الجـمهورية الجـزائرية الديمقراطـية الشـعبية
PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
وزارة التعليم العـالي و البحـث العـلمي
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
جـامعة عمار ثـليجي بالأغـواط
UNIVERSITY OF AMMAR TELIDJI LAGHOUAT

كـلية العلـوم
FACULTY OF SCIENCES
قـسـم الإعـلام الآلـي
DEPARTMENT OF COMPUTER SCIENCE

MASTER THESIS

Field : Mathematics and Computer Science

Option : Computer Science

Specialization : Information and Decision system

Submitted by : Chouikh Anfal

Theme

Using AI For Agriculture Crops Prediction and Crops Diseases Prediction

Jury members :

Mrs.	Zohra Abdelhafidi	MC(B)	(University of Laghouat)	President
Ms.	chellama BELKACEM	MC(A)	(University of Laghouat)	Examiner
Mr.	Taher BENDOUMA	MC(B)	(University of Laghouat)	Examiner
Ms.	Mustapha BOUAKKAZ	MC(B)	(University of Laghouat)	Advisor

Academic Year 2023/2024

بسم الله الرحمن الرحيم

Dedications

With heartfelt gratitude and profound appreciation,
I would like to dedicate this thesis
To the 40,000 martyrs since the 7th of October,
To those in tents and prisons
To all those who are oppressed
To those struggling for dignity
To the free souls of this world
To every man and woman fighting for freedom.

Anfal chouikh

Acknowledgments

First and foremost, no volume of words is enough to express praise to *Allah Almighty*. All praise is due to Allah, the Lord of the Worlds.

I extend my gratitude towards my supervisor Dr.Musatapha Bouakkaz , for his guidance and his advice which allowed me to carry out this thesis.

My sincere thanks go to the committee members : Dr.Zohra Abdelhafidi, Dr.chellama BELKACEM, and Dr.Taher BENDOUMA , for accepting to review my work.

My deepest appreciation goes to my beloved parents, to my beautiful sisters, and to my cherished brothers, whose love, encouragement, and sacrifices have been always my driving force. Your unwavering support and belief in me have given me the strength to overcome challenges and pursue my goals and be the person I am today.

I address my warmest thanks to my dearest friends, for their constant support, for our deep conversations, and for our laughs that have kept me going through the highs and lows of this academic pursuit.

Lastly, I do not forget my gratitude to all the professors of the Computer Science Department at Laghouat University and every professor pass in my academic career, and thanks to everyone who contributed to the development of this work from near or far. I will be forever grateful for all those who have made this journey possible.

Anfal CHOUIKH , September 2024

Using AI for Agriculture crops prediction

Abstract

In recent years, the significant effects of climate change on agriculture have become increasingly apparent. Rising temperatures, erratic weather patterns, and shifting climatic conditions have introduced challenges for agriculture which threaten crop yields due to altered rainfall patterns, heat stress, and increased occurrences of extreme weather events such as droughts, floods, and storms. Such disruptions not only impact crop productivity but also pose substantial risks to global food security and livelihoods,to address these challenges, our work focuses on leveraging AI and machine learning techniques for two critical applications: predicting suitable crops and detecting plant diseases. By using these advanced technologies, we aim to assist farmers in making well-informed decisions about crop selection, optimal planting times. Additionally, detecting plant diseases early can help mitigate their impact, ensuring more resilient and sustainable food production in the face of climate change.

Keywords : AI,ML,agriculture crops,climate change.

استخدام الذكاء الاصطناعي لتوقع المحاصيل الزراعة و أمراض النباتات

.

ملــخـــص

في السنوات الأخيرة، أصبحت التأثيرات الكبيرة لتغير المناخ على الزراعة أكثر وضوحاً حيث أدت الزيادة في درجات الحرارة، وأنماط الطقس غير المنتظمة، وتغير الظروف المناخية إلى ظهور تحديات للزراعة تهدد إنتاج المحاصيل بسبب تغير أنماط الأمطار وزيادة حدوث الأحداث المناخية المتطرفة مثل الجفاف، والفيضانات، والعواصف. هذه الاضطرابات لا تؤثر فقط على إنتاجية المحاصيل ولكنها تشكل أيضاً مخاطر كبيرة على الأمن الغذائي العالمي وسبل العيش. لمواجهة هذه التحديات، يركز عملنا على استغلال تقنيات الذكاء الاصطناعي والتعلم الآلي في تطبيقين حاسمين: التنبؤ بالمحاصيل المناسبة وأمراض المحاصيل الزراعية. من خلال استخدام هذه التقنيات المتقدمة، نهدف إلى مساعدة المزارعين في اتخاذ قرارات مدروسة بشأن اختيار المحاصيل وأوقات الزراعة المثلى. بالإضافة إلى ذلك، يمكن أن تساعد الكشف المبكر عن الأمراض النباتية في التخفيف من تأثيرها، مما يضمن إنتاجاً غذائياً أكثر مرونة واستدامة في مواجهة تغير المناخ.

Utilisation de l'IA pour la prédiction des cultures agricoles et des maladies des plantes

Résumé

Ces dernières années, les effets significatifs du changement climatique sur l'agriculture sont devenus de plus en plus évidents. L'augmentation des températures, les schémas météorologiques erratiques et les conditions climatiques changeantes ont introduit des défis pour l'agriculture, menaçant les rendements des cultures en raison des changements dans les régimes de précipitations, du stress thermique et des occurrences accrues d'événements climatiques extrêmes tels que les sécheresses, les inondations et les tempêtes. Ces perturbations affectent non seulement la productivité des cultures mais posent également des risques considérables pour la sécurité alimentaire mondiale et les moyens de subsistance,pour relever ces défis, notre travail se concentre sur l'utilisation des techniques d'IA et d'apprentissage automatique pour deux applications cruciales : la prédiction des cultures adaptées et la détection des maladies des plantes. En utilisant ces technologies avancées, nous visons à aider les agriculteurs à prendre des décisions éclairées sur le choix des cultures et les moments optimaux pour les planter. De plus, la détection précoce des maladies des plantes peut aider à atténuer leur impact, garantissant une production alimentaire plus résiliente et durable face au changement climatique.

Contents

0.1 General Introduction

Agriculture, being the backbone of many economies worldwide, faces numerous challenges, ranging from unpredictable weather patterns to evolving market demands. In response to these challenges, predictive modeling techniques have emerged as powerful tools to forecast agricultural outcomes and optimize crop production. By leveraging historical data, climatic variables, soil characteristics, and other relevant factors, predictive models can provide valuable insights into crop yields, pest outbreaks, and optimal planting strategies.

The prediction of agriculture crops is not merely a scientific endeavor but also a practical necessity for farmers, policymakers, and agricultural stakeholders. Accurate predictions enable farmers to make informed decisions about crop selection, irrigation, fertilization, and pest management, thereby maximizing yields and minimizing risks. Moreover, policymakers rely on crop predictions to formulate effective agricultural policies, ensure food security, and mitigate the impact of climate change on agriculture.

machine learning techniques enable farmers and policymakers to optimize resource allocation, minimize input costs, and maximize productivity, thereby enhancing agricultural efficiency and sustainability. From small-scale family farms to large commercial operations, the integration of machine learning into agricultural practices holds immense potential to revolutionize the way we cultivate crops and manage agricultural systems. in this study we aim to predict the suitable agriculure crops for different regions using a machine learning algorithm based on weather features, we have orgonised the study as follows :

- in the first chapter we explain the agriculture in algeria and the effects of climate change on agriculture crops production and how the prediction of crops might be the solution.
- Chapter two provides a comprehensive review of literature on using machine learning (ML) algorithms and multi-sensor remote sensing (MRS) for crop yield prediction.
- Chapter three introduces the K-Means Algorithm for clustering datasets in unsupervised machine learning. It outlines the algorithm's steps, discusses the dataset and tools used and the results of applying the algorithm to predict suitable agricultural crops based on weather data are discussed. Different clusters are identified for cities based on temperature, precipitation, and humidity.

Chapter 1

Introduction

1.1 Context

Agriculture in Algeria is characterized by a diverse range of crops due to its varied climatic regions, from the Mediterranean coast to the Sahara Desert. The northern regions, with their more favorable climate, primarily cultivate cereals such as wheat and barley, which are staple crops. Vegetables, including potatoes, tomatoes, and onions, are also widely grown, contributing significantly to both local consumption and export markets. Fruit production is notable, with citrus fruits, dates, and olives being prominent, especially in coastal and oasis areas. Despite challenges like limited arable land and water scarcity, Algeria's agricultural sector shows potential for growth through modernization and investment initiatives.

1.2 Problem Statement and Objective

Global warming can change the types of crops that thrive in a region known for a particular crop type, including Algeria. The suitability of a region for specific crops is closely tied to its climate, including factors like temperature, rainfall, and seasonal patterns. As global temperatures rise and weather patterns become more erratic, the climate conditions that favor certain crops may shift, making it more challenging for traditional crops to thrive in the same regions. For example, if a region has historically been suitable for growing wheat due to its moderate temperatures and sufficient rainfall, increasing temperatures and changes in precipitation patterns associated with global warming could make conditions less conducive for wheat cultivation in Algeria. In such a scenario, farmers may need to consider switching to alternative crops that are better adapted to the changing climate.

1.3 Solution

One solution to mitigating the impact of global warming on agriculture in Algeria is to predict regions with similar climate conditions that may support the same agricultural production. By utilizing climate modeling and data analysis, it's possible to identify areas with comparable environmental factors to those currently favorable for specific crops. This predictive method allows farmers and policymakers to anticipate shifts in crop suitability and plan accordingly by relocating cultivation to regions with analogous climates. By adopting this proactive strategy, agricultural stakeholders can better prepare for climate change, ensuring the continued productivity and resilience of Algeria's agricultural sector.

Chapter 2

Literature Review

2.1 Introduction

With advancements in machine learning (ML), new opportunities have emerged to enhance the precision of yield forecasts. This chapter provides the application of popular ML algorithms for predicting agricultural crop yields, focusing on their effectiveness in processing historical data versus future forecasting. Utilizing biophysical crop models and synthetic datasets, the study assesses the potential of ML in accurately predicting yields for crops such as sunflower and wheat. Additionally, the chapter provides insights and guidelines for the fair and effective application of ML in the context of yield prediction, highlighting its advantages and challenges in various agricultural settings.

2.2 Using machine learning for crop yield prediction in the past or the future

This paper examines the use of popular machine learning (ML) algorithms for yield prediction, essential for planning farm management and agricultural sectors,It hypothesizes that ML algorithms are more accurate for historical data than future forecasting,Using the biophysical crop models Ceres-Wheat and OilcropSun, the paper generates synthetic datasets to analyze ML's effectiveness in predicting sunflower and wheat yields and to propose guidelines for fair ML application in yield prediction [2].

2.2.1 Dataset

The study focused on farms across five regions in Spain, characterized by Mediterranean climates with varying rainfall levels and soil types. Each region had farms with different soil depths, cultivars, sowing dates, N fertilizer amounts, and supplementary irrigation plans.

Table 2.1: Farm characteristics of the different regions. [?]

Region	Soil type	Soil depth	Farms	Mean farm size	Range farm size
Lobon	Sandy loam	1.25,1.50,1.75	15	20	15-25
Belmez	Silt loam	0.75,1.0,1,25	12	30	25-35
Cordoba	clay laom	0.5,1,1.5	11	40	35-45
Baza	Silt loam	0.5,0.75,1.0	12	25	20-30
Casttelon	Sandy loam	1.25,1.50,1.75	15	20	15-20

Simulations were conducted for monoculture of sunflower and wheat crops using DSSAT 4.8, with simulations spanning from 2000 to 2020. Cultivar changes occurred within the simulation period, affecting planting density and yield potential.

Different algorithms were evaluated for predicting sunflower yield, with random forest (RF) showing the most reliability and success, particularly in forecasting.

2.2.2 Result

The study found that machine learning algorithms had limited success in predicting sunflower and wheat yields across various regions in Spain, performing only slightly better than a basic average-yield baseline. Random Forest (RF) algorithms consistently performed at least as well as the best guess estimate based on past farm yield data, which is not guaranteed by Artificial Neural Networks (ANN) or linear models unless sufficient data is available.

2.3 Crop monitoring by multimodal remote sensing

This paper reviews the literature on the application of MRS in various agricultural processes, including crop monitoring and classification, land cover mapping, yield estimation, and detection of plant stress due to climate change [3].

2.3.1 Yield Estimation in Agriculture

Accurate and timely crop yield estimation is vital for ensuring food security and helping farmers maximize profits. Various remote sensing methods are employed for crop yield estimation, including hyperspectral and satellite imaging. Researchers frequently use unmanned aerial vehicles (UAVs) equipped with different sensors to gather high-resolution spatiotemporal data. UAV-based imaging has proven to be highly correlated with crop yield. For example, UAV-based RGB and multispectral images are used for predicting rice yield, with time-series vegetation indices (VIs) like NDVI and VARI providing high accuracy [3] .

2.3.2 Integration of Multi-Sensor Data

Recent studies highlight the benefits of integrating different sensor modalities on UAVs for yield estimation. Combining data from RGB, multispectral, and thermal infrared cameras, researchers develop regression models that outperform single-modal approaches. This integration improves the estimation accuracy for various crops, including wheat and soybeans. Machine learning methods such as Random Forest Regression (RFR), Deep Neural Networks (DNN), and Support Vector Regression (SVR) enhance the predictive power when used with multimodal data.

2.3.3 Advantages and Challenges of Multisensor Remote Sensing (MRS) for Crop Monitoring

2.3.4 Advantages

Enhanced Accuracy and Reliability : MRS integrates data from various sensors, capturing multiple aspects of crop growth and development, leading to more accurate and reliable crop monitoring .

Overcoming Sensor Limitations : By combining data from different sensors, MRS provides more frequent and comprehensive data coverage, mitigating the limitations of individual sensors .

Comprehensive Crop Understanding : MRS captures diverse aspects of crop development, such as plant physiology, soil moisture, and nutrient status, offering a more holistic view of crop growth .

Improved Agricultural Resilience : Utilizing MRS with artificial intelligence enhances the resilience of agricultural systems by providing timely and precise information on crop conditions .

2.3.5 Challenges

Data Processing Requirements : The large volumes of data from multiple sensors require efficient processing and analysis methods to extract useful information for crop monitoring .

Integration Difficulties : Differences in spatial and temporal resolution, spectral range, and data formats among sensors complicate data fusion and analysis, affecting accuracy and reliability .

Cost Barriers : The high costs of acquiring and processing remote sensing data can limit its widespread use, especially in developing countries .

Data Quality Variability : Variations in the quality and consistency of remote sensing data due to factors like sensor calibration and atmospheric conditions pose challenges for data fusion and analysis .

Conclusion The paper emphasizes the significant advantages of integrating diverse data sources through MRS for crop monitoring over relying on single-mode data. MRS applications have been successfully implemented in various agricultural domains, including crop monitoring, land cover mapping, yield estimation, and stress detection due to climate change. Satellite imaging is favored for crop classification and land cover mapping, while UAV-based multimodal imaging is mainly used for yield estimation. Thermal infrared imaging is crucial for detecting drought and flood stress in crops. MRS proves to be a valuable tool for enhancing agricultural efficiency and addressing complex challenges in the agricultural landscape.

2.4 Approach

Our process in this study involves using clustering techniques, K-means clustering, to group regions with similar climatic conditions without using predefined labels. Since no labeled data (e.g., known crop types for each city) is used to guide the grouping, it is classified as an unsupervised learning approach.

The goal is to identify patterns and natural groupings within the weather data, which can then inform crop suitability predictions based on the identified clusters.

2.5 Conclusion

This chapter examines the use of machine learning (ML) and multi-sensor remote sensing (MRS) for crop yield prediction. ML algorithms, especially Random Forest (RF), show potential for reliable yield prediction but require high-quality data. Integrating MRS data enhances accuracy by providing a comprehensive view of crop growth, though challenges such as data processing, integration, and costs remain. Combining ML and advanced remote sensing methods can significantly improve crop yield prediction, aiding in agricultural planning and sustainability

Chapter 3

K-means algorithm

3.1 Introduction

This Chapter introduces the K-Means Algorithm, a powerful tool in unsupervised machine learning for clustering datasets [4]. It outlines the steps of the algorithm and provides a real-world example for better understanding. The chapter also introduces the dataset used for analysis, sourced from different regions of Algeria, and discusses the tools and libraries employed for implementation. Finally, it briefly summarizes the outcomes of applying the algorithm to predict suitable agricultural crops for various cities based on weather data and discusses the limitations of the study.

3.2 Definition

K-means clustering is a very famous and powerful unsupervised machine learning algorithm [4] . It is used to solve many complex unsupervised machine learning problems.

K-Means Clustering is an unsupervised machine learning algorithm used to group unlabeled datasets into distinct clusters. This process involves teaching a computer to handle unclassified data, allowing the algorithm to operate without supervision.

The algorithm's primary task is to organize unsorted data based on similarities, patterns, and variations, without prior training data.

3.3 K-means steps

The algorithm will categorize the items into k groups or clusters of similarity. To calculate that similarity, we will use the Euclidean distance as a measurement.

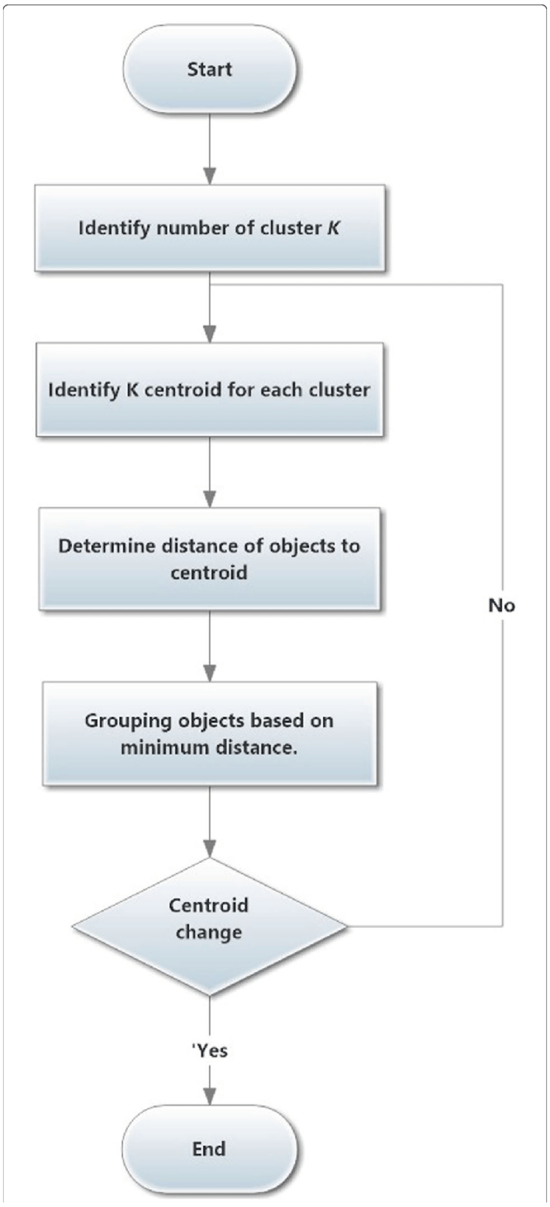

Figure 3.1: K-means steps

The algorithm works as follows :

1. Initialization : Randomly initialize k points, called means or cluster centroids. Assignment : Categorize each item to its closest mean.

2. Update : Update the mean's coordinates, which are the averages of the items categorized in that cluster so far.

3. Iteration : Repeat the assignment and update steps for a given number of iterations until convergence is achieved.

The "points" are called means because they represent the mean values of the items categorized in them.

3.4 Example

A bank wants to offer credit cards to its customers but manually reviewing each customer's details is impractical due to the large number of customers. Instead, the bank can segment its customers into groups based on factors like income and debt using K-means clustering.

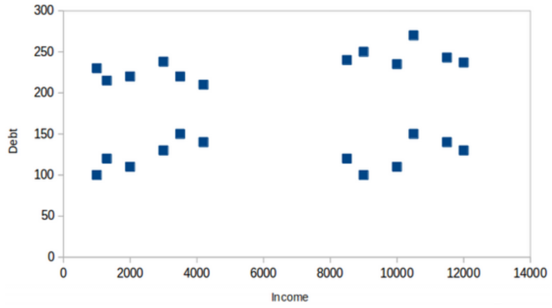

Figure 3.2: customers raw data.

The first figure shows data before applying the K-Means Clustering algorithm, with all three different categories mixed together. In the real world, such mixed data makes it difficult to distinguish between different categories.

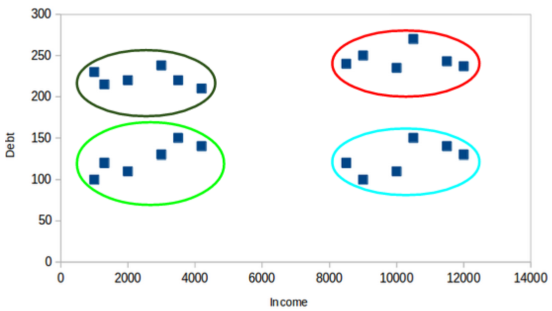

Figure 3.3: customers incomes after clustering.

the second figure shows the customers incomes after applying k-means clustering the bank identified four distinct clusters . These clusters help the bank to efficiently create targeted strategies and offer appropriate discounts to its customers. Clustering thus aids in grouping customers and streamlining the decision-making process for credit card offers.

3.5 Dataset

In this study,we used weather data from various cities across different regions of Algeria. The dataset included temperature, humidity, and precipitation data for each city from 1981 to 2022.

The cities were chosen to represent a diverse range of climatic conditions from different geographical areas :

North : Algiers and Blida

South : Laghouat and Ghardaia
East : Tebessa and Souk Ahras
West : Naama and Sidi Bel Abbes

3.6 Data Collection

The weather data used in this study was sourced from NASA, the National Aeronautics and Space Administration. NASA's comprehensive dataset, available as Excel sheets on their official website, provided detailed and reliable weather information for the selected cities in Algeria.

3.7 Libraries and platforms

A wide range of tools played a crucial role in successfully completing this work. In the following sections, we will provide a detailed overview of each tool, accompanied by a brief explanation of how it contributed to the overall process. These tools were essential in various stages of the project, from data processing and analysis to model development and evaluation, ensuring that each task was executed efficiently and accurately.

3.7.1 Python

Python is a high-level, object-oriented programming language that has the capability to handle various data structures, in addition to offering predefined packages and libraries. It is widely regarded as the most used language in the field of artificial intelligence (AI) due to its support for powerful libraries such as TensorFlow and Keras.

Python provides extensive functionalities for data manipulation, automation, and algorithm development, Its ease of integration with other programming languages and environments further enhances its appeal across diverse fields like machine learning, web development, and scientific computing.

Python's extensive ecosystem of libraries like NumPy, Pandas, and SciPy, as well as frameworks for deep learning and AI, make it indispensable for developing complex models and performing advanced data analysis. Its open-source nature and large community contribute to continuous improvements and a wide range of resources for developers. The following Python libraries were employed to cluster the data :

3.7.2 Pandas

For data manipulation and analysis, providing data structures and functions needed to work with structured data seamlessly.

3.7.3 Matplotlib

A plotting library used for creating static, animated, and interactive visualizations in Python. It was essential for visualizing the clusters and patterns in the data.

3.8 Results

To predict the agricultural crops suitable for various cities, we applied the K-Means Clustering algorithm to the weather data of these cities. Different numbers of clusters were chosen for each weather attribute.

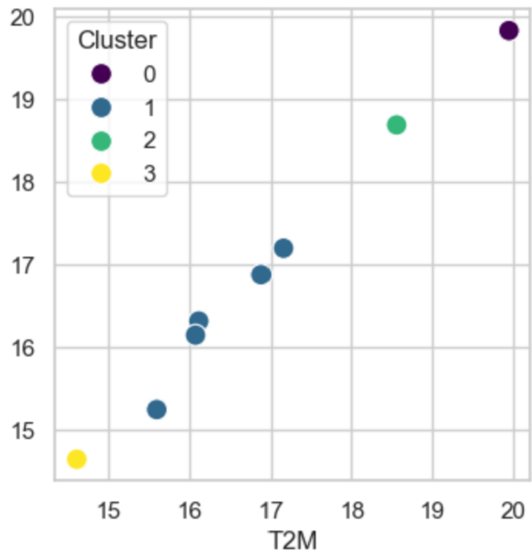

Figure 3.4: Cities clustered by Temperature.

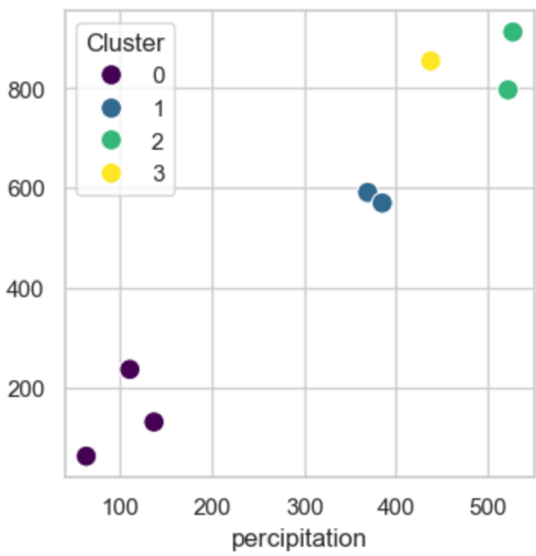

Figure 3.5: Cities clustered by Percipitation.

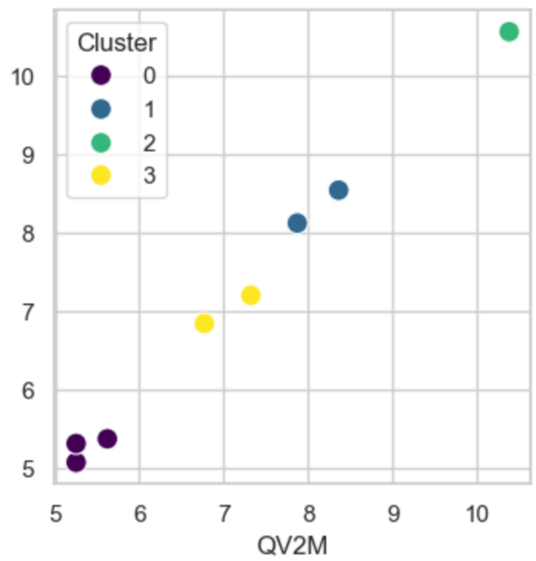

Figure 3.6: Cities clustered by Humidity.

Results for different numbers of clusters can be found in the appendix A.

3.9 Discussion

Medium Temperature and Humidity : Blida and Souk Ahrass are suitable for crops that require moderate conditions.

Suitable Crops :

Tomatoes : Thrive in moderate climates with sufficient moisture. Bell Peppers : Require moderate temperature and humidity for optimal growth. Citrus Fruits (e.g., oranges, lemons) : Prefer mild temperatures and consistent humidity.

High Rainfall and Medium Humidity : Tebessa and Sidi Bel Abbes can support crops that need more water and moderate humidity. Suitable Crops :

Rice : Requires abundant water and grows well in regions with high rainfall. Sugarcane : Needs a lot of water and thrives in humid conditions. Beans (e.g., kidney beans, black beans) : Prefer moderate humidity and high moisture levels.

Low Rainfall and Low Humidity : Laghouat, Ghardaya, and Naama are ideal for drought-resistant crops. Suitable Crops :

Dates : Highly drought-resistant and thrive in arid conditions. Barley : Adapted to dry conditions and can withstand low humidity. Carrots : Can grow in low moisture conditions, provided they are well-watered occasionally.

3.10 Conclusion

In conclusion, this study utilized the K-Means Clustering algorithm to predict suitable agricultural crops for various cities based on weather data. By clustering cities according to temperature, precipitation, and humidity, we identified regions conducive to different crop types.For instance, cities like Blida and Souk Ahrass were found to be suitable for crops requiring moderate temperature and humidity, such as tomatoes, bell peppers, and citrus fruits. Conversely, cities like Tebessa and Sidi Bel Abbes were identified as favorable for crops needing high rainfall and moderate humidity beans. Additionally, Laghouat, Ghardaya, and Naama were deemed ideal for drought-resistant crops like dates, barley, and carrots.

3.11 Future Perspectives

We aspire to implement in our forthcoming work . . .

- Expansion of Dataset : Future research could focus on expanding the dataset to include a more extensive range of cities and regions across Algeria. This would provide a more comprehensive understanding of the diverse climatic variations and their impact on crop suitability throughout the country.

- Utilization of Comprehensive Weather Data : Incorporating a more extensive set of weather variables and details, such as wind speed, solar radiation, soil moisture levels, and atmospheric pressure, can enrich the predictive models and improve their accuracy. Access to comprehensive weather data can provide deeper insights into the complex interactions between climatic factors and crop growth, enabling more precise predictions of crop yields and suitability across diverse geographic regions.

- Integration of Advanced Techniques : Incorporating advanced machine learning techniques beyond K-Means clustering could enhance the accuracy and robustness of crop yield predictions. Techniques such as deep learning algorithms or ensemble methods could be explored to improve the predictive power of the models.

- Decision Support Systems : Developing decision support systems (DSS) based on predictive models can empower farmers with actionable insights for crop planning, resource allocation, and risk management. Interactive dashboards or mobile applications could deliver tailored recommendations to farmers based on local weather conditions and soil characteristics.

- Integration with Agroecological Practices : Integrating crop yield prediction models with agroecological principles and sustainable farming practices can promote environmentally friendly agriculture while enhancing productivity and resilience against climate change impacts.

Chapter 4

Crops Diseases prediction

4.1 Introduction

This chapter explores the application of Convolutional Neural Networks (CNNs) in predicting agricultural crop diseases, with a particular focus on their effectiveness in enhancing disease detection and prediction accuracy. CNNs, a class of deep learning algorithms well-suited for image analysis, offer significant advantages in processing and interpreting complex visual data from agricultural environments. By leveraging large and diverse datasets, this study evaluates the potential of machine learning models to accurately forecast disease occurrences in crops such as corn and tomato.

4.2 Literature Review

4.2.1 Image-Based Wheat Fungi Diseases Identification by Deep Learning [1]

Wheat is a crucial global crop, and its yield is significantly impacted by fungal diseases like rust, septoria, and powdery mildew. These pathogens cause yield losses of 15-30% or more depending on regional and seasonal factors. The best way to mitigate these diseases is through early detection and timely application of fungicides. Traditional visual methods, which require expert knowledge, have been supplemented by advanced technologies such as molecular and spectral analysis, and, more recently, digital image analysis.

Machine learning and neural network approaches, particularly convolutional neural networks (CNNs), have shown great promise in diagnosing plant diseases by analyzing digital RGB images. Advanced architectures such as AlexNet, ResNet, and EfficientNet have significantly improved disease recognition accuracy. EfficientNet, with its scalable architecture, has outperformed previous models in various tasks, including wheat disease identification.

4.2.2 Dataset

The image dataset used in this research comes from various sources, including the Zindi platform, the Plant Disease Detection Platform (PDDP), Google Images, and images taken by the authors in labs and fields. Redundant images were removed using the aHash algorithm, leading to the exclusion of 612 images. The dataset was manually labeled, classifying images as healthy or diseased (six fungal diseases), with tags for plant development stages.

Table 4.1: Sources of images in the dataset Wheat Fungi Diseases.

Source	number of Images	Reference
Zindi.Africa	874	[5]
Google Images	259	-
PddP	121	[6] [7]
Saint Petersburg	367	this work
Novosibirsk	793	this work

The neural network architecture employed was EfficientNet-B0, trained using transfer learning from the ImageNet dataset. Two fully connected layers were added for classification, and the network was trained on 512x512 RGB images with dropout regularization. The images were split into training, validation, and test sets. Performance metrics like accuracy and F1-score were calculated to assess the model.

Three training strategies were tested : baseline, augmentation (image transformations like rotation, reflection, etc.), and style transfer (using FDA). Additionally, a Grad-CAM visualization technique was used to map network activations.

The model was deployed as a Telegram chatbot, allowing users to upload plant images for disease classification via a simple user interface. The chatbot was integrated with a server-side inference system using FastAPI, RabbitMQ, and PyTorch.

4.2.3 EfficientNet Strategy

- The EfficientNet-B0-baseline strategy refers to the use of the EfficientNet-B0 model in its basic form without any enhancements like data augmentation or additional techniques. In this strategy :

- The model is trained using a standard dataset without applying extra transformations or adjustments to the images. The primary goal is to establish a baseline for comparison, meaning it shows how well the model performs under normal conditions.

- EfficientNet-B0 is part of the EfficientNet family, which is designed to optimize accuracy and computational efficiency by balancing network depth, width, and resolution.

In the study, the EfficientNet-B0-baseline model achieved an accuracy of 0.933 on the test sample, serving as a reference point to measure the effectiveness of more advanced training strategies, like using data augmentation or FDA.

4.2.4 EfficientNet strategy with augmentation

The EfficientNet-B0-augm strategy involves training the model using data augmentation techniques to enhance its performance. The model was trained with a label-smoothing method that introduced 1% random noise. Image augmentations were applied using the Albumentations library with a 0.3 probability. These augmentations included :

- Rescaling using the IAAPerspective method.
- Random rotation (15 degrees).
- Vertical and horizontal flips.
- Brightness and contrast adjustments.
- Random masking of square areas with the Cutout method.
- These augmentations help the model generalize better by adding variability to the training data.

4.2.5 The strategy with augmentation and transfer of image styles

The EfficientNet-B0-FDA strategy builds on the EfficientNet-B0-augm approach, incorporating additional style transfer techniques using the FDA (Fourier Domain Adaptation) algorithm. In this method, the FDA algorithm transforms the image styles during training. The transformation was applied with a mask parameter $= 0.01$, using the FDA() method from the domain adaptation package of the Albumentations library.

Instead of randomly selecting style images for transfer, the process identifies several typical styles that represent the dataset. These characteristic styles are then used during training to improve model generalization by exposing it to different visual styles, making it more robust in real-world scenarios where styles may vary.

4.2.6 Results

The study assessed the image classification accuracy of the EfficientNet-B0 model for wheat disease detection, focusing on different training strategies. Three strategies were tested : baseline, with augmentations, and FDA (Feature Disentanglement Augmentation). The FDA strategy achieved the highest accuracy (0.942), with minimal overfitting, as seen by the small difference between validation and test accuracies (0.1%).

Augmentation during training increased accuracy by 0.6%, and adding FDA improved it by an additional 0.3%. Disease recognition accuracy was higher for seedlings (0.963) than for adult plants (0.926). Among disease labels, "septoria" had the highest recognition accuracy (0.956), while "leaf rust" had the lowest (0.890).

Chapter 5

Convolutional Neural Networks

5.1 Introduction

Convolutional Neural Networks (CNNs) are a specialized type of neural network particularly well-suited for processing data that has a grid-like structure, such as images. CNNs are widely used in image recognition, object detection, and other tasks where spatial relationships are important.

5.1.1 Components of CNNs

5.1.2 Convolutional Layers

The core of a CNN is the convolution operation ,In these layers small filters are applied to the input image, scanning over it to detect patterns like edges, textures, or more complex features. Each filter learns to detect a different feature in the image.

- Convolution Operation it begins with The filter (usually smaller than the input, e.g., 3x3 or 5x5) slides across the input image (or the output of a previous layer). At each position, it performs an element-wise multiplication between the filter's weights and the image's pixel values, then sums these values to create a feature map.

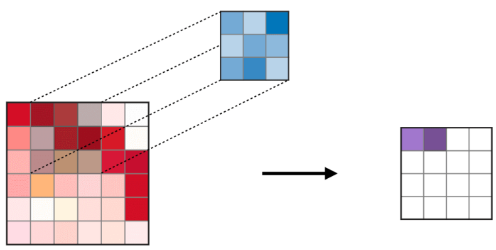

Figure 5.1: the convolution step

- Feature Maps it is The result of applying the filter across the input image , It captures the presence of the feature detected by that filter in different regions of the input.

5.1.3 Activation Function

After the convolution operation, a nonlinear activation function like ReLU (Rectified Linear Unit) is applied to the feature maps. ReLU introduces non-linearity into the model, which helps it learn complex patterns. ReLU outputs the input if it's positive and zero if it's negative, thus making the network faster and less prone to vanishing gradients.

5.1.4 Pooling Layers

Pooling layers are used to down-sample the feature maps, reducing their size and thus the number of parameters in the network. This also helps the network become invariant to small shifts and distortions in the image.

- Max Pooling : The most common pooling operation. It selects the maximum value from a small region (e.g., 2x2) of the feature map.

- Average Pooling : Instead of taking the maximum value, it averages the values in a small region.

5.1.5 Fully Connected Layers

After several convolutional and pooling layers, the high-level reasoning is done through fully connected layers, where each neuron is connected to all neurons in the previous layer. These layers resemble the layers in traditional neural networks and make the final classification decision.

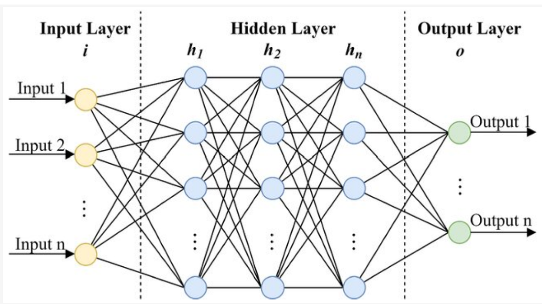

Figure 5.2: fully connected layers

5.1.6 Output Layer

The final fully connected layer outputs predictions, typically using a softmax activation function for classification tasks. The softmax function outputs a probability distribution over classes, helping the network decide which class the input belongs to.

5.2 Model Implementation

5.3 Data Collection

the dataset is quite comprehensive, encompassing a total of 18 distinct types of plants. Each plant type is accompanied by the diseases that it may be susceptible to. As a result, the dataset comprises a total of 38 categories, which include both the different plant types and the specific diseases associated with them.

the dataset used in this study is obtained from Kaggle it includes approximately 87,000 RGB images of healthy and diseased crop leaves, organized into 38 distinct classes.

The dataset is split into training and validation sets while maintaining the original directory structure. Additionally, a separate directory containing 33 test images was created later for prediction purposes.

5.4 Libraries and platforms

A diverse set of tools was vital for the successful completion of this project. The following sections will outline each tool and explain its role in the process. These tools were key throughout different stages, including data processing, analysis, model development, and evaluation, ensuring efficiency and accuracy in every task.

5.4.1 TensorFlow

Tensorflow is an open-source machine learning library developed by Google that brings together a vast array of machine learning and deep learning models and algorithms. It is designed to simplify the process of building and training neural networks and is widely used for tasks ranging from natural language processing (NLP) to image recognition.

Tensorflow provides a comprehensive toolkit for solving highly complex mathematical problems with ease. It allows researchers and developers to quickly translate experimental machine learning architectures into software applications.

Tensorflow flexible architecture enables deployment across a variety of platforms, from mobile devices to large-scale distributed systems.

TensorFlow's versatility, coupled with its vast range of tools and extensive documentation, has solidified its position as a foundational tool in the AI and deep learning communities.

5.4.2 Keras

Keras is a deep learning API written in Python that runs on top of the TensorFlow platform, Microsoft Cognitive Toolkit (CNTK), and Theano.

Keras is designed to be simple, flexible, and powerful, making it accessible to both beginners and advanced users.

Keras modular structure allows for easy customization, and it supports both convolutional networks (CNNs) and recurrent networks (RNNs), among others. Due to its integration with TensorFlow, Keras is able to leverage TensorFlow's efficiency while maintaining its simplicity.

5.5 Metrics

When evaluating the performance of a classifier, the used metrics are crucial. For instance the Confusion matrix is used to describe classifier prediction performance in two class problems. By comparing the actual and expected values of all samples, The confusion matrix visualizes properly or erroneously predicted samples of each class. It's especially beneficial in supervised learning, when each sample has a target class assigned to it. In addition there are other important metrics which are described bellow.

— **TruePositives** (TP) : These are the correctly predicted positive values which means that the value of actual class is yes and the value of predicted class is also Yes.

— **True Negatives** (TN) :These are the correctly predicted negative values which means that the value of actual class is no and the value of predicted class is also No.

— **F**alse positives (FP) and False negative (FN) occur when our actual class contradicts with the predicted class.

— **False Positives** (FB) When actual class is no and the predicted class is yes.

— Negatives (FN) When actual class is yes but the predicted class is No.

5.5.1 Precision

Precision measures the proportion of positive cases that are correctly identified by the model (True Positives) among all cases predicted as positive (True Positives + False Positives). It is also known as Positive Predictive Value.

$$\text{Precision} = \frac{TP}{TP + FP}$$

5.5.2 Recall

(Sensitivity or True Positive Rate) Recall represents the proportion of actual positive cases that are correctly identified by the model. It is also known as sensitivity or the True Positive Rate (TPR).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Low recall indicates that the model is missing a high number of actual positive cases (high number of false negatives).

5.5.3 F1 Score

The F1 score is the harmonic mean of Precision and Recall, providing a balance between them. It is useful when you want to consider both false positives and false negatives.

$$\text{F1 Score} = 2\frac{\text{Precision}\text{Recall}}{\text{Precision} + \text{Recall}}$$

This metric is particularly useful for imbalanced datasets where one class is more common than the other.

5.6 CNN Architecture Construction

The first layer is a convolutional layer with 32 filters of size (3,3), and the input images are of size 128x128 to be compatible with the images in the dataset used.

The second layer we added is a MaxPooling layer with a filter size of (2,2).

We then added two more convolutional layers with 128 , 256,256 filters respectively, followed by MaxPooling layers.

Next, 25% of the neurons from the previous layer will be randomly deactivated during training by adding a Dropout layer with a rate of 0.25.

Following that, a Flatten layer is added, and then a Dense layer with 1500 neurons.

Another Dropout layer with a rate of 0.25 is added. Finally, a Dense layer with 38 neurons is included.

5.7 Results

The accuracy of a classification model can be expressed with the following equation :

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

Table 5.1: resluts.

model set	Accuracy	loss
training	0.981037082879	0,06002769619226456
validation	0.19692450761795044	0.9507739543914795

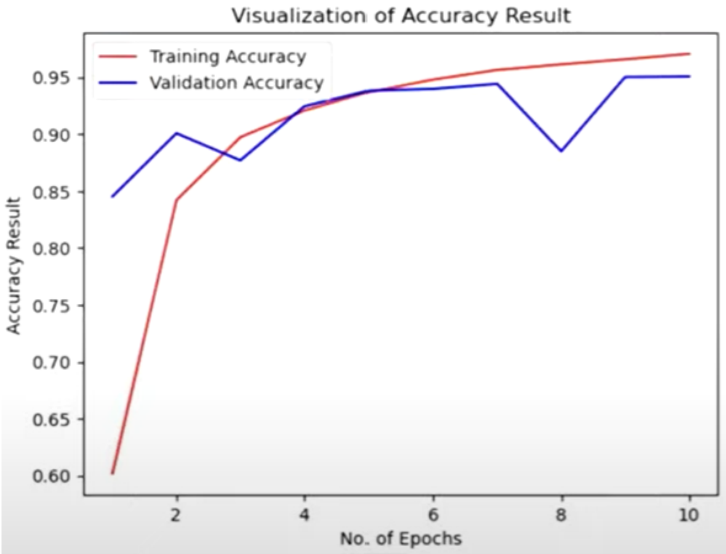

Figure 5.3: Performance of the Generated Model in term of accuracy

The reached results in term of accuracy for both predicting model are promising more than 80

5.8 Dicussion

As we can see in figure 5.3 the accuracy increases with the increase of the number of epochs. After reaching 4 epochs , the system performs an accuracy of 0.93, and it starts to slowly increase until reaching 0.98 at 10 epochs.

During the study, we ended up having a large number of possible configurations. Testing each of them was not feasible due to time constraints.

5.9 Conclusion

In this chapter, We have described the proposed approach, its implementation and the experimental phase. In addition, we have demonstrated that our Machine learning based models using the CNNS are very suitable for the prediction of plants diseases. In fact, we get a prediction with 98 to predict plants disease.

5.10 Relation between crops prediction and plants diseases prediction

5.10.1 Crops Affected by Diseases Risk

The likelihood of diseases can influence crop selection. For example, certain crops may be more susceptible to specific diseases in particular regions. Predicting the occurrence of diseases can thus feed into crop selection models to recommend crops that are less likely to be impacted by diseases in a given environment.

5.10.2 Yield Optimization

Both crop prediction and disease detection aim to maximize agricultural yield. While crop prediction ensures that the right crops are planted, disease detection ensures that those crops remain healthy throughout the growing season.

5.10.3 Impact of Diseases on Crop Prediction

Disease outbreaks in previous seasons may influence future crop predictions. For example, if a particular disease becomes endemic in an area, AI models may predict that certain crops should be avoided or rotated to reduce the risk of further disease spread.

General Conclusion

In this work we have investigated using Machine learning based approach to reach a Smart-agriculture application. In fact, Plant diseases adversely affected the quality and productivity of the Algerian crop but by applying CNNs we reached some promising result in order to predict some of the agriculture crops diseases caused by weather change.

This smart-agriculture application is performed to make farmers aware of the need to prevent them and avoid their dangers, especially plants that are not resistant such as tomatoes, potatoes and grapes. This project is in its beginnings. It can be promoted using perspectives :

Real-time Processing the system once trained, it can provide real-time disease detection from images taken by a smartphone or drone.

This can be a game-changer for farmers and agricultural businesses, enabling them to receive instant feedback on plant health without needing expertise in plant pathology.

IoT and Drones Explain the potential for integration with Internet of Things (IoT) devices or drones for automated, large-scale monitoring. For instance, drones equipped with cameras could capture plant images across a wide area, and our CNN-powered system could analyze these images for disease detection in real-time.

Bibliography

[1] Mikhail A Genaev, Ekaterina S Skolotneva, Elena I Gultyaeva, Elena A Orlova, Nina P Bechtold, and Dmitry A Afonnikov. Image-based wheat fungi diseases identification by deep learning. *Plants*, 10(8) :1500, 2021.

[2] Alejandro Morales and Francisco J Villalobos. Using machine learning for crop yield prediction in the past or the future. *Frontiers in Plant Science*, 14 :1128388, 2023.

[3] Priyabrata Karmakar, Shyh Wei Teng, Manzur Murshed, Shaoning Pang, Yanyu Li, and Hao Lin. Crop monitoring by multimodal remote sensing : A review. *Remote Sensing Applications : Society and Environment*, page 101093, 2023.

[4] Mohiuddin Ahmed, Raihan Seraj, and Syed Mohammed Shamsul Islam. The k-means algorithm : A comprehensive survey and performance evaluation. *Electronics*, 9(8) :1295, 2020.

[5] Iclrworkshop challenge 1 : Cgiar computer vision for crop. Accessed : 2021-04-10.

[6] Alexander Uzhinskiy, Gennady Ososkov, Pavel Goncharov, and Andrey Nechaevskiy. Multifunctional platform and mobile application for plant disease detection. In *CEUR Workshop Proc*, volume 2507, pages 110–114, 2019.

[7] Plant diseases detection platform image database. Accessed : 2021-04-10.

Appendices

Appendix A

Results

A.1 Values of parameter K

k=3 :

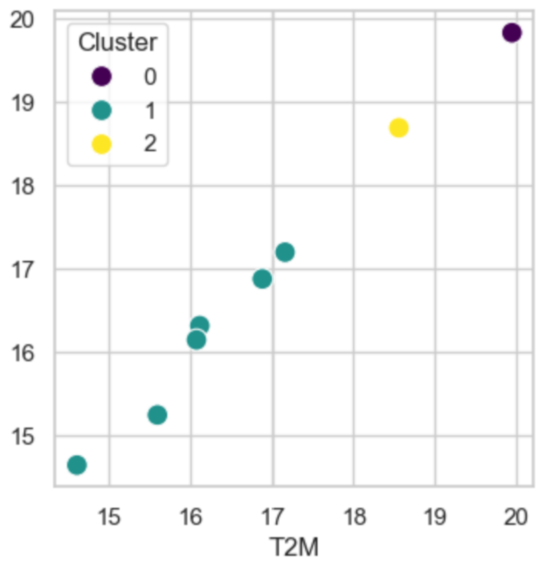

Figure A.1: Cities clustered by temperature.

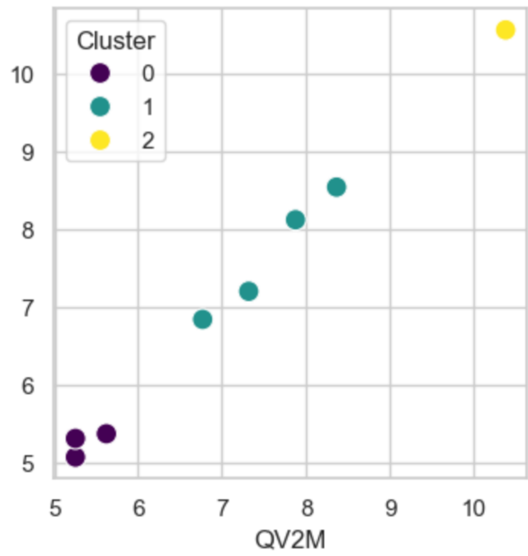

Figure A.2: Cities clustered by Humididy.

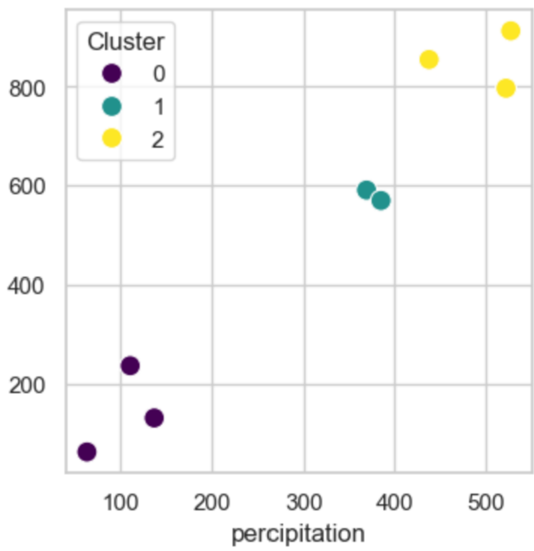

Figure A.3: Cities clustered by percipitation.

k=5 :

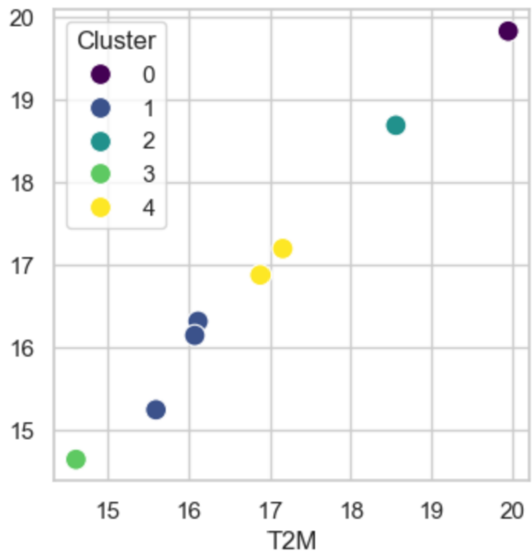

Figure A.4: Cities clustered by temperature.

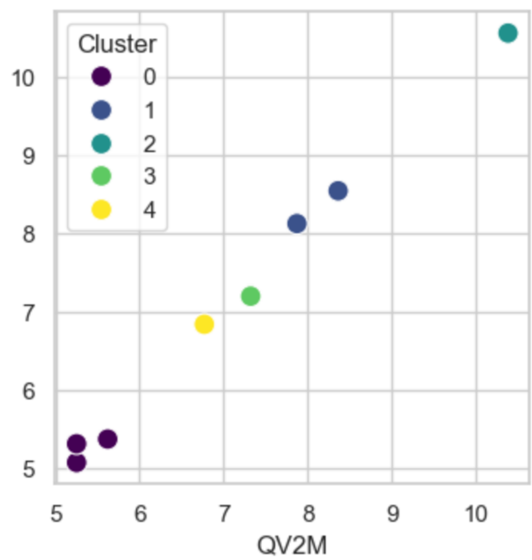

Figure A.5: Cities clustered by Humididy.

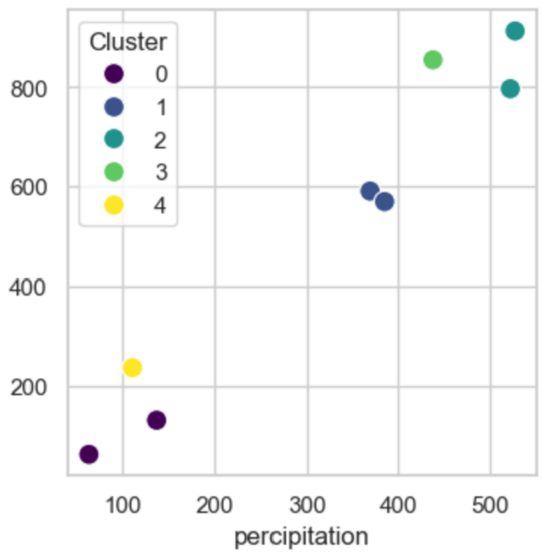

Figure A.6: Cities clustered by percipitation.

k=6 :

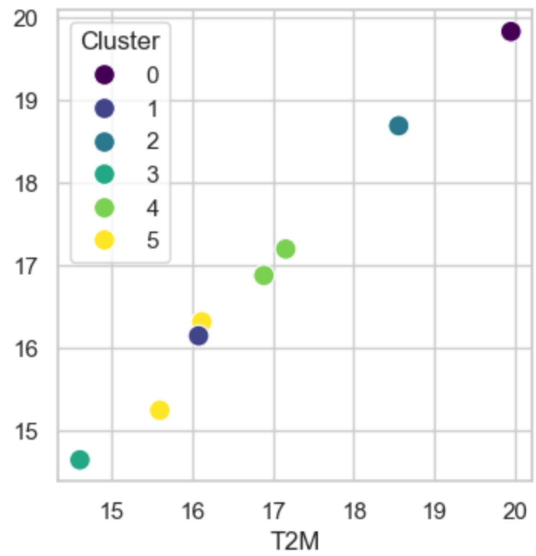

Figure A.7: Cities clustered by temperature.

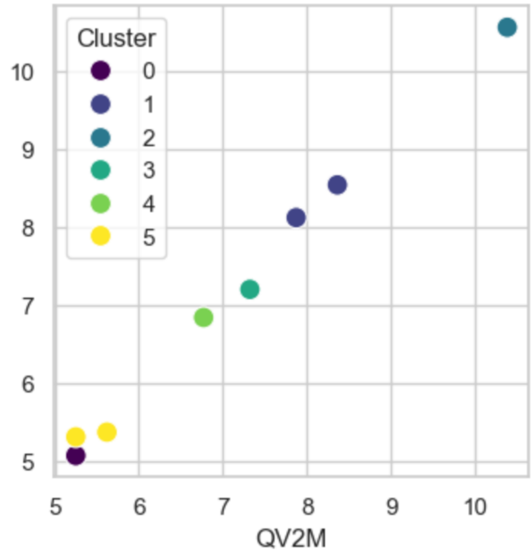

Figure A.8: Cities clustered by Humididy.

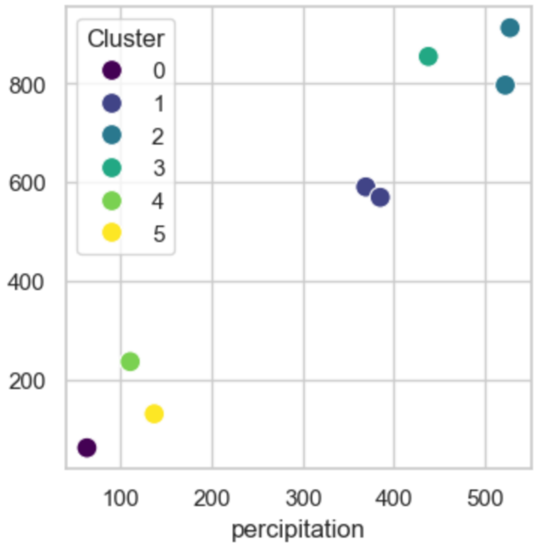

Figure A.9: Cities clustered by percipitation.