الجمهورية الجزائرية الديموقراطية الشعبية

People's Democratic Republic of Algeria
Ministry of High Education and Scientific Research

Ibn-Khaldun University of Tiaret

Faculty of Natural and Life Sciences
Department of Nutrition and Agro-Food Technology
Thesis in view of obtaining the Master degree

Field: "Agronomy Sciences"
Specialty: "Precision Agriculture"

The topic:

Crops growth prediction using machine learning techniques

Presented by:

- Mehdi Youssra Yamena
- Mazouzi Wahiba

Jury Member:

Dr Mansouri Dou El Kefel	MCA	Supervisor
Dr Bouacha Mohamed Islem	MCA	Examiner
Pr Rezzoug Wafaa	Professor	President

2022/2023

Thanks

We first express our deep gratitude to God who has given us the courage and the will to complete this work.

We would like to extend our heartfelt thanks to our parents and families; our success is undoubtedly the result of their years of sacrifices and education.

Our feelings of appreciation and gratitude also go to our supervisor, Mr. Mansouri Dou El Kefel, for his advice, encouragement, patience, expertise, and remarkable human and scientific qualities. He has been an exemplary figure who guided us throughout this work. The support and guidance we have received from him have been invaluable. We are constantly amazed at how he pushed us to achieve our goals and identified and nurtured our potential.

We would like to express our gratitude to the jury president, Mrs Rezoug Wafaa, for honoring us by accepting to chair this jury.

We also extend our heartfelt thanks to Mr. Bouacha Mohamed Islem, who kindly agreed to examine our work and for which we feel greatly honored.

Dedications

I dedicate this humble work.

The first is to the late my beloved grandmother Maryam,

who I hoped would be present in all my joys Paradise, God willing, then to all those who carried Mazouzi's name, beginning with my father Al arbi and my mother, Mrs. Umm al-khayr Mazouzi Arab, for all the hard work for me and a sacrifice because you were the reason for my success and then my brothers first, any of my beloved and only lovely sisters Aya and my brothers dear Faisal and Tawfiq, Omar and Muhammad and Basil my dear baby , and then their friends, who are like my brothers Mustafa and Muhammad, and the whole family in general, and finally I thank Saida, Laboratory engineer, for her goodness and kindness. And everyone who had a role as professors in our university academic career And everyone who had efforts in our specialty and our master.

Mazouzi Wahiba

Dedications

From the bottom of my heart, I dedicate this work to all those who are dear to me, to my dear parents,

I lack words to express my undying respect, love, and gratitude for the sacrifices you have made for my education and well-being.

I thank you for all the support and love you have given me since my childhood, and I hope that this blessing will always accompany me that this humble work is the fulfillment of your desires, thus the fruit of your countless sacrifices has been formulated, may Almighty God give to you. Happiness, Heath and long life.

I dedicate this work to my brothers and sisters may God keep them.

In the end, everyone who helped me supported me I also thank my friends, especially my classmates, for whom I appreciate and respect, and I wish them success.

Mehdi Youssra Yamena

Contents

Abreviations list

% : percentage.

°C : degree Celsius.

N : nitrogen.

P : phosphorus.

K : potassium.

Kg /ha : kilo grame per hectare.

PH : Hydrogen potential.

Mg/l : milligram per liter.

Ml : milliliter.

KNN : k-nearest neighbor.

+ : positive.

- : negative.

TP : True Positives.

TN : True Négatives.

FP : False Positives.

FN : False Négatives.

List of figures

List of tables

Summary:

This study explores machine learning for crop growth prediction in Algeria using soil, and crop data. The technique has potential for accurate and efficient yield prediction for effective agricultural planning. Optimal parameter values were identified, emphasizing their importance for different datasets. Results vary depending on data characteristics. Implications for farmers, policymakers, and organizations include improved decision-making and resource allocation for increased production and food security. Further research is needed to evaluate other machine learning techniques for sustainable and resilient agriculture.

Keywords:
Machine learning, crop growth, yield prediction.

1

Introduction

Agriculture is a fundamental sector for humanity, playing a crucial role in global food security for almost all cultures in the world. Additionally, it is an important economic engine for several countries, providing vital jobs, incomes, and exports.

In Algeria, the agricultural sector is complex, including a wide variety of crops and livestock industries, and is facing challenges such as inadequate irrigation, lack of modern technologies, and insufficient labor force. These latter, are hindering the growth of the agricultural sector. In this context, government support farmers as well as the researchers to find solutions that can overcome these challenges and improve agricultural production. This can improve the efficiency, profitability, and sustainability of agriculture, and leading to broader economic development in Algeria and also the all Africa country.

Crop forecasting can play a crucial role by enabling the evaluation of future yields using climatic, soil, and resource data. Optimizing crop forecasting using innovative technologies has received a great importance in these recent years. In particular, Machine learning has become a valuable tool determining the most important factors affecting yields and use them to predict future results with increased accuracy. Furthermore, it can help analyze soil, climate, and crop data for better understanding of current conditions and planning for the future [7], [8], [9], [10], [11]. In this context, several works have been proposed, for instance, Suruliandia et al. [1], conducted a comparative study of different machine learning methods carried out for crop prediction. Paudelet al. [2], introduced a machine learning baseline for large-scale early and end of season crop yield forecasts. The same authors suggested a framework for regional crop yield forecasting in Europe [3]. Duboiset al. [4], introduced a machine learning method in order to model the problem of soil water potential prediction, in the context of potato farming. In Algeria, Meroni et al. [5], presented a machine learning system for predicting crop yield of barley, soft wheat and durum.

Objectives of the research:

The main objective of this research is to conduct a comparative study between the classical crop prediction technique (Physico-chemical parameters) often performed in laboratories vs a machine learning technique. The comparison is made using data obtained from the Algerian regions. Specifically, the study will address the following research questions:

1-What are the challenges facing the agricultural sector in Algeria ?

2-How can machine learning techniques be utilized to optimize crop forecasting and improve agricultural production ?

3-What are the potential benefits of utilizing machine learning techniques in crop forecasting for Algeria ?

4-Various experiments are conducted to answer these questions and validate the effectiveness of machine learning techniques.

Organization of the manuscript:

The manuscrit is organized as follow:

In section 1, we start by an introduction that presents the problematic and reviews some related works on crop forecasting using machine learning techniques. Section 2, presents the materiels and methods. Section 3 provides the experimental results. Section 4 draws conclusion and some future directions from this work.

2

Materiels and methods

2.1. Description of the site:

The experiments are conducted on the following places:

A. Cherif Eddine: is a pilot farm located in a rural area in the northern part of Sougueur, between the cities of Tiaret and Sougueur on National Road 23. The farm is situated at the approximate coordinates of 35.181751, 1.918449 and is known for its cereal culture. The surrounding area is primarily agricultural, with a number of farms and fields visible on satellite imagery. Overall, the Sid El Abed farm is located in a distinct agricultural area with a focus on cereal crops and is easily accessible via National Road 23.Figure 1 shows the location of the pilot farm using Google map.

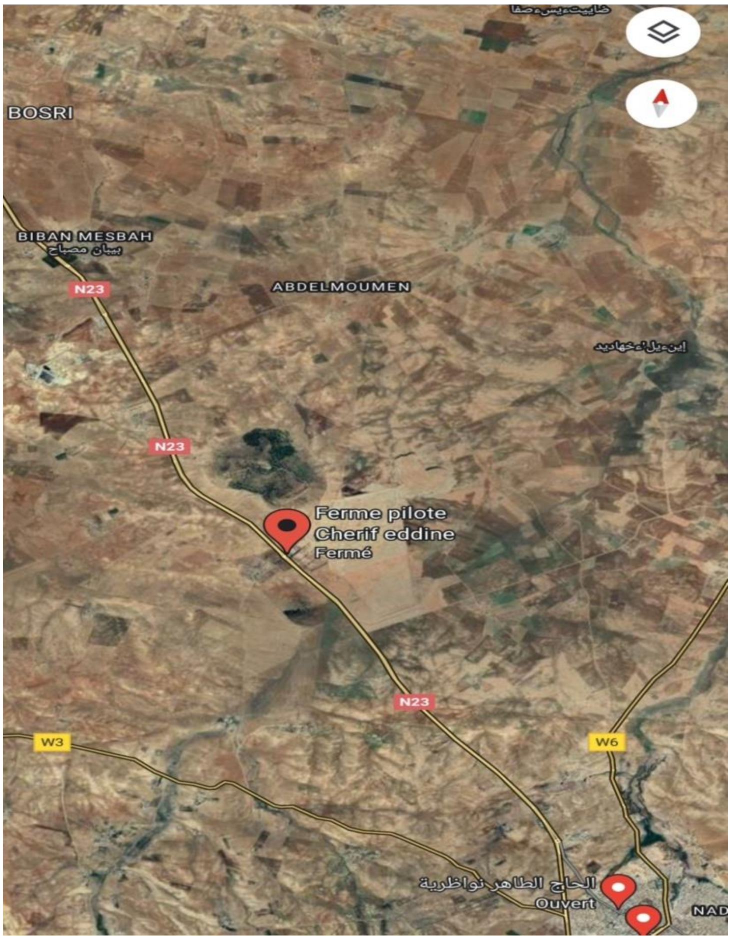


Figure 1 : Charife Eddine souguer Sid El Aed

B. Ain Metnene: is a small town located in the Tiaret province of Algeria, between the towns of Souguer and Sidi Abdelghani, at the geographic coordinates of 35.233718 latitude and 1.557125 longitude. The area is known for its fertile agricultural land, where crops such as wheat, barley, olives, fruits, and vegetables are grown. The land is also used for grazing livestock such as sheep and goats. Agriculture is an important contributor to the local economy in this region, providing employment opportunities for residents and contributing to the production of food and other agricultural products for

The area. Figure 2 shows the location of agriculture field in Ain Metnene (Tiaret) using Google maps.

Figure 2 : Ain Metnene Si Abdelghani

C. Ain Said: located near Ain d'hab and Naima, in western Algeria, at the geographic coordinates of 35.054986 latitude and 1.508192 longitude. The area is known for its rolling hills and fertile valleys, which make it ideal for farming. The land is used for growing various crops such as wheat, barley, olives, fruits, and vegetables, and is also suitable for grazing livestock. Agriculture plays a vital role in the local economy of Ain Said, providing employment opportunities for residents and contributing to the production of food and other agricultural products for the region. Figure 3 shows location of agriculture field in Ain Saïd (Naima) using Google maps.

Figure 3 : Ain Said Naima

2.2. Description of the data:

The data used for the analysis are mainly meteorological data, soil data, and crop data, collected from the above different sources. The collected data is then preprocessed to remove missing values, outliers, and duplicates. Figure 4 shows an example of data set used for crop forecasting.

N	P	K	humidity	pH	label
90	42	43	82,00274423	6,502985292	1
85	58	41	80,31964408	7,038096361	1
60	55	44	82,3207629	7,840207144	1
74	35	40	80,15836264	6,980400905	1
78	42	42	81,60487287	7,628472891	1
69	37	42	83,37011772	7,073453503	1
69	55	38	82,63941394	5,70080568	1
94	53	40	82,89408619	5,718627178	1

Figure 4 : Used data set for crop forecasting

2.3. Physico-chemical parameters:

Physico-chemical analyses were conducted in the laboratory at the Faculty of Natural and Life Sciences Ibn-Khaldoun in Tiaret. We measure the levels of Humidity, ph, and major elements (nitrate, phosphate, and potassium) in the soil. Figures 5, 6 and 7 illustrate the different instruments used during the analysis.

A. Major elements:

(a) Balanc

(b) Glass Beakers (water + soil)

(c) Solution water+ soil

(d) Stirrer

(e) N P K Sensors

Figure 5 : Instruments and analysis process of N, P and K.

N (nitrogen): an essential element for plant growth, which is often supplied to plants through fertilizer applications or through biological processes such as nitrogen fixation.

P (phosphorus): another essential element for plant growth, which is often limiting in soils and must be supplied through fertilization.

K (potassium): a third essential element for plant growth, which is important for many plant functions such as regulating water balance, enzyme activation, and protein synthesis.

Measuring the levels of these three elements in soil is crucial for maximizing crop yields and minimizing the use of fertilizers. One method of measuring these elements is by using soil sensors, which are capable of detecting nutrient levels in soil solutions. To measure N, P, and K using a sensor, soil samples are first mixed with water to create a soil solution. Specific ion-selective electrodes are then utilized to measure the concentration of the nutrient in the soil solution. For instance, nitrate ion-selective electrodes can be used to measure nitrate concentration for N, while specific ion-selective electrodes can be employed for P and K. While soil moisture and ph levels can impact the accuracy of the sensor method, calibrating the sensors before use and using them in combination with other soil testing methods can provide a more comprehensive analysis of soil nutrient levels.In the following, we present the procedure for measuring N, P and K.

Procedure:

1. Remove the protective cap from the probe.

2. Press the OFF/ON button to turn on the tester, and the display should read 0.

3. For a quick test, water the soil with distilled water, ensuring it reaches a depth of at least 10 cm. The ideal humidity range is 70%-80%. Insert the metal probe vertically and clockwise into the soil to a depth of 6-8 cm, ensuring full contact with the soil. After approximately 6 seconds, the tester will display the measured value. The soil's compactness will affect the accuracy of the result, so multiple tests should be performed at various points and the average taken as the final result. For accurate testing, mix dried soil with distilled water in a 200ml beaker container until it forms a mud-like consistency. Insert the metal probe into the mud to a depth of 6-8 cm, and after approximately 6 seconds, the tester will display the measured value.

4. After each test, wipe the metal probe with a towel or tissue paper until it is clean and dry, and the display reads 0.

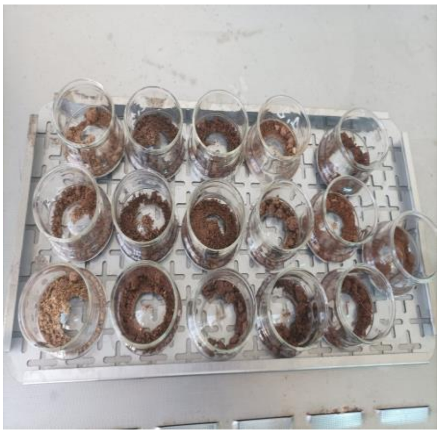
.

B. Humidity:

(a) Electronic scale

(b) Laboratory oven **(étuve)**

(c) Electronic scale



(d) After drying

Figure 6 : Instruments and analysis process of Humidity.

Analysis of humidity in soil samples should be performed on the same day as sample collection to obtain accurate information on soil water content. Note that the analysis does not require crushing and sieving.

The following protocol is recommended:

- Using a precision scale, weigh an empty glass capsule.
- Record the weight of the capsule and then weigh 20 g of soil sample.
- Place the capsule with the soil sample in an oven at 105°C for 24 hours.
- After allowing it to cool to room temperature, transfer the capsule containing the dried sample to a desiccator.
- Calculate the moisture (Humidity) content (H) in percentt using the following formula:
  H = (pair – p105°C) / pair x 100

Where pair is the weight of the air-dried soil and p105°C is the weight of the soil after drying in the oven.

C. PH:

(a) Stirrer

(b) PH meter

Figure 7 : Instruments and analysis process of ph.

Ph is a measure of the acidity or alkalinity of a solution, ranging from 0 to 14. A ph of 7 is considered neutral, below 7 is acidic, and above 7 is alkaline. Ph meters are commonly used to measure ph and consist of a glass electrode and a reference electrode. The accuracy of ph measurements depends on proper calibration using buffer solutions of known ph values. Temperature can also affect ph readings, so it is important to consider and compensate for temperature variations during measurements. Ph measurements have broad applications in water and soil analysis, food and beverage industry, and medical and pharmaceutical research.

To measure ph, we follow the steps below:

- Weigh 10 g of fine soil into a 250 ml beaker.
- Add 50 ml of distilled water to the soil.
- Shake the mixture for 5 minutes to ensure thorough mixing.
- Allow the solution to stand for 30 minutes.
- Turn on the ph meter and perform calibration according to the manufacturer's instructions.
- Immerse the ph meter electrode in the supernatant solution.

Read and record the ph value displayed on the ph meter.

2.4. Machine learning technique:

In this section, we test the effectiveness of a machine learning technique in crop forecasting. The choice of the technique is based on its effectiveness in the field of prediction.

2.4.1. k-Nearest Neighbors (KNN):

KNN is a supervised classification algorithm that assigns a new observation (point) to the most appropriate class based on a calculated distance between its k neighbors in the training data, where k is a positive integer [6]. Figure 8 presents the KNN concept.

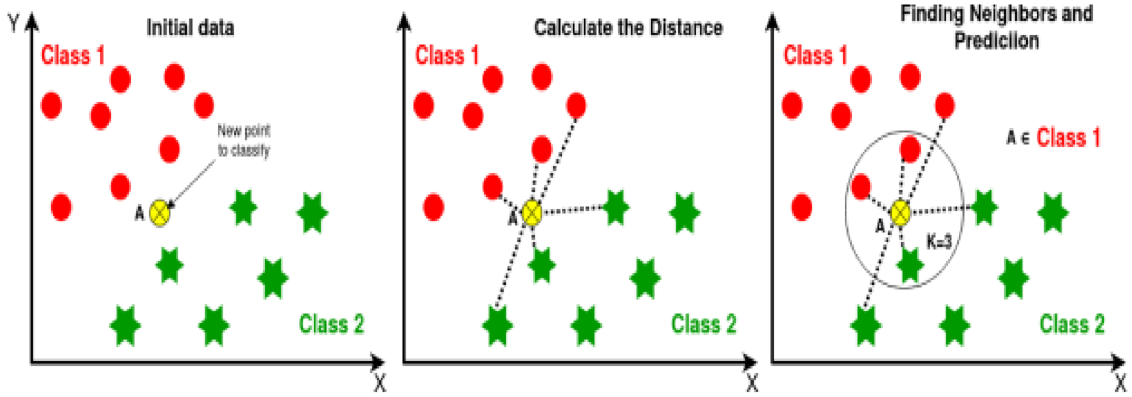


Figure 8 : Concept of KNN

Figure 9 presents the algorithm of KNN written in mattlab.

2.4.1.1. Evaluation of KNN:

To evaluate the performance of our proposed model, weused the "Accuracy" metric. This is one of the most commonly used measures for ensuring a thorough comparison between methods. Accuracy represents the number of correct predictions made by a model over all observed values. It is given by the following equation.

Accuracy = (TP + TN)/ (TP + FP + TN + FN).

Where: TP (True Positives) refers to a result where the model correctly predicts the positive class. TN (True Negatives) refers to a result where the model correctly predicts the negative class. FP (False Positives) refers to a result where the model

Incorrectly predicts the positive class. FN (False Negatives) refers to a result where the model incorrectly predicts the negative class.

```
%import data

Data =readtable('\dataset.csv');
% Cross varidation (train: 70%, test: 30%)
Cv= cvpartition(size(data,1),'holdout',0.3);
Idx = cv.test;
% Separate to training and test data
Train_data = data(~idx,:);
Test_data = data(idx,:);
Knn_result = knnf(train_data,test_data,5);
KNN = Accuracy(knn_result);
Function result = knnf(train_data,test_data,no_of_neighbours)
% the following code converts training data files into training variables which are used to train
the model;
X_train = train_data(:,1:end-1);
Y_train = train_data(:,end);
% the following code converts testing data files into testing variables
 % which are used to test the model
X_test = test_data(:,1:end-1);
Y_test = test_data(:,end);
 %testing
Eval_matrix = zeros(size(x_train,1),2);
Resultant_matrix = zeros(size(x_test,1),1);
For i =1 : size(x_test,1)
For j = 1 : size(x_train,1)
Eval_matrix(j,1) = norm(x_test(i,:) - x_train(j,:));
Eval_matrix(j,2) = y_train(j);
End
Eval_matrix = sortrows(eval_matrix);
Res = eval_matrix(1:no_of_neighbours,2);
Resultant_matrix(i) = mode(res);
Eval_matrix = zeros(size(x_train,1),2);
End
 Predicted_Values = resultant_matrix;
 Expected_Values = y_test;
Result = table(Predicted_Values,Expected_Values);
End
Function result = Accuracy(output_table)
P_val = output_table.Predicted_Values;
E_val = output_table.Expected_Values;
Acc =0;
For i = 1 : size(p_val,1)
If(p_val(i) == e_val(i))
Acc =acc+1;
End
End
Acc= acc/size(p_val,1);
Result =acc;
End
```

Figure 9 : Algorithm of KNN in Matlab.

2.4.1.2. Cross validation:

Cross-validation is a widely used statistical method for assessing the performance of machine learning models. It involves splitting the available data into two parts – one for training and one for validation – to ensure that the model is evaluated on data that it hasn't seen before. In this study, we used the Train-Test Split method, which randomly divides the dataset into 70% training data and 30% testing data. The dataset consisted of 100 samples, with the 30% of testing data assumed to be new and unseen by the model. We trained the proposed model on the training data and evaluated its performance on the testing data using the Accuracy metric. If the model accurately predicted the class labels of the testing data, it was deemed to be reliable. This approach allowed us to thoroughly evaluate the effectiveness of our proposed model and make any necessary adjustments.

2.4.1.3. Configuration:

We used MATLAB as a tool to conduct the experiment. MATLAB (an abbreviation of "matrix laboratory") is a high-level programming language and interactive environment that is widely used in science and engineering fields. It is an abbreviation for "Matrix Laboratory" because it was originally designed to perform matrix operations. With MATLAB, you can perform complex numerical computations and visualize data with high-quality graphics.

It has a vast range of built-in functions and toolboxes that provide solutions for a variety of applications, including signal processing, image and video processing, control systems, optimization, machine learning, and more. MATLAB is also known for its easy-to-learn syntax, making it accessible for users with varying levels of programming experience. It has a variety of features, such as its powerful plotting and visualization tools, interactive debugging tools, and a comprehensive help system, which make it a valuable tool for research and development.

Furthermore, MATLAB's compatibility with other programming languages such as C, C++, and Python, as well as its ability to interface with hardware such as sensors and cameras, makes it a versatile platform for developing an implementing complex algorithms and systems.

Overall, MATLAB is a powerful tool for numerical computing and scientific programming that has gained popularity due to its rich set of features, flexibility, and user-friendly interface.

3

Results and discussion

3.1. Physico-chemical results:

In this section, we present the main results obtained from experiments. Table 1 presents results on the physical and chemical parameters obtained from the Cherif Eddine Farm. The table includes data on nitrogen (N), phosphorus (P), potassium (K), ph, and humidity levels for different crops.

Table 1 : Physico-chimical parameters obtaind from the Cherif Eddine farme .

Soil		N kg/ha	P Kg/ha	K Kg/ha	PH	Humidity %
Barley	B1	25.2	21.6	64.8	8.77	3.25
	B2	32.4	28.8	93.6	8.62	2.35
	B3	25.2	25.2	75.6	8.64	2.51
	B4	36	43.2	122.4	7.94	2.48
Durum wheat	D1	158.4	198	500.4	8.29	2.76
	D2	79.2	79.2	198	7.82	2.00
	D3	72	75.6	198	8.50	2.18
soft wheat	C1	32.4	28.8	90	8.64	2.47
	C2	28.8	25.2	79.2	8.72	2.68
	C3	32.4	32.4	90	8.84	2.32
	C4	28.8	28.8	79.2	8.67	2.47

1- Barley (From sample B1 to B4):

- Nitrogen (N) : The nitrogen levels range from 25.2 to 36.
- Phosphorus (P) : The phosphorus levels range from 21.6 to 43.2.
- Potassium (K) : The potassium levels range from 64.8 to 122.4.
- Ph : The ph values range from 7.94 to 8.77.
- Humidity : The humidity levels range from 2.35% to 3.25%.

2- Durum Wheat (D1-D3):

- Nitrogen (N) : The nitrogen levels range from 72 to 158.4.
- Phosphorus (P) : The phosphorus levels range from 75.6 to 198.
- Potassium (K) : The potassium levels range from 198 to 500.4.
- Ph : The ph values range from 7.82 to 8.50.
- Humidity : The humidity levels range from 2.00% to 2.76%.

3- Soft Wheat (C1-C4):

- Nitrogen (N) : The nitrogen levels range from 28.8 to 32.4.
- Phosphorus (P) : The phosphorus levels range from 25.2 to 32.4.
- Potassium (K) : The potassium levels range from 79.2 to 90.
- Ph : The ph values range from 8.64 to 8.84.
- Humidity : The humidity levels range from 2.32% to 2.68%.

Interpretation:

These observations provide insights into the nutrient composition and soil conditions for each crop on the Cherif Eddine Farm. The data can be utilized to make informed decisions about fertilizer application, soil amendments, and ph adjustments to optimize crop growth and yield.

Table 2 presents results on the physical and chemical parameters obtained from the land of Ain Metnene. The table includes data on nitrogen (N), phosphorus (P), potassium (K), ph, and humidity levels for different crops.

Table 2 : Physico-chimical parameters obtaind from the land of Ain metnene.

Soil		N Kg/ha	P Kg/ha	K Kg/ha	PH	Humidity %
Potato	P1	50.4	46.8	133.2	8.90	2.83
	P2	50.4	50.4	129.6	8.62	2.61
	P3	50.4	61.2	140.4	8.56	2.37
Beans	B1	68.4	68.4	183.6	8.28	2.65
	B2	50.4	46.8	133.2	8.52	2.35

1- Potatoes (from sample P1 to P3):

- Nitrogen (N) : The nitrogen levels are consistent at 50.4 across all three varieties.
- Phosphorus (P) : The phosphorus levels range from 46.8 to 61.2.
- Potassium (K) : The potassium levels range from 129.6 to 140.4.
- Ph : The ph values range from 8.56 to 8.90.
- Humidity : The humidity levels range from 2.37% to 2.83%.

2- Beans (B1 - B2):

- Nitrogen (N) : The nitrogen levels are consistent at 50.4 for both bean varieties.
- Phosphorus (P) : The phosphorus levels are consistent at 68.4 for both bean arieties.

- Potassium (K) : The potassium levels are consistent at 133.2 for both bean varieties.

- Ph : The ph values range from 8.28 to 8.52.

- Humidity : The humidity levels range from 2.65% to 3.35%.

Interpretation:

The data from Table 2 provides insights into the nutrient composition and soil conditions for each crop in the land of Ain Metnane. This information can be valuable for making informed decisions regarding fertilizer application, soil amendments, and ph adjustments to optimize crop growth and yield on this specific land.

Table 3 presents results on the physical and chemical parameters obtained from the land of Ain Said. The table includes data on nitrogen (N), phosphorus (P), potassium (K), ph, and humidity levels for two crops: Barley and Wheat.

Table 3 : Physico-chimical parameters obtaind from the land of ain said.

Soil		N Kg/ha	P Kg/ha	K Kh/ha	PH	Humidity %
Barely	B1	36	28.8	90	7.49	3.15
	B2	21.6	25.2	61.2	7.98	3.46
	B3	18	21.6	72	7.99	3.11
	B4	18	21.6	64.8	7.92	1.95
Wheat	W1	36	36	27	7.89	2.42
	W2	140.4	154.8	514.8	7.77	3.29
	W3	25.2	25.2	75.6	7.67	3.11

1- Barley (B1 - B4):

- Nitrogen (N) : The nitrogen levels range from 18 to 36.

- Phosphorus (P) : The phosphorus levels range from 21.6 to 28.8.

- Potassium (K) : The potassium levels range from 64.8 to 90.

- ph : The ph values range from 7.49 to 7.99.

- Humidity : The humidity levels range from 1.95% to 3.46%.

2- Wheat (W1 - W3):

- Nitrogen (N) : The nitrogen levels range from 25.2 to 140.4.

- Phosphorus (P) : The phosphorus levels range from 25.2 to 154.8.

- Potassium (K) : The potassium levels range from 27 to 514.8.

- ph : The ph values range from 7.67 to 7.89.
- Humidity : The humidity levels range from 2.42% to 3.29%.

Interpretation:

The data from Table 3 provides insights into the nutrient composition and soil conditions for Barley and Wheat crops in the land of Ain Said. It suggests variations in nitrogen, phosphorus, potassium, ph, and humidity levels, which can impact the growth and productivity of these crops. Farmers can utilize this information to adjust fertilizer application, soil amendments, and irrigation practices to optimize crop health and yield on the specific land of Ain Said.

Table 4 presents an analysis of physico-chemical parameters obtained from the “Other Land” for various crops. The table includes information on nitrogen (N), phosphorus (P), potassium (K), ph, and humidity levels for different crops such as Rice, Maize, Chickpea, Kidney Beans, Pigeon Peas, Moth Beans, Banana, Mango, and Grapes. The data provides insights into the nutrient composition and soil conditions for these crops, which are essential factors influencing their growth and productivity.

Table 4 : Physico-chimical parameters obtaind from other land.

Soil		N kg/ha	P kg/ha	K kg/ha	PH	Humidity %
Rice	P1	90	42	43	6.50	82.00
	P2	85	58	41	7.03	80.31
	P3	60	55	44	7.84	82.32
	P4	74	35	40	6.98	80.15
Maize	P1	71	54	16	5.74	63.69
	P2	61	44	17	6.93	71.57
	P3	80	43	16	6.65	71.59
	P4	73	58	21	6.59	57.68
Chickpea	P1	23	72	84	6,92	17,13
	P2	39	58	85	5,99	15,40
	P3	22	72	85	6,39	15,65
	P4	36	67	77	7,15	19,56
Kidneybeans	P1	13	60	25	5,68	20,59

	P2	25	70	16	5,75	18,90
	P3	31	55	22	5,87	21,33
	P4	40	64	16	5,92	24,24
Pigeonpeas	P1	3	72	24	6,03	57,92
	P2	40	59	23	5,26	62,73
	P3	33	73	23	5,98	59,38
	P4	27	57	24	6,09	43,35
Mothbeans	P1	3	49	18	3,69	64,70
	P2	22	59	23	4,37	51,27
	P3	36	58	25	8,39	59,31
	P4	4	43	18	8,84	61,09
Banana	P1	91	94	46	6,14	76,24
	P2	105	95	50	5,84	83,67
	P3	108	92	53	6,27	82,96
	P4	86	76	54	5,92	80,11
Mango	P1	39	24	31	4,75	53,72
	P2	21	26	27	5,69	47,67
	P3	25	22	25	5,97	45,53
	P4	0	21	32	6,43	54,25
Grapes	P1	24	130	195	6,11	81,54
	P2	13	144	204	6,09	82,42
	P3	22	123	205	5,89	83,88
	P4	36	125	196	6,15	80,65

1- Rice (P1 - P4) :

- Nitrogen (N): The nitrogen levels range from 60 to 90.
- Phosphorus (P): The phosphorus levels range from 35 to 58.
- Potassium (K): The potassium levels range from 40 to 44.
- ph: The ph values range from 6.50 to 7.84.

- Humidity: The humidity levels range from 80.15% to 82.32%.

2- Maize (P1 - P4) :

- Nitrogen (N) : The nitrogen levels range from 61 to 80.
- Phosphorus (P) : The phosphorus levels range from 43 to 58.
- Potassium (K) : The potassium levels range from 16 to 21.
- ph : The ph values range from 5.74 to 6.93.
- Humidity : The humidity levels range from 57.68% to 71.59%.

3- Chickpea (P1 - P4) :

- Nitrogen (N) : The nitrogen levels range from 22 to 39.
- Phosphorus (P) : The phosphorus levels range from 58 to 72.
- Potassium (K) : The potassium levels range from 77 to 85.
- ph : The ph values range from 5.99 to 7.15.
- Humidity : The humidity levels range from 15.40% to 19.56%.

4- Kidney Beans (P1 - P4) :

- Nitrogen (N) : The nitrogen levels range from 13 to 40.
- Phosphorus (P) : The phosphorus levels range from 55 to 70.
- Potassium (K) : The potassium levels range from 16 to 25.
- ph : The ph values range from 5.68 to 5.92.
- Humidity : The humidity levels range from 18.90% to 24.24%.

5- Pigeon Peas (P1 - P4) :

- Nitrogen (N) : The nitrogen levels range from 3 to 40.
- Phosphorus (P) : The phosphorus levels range from 57 to 73.
- Potassium (K) : The potassium levels range from 23 to 24.
- ph : The ph values range from 5.26 to 6.09.
- Humidity : The humidity levels range from 43.35% to 62.73%.

6- Moth Beans (P1 - P4) :

- Nitrogen (N) : The nitrogen levels range from 3 to 36.
- Phosphorus (P) : The phosphorus levels range from 43 to 59.
- Potassium (K) : The potassium levels range from 18 to 25.
- ph : The ph values range from 3.69 to 8.84.
- Humidity : The humidity levels range from 51.27% to 64.70%.

7- Banana (P1 - P4) :

- Nitrogen (N) : The nitrogen levels range from 86 to 108.
- Phosphorus (P) : The phosphorus levels range from 76 to 95.
- Potassium (K) : The potassium levels range from 46 to 54.
- ph : The ph values range from 5.84 to 6.27.
- Humidity : The humidity levels range from 76.24% to 83.67%.

8- Mango (P1 - P4) :

- Nitrogen (N) : The nitrogen levels range from 0 to 39.
- Phosphorus (P) : The phosphorus levels range from 21 to 26.
- Potassium (K) : The potassium levels range from 25 to 32.
- ph : The ph values range from 4.75 to 6.43.
- Humidity : The humidity levels range from 45.53% to 54.25%.

9- Grapes (P1 - P4) :

- Nitrogen (N) : The nitrogen levels range from 13 to 36.
- Phosphorus (P) : The phosphorus levels range from 123 to 144.
- Potassium (K) : The potassium levels range from 195 to 204.
- ph : The ph values range from 5.89 to 6.15.
- Humidity : The humidity levels range from 80.65% to 83.88%.

Interpretation:

The data for each crop provides insights into the nutrient composition, ph levels, and humidity conditions. These factors play crucial roles in determining the growth and productivity of the crops. Farmers can utilize this information to make informed decisions regarding fertilizer application, soil management, and irrigation practices specific to each crop to optimize their growth and yield in the given land.

The measured values in the tables indicate that the soil parameters are within a favorable range for plant growth, as they comply with the established standards for soil analysis. This suggests that the growing conditions are favorable for crops and that plants can access the necessary nutrients for healthy growth. These results are encouraging for farmers, as they may indicate that the soils are healthy and that crops are likely to develop optimally. However, it is important to remember that standards may vary depending on the region, soil type, and cultivated crops. It is therefore recommended to refer to specific standards established for the region for a more accurate assessment of the soil health and crop quality.

3.2. Result of KNN:

Figure 10 and 11 present the main results obtained from the proposed machine learning method.

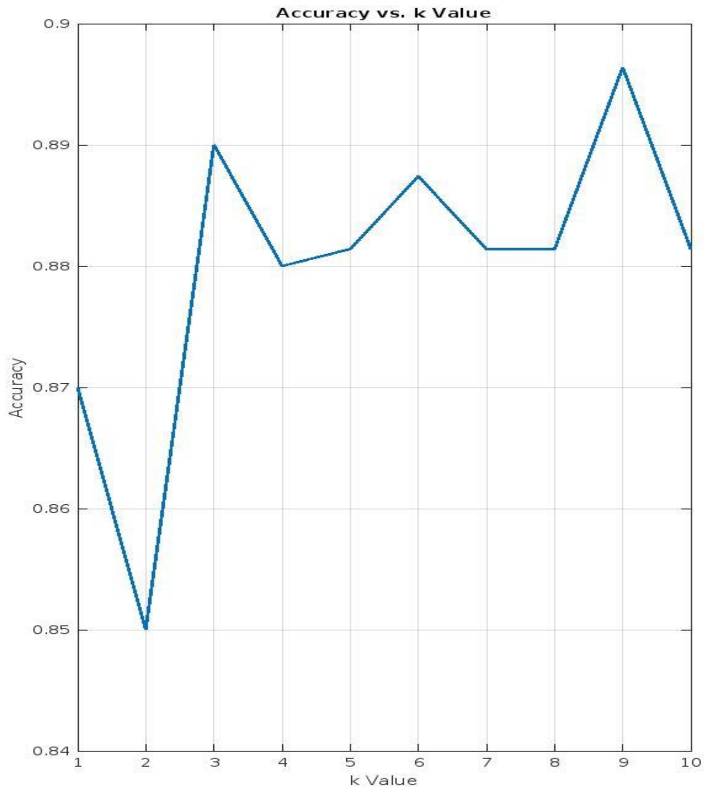


Figure 10 : Accuracy of the KNN vs K-value using 30% of test data.

Figure 10 shows the relationship between the value of k and the accuracy of the model, with accuracy measured on the test data (30% of the data set). The x-axis represents the different values of k, while the y-axis represents the corresponding accuracy values.

From the figure, we can see that the accuracy of the model generally fluctuates as the value of k changes. The highest accuracy is achieved at k = 9, with a value of 0.8964. However, there are several other values of k that have accuracy values that are almost as high, such as k = 3 and k = 6.

Overall, the figure suggests that the optimal value of k for this particular model is not as clear-cut as it was for the previous data set. Rather, there are several values of k that could be considered optimal, depending on the specific needs and constraints of the model. However, it is worth noting that the accuracy of the model is generally quite high for all values of k, indicating that the model is likely to perform well regardless of the specific value chosen for k.

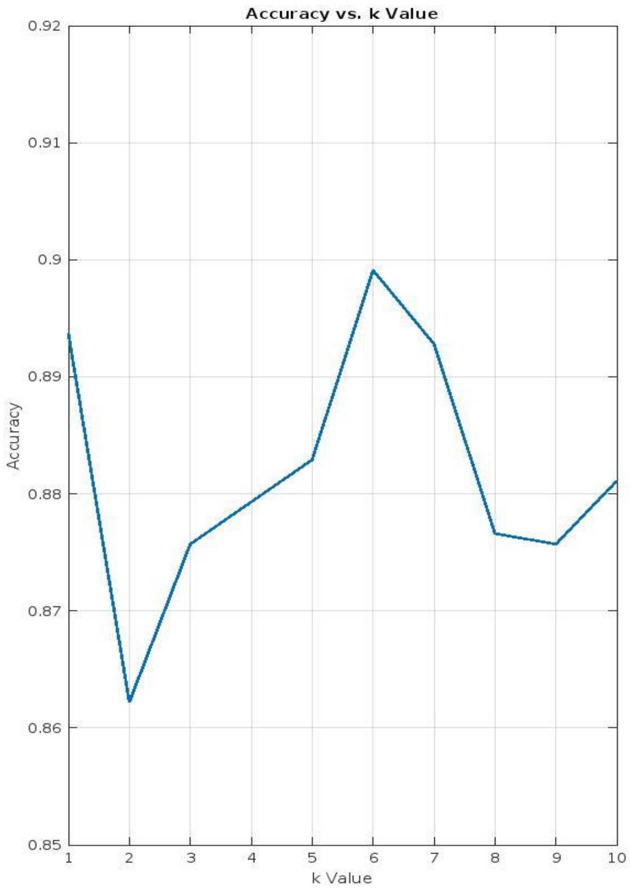


Figure 11 : Accuracy of the KNN vs K-value using 50% of test data .

The figure 11 shows the relationship between the value of k and the accuracy of a model, with accuracy measured on the test data (50% of the data set). The x-axis represents the different values of k, while the y-axis represents the corresponding accuracy values.

From the figure, we can see that the accuracy of the model generally increases as the value of k increases, up to a certain point, after which it decreases. The highest accuracy is achieved at k = 6, with a value of 0.8991. However, the accuracy dips at k = 2 and k = 8, which have the lowest accuracy values of 0.8622 and 0.8766, respectively.

Overall, the figure suggests that the optimal value of k for this particular model is around 6, as this is where the highest accuracy is achieved. However, it is also important to note that the accuracy of the model can vary significantly depending on the choice of k, and that some values of k may lead to much lower accuracy values.

3.3. Comparative study between the Physico-chemical technique of parameters and the KNN:

The main difference between the physico-chemical results of parameters and the KNN model is the type of information they provide and the purpose they serve in crop management. The physico-chemical results of parameters provide information about the nutrient composition, soil conditions, ph levels, and humidity conditions of the soil in which crops are grown. This information is based on direct measurements and observations of the physical and chemical properties of the soil, and can be used to identify potential deficiencies or excesses of nutrients or other factors that may impact crop growth and yield. The primary purpose of the physico-chemical analysis is to provide farmers with a better understanding of the soil conditions on their farm, and to inform decisions about fertilization, soil amendments, and irrigation practices to optimize crop growth and yield.
8787On the other hand, the KNN model is a machine learning algorithm that uses historical data to make predictions about future crop yields based on the physical and chemical properties of the soil. The KNN model is based on the assumption that crops with similar physical and chemical properties tend to have similar yields, and it is used to classify new data points based on their similarity to historical data. The primary purpose of the KNN model is to provide farmers with a tool to predict crop yields and to identify areas of the farm that may require special attention or management practices to improve yield.

The table 5 compares the time required for physico-chemical analysis and the KNN machine learning technique for predicting crop yields. The physico-chemical analysis involves direct measurement and observation of soil properties such as nutrient composition, pH levels, and humidity conditions, which takes several hours to days to complete. On the other hand, the KNN model uses historical data to make predictions about future crop yields, and takes only a few minutes to generate results. Despite their differences in time requirements, both techniques provide valuable information for optimizing crop management practices. The physico-chemical analysis provides direct information about the soil conditions, while the KNN model predicts future yields based on historical data. By combining the information from both tools, farmers can make more informed decisions about fertilization, soil amendments, and irrigation practices to optimize crop growth and yield.

Table 5 : Comparison table between the techniques used by time

Analysis technique	Time required
Physical-chemical analysis	N, p, k: *2 hours*
	Humidity: *24h*
	Ph: *3 days after soil sampling dry soil*
Prediction of classical analysis	*3 to 5 minutes*
KNN machine learning technique	*2 minutes*

According to the table, physical-chemical analysis for n p k analysis takes around 2 hours, while humidity analysis takes up to 24 hours, and ph analysis requires 3 days after soil sampling with dry soil. On the other hand, prediction techniques take only a few minutes, ranging from 3 to 5 minutes. Knn machine learning technique is the fastest method, taking only 2 minutes. Based on this table, it is evident that using prediction techniques and machine learning models can significantly reduce the time and effort required for soil analysis compared to traditional physical-chemical analysis.

4

Conclusion

In this study, we explored the use of a machine learning techniques for predicting crop growth in Algeria, using a combination of soil and crop data. Our analysis demonstrated the potential of such techniques to improve the accuracy and efficiency of crop yield prediction, which is critical for effective agricultural planning and resource allocation. By comparing the performance of the proposed technique using a range of parameters, we identified the optimal parameter values for our data set, highlighting the importance of careful selection of this parameter. Moreover, by examining the results of our model for different datasets, we have shown that the optimal value of parameters can vary depending on the specific context and characteristics of the data being used. Our findings have important implications for farmers, policymakers, and agricultural organizations in Algeria and beyond. By leveraging machine learning techniques to more accurately predict crop yields, we can support more effective decision-making and resource allocation in agriculture, improving both production and food security. However, further research is needed to explore the use of other machine learning techniques and evaluate their performance in different contexts, in order to continue improving our ability to predict crop yields and support more sustainable and resilient agricultural systems.

References

[1] Suruliandi, A., G. Mariammal, and S. P. Raja. "Crop prediction based on soil and environmental characteristics using feature selection techniques." Mathematical and Computer Modelling of Dynamical Systems 27.1 (2021): 117-140.

[2] Paudel, Dilli, et al. "Machine learning for large-scale crop yield forecasting." Agricultural Systems 187 (2021): 103016.

[3] Paudel, Dilli, et al. "Machine learning for regional crop yield forecasting in Europe." Field Crops Research 276 (2022): 108377.

[4] Dubois, Amaury, Fabien Teytaud, and Sébastien Verel. "Short term soil moisture forecasts for potato crop farming: A machine learning approach." Computers and Electronics in Agriculture 180 (2021): 105902.

[5] Meroni, Michele, et al. "Yield forecasting with machine learning and small data: What gains for grains?." Agricultural and Forest Meteorology 308 (2021): 108555.

[6] Soucy, Pascal, and Guy W. Mineau. "A simple KNN algorithm for text categorization." Proceedings 2001 IEEE international conference on data mining. IEEE, 2001.

[7] Aghdam, M. S., Minaei, S., & Moridpour, S. (2019). Crop yield prediction under climate change using machine learning : A systematic literature review. Agricultural Systems, 176, 102656. https://doi.org/10.1016/j.agsy.2019.102656.

[8] Oyeyemi, O. J., & Bishop, I. D. (2020). Machine learning approaches for crop yield prediction and climate change impact studies : A review. Journal of Agricultural and Biological Science, 15(2), 39-50. https://doi.org/10.5281/zenodo.3712213.

[9] Gangwar, S., Singh, S. K.,& Srivastava, A. K. (2020). A review of crop yield prediction models and applications using machine learning techniques. Archives of Computational Methods in Engineering, 27(4),11951214.https://doi.org/10.1007/s11831-019-09368-7.

[10] Pham, T., Tran, D.,& Nguyen, T. (2021). A review of crop yield prediction models and techniques using machine learning. Journal of Open Innovation: Technology, Market, and Complexity, 7(2),108. https://doi.org/10.3390/joitmc7020108.

[11] Zhang, C., & Lobell, D. B. (2019). Combining satellite imagery and machine learning to predict soybean yield in the United States. Environmental Research Letters, 14(9), 094005. https://doi.org/10.1088/1748-9326/ab360d.