	9
Table 1.8: Average number of hours of sunshine Source : ONM M'sila (1990-2019)	9
Table 1.9: Is Values [9]	11

List of Figures

Figure 1.1: map of m'sila	2
Figure 1.2: Agricultural region of th wilaya of msila	3
Figure 1.3 : Histogram of monthly precipitation variation (mm)	5
Figure 1.4: Seasonal distribution of precipitation in (%)	6
Figure 1.5 :Mean seasonal pattern of precipitation	6
Figure 1.6: Monthly variation of minimum, maximum and average temperatures	7
Figure 1.7: Distribution of the humidity rate by month	9
Figure 1.8: Location of the study region in the EMBERGER climagram	11
Figure 1.9 : profil N° 1 and its environnement	13
Figure 1.10 : Profile Descriptions n°1. [12]	14
Figure 1.11: analytical sheet for profile n°1. [12]	15
Figure 1.12: Districts of M'sila	16
Figure 1.13: map of BoussaAda	16
Figure 1.14:map of Magra	17
Figure 1.15: map of Sidi Amer	17
Figure 2.1: Climate conditions application	21
Figure 2.2 : Precision Farming application	22
Figure 2.3: Smart Greenhouse application	23
Figure 2.4: Great demand for the digitization agricultural sector	25
Figure 3.1 : KNN Algorithm	43
Figure 3.2 : Euclidean Distance between A1 and B1.[17]	44
Figure 3.3 : Project framework	46
Figure 3.4: Characteristics of the region of M'sila geographically on 8 areas	47
Figure 3.5: seven factors influencing growth plants	48
Figure 3.6: Dataset dataset construction	49
Figure 3.7: Dataset csv upload	50

Figure 3.8: Splitting attributes and labels	50
Figure 3.9: Dataset of plants characteristics (part 1)	51
Figure 3.10: Dataset of plants characteristics (part 2)	51
Figure 3.11 : training the model for KNN	52
Figure 3.12 : Euclidean distance code	52
Figure 3.13: training the model (test-train)	52
Figure 3.14 : vector for Characteristics	53
Figure 3.15 : testing the model	54
Figure 3.16: developement tools	55
Figure 3.17: building machine learning model with python	56
Figure 3.18: dataset containing the characteristics of regions of Msila	57
Figure 3.19: Python code of finding the first land similar to the given area of the investor.	58
Figure 3.20: FInding the first two plants of the characteristic lands given by investor.	58

Table of Contents

INTRODUCTION.....	1
1. Context of the study	1
2. Statement of the problem	1
3. Objective	2
4. Motivation.....	2
5. Methodology.....	3
6. Report outline	3
CHAPTER 1 : THE WILAYA OF M'SILA	1
1. Introduction	1
2. The agricultural sector in M'sila area :.....	1
2.1 Geographic :	1
2.2 Agricultural	2
2.3 Climate:	3
2.3.1 Reference station:.....	3
2.3.2 Precipitation (rainfall):	4
2.3.3 Seasonal regime :	5
2.3.4 Temperatures.....	6
2.3.5 Wind:.....	8
2.3.6 Relative humidity:	8
2.3.7 ETP (Evapotranspiration):	9
2.3.8 Sunstroke:	9
2.3.9 Climate summary :	9
2.4 Soil :	11
2.4.1 Soil Definition :.....	11
2.4.2 Main properties of the soil:.....	12
2.4.3 The different soil analyses :	12
3. Studied sections in M'sila :.....	15
3.1 M'sila :	15
3.2 Bousaada (Al-Maathar) :	16
3.3 Magra :	16
3.4 Sidi ammer :	17
4. Conclusion.....	17
CHAPTER 2 : AGRICULTURAL INVESTMENT	19
1. Introduction:	19
2. Agri-Investment:	19

3. Objectives of agricultural investment :.....	20
3.1 Internet-based agriculture for Economic growth:	20
3.2 Smartphone-based agriculture for Economic growth:	20
4. Practical Applications of AI in Agriculture :.....	24
5. Application of machine learning techniques	25
6. Conclusion :.....	26
CHAPTER 3: A PLANT PREDICTION FRAMEWORK USING KNN	28
1. Introduction	28
2. Methodology Steps.....	28
3. K Nearest Neighbours (KNN).....	28
3.1 Why do we need a KNN Algorithm?	29
3.2 How does K-NN work?	30
3.3 Euclidean distance :.....	30
3.4 How to select the value of K in the K-NN Algorithm?[19].....	31
3.5 Advantages of KNN Algorithm:[19]	32
3.6 Disadvantages of KNN Algorithm:[19].....	32
3.7 Steps to implement the K-NN algorithm:	33
4. Dataset construction	34
4.1 Study of Area characteristics and trait extraction.....	34
4.2 Building dataset :.....	35
5. Building machine learning :	36
5.1 Loading dataset.....	36
6. KNN algorithm:.....	38
7. Application Modeling	41
8. Implementation :.....	42
8.1 Development Tools.....	42
8.1.1 Anaconda environment.....	42
8.1.2 Jupyter lab:.....	42
CONCLUSION	46



GENERAL INTRODUCTION

Introduction

GENERAL INTRODUCTION

1. Context of the study

The role of Agriculture is important and very large in the overall economic development, as it played a large role in the prosperity and establishment of different civilizations, and agriculture is of great importance in pushing economic development in general and industrial development in particular in poor countries, especially as it is a source of income for the majority of these countries with its large proportions of employment in terms of providing the necessary goods of food and needs.

Investment is the driving force of economic activity, it is a clear and prominent area in most economic policies for what it produces in economic development, based on investment, the region of M'sila can exploit what is available to it, knowing that our studied state is full of many agricultural possibilities and advantages .therefore, it is not possible to promote its economy without focusing and caring for agricultural investment, because of its importance in obtaining food security for the state through the development of agricultural and food production by developing domestic and external demand. It is also possible to use primary agricultural crops as raw materials in industry and to reduce the unemployment rate. its great importance is to try to get Algeria as a whole out of the crisis of food dependence by increasing production and improving the product.

We see recently that Algeria committed a lot of banks to finance agricultural investments and actually succeeded to be a good base for the advancement of the economy and keep pace with progress, but what we see is that there are vast areas without agriculture and lack of diversity of agricultural products, which indicates that this renaissance is just a prelude, and a lot of work will be performed in the future.

2. Statement of the problem

Despite the large spread of land and there is plant diversity, it is not applied in these lands, because identifying the features of the area costs and what can be planted in it costs more .

As the use of internet services and technologies gives new social life with new opportunities, our research question is: how to take advantage of this opportunity for investors

Introduction

who are faced with ignorance of information about land and ignorance of the required plants in the market limits access to the productive diversity of the state of M'sila .

What we can produce in the state of M'sila based on the characteristics of the Wilaya?

What is the addition that helps investors and the interests of agriculture and the market at the same time?

What is the algorithm used to find compatibility between the characteristics of M'sila and plant species?

3. Objective

In this work, we want to help investors choose plants for the studied area to track what is needed in the market and what they can invest based on the characteristics of the plants and the land.

4. Motivation

During my training in the Directorate of Agriculture for a month and a half , we were struck by the arrival of many investors, not a few of whom are consulting what plants they can invest in and what areas are suitable for it, but the Directorate of Agriculture do not know the accurate response. We were also asked if there is a database of lands and what they can produce (characteristics of soil analysis, climate and characteristics of plants), unfortunately we were answered that such a database is not available.

The ministry of agriculture stressed that this interesting idea comes within the framework of the strategy of the public authorities aimed at encouraging investments for the development of agricultural production, especially strategic crops (cereals, corn, fodder and sugar beet)...), with the aim of enhancing the country's food security and increasing the supply of widely consumed food products.

As for the abundance of real estate, the Ministry of Agriculture and Rural Development stated that it has begun to restore 750 thousand unused hectares to make them available to potential investors.[1] .

Introduction

5. Methodology

Our methodology to develop the framework for predicting preferred plants in a given land is based on the following steps:

1. Study the areas of the state of M'sila to derive the main features of each region of the state and the appropriate plants of the studied area .
2. Obtain the dataset of all the plants and their land characteristics.
3. Choose some land characteristics given by the investor based on the characteristics of the state of M'sila (climate, soil) .
4. Predict the set of lands similar to the land characteristics given by the investor using Machine Learning techniques.
5. Determine the plants cultivated in the set of lands given in 4.

6. Report outline

This document was divided into four chapters:

- ✓ **The First chapter** presents the important concepts: the presentation of M'sila geographically and agriculturally, investment, criteria for the selection of plants and areas; that we need in our work.
- ✓ **The second chapter** In this chapter, we present the types of methods used to predict plants that can be invested in , and mention some previous works proposed in the field of electronic investment to detect plants that can be invested in.
- ✓ **The Third chapter** 1In this chapter we explain our methodology to determine the type of plants. We describe our machine learning model that predicts the type of plants through extracted informations and we present the implementation and experimentation procedures that we will take to develop and test our application in order to determine the type of plants.
- ✓ **And finally**, a conclusion of our work is presented, that contains the important points of our project and its perspectives.

Introduction



Chapter 1: The Wilaya Of M'sila

CHAPTER 1 :

THE WILAYA OF M'SILA

1. Introduction

The region of M'sila is considered one of the most important regions in the country, covering 20 million hectares. M'Sila is part of the steppe region.

2. The agricultural sector in M'sila area :

2.1 Geographic :

The region of M'sila emerged from the administrative division of 1974, according to which in Algeria there were 31 states after there were 15 states, which before this date belonged to the state of Setif, like the state of Bejaia and the state of Bourj bouarrij, while boussaada, Sidi Issa and Ain Al-haggad were subordinate to the titiri (Medea, which in turn became the state of 1984.

M'sila is a connecting point between East and west, north and South. It is bordered on the north by Bourj bouarrij and Al-Boira States, on the Northeast by Setif State, on the northwest by Medea state, on the East by Batna State, on the West and on the southwest by djelfa State, on the southeast by Biskra state. Most of the state is flat, with an altitude of 200-300 m above sea level. It is called the capital of brood, which was an independent Berber Kingdom during the reign of the Romans and was so named for its embrace between the tall Atlas and the desert Atlas.



FIGURE 1.1: MAP OF M'SILA

2.2 Agricultural

M'sila is primarily agricultural, and the Ma'athar region of Bousaada (60 km from the state headquarters and the Maarif municipality, 45 km from the state headquarters) is one of the largest agricultural areas in the state of M'sila and the Algerian country as a whole. Many plants like carrots, turnips, potatoes, lettuce, cauliflower, apricots, pomegranates, grapes, peaches, and apples as a new experience cultivation, among others.

Wheat and barley are among the most important agricultural crops in the region of M'sila, in addition to many types of fruits, such as apricots in the Massif region and pomegranates in the Tarmont region. Kosidar, which specializes in the production of cotton, vegetables and fruits, is 97 km away from the state headquarters.

The state of M'sila is confined between a desert Atlas and a hill, where it derives its characteristic from flat and high terrain, its climate is characterized by weak rainfall (less than 250 mm per year) with high temperature in summer and low in winter and low soil resistant to drift.

The state of M'sila is a pastoral-peasant region :

Pastoral area estimated at 1,185,947 hectares, of which 209863 hectares are allied plants.

The plains area It represents 26% of the area of the Wilayat of M'sila dedicated to the cultivation of grains, vegetables and fruits due to its availability of ground and surface water.

Mountainous area 8% of the total area is covered with forest trees and some major crops.

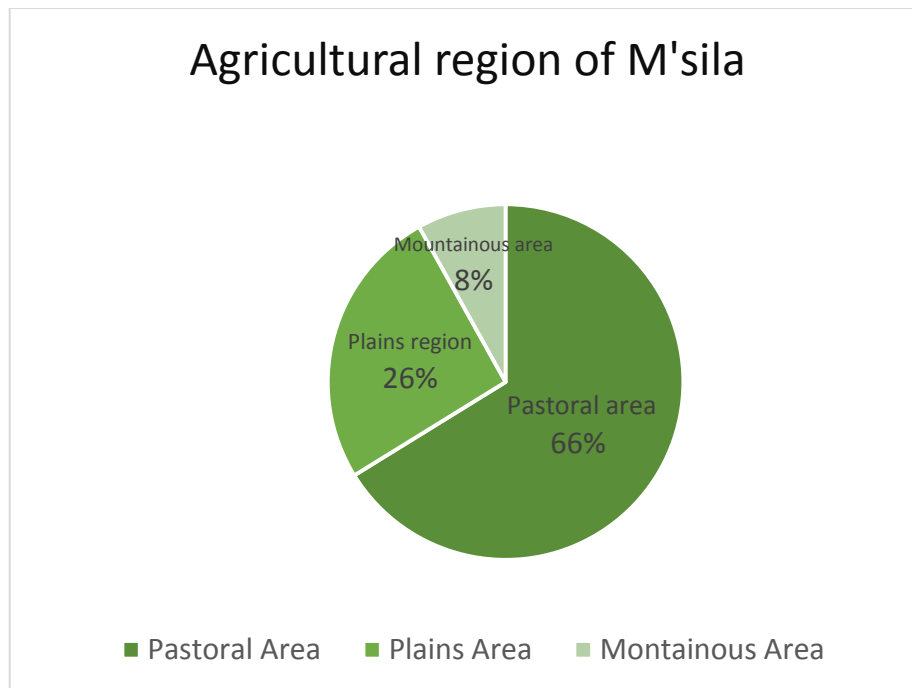


FIGURE 1.2: AGRICULTURAL REGION OF TH WILAYA OF MSILA

2.3 Climate

2.3.1 Reference station

The main climatic characteristics of the study area were identified from the exploitation of data from the nearest meteorological station, that is to say that of M'sila (Source ONM, moyenne1990–2019) [4]

To properly characterize the climate of our study area, we used a series of observations over a reference period of 29 years from 1990 to 2019, recorded in the M'sila meteorological station; which has the coordinates:

Name Station	Lat_deg	Lat_min	Lon_deg	Lon_min	Altitude
M'sila	35	42	4	32	477

TABLE 1.1 COORDINATES OF M'SILA

The study of the climate makes it possible to establish the calculations of the water deficit in relation to each month of the year and each proposed crop.

This is even more important for irrigated crops insofar as climatology allows us to quantify the irrigation water supply for each crop and therefore to know the areas that could be irrigated according to the quantity of water mobilized for irrigation and crops that are proposed.

The climatic factors which very closely condition crop development conditions during the year are analyzed below in order to detect favorable or unfavorable factors in the study area and to quantify their probable effects on crop development.

2.3.2 Precipitation (rainfall)

Monthly and annual rainfall : Rain is among the most important factors due to the beneficial or harmful influence it exerts on plantations and poultry farming.

Mois	Sep	Oct	Nov	Dec	Jan	Feb
Rainfall (mm)	22,13	23,64	22,62	17,67	19,59	43,96
Mar	Apr	Mai	Jun	Jul	Aug	Total
15,81	20,13	23,41	8,90	3,90	6,71	228,47

TABLE 1.2 : MONTHLY RAINFALL DISTRIBUTION. SOURCE : ONM M'SILA (1990-2019)

The annual average rainfall recorded for 29 years (1990/2019) is 228.47 mm with a monthly average of 19.04 mm.

The rainfall in this region is irregular, and we have the following results:

- 1) The rainiest month is February (43.96 mm)
- 2) The driest month is July with (3.90 mm).

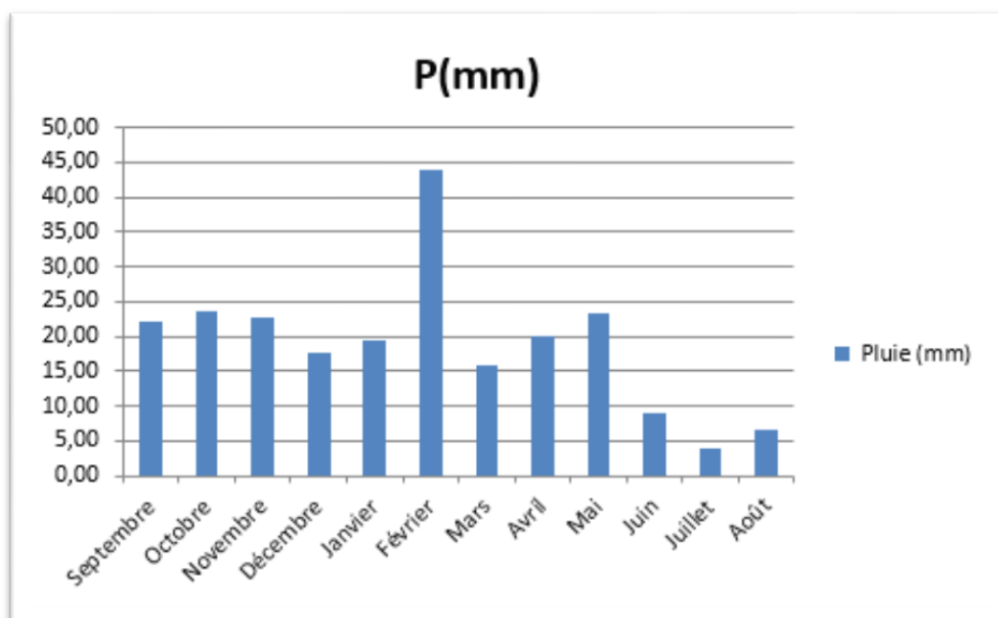


FIGURE 1.3 : HISTOGRAM OF MONTHLY PRECIPITATION VARIATION (MM)

2.3.3 Seasonal regime

The development of perennial and herbaceous crops is closely linked to the seasonal rainfall pattern. In summer, the rains recorded are low, it is necessary to provide irrigation.

Season	Autumn	Winter	Spring	Summer	Annual average
P (mm)	68,39	81,22	59,35	19,51	228,47
P (%)	30%	36%	26%	9%	100%

TABLE 1.3: SEASONAL DISTRIBUTION OF RAINFALL. SOURCE : ONM M'SILA (1990-2019)

From the table we can see that the distribution of precipitation varies from one season to another.

The rainy period begins in autumn with 68.39mm representing 30%, winter with 81.22mm representing 36% and spring with 59.51mm representing 26%. The summer period is marked by 19.51 mm representing 9% of the annual rainfall. The seasonal regime is A-H-P-E (Autumn -Winter-Spring-Summer).

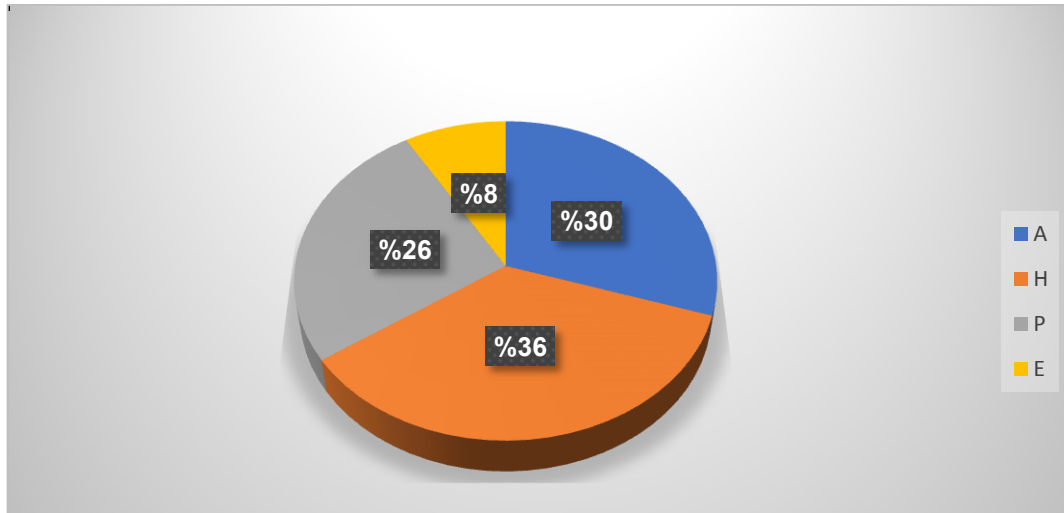


FIGURE 1.4: SEASONAL DISTRIBUTION OF PRECIPITATION IN (%)

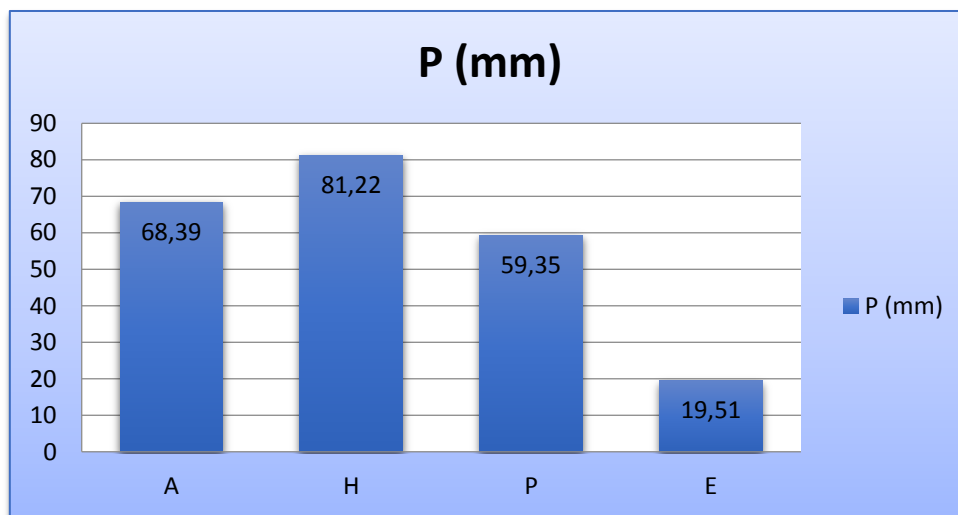


FIGURE 1.5 :MEAN SEASONAL PATTERN OF PRECIPITATION

2.3.4 Temperatures

Temperature is a very important factor in the distribution of climate. It is the extreme temperatures and their frequencies that play the primary roles, in particular for high evapotranspiration at high temperatures and for the risk of freezing at low temperatures.

The temperature of a place depends above all on the location of this place on the earth (its altitude); the climatic zones are determined according to this situation, as there is also a more local variation which depends for example on the shape of the relief, the exposure to the sun, the vegetation cover. The monthly climatic parameter (temperature) recorded at the M'sila station is shown in the following table:

Mois	J	F	M	A	M	J	J	A	S	O	N	D	Avrg
T min (C°)	3,47	4,24	7,56	10,72	16,11	21,02	24,47	23,92	19,37	14,59	8,44	4,58	13,21
T max (C°)	14,32	16,37	19,7	23,66	29	35,2	38,56	38,11	32,18	26,32	19,27	14,7	25,62
T moy (C°)	8,48	9,89	13,54	17,09	22,56	28,26	32,06	31,22	25,61	20,23	13,17	9,16	19,27

TABLE 1.4 :BREAKDOWN OF MINIMUM (M), MAXIMUM (M) AND AVERAGE TEMPERATURES PER MONTH Source : ONM M’sila (1990-2019)

Average temperatures

Temperature has an important role in the development of vegetation. In M'sila the average annual temperature is 19.27°C, which corresponds to a cold region ($T^{\circ} < 20^{\circ}C$). The hot season is spread over 4 months.

More important are the extreme temperatures especially the summer temperatures whose high values.

The two essential factors to consider for plant growth and development relate:

- The intensity and duration of winter cold (winter dormancy period).
- To summer drought.

In January and February the average temperature oscillates between 8.48 °C and 9.89°C, on the other hand in summer it is high ($> 32^{\circ}C$).

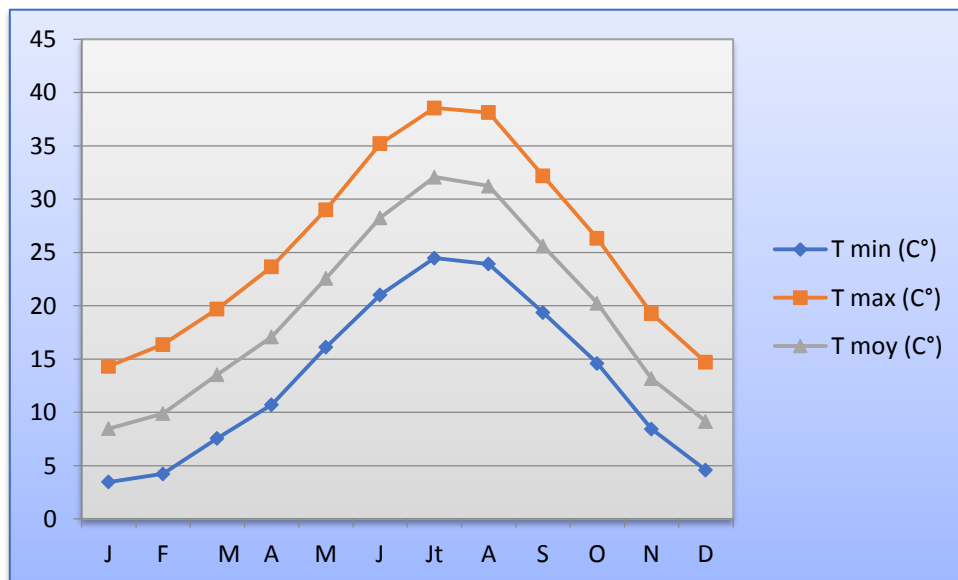


FIGURE 1.6: MONTHLY VARIATION OF MINIMUM, MAXIMUM AND AVERAGE TEMPERATURES

Extreme average temperatures

To better understand the thermal variations in our region, it is useful to look at the monthly extreme values because it is their effects that act directly on plant life. The lowest minimum temperature is observed in January (3.47 °C). As for the maximum temperature, the highest is 38.56 °C observed in July.

2.3.5 Wind:

The wind is the most striking climatic component in the semi-arid region, it is an important factor to consider in agriculture, it plays an essential role in the phenomenon of pollination, as it can cause the withering of certain plant species sensitive to drought. M'sila by its topography, is exposed to North-West winds in South-East direction.

Mois	J	F	M	A	M	J	Jt	A	S	O	N	D	Moy.
Vent (m/s)	32	35	34	37	31	40	41	39	44	38	28	37	36.33

TABLE 1.5: MONTHLY AVERAGE WIND SPEED (M/S) SOURCE : ONM (1990-2019)

Reading the table above shows that throughout the year the wind speed is greater than or equal to 36.33 m/s (Station M'sila), and therefore our region is exposed to these winds and the silting generated by the force of the winds.

2.3.6 Relative humidity:

It is the percentage ratio between the water vapor pressure actually measured in air at a temperature of T°C and the saturated quantity at the same temperature.

Mois	J	F	M	A	M	J	Jt	A	S	O	N	D	Moy
H (%)	69.87	65.03	59.67	53.55	45.52	37.66	32.72	34.83	49.43	56.79	69.14	74.03	54.02

TABLE 1.6: MONTHLY AVERAGE RELATIVE AIR HUMIDITY SOURCE : ONM M'SILA (1990-2019)

The data in the table show that the relative humidity is low and varies greatly due to the effects of high temperatures and significant thermal amplitudes. They generally vary from 32.72% in July to 74.03% in December for the period 1990 to 2019.

2.3.7 ETP (Evapotranspiration)

It is the emission of water vapor by a liquid surface or by a humid environment (soil, plant) at a temperature below the boiling point. Its approximate determination is made with the evaporimeter placed under shelter.

In the absence of direct measurement of the ETP, only the calculation allows an estimate very close to reality. The Penman-Montheith formula incorporates the most climatic factors directly influencing ETP.

Mois	Sep	Oct	Nov	Dec	Jan	Feb	Mar
ETP	156,93	104,79	49,9	30,97	37,43	57,4	104,9
Apr	May	Jun	July	Aug	cumul		
146,41	195,9	236,55	256,9	224	1602,08		

TABLEAU 1.7: CALCULATION OF EVAPOTRANSPIRATION SOURCE : ONM M’SILA (1990-2019)

Evapotranspiration in the region is very high, its peak reaching an average of 256.9 mm in July. However, its minimum was observed an average of 30.97 mm in the month of December.

2.3.8 Sunstroke

It is the duration during which the sun shines on the ground, it conditions the photosynthetic activity.

Months	J	F	M	A	M	Jun	Jul	A	S	O	N	D	Moy.
Sunstroke (h/j)	6.12	7.59	7.98	9.54	10.03	10.84	10.83	9.96	9.09	7.87	6.54	6.1	8.54

TABLE 1.8: AVERAGE NUMBER OF HOURS OF SUNSHINE SOURCE : ONM M’SILA (1990-2019)

The duration of insolation varies regularly during the year, it increases in summer and reaches its maximum in June with a daily average of 10.84 h/d, then it gradually descends in winter where it reaches the minimum value of 6.12 h/d in January. The monthly average is around 8.54 h/d.

2.3.9 Climate summary

GAUSSEN Ombrothermic Diagram: This diagram is a graphic method where the months are shown on the abscissa and the precipitation on the right and on the left the temperatures on a scale twice that of the precipitation ($P = 2T$).

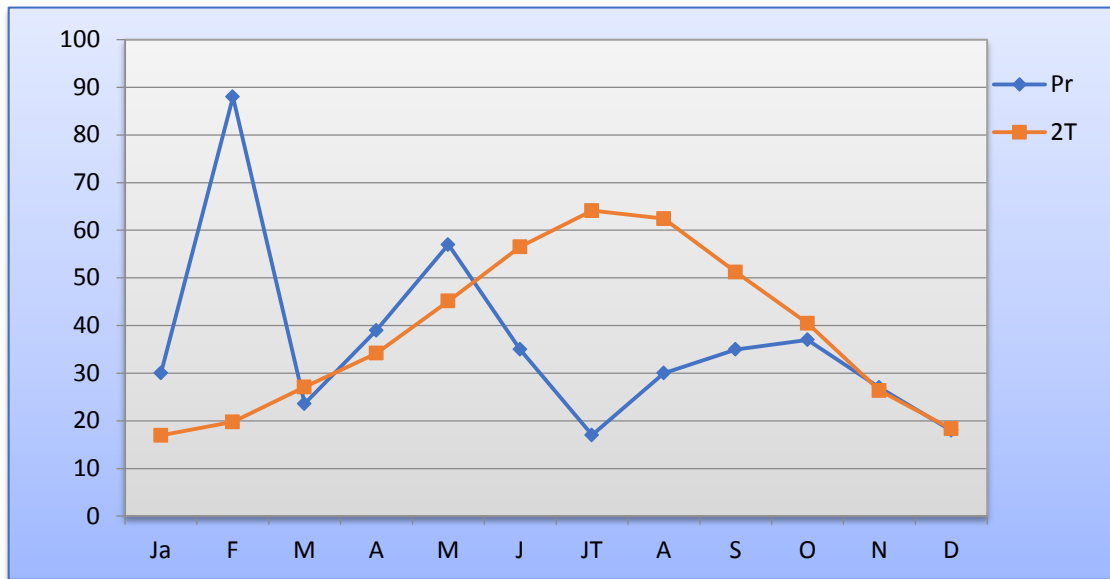


FIGURE 1.7: OMBRO-THERMAL DIAGRAM

For the M'sila station, the diagram thus drawn up shows that the dry period is spread over 5 months (May - October).

The wet season occupies the rest of the year, which translates into a demand for water during this cycle.

Emberger rainfall diagram: To determine the bioclimatic stage of the study area, we apply the formula of Emberger (1932) modified by Stewart in 1969

$$I_s = 3,43. P / M - m$$

- I_s : Stewart index.
- P : Annual precipitation in (mm).
- M : Average maximum temperatures of the hottest month in degrees Kelvin ($^{\circ}K$).
- m : Average minimum temperatures of the coldest month in degrees Kelvin ($^{\circ}K$).

Valeurs d' I_s	Type of climate
$I_s > 100$	Wet
$100 > I_s > 50$	Temperate
$50 > I_s > 25$	Semi-arid
$25 > I_s > 10$	Arid
$10 > I_s$	Desert

TABLE 1.9: IS VALUES [9]

According to data from the M'sila station

- P = 228.47 (mm),
- M = 38.56°C
- m = 3.47°C

We obtained a value of Is equal to 15.82 indicating that our study region has an arid climate.

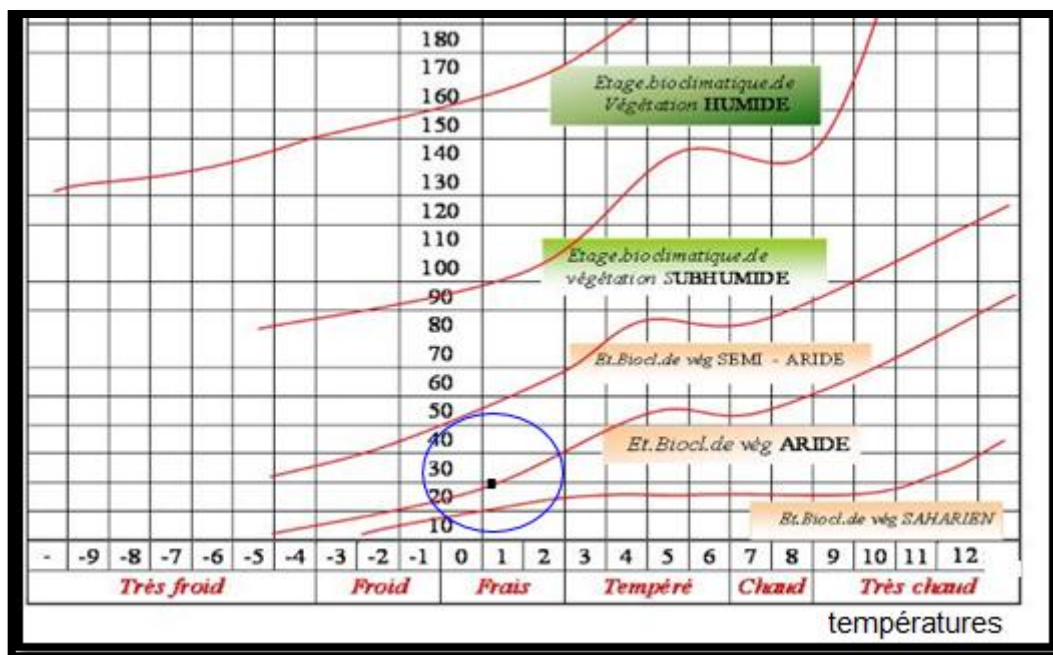


FIGURE 1.8: LOCATION OF THE STUDY REGION IN THE EMBERGER CLIMAGRAM

However, it appears from the location of this index on the Emberger climagram, that the region studied is located in the arid bioclimatic stage with Cool winter. Therefore, it is a climate that presents favorable conditions for agriculture.

2.4 Soil

2.4.1 Soil Definition

The soil is an independent and varying natural body, it is the set of outer horizons of rocks naturally modified by the mutual influence of water, air and living and dead organisms. [11]

It is also a divided medium; the solid phase consists of mineral grains of variable size, and a minority organic fraction. These constituents are not randomly arranged; at different scales ranging from that of the arrangement of solid particles, to that of a section of soil or landscape, we can recognize morphological units, it is an organized environment (Stengel and Gelin, 1998).

2.4.2 Main properties of the soil

The soil is characterized by several properties, the most important of which to determine are :

- texture .
- structure.
- porosity and permeability.
- PH.
- the EC.
- the CEC.
- the rate of the total limestone as well as active.
- the rate of organic matter.

2.4.3 The different soil analyses

Soil analysis is a procedure aimed at characterizing the composition and physico-chemical qualities of a soil.

Analysis

- a. Physical analyses.
- b. Physico-chemical analyses.
- c. Biological.

EX

Discussion

This profile is characterized by an alkaline pH, which varies from 7.51 to 8.00. It is very poor in organic matter throughout the profile, the level of limestone moderate.gypsum, non-saline with an eclectic conductivity of less than 0.5 dS /m. It is a light soil, is characterized by very high sand contents, low silt and clay the latter increases by going deep, the natural fertility is low.



FIGURE 1.9 : PROFIL N° 1 AND ITS ENVIRONNEMENT

Profile Descriptions #1

Date: 9/March/ 2015

Time:sunny

Landform: plain

Venue : forest house

Altitude: 605m

Contact information

X : 2° 56' 11.26'' Est/ Y : 35° 36' 11.16'' Nord

External drainage: good

Occupation/Use: Aleppo pine

Slope: flat

Magnified surface features: some gravel

Parent material: alluvium

Erosion: no erosion

Classification

C.P.C.S: little evolved soil with modal alluvial contribution

0 – 10 cm: dry; 10YR 7/6 (Bright Yellowish Brown); sandy-loamy texture; particle structure; porous; strong reaction to HCl; moderately organic; low biological activity; the roots few; few coarse elements; gradual transition. 10 – 38 cm: fresh; 10YR 5/8 (Yellowish Brown); sandy-loamy texture; particle structure; pores; strong reaction to HCl; low organic matter; low biological activity; very few roots; no coarse elements; clear and gradual transition.

38 – 50 cm: dry; 10YR 5/8 (Brownish yellow); sandy texture; particle structure; many more pores; strong reaction to HCl; low organic matter; low biological activity; very little root; no coarse elements; clear and gradual transition. 50-75cm: dry; 10YR 6/8 (Bright Yellowish Brown); sandy texture; particle structure; many pores; strong effervescence with HCl; low organic matter; no biological activity; very few roots; no coarse elements, clean transition. >75 cm alluvial

Figure1.11:Profile Descriptions n°1. [12]

Horizon		H1	H2	H3	H4
Profondeur (cm)		0 – 10	10 – 38	38 – 50	50-75
Analyses chimiques					
humidity (%)		0,81	1,17	0,63	0,58
pH eau 1/ 2.5		7,51	7,7	7,53	8
C.E dS/m (%) à 25 °C		0,20	0,13	0,29	0,29
Gypse %		12,24	12,94	12,62	24,89
Calcaire totale %		5,18	4,56	4,15	0,21
Complexe adsorbant (méq/100g de terre)	K ⁺	0,52	2,12	0,45	0,52
	Na ⁺	0,51	0,44	0,44	0,48
	Ca ⁺	0,50	0,48	0,34	0,30
	CEC	6,5	6	5	5
ESP (Na/CEC)		7.8	7.3	8.8	9.7
Analyses biologiques					
C %		0,28	0,31	0,15	0,06
MO %		0,48	0,54	0,27	0,11
Analyses physiques					
Granulometrie en %	SG	29,6	13,8	49,06	33,53
	SF	42,99	65,04	31,99	48,45
	LF	14,77	9,45	5,82	2,14
	LG	4,58	5,64	8,6	6,32
	A	8,06	6,07	4,53	9,56

FIGURE 1.11: ANALYTICAL SHEET FOR PROFILE N°1. [12]

3. Studied sections in M'sila

3.1 M'sila

All districts except Bousaada, Sidi Amer and Makara, the district of Ain Hajl was chosen in this study.



FIGURE 1.12: DISTRICTS OF M'SILA

3.2 Bousaada (Al-Maathar)



FIGURE 1.13: MAP OF BOUSSAADA

3.3 Magra

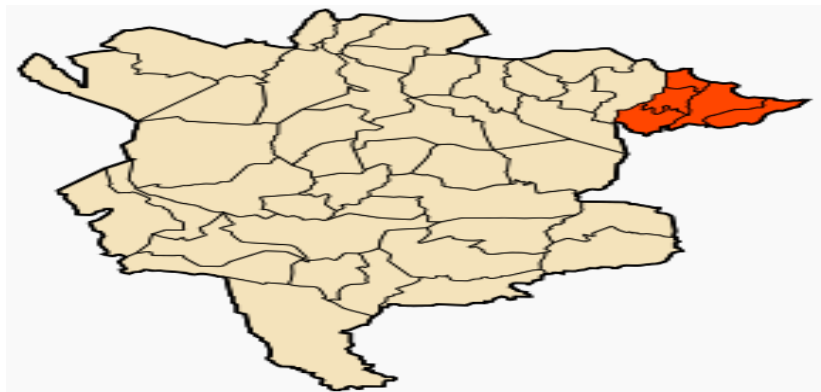


FIGURE 1.14:MAP OF MAGRA

3.4 Sidi ammer



FIGURE 1.15: MAP OF SIDI AMER

4. Conclusion

In this chapter, we touched on the wilaya of M'sila, based on the advice of agricultural managers and some engineers in the field of agriculture. In the next chapter, we will talk about previous work in the field of agriculture.



Chapter 2: agricultural investment

CHAPTER 2 :

AGRICULTURAL INVESTMENT

1. Introduction

In this chapter, we will clarify the concept of agricultural investment and its role in land afforestation. We will also clarify the types of methods that are used to detect Suitable plants for investment, and we will mention some proposed methods in the field of e-agricultural investment to detect plants ..

2. Agri-Investment

Agri-Investment is one of the fastest growing areas of Agri-Investment Research and can be defined as 'the use of digital technologies and new media to Economic growth, Agricultural sector, prevention or treatment as well as to improve agricultural delivery and vocational education.

The use of technology may hold great promise. In particular, it has been suggested that internet-based Economic growth interventions have the potential to overcome traditional barriers to agriculture, and also to reduce the demand on Agriculture Directorates at lower costs. Currently, computerized interventions are available for a broad range of Investments, particularly common economic investment such as agricultural investment and Land Reclamation, Ranching and Irrigation and Beekeeping and Forests and Marine fishing and Poultry farming

The act of investment is a decisive moment in man's relationship with nature: man uses his labor force, science and technology to destroy nature and rebuild it in such a way that he can get a better satisfaction of his needs from it.

what's Investing in agriculture: Investing in agriculture means putting your money

feed a growing population and with less land, interest in agriculture production as an investment has grown right along with the world population [5].

3. Objectives of agricultural investment

The objectives of agricultural investment are many and integrated, so that total or partial goals can be achieved in the course of an investment cycle to achieve small goals, and this integration means a lot of social and economic goals are achieved even when they are not targeted for the presence of vertical integration in agriculture ,this integration ends with:

- 1 / providing raw materials for industry and industrial agriculture.
- 2 / development of geographical areas or intended groups of people.
- 3 / the necessary commodity and monetary flows for research and development of markets and the environment .
- 4 / Configure nutrition and improve their levels..

3.1 Internet-based agriculture for Economic growth

When the internet began to spread to our home, workplace and government institutions in the mid-1990s, new opportunities began to emerge in the field of agriculture. For in agriculture have been used so far: Provision and transmission Information. This feature has been designed across programs in more or less user-friendly ways. It was done in the best cases. A written communication between a Investors and agricultural directorates.

3.2 Smartphone-based agriculture for Economic growth

Agriculture has gained more importance with the rapid uptake and utilization of the smartphones, which are powerful and monitor the real-time data of the user. Many researchers have claimed that agriculture is a mechanism which can deliver more effective and more accessible agricultural investment care. Since mobile phones have become an extension of individual's behavior, agriculture can be greatly affected by the use of smartphones. Smartphone devices run various software applications (mobile apps) which tap the hardware features of the phone and can provide the

Investor with valuable information in a user-friendly format. When these apps are built with Agriculture Directorates orientation, they can provide specific and Investor-centered and peasants information which is meaningful, accurate, relevant and up-to-date. By the use of mobile apps, Information Providers are able to target the investor, based on their condition and response to requirements.

Regardless of what farming technology offers, mobile devices can perform functions that can track the agricultural condition in real time which has the capability to correlate the Agriculture requirements data to the real-time data provided by the smartphone. This allows monitoring the user's status in real time and providing apt information to the user's.

Climate conditions

Climate plays a very critical role for farming. And having improper knowledge about climate heavily deteriorates the quantity and quality of the crop production. But IoT solutions enable you to know the real-time weather conditions. Sensors are placed inside and outside of the agriculture fields. They collect data from the environment which is used to choose the right crops which can grow and sustain in the particular climatic conditions.

The whole IoT ecosystem is made up of sensors that can detect real-time weather conditions like humidity, rainfall, temperature and more very accurately. There are numerous no. of sensors available to detect all these parameters and configure accordingly to suit your smart farming requirements. These sensors monitor the condition of the crops and the weather surrounding them. If any disturbing weather conditions are found, then an alert is send. What gets eliminated is the need of the physical presence during disturbing climatic conditions which eventually increases the productivity and help farmers to reap more agriculture benefits [7].



FIGURE 2.1: CLIMATE CONDITIONS APPLICATION

Precision Farming

Precision Agriculture/Precision Farming is one of the most famous applications of IoT in Agriculture. It makes the farming practice more precise and controlled by realizing smart farming applications such as livestock monitoring, vehicle tracking, field observation, and inventory monitoring.

The goal of precision farming is to analyze the data, generated via sensors, to react accordingly. Precision Farming helps farmers to generate data with the help of sensors and analyze that information to take intelligent and quick decisions. There are numerous precision farming techniques like irrigation management, livestock management, vehicle tracking and many more which play a vital role in increasing the efficiency and effectiveness.

With the help of Precision farming, you can analyze soil conditions and other related parameters to increase the operational efficiency. Not only this you can also detect the real-time working conditions of the connected devices to detect water and nutrient level.[7]



FIGURE 2.2 : PRECISION FARMING APPLICATION

Smart Greenhouse

To make our greenhouses smart, IoT has enabled weather stations to automatically adjust the climate conditions according to a particular set of instructions. Adoption of IoT in Greenhouses has eliminated the human intervention, thus making entire process cost-effective and increasing accuracy at the same time.

For example, using solar-powered IoT sensors builds modern and inexpensive greenhouses. These sensors collect and transmit the real-time data which helps in monitoring the greenhouse state very precisely in real-time. With the help of the sensors, the water consumption and greenhouse state can be monitored via emails or SMS alerts. Automatic and smart irrigation is carried out with the help of IoT. These sensors help to provide information on the pressure, humidity, temperature and light levels.[8]



FIGURE 2.3: SMART GREENHOUSE APPLICATION

4. Practical Applications of AI in Agriculture

From automated pest and plant disease detection to intelligent spraying and produce sorting here's how Computer Vision is transforming the food and agriculture sector.

- 1/ Crop and soil monitoring
- 2/ Insect and plant disease detection
- 3/ Livestock health monitoring
- 4/ Intelligent spraying
- 5/ Automatic weeding
- 6/ Aerial survey and imaging
- 7/ Produce grading and sorting
- 8/ The future of AI in Agriculture: Farmers as AI engineers?

Throughout human history, technology has long been used in agriculture to improve efficiency and reduce the amount of intensive human labor involved in farming. From improved plows to irrigation, tractors to modern AI, it's an evolution that humans and agriculture have undergone since the invention of farming

The growing and increasingly affordable availability of computer vision stands to become another significant step forward here.

With considerable changes occurring in our climate, environment, and global food needs, AI has the ability to transform 21st century agriculture by:

- Increasing efficiency of time, labor, and resources.
- Improving environmental sustainability.
- Making resource allocation “smarter”.
- Providing real-time monitoring to promote greater health and quality of produce[9].

5. Application of machine learning techniques

There are several applications of Data Mining techniques in the field of agriculture. The K Nearest Neighbour (KNN) is applied for simulating daily precipitations and other weather variables [15], and different possible changes of the weather scenarios are analyzed using SVMs [16].

Data mining techniques are applied to study sound recognition problems. For instance, Fagerlund.S [17] uses SVMs to classify the sound of bird and other different sounds. There are unit 2 major kinds of predictions: one will either try and predict some inaccessible information values or unfinished trends, or predict a category label for a few information. The latter is tied to classification. Once a classification model is constructed supported a coaching set, the category label of associate object may be predicted supported the attribute values of the article and therefore the attribute values of the categories [15]. Prediction is but additional typically remarked the forecast of missing numerical values, or increase/ decrease trends in time connected information. The foremost plan is to use an outsized range of past values to think about probable future values [10].

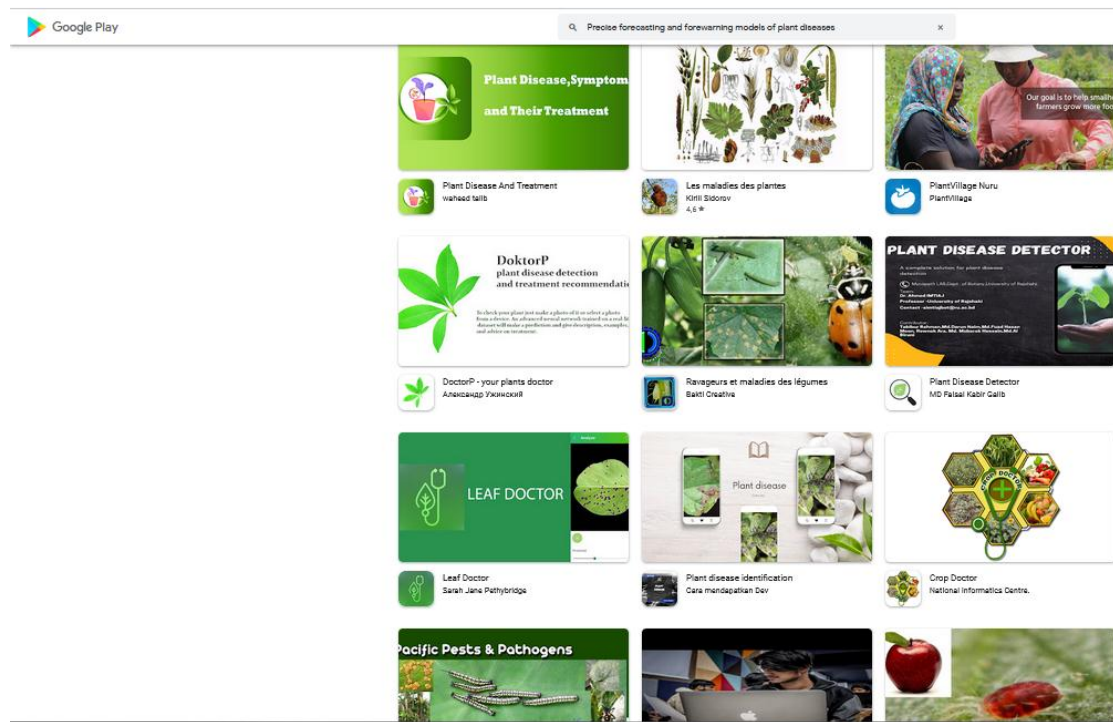
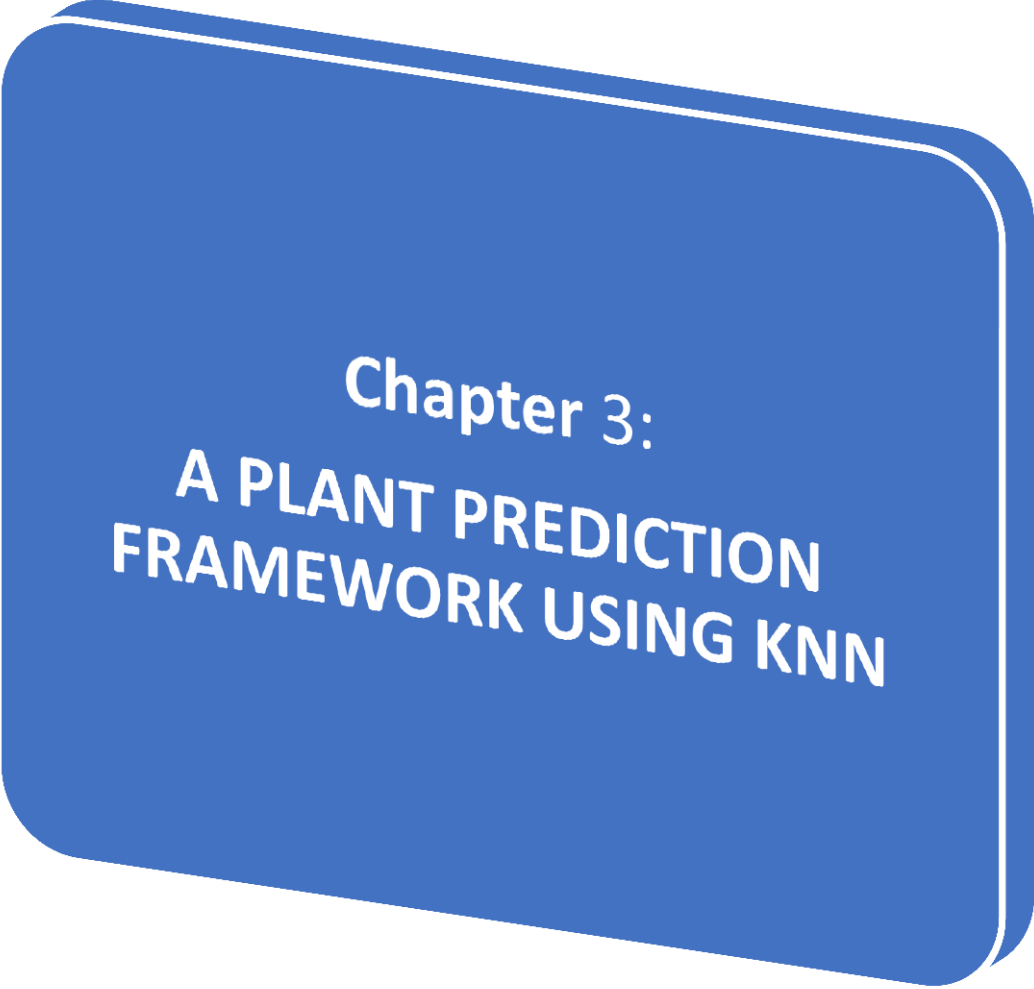


FIGURE 2.4: GREAT DEMAND FOR THE DIGITIZATION AGRICULTURAL SECTOR

6. Conclusion

In this chapter, we have presented different concepts of e-Agri-Investment and some existing works for plant diagnostics through smart phone applications. In the following chapter we will present our own method to implement a smartphone application to diagnosis plants type.



**Chapter 3:
A PLANT PREDICTION
FRAMEWORK USING KNN**

CHAPTER 3:

A PLANT PREDICTION FRAMEWORK USING KNN

1. Introduction

In this chapter we explain our methodology for predicting adjacent areas and determining the type of plants.

1. Building a machine learning model that predicts neighboring areas and determines the type of plants through the extracted information.
2. We present the implementation and trial procedures that we will take to develop and test our application in order to determine the type of plants or select the appropriate area for the type of plants based on the requirements of the investor.

2. Methodology Steps

Our work consists of the following steps:

1. A theoretical and an applied study of the M'sila regions to extract the main characteristics of each region.
2. Data processing and feature selection.
3. Building data sets.
4. Training phase: Building machine learning models from the data set.
5. Test phase.
6. Finding the type of plants
7. Finding the right areas

3. K Nearest Neighbours (KNN)

The purpose of the K nearest neighbors (KNN) method is to classify target points (unknown class) according to their distances with respect to points constituting a learning sample (i.e. whose class is known a priori).

KNN is an intuitive supervised classification approach often used in machine learning. This is a generalization of the nearest neighbor (NN) method. NN is a special case of KNN, where $k = 1$.

The KNN classification approach is based on the assumption that each case of the training sample is a random vector from R^n . Each point is described as $x = \langle a_1(x), a_2(x), a_3(x), \dots, a_n(x) \rangle$ where $a_r(x)$ corresponds to the value I of the r th attribute. $a_r(x)$ can be either a quantitative variable or a qualitative variable.

In order to determine the class of a target point, each each of the k points closest to x_q proceed to a vote. The class of x_q corresponds to the majority class.[17]

- K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.
- K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
- K-NN is a **non-parametric algorithm**, which means it does not make any assumption on underlying data.
- It is also called a **lazy learner algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.
- KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

3.1 Why do we need a KNN Algorithm?

Suppose there are two categories, i.e., Category A and Category B, and we have a new data point x_1 , so this data point will lie in which of these categories. To solve this

type of problem, we need a KNN algorithm. With the help of K-NN, we can easily identify the category or class of a particular dataset. Consider the below diagram: [18]

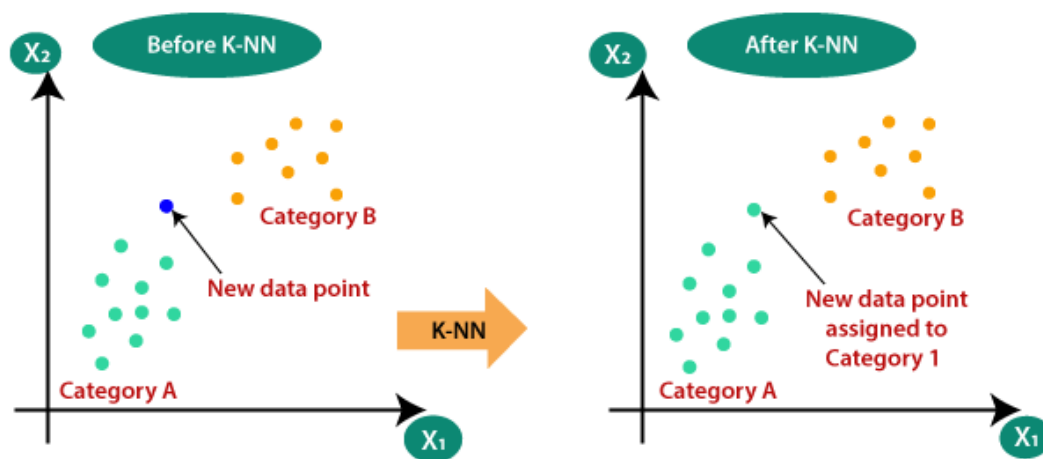


FIGURE 3.1 : KNN ALGORITHM

3.2 How does K-NN work?

The K-NN working can be explained on the basis of the below algorithm:

- **Step-1:** Select the number K of the neighbors
- **Step-2:** Calculate the Euclidean distance of **K number of neighbors**
- **Step-3:** Take the K nearest neighbors as per the calculated Euclidean distance.
- **Step-4:** Among these k neighbors, count the number of the data points in each category.
- **Step-5:** Assign the new data points to that category for which the number of the neighbor is maximum.
- **Step-6:** Our model is ready

3.3 Euclidean distance :

For Euclidean distance between the data points. The Euclidean distance is the distance between two points, which we have already studied in geometry. It can be calculated as: $\sqrt{(x - x')^2 + (y - y')^2}$. [18]

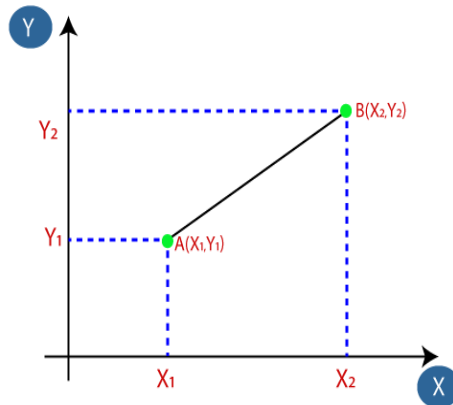


FIGURE 3.2 : EUCLIDEAN DISTANCE BETWEEN A1 AND B1.[17]

3.4 How to select the value of K in the K-NN Algorithm?[19]

There is no straightforward method to calculate the value of K in KNN. You have to play around with different values to choose the optimal value of K. Choosing a right value of K is a process called Hyperparameter Tuning.

The value of optimum K totally depends on the dataset that you are using. The best value of K for KNN is highly data-dependent. In different scenarios, the optimum K may vary. It is more or less hit and trail method.

You need to maintain a balance while choosing the value of K in KNN. K should not be too small or too large.

A small value of K means that noise will have a higher influence on the result. Larger the value of K, higher is the accuracy. If K is too large, you are under-fitting your model. In this case, the error will go up again. So, at the same time you also need to prevent your model from under-fitting. Your model should retain generalization capabilities otherwise there are fair chances that your model may perform well in the training data but drastically fail in the real data. Larger K will also increase the computational expense of the algorithm.

There is no one proper method of estimation of K value in KNN. No method is the rule of thumb but you should try considering following suggestions:

1. Square Root Method Take square root of the number of samples in the training dataset.

2. Cross Validation Method We should also use cross validation to find out the optimal value of K in KNN. Start with K=1, run cross validation (5 to 10 fold), measure the accuracy and keep repeating till the results become consistent. K=1, 2, 3... As K increases, the error usually goes down, then stabilizes, and then raises again. Pick the optimum K at the beginning of the stable zone. This is also called Elbow Method.

3. Domain Knowledge also plays a vital role while choosing the optimum value of K. K should be an odd number.

3.5 Advantages of KNN Algorithm[19]

- 1. No Training Period-** KNN modeling does not include training period as the data itself is a model which will be the reference for future prediction and because of this it is very time efficient in term of improvising for a random modeling on the available data.
- 2. Easy Implementation-** KNN is very easy to implement as the only thing to be calculated is the distance between different points on the basis of data of different features and this distance can easily be calculated using distance formula such as- Euclidian or Manhattan
- 3.** As there is no training period thus new data can be added at any time since it wont affect the model.

3.6 Disadvantages of KNN Algorithm[19]

- 1. Does not work well with large dataset** as calculating distances between each data instance would be very costly.
- 2. Does not work well with high dimensionality** as this will complicate the distance calculating process to calculate distance for each dimension.
- 3. Sensitive to noisy and missing data**
- 4. Feature Scaling-** Data in all the dimension should be scaled (normalized and standardized) properly .

3.7 Steps to implement the K-NN algorithm

- Data Pre-processing step
- Fitting the K-NN algorithm to the Training set
- Predicting the test result
- Test accuracy of the result(Creation of Confusion matrix)
- Visualizing the test set result..

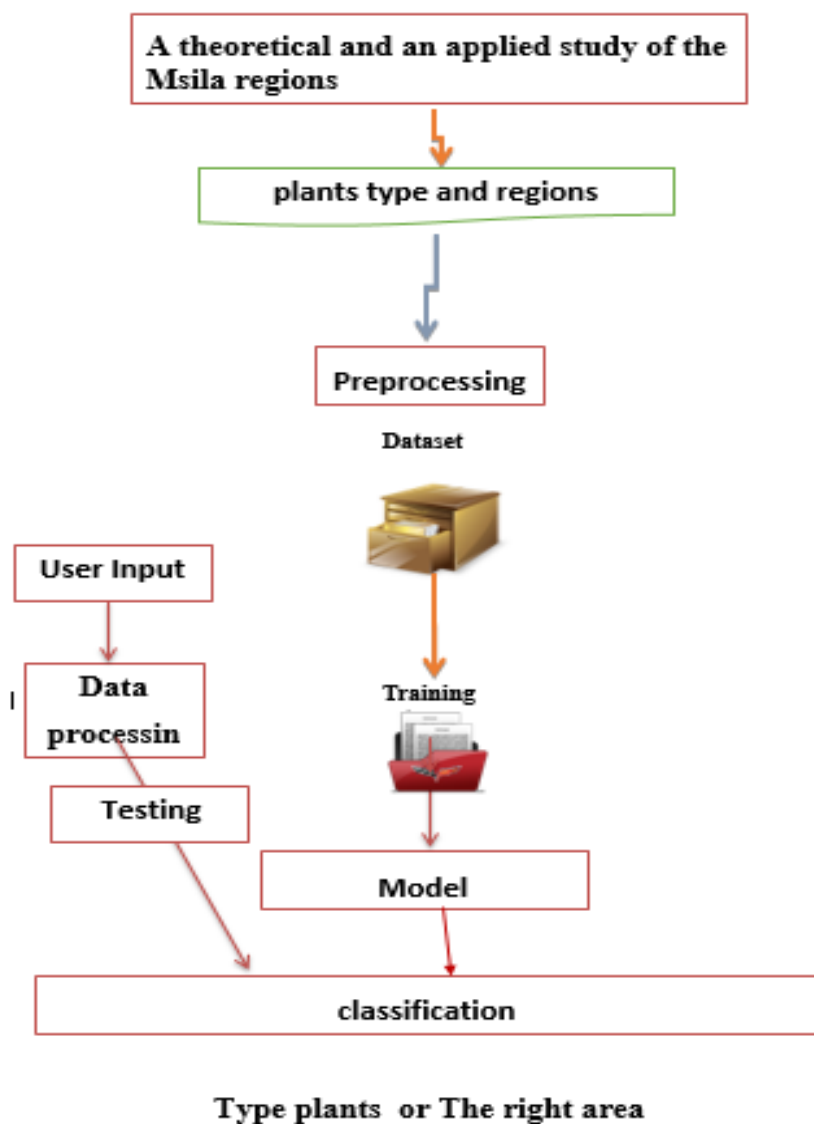


FIGURE 3.3 : PROJECT FRAMEWORK

Chapter 3- A plant prediction framework using KNN

A theoretical and an applied study of the Msila regions to extract the main characteristics of each region.

After studying and extracting the characteristics as we discussed in the first chapter, we have created a data set for the characteristics of certain areas in the city of msila

	P	K	temperature	humidity	PH	rainfall
1	60	52	19.27	54.02	7.51	228.47
2	30	74	19.27	54.02	7.82	228.47
3	6	62	19.27	54.02	7.8	228.47
4	28	98	19.27	54.02	7.77	228.47
5	72	22	19.27	54.02	7.5	228.47
6	52	37	19.27	54.02	7.26	228.47
7	85	23.2	19.27	54.02	7.95	228.47
8	45	22.7	19.27	54.02	7.8	228.47
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						

FIGURE 3.4: CHARACTERISTICS OF THE REGION OF M'SILA GEOGRAPHICALLY ON 8 AREAS

4. Dataset construction

In order to classify the type of plant that I can plant in a particular area, we first need to study what conditions plants need in order to succeed in growing

4.1 Study of Area characteristics and trait extraction

When studying the factors influencing the growth and success of plants after consulting experts in the field of agriculture, we have deduced seven factors influencing growth :

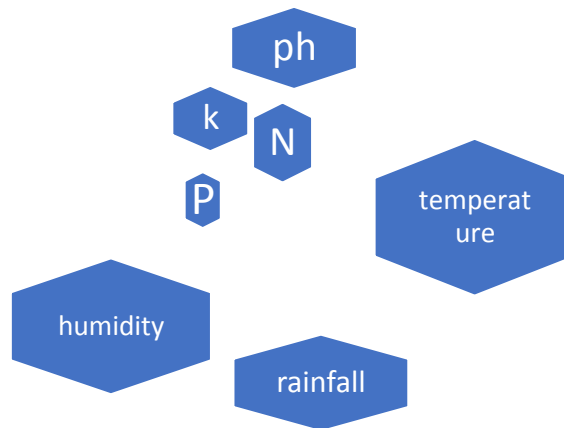


FIGURE 3.5: SEVEN FACTORS INFLUENCING GROWTH PLANTS

4.2 Building dataset

After extracting all the necessary elements for growth, we studied and built a data containing the growth prospects for 22 species of plants and building a datasat.

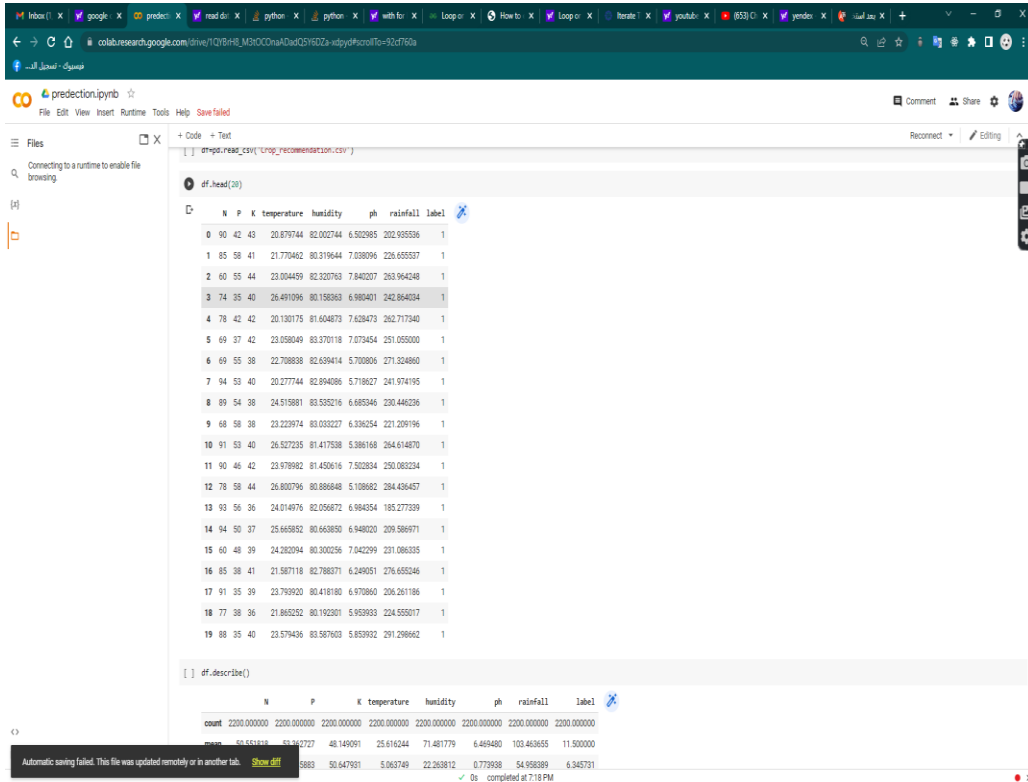


FIGURE 3.6: DATASET DATASET CONSTRUCTION [21]

5. Building machine learning

In order to build the model, we need data sets and machine learning algorithm, after we have created datasets, we used KNN algorithm to perform the prediction and determine the appropriate plants in a given area.

5.1 Loading dataset

The first thing we do after creating the dataset is to upload it to the work file:

```
+ Code + text
[ ] import numpy as np
import pandas as pd

[ ] df=pd.read_csv('Crop_recommendation.csv')

[ ] df.head(20)
```

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	1
1	85	58	41	21.770462	80.319644	7.038096	226.655537	1
2	60	55	44	23.004459	82.320763	7.840207	263.964248	1
3	74	35	40	26.491096	80.158363	6.980401	242.864034	1
4	78	42	42	20.130175	81.604873	7.628473	262.717340	1
5	69	37	42	23.058049	83.370118	7.073454	251.055000	1
6	69	55	38	22.708838	82.639414	5.700806	271.324860	1
7	94	53	40	20.277744	82.894086	5.718627	241.974195	1
8	89	54	38	24.515881	83.535216	6.685346	230.446236	1
9	68	58	38	23.223974	83.033227	6.336254	221.209196	1
10	91	53	40	26.527235	81.417538	5.386168	264.614870	1
11	90	46	42	23.978982	81.450616	7.502834	250.083234	1
12	78	58	44	26.800796	80.886848	5.108682	284.436457	1
13	93	56	36	24.014976	82.056872	6.984354	185.277339	1
14	94	50	37	25.665852	80.663850	6.948020	209.586971	1
15	60	48	39	24.282094	80.300256	7.042299	231.086335	1
16	85	38	41	21.587118	82.788371	6.249051	276.655246	1
17	91	35	39	23.793920	80.418180	6.970860	206.261186	1
18	77	38	36	21.865252	80.192301	5.953933	224.555017	1
19	88	35	40	23.579436	83.587603	5.853932	291.298662	1

FIGURE 3.7: DATASET CSV UPLOAD

For training: two tasks will be performed in this step. The first task is to divide data into 'attributes' and 'label' sets. The resultant data is then divided into training and test sets. For split data to X and Y:

```
[5] #split data
X=dataset.drop("label",axis=1)
Y=dataset["label"]
```

FIGURE 3.8: SPLITTING ATTRIBUTES AND LABELS

```
[6] print(X)

      N  P  K  temperature  humidity      ph  rainfall
0     90 42 43   20.879744  82.002744  6.502985  202.935536
1     85 58 41   21.770462  80.319644  7.038096  226.655537
2     60 55 44   23.004459  82.320763  7.840207  263.964248
3     74 35 40   26.491096  80.158363  6.980401  242.864034
4     78 42 42   20.130175  81.604873  7.628473  262.717340
...   ... .. ..      ...      ...      ...      ...
2195 107 34 32   26.774637  66.413269  6.780064  177.774507
2196  99 15 27   27.417112  56.636362  6.086922  127.924610
2197 118 33 30   24.131797  67.225123  6.362608  173.322839
2198 117 32 34   26.272418  52.127394  6.758793  127.175293
2199 104 18 30   23.603016  60.396475  6.779833  140.937041

[2200 rows x 7 columns]
```

FIGURE 3.9: DATASET OF PLANTS CHARACTERISTICS (PART 1)

```
[7] print(Y)

0      rice
1      rice
2      rice
3      rice
4      rice
...
2195  coffee
2196  coffee
2197  coffee
2198  coffee
2199  coffee
Name: label, Length: 2200, dtype: object
```

FIGURE 3.10: DATASET OF PLANTS CHARACTERISTICS (PART 2)

6. KNN algorithm

After load the dataset and converting it to a data frame, we divide it into two sections, the first section for training and the second section for testing.

After that, we make a call to the KNN algorithm in order for it to perform its turn based on the divisors, and then a Moodle is produced for us by making a classification based on the rules he deduced.

```
[ ] def KNN(X_train,X_test,Y_train,Y_test,K_value):
    #y_hat list for save preeduct
    y_hat=[]
    for test_pt in X_test.to_numpy():
        distances=[]
        for i in range(len(X_train)):
            # هنا نأخذ المسافة بين نقطة التمرين و نقطة التامت
            distances.append(euclidean_distance((np.array(X_train.iloc[i])), test_pt))
            # ثم بعد ذلك نقوم بتحويل هاته المسافة الى داتا فريم لكي نستطيع استغلالها

        distances_data=pd.DataFrame(data=distances,columns=['distance'], index=Y_train.index)
        # هنا نقوم بعمل سورت للداتا وتأخذ الجيران القريبين
        K_neighbors_list = distances_data.sort_values(by=['distance'],axis=0)[:K_value]
        # هنا نعرف احيرا الجيران لتايناز
        labels=Y_train.loc[K_neighbors_list.index]
        voiting = mode(labels).mode[0]

        y_hat.append(voiting)
    return y_hat

///
```

FIGURE 3.11 : TRAINING THE MODEL FOR KNN

We use Euclidean distance between the data points :

```
def euclidean_distance(pt1,pt2):
    distance=np.sqrt(np.sum(pt1-pt2)**2)
    return distance
```

FIGURE 3.12 : EUCLIDEAN DISTANCE CODE

```
[ ] # Splitting the dataset into training and test set.
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test= train_test_split(X, Y, test_size= 0.25, random_state=0)

#feature Scaling
from sklearn.preprocessing import StandardScaler
st_x= StandardScaler()
x_train= st_x.fit_transform(x_train)
x_test= st_x.transform(x_test)

[ ] #Fitting K-NN classifier to the training set
from sklearn.neighbors import KNeighborsClassifier
classifier= KNeighborsClassifier(n_neighbors=5, metric='minkowski', p=2 )
classifier.fit(x_train, y_train)
```

FIGURE 3.13: TRAINING THE MODEL (TEST-TRAIN)

In order to determine the type of plants, the built model needs numeric vector to classify types, we process and analyze the text entered by the user related to his characteristics state in order to extract the features in the form of numeric vector for that we use language processing :

```
dfmsila.head(10)
```

	R	N	P	K	temperature	humidity	PH	rainfall	label
0	R1	51.0	60	52.0	19.27	54.02	7.51	228.47	NaN
1	R2	73.0	30	74.0	19.27	54.02	7.82	228.47	NaN
2	R3	39.0	6	62.0	19.27	54.02	7.80	228.47	NaN
3	R4	53.0	28	98.0	19.27	54.02	7.77	228.47	NaN
4	R5	24.1	72	22.0	19.27	54.02	7.50	228.47	NaN
5	R6	81.2	52	37.0	19.27	54.02	7.26	228.47	NaN
6	R7	90.0	85	23.2	19.27	54.02	7.95	228.47	NaN
7	R8	48.0	45	22.7	19.27	54.02	7.80	228.47	NaN

```
[ ] for (columnName, columnData) in dfmsila.iteritems():  
    print('Column Name : ', columnName)  
    print('Column Contents : ', columnData.values)
```

```
Column Name : R  
Column Contents : ['R1' 'R2' 'R3' 'R4' 'R5' 'R6' 'R7' 'R8']  
Column Name : N  
Column Contents : [51.  73.  39.  53.  24.1 81.2 90.  48. ]  
Column Name : P  
Column Contents : [60 30  6 28 72 52 85 45]  
Column Name : K  
Column Contents : [52.  74.  62.  98.  22.  37.  23.2 22.7]  
Column Name : temperature  
Column Contents : [19.27 19.27 19.27 19.27 19.27 19.27 19.27 19.27]  
Column Name : humidity  
Column Contents : [54.02 54.02 54.02 54.02 54.02 54.02 54.02 54.02]
```

FIGURE 3.14 : VECTOR FOR CHARACTERISTICS

7. Application Modeling

```

[ ] from pandas.core.indexes.base import Index
    for i in dfmsila.index:
        print("-----")
        print("region:"+str(i))
        list_neighbors_predict=[]

        print(list_neighbors_predict.append(classifier.kneighbors([(dfmsila["N"][i],dfmsila["P"][i],dfmsila["K"][i],dfmsila["temperature"][i],dfmsila["humidit
        print(classifier.kneighbors([(dfmsila["N"][i],dfmsila["P"][i],dfmsila["K"][i],dfmsila["temperature"][i],dfmsila["humidity"][i],dfmsila["PH"][i],dfmsila
        print(list_neighbors_predict)
        for j in list_neighbors_predict:
            index1=j[0][0]
            print("    firts kneighbors 1 ")
            print("-----")
            print(" N= "+str(dataset["N"][index1])+" P= "+str(dataset["P"][index1])+" K= "+str(dataset["K"][index1])+" temperature= "+str(dataset["temperature"
            vect=[]
            vect.append(dataset["N"][index1])
            vect.append(dataset["P"][index1])
            vect.append(dataset["K"][index1])
            vect.append(dataset["temperature"][index1])
            vect.append(dataset["humidity"][index1])

            vect.append(dataset["rainfall"][index1])
            print("lable of first region ")
            print(classifier.predict([vect]))
            index2=j[0][1]
            print("    Secund 2 kneighbors 2 ")
            print("-----")
            print(" N= "+str(dataset["N"][index2])+" P= "+str(dataset["P"][index2])+" K= "+str(dataset["K"][index2])+" temperature= "+str(dataset["temperature"
            vect1=[]
            vect1.append(dataset["N"][index2])
            vect1.append(dataset["P"][index2])
            vect1.append(dataset["K"][index2])
            vect1.append(dataset["temperature"][index2])
            vect1.append(dataset["humidity"][index2])
            vect1.append(dataset["ph"][index2])
            vect1.append(dataset["rainfall"][index2])
            print("lable of SECUND region 2")
            print(classifier.predict([vect1]))
            Index3=j[0][2]

```

FIGURE 3.15 : TESTING THE MODEL

8. Implementation

In order to create a machine learning to build a model that classify Neighboring areas and arable plants according to the characteristics of the area, we used the following tools.

8.1 Development Tools

To build machine learning model we use python and an anaconda environment.

8.1.1 Anaconda environment

Anaconda environment supports multiple versions of Python and associated packages. An environment generally includes one version of Python or R language and some packages[15]

8.1.2 Jupyter lab

JupyterLab is a web-based interactive development environment for Jupyter notebooks, code, and data. JupyterLab is flexible: configure and arrange the user interface to support a wide range of workflows in data science, scientific computing, and machine learning. JupyterLab is extensible and modular: write plugins that add new components and integrate with existing ones.



Figure 3.16: development tools

8.2 Implementing the framework

We use jupyter lab in anaconda environment to build machine learning model .

```

[ ] import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
from scipy.stats import mode

[ ] #load data
dataset=pd.read_csv('dataset.csv')

[ ] dataset.head()

```

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

```

[ ] #split data
X=dataset.drop("label",axis=1)
Y=dataset["label"]

```

FIGURE 3.17: BUILDING MACHINE LEARNING MODEL WITH PYTHON

8.3 Experimentation

- In order to create a machine learning to build a model that classify Neighboring areas and arable plants according to the characteristics of the area We perform the following procedures
- Effective machine learning model
- Classification

8.3.1 Classification

We build the model, and then we apply it to the dataset containing the characteristics of wilaya of Msila.

Chapter 3- A plant prediction framework using KNN

	B	C	D	E	F	G	H	I	J
1	P	K	temperature	humidity	PH	rainfall			
2	60	52	19.27	54.02	7.51	228.47			
3	30	74	19.27	54.02	7.82	228.47			
4	6	62	19.27	54.02	7.8	228.47			
5	28	98	19.27	54.02	7.77	228.47			
6	72	22	19.27	54.02	7.5	228.47			
7	52	37	19.27	54.02	7.26	228.47			
8	85	23.2	19.27	54.02	7.95	228.47			
9	45	22.7	19.27	54.02	7.8	228.47			
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									

FIGURE 3.18: DATASET CONTAINING THE CHARACTERISTICS OF REGIONS OF MSILA

In this way, we apply the model to the dataset to classify neighboring areas and determine the appropriate plant type for each area.

```
for i in dfmsila.index:
    print("-----")
    print("region:"+str(i))
    list_neightbors_predect=[]

    print(list_neightbors_predect.append(classifier.kneighbors([(dfmsila["N"][i],dfmsila["P"][i],dfmsila["K"][i],dfmsila["temperature"][i],dfmsila["humidity"][i],dfmsila["PH"][i],dfmsila["rainfall"][i])],n_neighbors=3, return_distance=False))) ;

print(classifier.kneighbors([(dfmsila["N"][i],dfmsila["P"][i],dfmsila["K"][i],dfmsila["temperature"][i],dfmsila["humidity"][i],dfmsila["PH"][i],dfmsila["rainfall"][i])],n_neighbors=3, return_distance=False)) ;
print(list_neightbors_predect)
for j in list_neightbors_predect:
    index1=j[0][0]
    print("      firts kneighbors 1 ")
    print("-----")
    print(" N= "+str(dataset["N"][index1])+" P= "+str(dataset["P"][index1])+" K= "+str(dataset["K"][index1])+" temperature= "+s
```

```
tr(dataset["temperature" ][index1])+" humidity= "+str(dataset["hu
midity" ][index1])+" PH= "+str(dataset["ph" ][index1])+" rainfall=
"+str(dataset["rainfall" ][index1]))
    vect=[]
    vect.append(dataset["N" ][index1])
...ect
```

FIGURE 3.19: PYTHON CODE OF FINDING THE FIRST LAND SIMILAR TO THE GIVEN AREA OF THE INVESTOR.

Prediction results

```
None
[[ 926 1392 344]]
[array([[ 926, 1392, 344]])]
firts kneighbors 1
-----
N= 2 P= 21 K= 44 temperature= 18.92157197 humidity= 87.31290342 PH= 6.56893406 rainfall= 102.8013275
lable of first region
['papaya']
Secund 2 kneighbors 2
-----
N= 83 P= 22 K= 54 temperature= 25.89762315 humidity= 81.96664832 PH= 6.277245253999999 rainfall= 54.49960057
lable of SECUND region 2
['rice']
-----
region:5
None
[[ 926 344 1392]]
[array([[ 926, 344, 1392]])]
firts kneighbors 1
-----
N= 2 P= 21 K= 44 temperature= 18.92157197 humidity= 87.31290342 PH= 6.56893406 rainfall= 102.8013275
lable of first region
['papaya']
Secund 2 kneighbors 2
-----
N= 27 P= 56 K= 22 temperature= 19.91853092 humidity= 20.70099804 PH= 5.833010957999999 rainfall= 108.6434544
lable of SECUND region 2
['rice']
-----
region:6
None
[[926 344 71]]
[array([[926, 344, 71]])]
firts kneighbors 1
-----
```

FIGURE 3.20: FINDING THE FIRST TWO PLANTS OF THE CHARACTERISTIC LANDS GIVEN BY INVESTOR.

Result:

For profile 00

Nearby areas for profile 00 is [[926 1392 344]].

Crops suitable for the area is (papaya) and (rice).

For profile 01

Nearby areas for profile 01 is [[926 1392 344]].

Crops suitable for the area is (papaya) and (rice).

As for the remaining areas (6), their neighboring areas and suitable plants are detected in the same way as the KNN algorithm is used in File 1 and file 2, and we can also enter other files by investors (from the state of M'sila or other wilayas) to find out the plants suitable for cultivation, but this depends on the availability of informations.

9. Conclusion

In this chapter we presented the methodology and all techniques and tools that we have exploited to build the framework predicting plants of a given land introduced by the investor. The framework found the first two plants closer to the characteristics of investor's land.

CONCLUSION



GENERAL CONCLUSION

CONCLUSION

GENERAL CONCLUSION

The agricultural sector plays a great role in the development of the rest of the various sectors, so the state must pay great attention to investment in this sector and provide all the necessary facilities for investors by increasing their demand for investment, especially in the field of agriculture.

Therefore, agricultural investment plays a prominent role in achieving economic development as it contributes to the provision of raw materials. It provides job opportunities and develops industry, as it works on the development of geographical areas or the intended groups of society.

We thought of developing a framework to make it easier for investors to know what plants suit their region. This study was concerned with knowing what plants are suitable for any region of the Wilaya of M'sila, whether it is famous for it or not, without the need to visit the Agricultural Chamber and the Agriculture Directorate.

Our framework is based on the following points:

1. A theoretical and an applied study of the Msila regions to extract the main characteristics of each region.
2. Data processing and feature selection.
3. Building data sets.
4. Training phase: Building machine learning models from the data set.
5. Test phase.
6. Classification the type of plants
7. Classification the adjacent areas

In our project we have built our own dataset to determine the type of plant. During the project, we tested our framework with the data we generated. Experimental results show that our application identifies the area and detects the types of plants suitable for it, which

CONCLUSION

proves the effectiveness of the list of selected features to discover the right and best type of plant.

Perspectives

We hope that everyone will benefit from this framework and extend it for more functionalities to be integrated in the agricultural investor space. I focused in this application on the state of M'sila only because of the difficulty of such important information, and it can be improved by adding a database for the various states.



References

- [1] https://www.aps.dz/ar/economie/121102-2022-02-10-18-53-48_
- [2] <https://www.investopedia.com/Investing/Commodities>.
- [3] ONM M'sila (1990-2019)(2022 ,
- [4] (What inventions helped grow the agricultural economy?)
- [5] <https://www.biz4intellia.com/blog/5-applications-of-iot-in-agriculture>.
- [8] <https://www.biz4intellia.com/blog/5-applications-of-iot-in-agriculture>.
- [9] <https://www.v7labs.com/blog/ai-in-agriculture#h1>.
- [10] (APPLICATION OF DATA MINING IN AGRICULTURE)
- [11] Qualite des sols de la wilaya de Msila.
- [12] mém inge chaficho.
- [13] https://en.wikipedia.org/wiki/Data_mining_in_agriculture .
- [15] <https://docs.anaconda.com/ae-notebooks/4.0/user/anaconda/> visited on June 2021.
- [16] <https://medium.com/@mardiansptr> visited on June 2021.
- [17] https://www.xlstat.com/fr/solutions/fonctionnalites/k-nearest-neighbors-knn_
- [18] <https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning?>
- [19] <http://theprofessionalspoint.blogspot.com/2019/02/how-to-choose-optimal-value-of-k-in-knn.html>.
- [20] <https://medium.com/@anuuz.soni/advantages-and-disadvantages-of-knn-ee06599b9336>.
- [21])ONM M'sila (1990-2019)(2022 ,

ملخص

تهدف الدراسة إلى إنجاز برنامج والذي هو عبارة عن نقطة وصل بين الغرفة الفلاحية والمستثمر والسوق . حيث تم الاعتماد على معلومات ولاية المسيلة . واعتمدنا في الدراسة على آلية التصنيف KNN من أجل الاستثمار الأمثل في المنتج الأنسب لولاية المسيلة ، وذلك بعد جمع البيانات والتي تتمثل في مميزات المناخية ومميزات التربة لولاية المسيلة من دراسات سابقة، حيث قسمت الدراسة إلى جانب نظري وآخر تطبيقي الذي تم من خلاله توقع المحاصيل الزراعية التي يمكن للمستثمر انتاجها طبقا لخصائص أرضه. وقد توصلت الدراسة إلى مجموعة من النتائج أهمها أننا نستطيع أن ننوع في المنتجات الزراعية والاستثمار فيها وبشكل اقتصادي.

الكلمات المفتاحية فضاء المستثمر، تصنيف البيانات ، الزراعة ، الشبكات العصبية الاصطناعية ، الاستثمار الزراعي ، تنوع نباتي جديد في ولاية المسيلة.

Abstract

The study aims to find a framework which is a link between the Chamber of Agriculture, the investor and the market. Where it was relied on the information of the state of M'sila. In the study, we relied on the KNN prediction mechanism in order to optimally invest in the most suitable product for wilaya of M'sila , after collecting data, which are the climatic and soil characteristics of wilaya M'sila from previous studies, where the study was divided into a theoretical and an applied side, in which the framework was able to find the most plants to be cultivated in a given investor land. The study reached a set of results, the most important of which is that we can diversify and invest in agricultural products in an economic way.

KEY WORDS: Investor space, classification of data , agriculture, KNN, agricultural investment , new plant diversity in the state of M'sila .

Résumé

L'étude vise à trouver un framework, qui soit un lien entre la chambre d'agriculture, l'investisseur et le marché. D'où l'on s'est fié aux informations de l'état de M'sila. Dans l'étude, nous nous sommes appuyés sur le mécanisme de classification de KNN afin d'investir de manière optimale dans le produit le plus adapté à la wilaya de M'sila, après avoir collecté des données, qui sont les caractéristiques climatiques et pédologiques de la wilaya de M'sila à partir d'études précédentes, où l'étude a été divisée en une partie théorique et une partie appliqué, dans lesquelles le projet à atteint ses objectifs, à savoir la détermination des produits agricoles appropriés aux caractéristiques du terrain de l'investisseur. L'étude a abouti à un ensemble de résultats, dont le plus important est que nous pouvons diversifier et investir dans les produits agricoles de manière économique.

MOTS CLÉS : Espace de l'investisseur, classification des données, agriculture, KNN, investissement agricole, nouvelle diversité végétale dans l'état de M'sila.