# Classification of Plant Species by Similarity Using Automatic Learning

Zacrada Françoise Odile Trey(✉), Bi Tra Goore, and Brou Marcellin Konan

Institut National Polytechnique Houphouët-Boigny, Yamoussoukro, Côte d'Ivoire
mariefranceodiletrey@gmail.com, bitra.goore@gmail.com, konanbroumarcellin@yahoo.fr

**Abstract.** The classification methods are diverse and variety from one field of study to another. Among botanists, plants classification is done manually. This task is difficult, and results are not satisfactory. However, artificial intelligence, which is a new field of computer science, advocates automatic classification methods. It uses well-trained algorithms facilitating the classification activity for very efficient results. However, depending on the classification criterion, some algorithms are more efficient than others. Through our article, we classify plants according to their type: trees, shrubs and herbaceous plants by comparing two types of learning meaning the supervised and unsupervised learning. For each type of learning, we use these corresponding algorithms which are K-Means algorithms and decision trees. Thus we developed two classification models with each of these algorithms. The performance indicators of these models revealed different figures. We have concluded that one of these algorithms is more effective than the other in grouping our plants by similarity.

**Keywords:** Automatic learning · Classification · Algorithm

## 1 Introduction

### 1.1 Context

Biodiversity conservation is based on a precise, science-based classification (i.e. a system for designating organisms). Without this classification, it will be unable to describe the multitude of species inhabiting tropical forests and compare them to the small number that live in these tropical countries. Also, without such classification, it would be impossible to identify plant species in our environment [1]. However, many of these species are either threatened with extinction or have already disappeared due to pollution and natural disasters, and others are still waiting to be discovered [2]. Plant species are important for nature and ecological balance, and many of them are raw materials for the chemical and wine industries for example… Therefore, their classification is of interest not only to botanists but also for other actors in different fields such as agronomists, environmental protectors, foresters, land managers and even amateurs or non-experts [3]. For a long time, this classification was done manually by


R. Zitouni et al. (Eds.): AFRICOMM 2019, LNICST 311, pp. 186–201, 2020.
https://doi.org/10.1007/978-3-030-41593-8_14

botanists with their own identification keys. Thus this process was slow and difficult. However, with the development of new computing tools such as Artificial Intelligence, several automatic classification methods have emerged. Artificial intelligence is defined as the set of means, theories, rules, techniques used to create machines' automatons, robots capable of simulating human behavior. These machines or artificial agents or non-human agents are effective, tireless and docile for performing repetitive tasks [4].

The main objective of this article is to reproduce all the facets of artificial intelligence in the field of botany through automatic learning by studying and training algorithms for making predictions on a large amount of botanical data. To achieve this, we firstly, analyzed the traditional classification of systematists. Secondly we used two types of learning, the clustering for unsupervised learning and decision trees for supervised learning to group plants into types, i.e. trees, shrubs, herbaceous plants. Finally, we compare the two classification methods to see which one has the best accuracy.

### 1.2 Motivation

For decades, the botanical field has remained in its traditional manual practices probably due to a lack of information and/or lack of computing skills by its specialists. In such way, for the morphological classification of their plants, a usual, flagship, popular activity, the botanists still work with dichotomous keys that are used by visual inspection of the systematist: a botanist, specialized in plant identification. The latest is quite used to this identification, which he could say, without consulting these keys, the type of plants. Would it be credible enough for novices who do not know anything about plant identification and who ask for its service? It sometimes happens that in countries such as Ivory Coast, these systematicians can be counted at their fingertips. How to transmit all these empirical knowledge to next generations?

In a context where new information and communication technologies are booming, it would be wise for the main actors of this technology, namely computer scientists, to be able to convey their knowledge. They must ensure that fields of study that do not yet demonstrate this technology can use it. Since the advent of IT in everyday practices has several advantages. Firstly, it helps saving time with the use of PLCs in manual activities making work easier and allowing results to be obtained in a short time. Then it promotes better results and automatic learning in this kind of human activity uses well-trained algorithms. However, it is their role to imitate human behavior, their results can far exceed those obtained manually. Therefore our article shows how to translate all this empirical knowledge of the systematist into algorithms. This will have the advantage of: (1) automating the classification of plants, (2) perennializing all its information and (3) allowing better results in a minimal time.

This system could therefore be used not only by experts such as the systematist but also by non-experts, such as florists, agronomists, etc.

## 2 Materials and Methods

### 2.1 State of the Art

Traditionally, botanists use identification keys to classify plants. These identifications generally concern the morphological characteristics of plants and are used manually. Authors have created a key that identifies sixteen genera in the *cyperraceas* family. To recognize the type of plant, they successively describe the characteristics of the leaves of these plants [5]. When these leaf descriptions fail to identify the plant, systematicians proceed to describe the flower of this species [6]. Others go as far as describing the fruits and even the roots. This is the case with cruciferous keys. They describe all the vegetative aspects of the plant, namely its leaves, roots, fruit and flowers [7]. For all these dichotomous keys, the operating mode remains a difficult activity. To compensate this lack, some actors are digitizing these keys. They created an automatic botanical document analysis tool, based on XML, corrected artifacts resulting from the digitization from written documents, performed a morpho-syntactic analysis for identification and finally an extraction of knowledge [8]. It is clear that this tool has not been validated by experts, which makes it unreliable. Raymond Boyd et al. have created a tool that includes a transdisciplinary database. They identified plant species by botanical nomenclature and names of these species in the *Sémé* language of Burkina Faso. Using XML, They made a semi-structured questionnaire describing the Semitic language. Then, they drew up a directory of scientific names of plants and made them correspond to their morphological representation, including a vocabulary of Semitic language [9]. Moreover, in the purpose of this idea perfection, scientists are setting up a key to determine plants by flower type. They also classify plants according to the criteria of the flower as well as identifying for each plant, the lengths and widths of the petals and sepals. Using the machine learning, they created four zones corresponding to these types of flowers, being able to be used from a computer. It is sufficient to harvest the plant, enter the different values of the petals and sepals into the system, and get automatically the result on the flower type. However, not all plants necessarily have flowers [10]. In addition, in their work, the vegetative characteristics of the plant, which are the roots, stem, leaf and flower are used. They obtained more than thirty descriptors per character. As a result, they ended up with many descriptors, to be entered which was tedious task. Nevertheless, they obtain a knowledge base of these plants. To classify plants by type, their system compares characteristics until a significant match rate is reached [11]. Sometimes the type of plant returned by the system does not correspond to reality. Other scientists have dealt with the recognition of weeds in a field. They used the leaf of the plant at its evolutionary stage. They identified the descriptors of shape, density and color. After segmentation, they pre-process the images of the leaves and apply neural networks on them to make the classification. They are able to develop an approach for the design and formation of deep

convolutional neural networks for identification a large number of plant species [12]. However, this system encounters some difficulties due the use of this system, the harvested plant must have a size that conforms to the size of the images in the database, if necessary the result would be distorted; moreover, during segmentation, the removal of the stem from the leaf eliminates certain parts, which results in a defect in the extraction of the characteristics. The omission of segmentation or retention of the stem should also increase the accuracy of classification, as the regions of the stem removed by segmentation, will be retained. In this work, it determined the name of a plant from its leaf. The methodology adopted is as follows: on a given database of 126 plants, a segmentation of the images was carried out. For a better appreciation of the characteristics, an extraction of the texture, color and shape of the leaf is done. Finally, an Android-based plant leaf identification system is built [13]. However it must be noted that in its application, the name did not necessarily correspond to the plant; since semantics such as texture, color and shape do not give enough information. Therefore, the search for new descriptors to improve identification is in prospect.

### 2.2 Dataset

The classification of plants is a very important area. Several authors based their classification on the vegetative characteristics of the plant such as roots, leaves, flower and fruit. However, when all these characters are taken into account, it ends up with an infinite number of descriptors, which often skews the results. In recent years, other authors have focused their research on describing the flower. They identified different sizes of sepals and petals. For example, flowers do not appear in all plants throughout the year, Although roots, flowers and fruits are vegetative characteristics of plants, they are not present all year round, but the leaf and stem are still present [14]. Other authors have therefore based the classification of plants on the description of the plant leaf. However, leaves alone cannot identify the plant. And need to be combined with the leaf and stem to classify plants into trees, shrubs and herbaceous types, using botanical book databases such as flora of West Tropical Africa. These books contain all the plants with their morphological descriptions accepted by the botanist community in general, more precisely that of West Africa [15]. Minimum and maximum stem heights and the maximum and minimum leaf lengths were considered. The present dataset contains 412 elements from different plants. We evaluate our system with plants harvested in the forest.

### 2.3 Proposed Method

We propose an algorithm with four distinct steps from dataset development to classification. The details of the different steps are described in the following sections (Fig. 1):

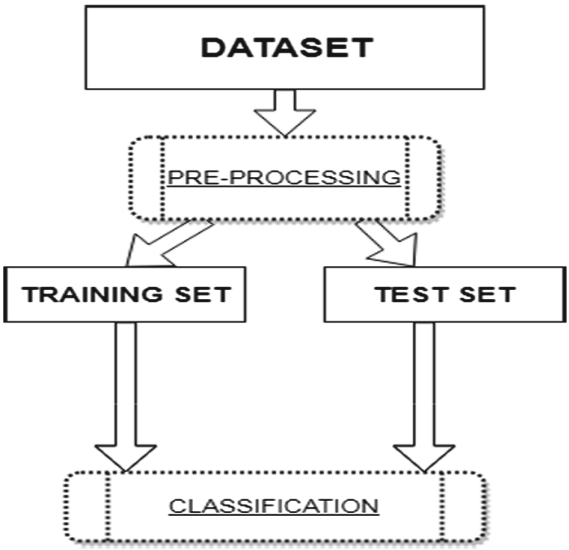


**Fig. 1.** Process for identifying type of plant

**Elaboration of Dataset**

To create our dataset, we acquire data from different sources and purge them. Figure 2 below shows how it works. According to our sources of acquisition of the different plants [15–17], we build a database. t the name of the plant with the maximum (**TMAX**) and minimum (**TMIN**) size of its stem in millimeters and the maximum length (**LMAX**) (**LMIN**) of its leaf in millimeters are reported. This is the redundant database as shown in Fig. 2.

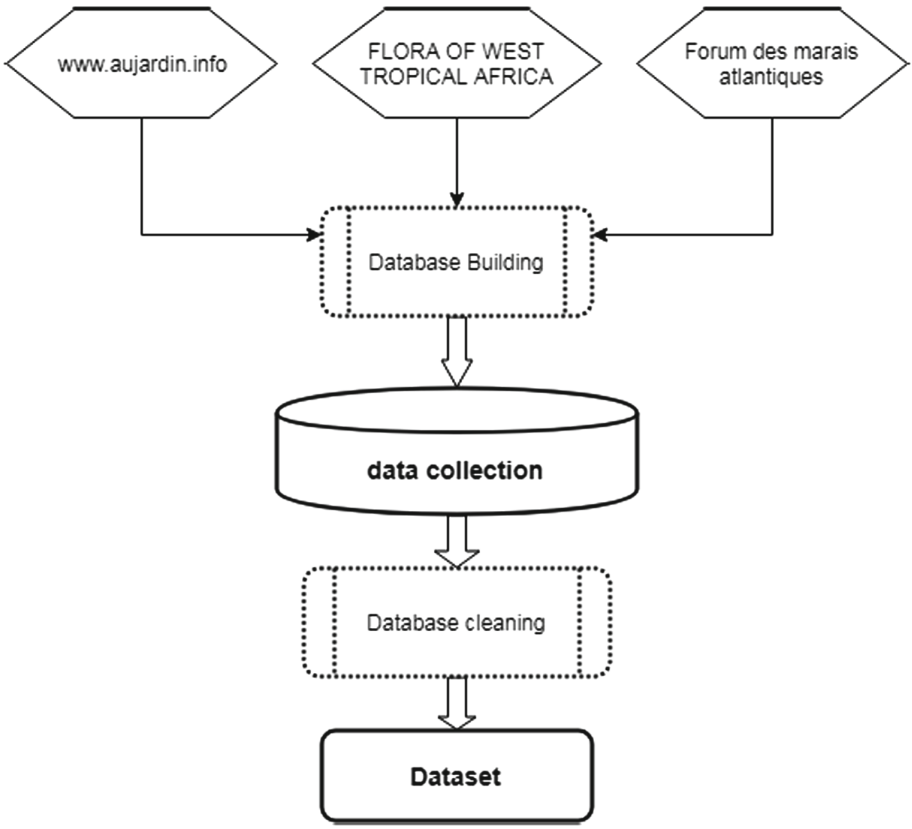


**Fig. 2.** Development of the dataset

**Preprocessing**

We remove redundant data from our database and put them all on the same scale (Table 1).

**Table 1.** Redundant data

| PLANTES | TMAX | TMIN | LMAX | LMIN |
|---|---|---|---|---|
| Thomsonii | 13716 | 13716 | 150 | 100 |
| Xylopia Africana | 12192 | 9140 | 160 | 90 |
| Staudtii | 45720 | 45720 | 160 | 90 |
| Ruberscens | 27432 | 27432 | 240 | 90 |
| Eliotii | 9144 | 9144 | 90 | 50 |
| … | …. | … | …. | …. |
| Sofa | …. | …. | …. | …. |
| Afzelii | …. | …. | …. | …. |
| Margaritaceus | …. | | …. | …. |
| Aucheri | 1500 | 1500 | 5 | 4 |
| UvariaScabrida | 9144 | 9144 | 180 | 100 |

After cleaning this database, we obtain the dataset following (Fig. 3).

| Index | TMIN | TMAX | LMIN | LMAX |
|---|---|---|---|---|
| 0 | 1.37e+04 | 1.37e+04 | 100 | 150 |
| 1 | 9.14e+03 | 1.22e+04 | 90 | 160 |
| 2 | 4.57e+04 | 4.57e+04 | 90 | 160 |
| 3 | 2.74e+04 | 2.74e+04 | 90 | 240 |
| 4 | 2.74e+04 | 2.74e+04 | 130 | 130 |
| 5 | 9.14e+03 | 9.14e+03 | 50 | 90 |
| 6 | 2.44e+04 | 2.44e+04 | 60 | 130 |
| 7 | 1.83e+04 | 1.83e+04 | 150 | 150 |
| 8 | 914 | 2.44e+03 | 40 | 80 |
| 9 | 3.05e+03 | 3.05e+03 | 45 | 150 |
| 10 | 6.1e+03 | 6.1e+03 | 70 | 120 |
| 11 | 6.1e+03 | 6.1e+03 | 60 | 90 |
| 12 | 4.57e+03 | 4.57e+03 | 70 | 250 |
| 13 | 9.14e+03 | 9.14e+03 | 100 | 180 |

**Fig. 3.** DatasetDataPlant

**Dataset Division**

Once the processing is complete, we divide the dataset into training and test sets (Figs. 4 and 5).

| | 0 | 1 |
|---|---|---|
| 0 | 1500.00 | 3.50 |
| 1 | 20.00 | 10.00 |
| 2 | 36576.00 | 240.00 |
| 3 | 1000.00 | 2.00 |
| 4 | 1500.00 | 0.90 |
| 5 | 9144.00 | 180.00 |
| 6 | 10.00 | 2.50 |
| 7 | 1000.00 | 2.00 |
| 8 | 160.00 | 7.00 |
| 9 | 9144.00 | 250.00 |

**Fig. 4.** Training set

| | 0 | 1 |
|---|---|---|
| 0 | 2438.40 | 80.00 |
| 1 | 45.00 | 1.50 |
| 2 | 1200.00 | 8.00 |
| 3 | 1300.00 | 2.00 |
| 4 | 450.00 | 90.00 |
| 5 | 9144.00 | 500.00 |
| 6 | 130.00 | 5.00 |
| 7 | 27432.00 | 160.00 |
| 8 | 170.00 | 40.00 |
| 9 | 12192.00 | 180.00 |

**Fig. 5.** Test set

### 2.4 Classification

Automatic learning is subdivided into two types: supervised and unsupervised learning. This step of our model highlights the different classifiers capable of classifying our plants. We clarify both types of classification methods. Based on the characteristics extracted from our different plants, we analyze the supervised and unsupervised classification methods. As part of our work, it will be clustering and decision trees. We will see both classifiers, which more accurately identifies our plants.

**K-Means**

Clustering is an integral part of unsupervised learning. And this type of method dedicated to unsupervised classification refers to a corpus of methods whose objective is to establish or find an existing typology, characterizing a set of n observations, based on p characteristics, measured on each of the observations [18]. By typology, we mean that the observations, although collected during the same experiment, are not all from the same homogeneous population, but rather from K different populations. In unsupervised classification, the affiliation of observations to one of the K populations is not known. It is precisely this belonging that must be found from the available descriptors.

*Let's Formalize the Problem*

The purpose of unsupervised classification is to determine groups. These groups will be referred to as homogeneous and distinct clusters. To formalize this, we start by defining the inertia of our plant cloud. Given a set of plants represented by *n* points (*P1, P2, ..., Pn*), PG is referred to as the barycenter of the cloud of these points.

$$\mathrm{P_G} = \frac{1}{n}\sum_{i=1}^{n} Pi \tag{1}$$

The total inertia is defined as follows.

$$\mathrm{I\,T} = \sum_{i=1}^{n} d^2(P_i + P_G) = \sum_{i=1}^{n} d^2 \|P_i + P_G\|^2 \tag{2}$$

where the chosen distance is the Euclidean distance. In reality, the point cloud is composed of K classes (cluster) of different points $C_1$, $C_2$, ..., $C_k$, each of these classes having for barycenter P G k, the total inertia is broken down as follows:

$$\begin{aligned} It &= \sum_{i=1}^{n} \|P_i + P_G\|^2 \\ &= \sum_{k=1}^{K} = \sum_{k=1}^{K} \sum_{i \in Ck} \|P_i - P_{Ck} - P_G\|^2 = \sum_{k=1}^{K} \sum_{i \in Ck} \|P_i - P_{Ck}\|^2 \\ &+ \|P_{Ck} - P_G\|^2 (\text{Hygens}' \text{ theorem}) \end{aligned} \tag{3}$$

$$\mathrm{IT} = \sum_{k=1}^{K} \sum_{i=1}^{n} d^2(P_i - P_{Ck}) + \sum_{k=1}^{K} n_k (P_{Ck} - P_G) \tag{4}$$

where is the number of observations of class C k

$$\mathrm{I_W} = \sum_{k=1}^{K} \sum_{i=1}^{n} d^2(P_i - P_{Ck}) \tag{5}$$

which is the sum of the distances between the points of a class and their center of gravity = intra-class inertia

$$I_{\mathrm{B}} = \sum_{k=1}^{K} n_k (P_{Ck} - P_G) \tag{6}$$

This term measures how far apart the classes are from each other = intra-class inertia, so if there are K well-identified classes, it is theoretically possible to find them

by trying all the possible groupings in k classes and choosing the one that minimizes intra-class inertia, which is the same as

$$I_T = I_w + I_B \tag{7}$$

and that $I$ T does not depend on classes. From a formal point of view, the optimal partition $C^*_k$, of observations in K classes is therefore defined as follows:

$$C_k^* = arg\,min \sum_{k=1}^{K} \sum_{i=1}^{n} d^2(P_i - P_{Ck}) \tag{8}$$

where $C_k^*$ is the set of possible partitions of the n observations in k classes. To meet our classification objective, all that remains is to identify the optimal partition.

Determines the optimal partition that will be represented by the number of Ks of classes or clusters. The difficulty of any unsupervised classification method lies in the choice of the number of class K. In most cases, this number is unknown. Concerning the use of the K-means algorithm, it is possible to plot the curve of intra- class WSS inertia as a function of $K$ (Fig. 6).

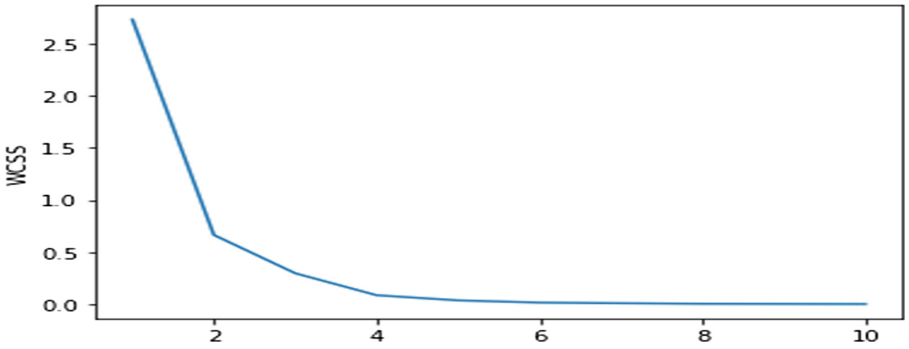


**Fig. 6.** Determination curve of the number K of classes

We try to identify the steps where we observe a break in this curve, synonymous with a strong degradation of inter-class inertia. This deterioration is the result of the strong heterogeneity of the two classes combined at the stage under consideration. It is therefore natural to consider a higher number of classes than the one for which the failure occurs, referred to as the elbow criterion, gives satisfactory results [19] (Figs. 7 and 8).

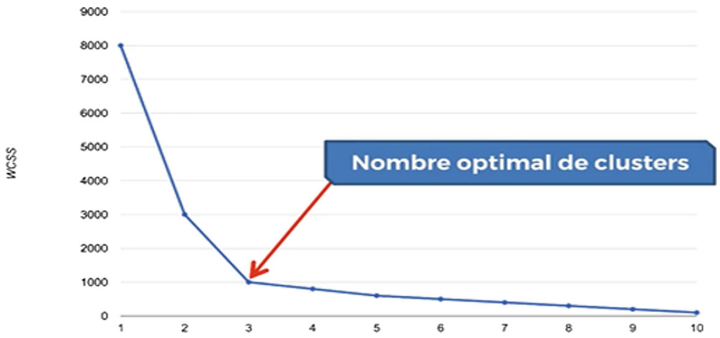


**Fig. 7.** Optimal number of classes K = 3

*Let's Apply the Model to Our Plants to Be Classified*

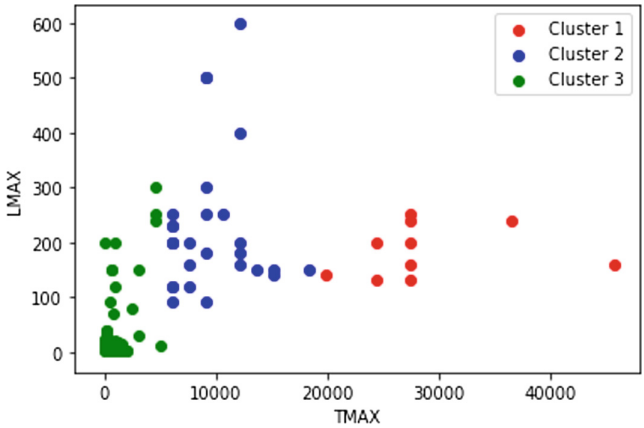


**Fig. 8.** Classified plants by K-Means

*Model Performance*

We evaluate the performance of our model. The table below lists the plants predicted from those observed (Table 2).

**Table 2.** Confusion matrix

| | Predicted plants | | | |
|---|---|---|---|---|
| Plants observed | CLUSTERS | Tree | Shrub | Herb |
| | Tree | 6 | 0 | 2 |
| | Shrub | 0 | 8 | 2 |
| | Herb | 2 | 0 | 7 |

In the following table, we calculate the basic indicators of the quality of prediction on the different clusters (Table 3).

**Table 3.** Performance indicators

| | Precision | Recall | F-measure |
|---|---|---|---|
| Tree | 0.857142 | 0.75 | 0.8 |
| Shrub | 0.8 | 1 | 0.88889 |
| Herb | 0.777778 | 0.7 | 0.736842 |

**Decision Trees**

The general purpose of a decision tree is to explain a value from a series of discrete or continuous variables. We are therefore in a very classical case of matrix $X$ with $m$ observations and $n$ variables, associated with a vector $Y$ to explain. The values of $Y$ can be of two kinds: continuous, we speak of a regression tree; or if the values of $Y$ are qualitative, we speak of a classification tree.

In our case, the values of $Y$ are either tree, shrub or herbaceous plants, the values of $Y$ are then qualitative. We are in the case of a classification decision tree. This inductive classification method has two advantages: it is quite efficient, non-parametric and linear. In principle, it will partition, by producing groups of plants, as homogeneous as possible from the point of view of the tree, shrub or herbaceous plants to be predicted, and taking into account a hierarchy of the predictive capacity of the variables stem size and leaf length considered [21].

*Formalization of the Problem*

The main principles for defining explicit explanatory rules are as follows: several iterations are necessary, at each iteration, the plants are divided. This division defines sub-populations represented by the "nodes" of the tree. Each node is associated with an output variable. The operation is repeated for each sub-population until no further separation is possible. Each leaf is characterized by a specific path through the tree called a rule. The set of rules for all sheets is the template (Fig. 9).

*Let's Apply the Model to Our Plants to Be Classified*

```
Rule 1:
If TMAX == {10668, 12181.9, 12192, 13716, 15240, 18288,
19812, 2438.4, 24384, 27432, 3048, 36576, 4572, 45720, 6096,
7620, 9144};
Then result = 100% trees
Rule 2:
If TMAX == {10, 100, 1000, 120, 1200, 130, 1300, 140, 150,
1500, 160, 1600, 170, 20, 200, 2000, 230, 250, 30, 300, 3000,
350, 45, 450, 5000, 60, 600, 700, 800, 90, 900} AND
TMIN == {10, 100, 120, 130, 140, 150, 160, 20, 200, 230, 30,
300, 350, 400, 450, 50, 60, 600, 700, 800} AND
LMAXF == {1, 1.9, 11, 120, 17, 20, 25, 3.2, 30, 6.5, 70, 90}
```

```
Then result = 90% Grass and10% Shrub
Rule 3
If TMAX == {10, 100, 1000, 120, 1200, 130, 1300, 140, 150,
1500, 160, 1600, 170, 20, 200, 2000, 230, 250, 30, 300, 3000,
350, 45, 450, 5000, 60, 600, 700, 800, 90, 900} AND
TMIN == {10, 100, 120, 130, 140, 150, 160, 20, 200, 230, 30,
300, 350, 400, 450, 50, 60, 600, 700, 800} AND
LMAXF == {1.5, 10, 150, 2, 2.5, 200, 4, 4.5, 40, 5, 7}
Then result = 100% grass
Rule 4
If TMAX == {10, 100, 1000, 120, 1200, 130, 1300, 140, 150,
1500, 160, 1600, 170, 20, 200, 2000, 230, 250, 30, 300, 3000,
350, 45, 450, 5000, 60, 600, 700, 800, 90, 900} AND
TMIN == {1000, 1200, 1300, 1500, 1600, 2000, 3000, 5000}
Then result = 100% shrub
```

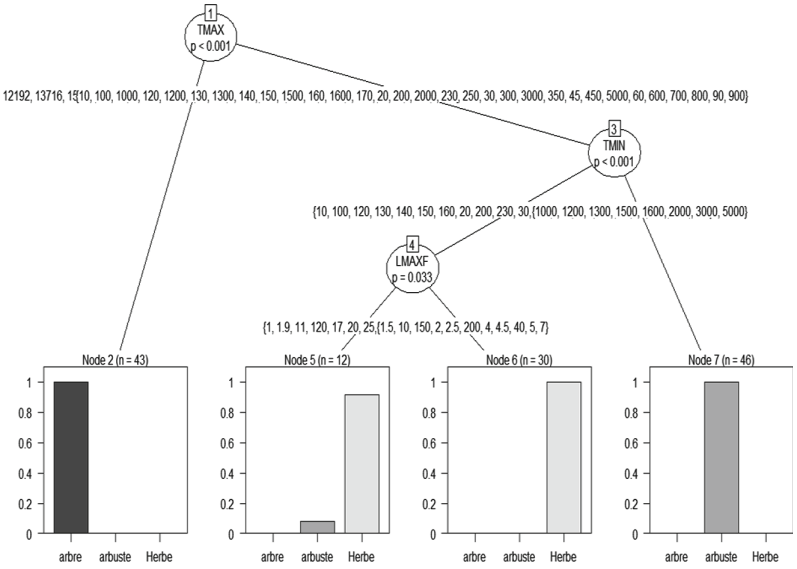


**Fig. 9.** Classified plants by decision trees

*Model Performance*

We evaluate the performance of our model. The table below lists the plants predicted from those observed (Table 4).

**Table 4.** Confusion matrix

| | Predicted plants | | | |
|---|---|---|---|---|
| Plants observed | CLUSTERS | Tree | Shrub | Herb |
| | Tree | 43 | 0 | 0 |
| | Shrub | 0 | 46 | 0 |
| | herb | 0 | 0 | 41 |

In the following table, we calculate the basic indicators of the quality of prediction on the different clusters (Table 5).

**Table 5.** Performance indicators

| Clusters | Tree | Shrub | Herb |
|---|---|---|---|
| Sensitivity | 1.0000 | 0,9787 | 1.0000 |
| Specificity | 1.0000 | 1.0000 | 0,9889 |

## 3 Discussion

We worked on both types of learning, supervised and unsupervised. The two classification methods were simulated with the same dataset called DataPlant; According to the criterion of the maximum leaf length and the maximum stem size of the plants, we obtain three classes of different plants, namely the tree, shrub and herb classes. We simulated the algorithms on the same computer. With the K-means algorithm, results are obtained after forty-five seconds, while with decision trees, results are obtained in only fifteen seconds. We tested the models with a base of harvested plants, and the results obtained are in accordance with the training results. However, for the K-Means, the total accuracy of the model is 81.16% with a margin of error of 18.83% and for the decision trees, we have an accuracy of 99.24% with a margin of error of 0.76%. We conclude that the decision trees are more appropriate in this type of classification. This table summarizes the comparison of the algorithms (Table 6):

**Table 6.** Algorithm comparison

| | Precision | Margin of error | Execution time (seconds) |
|---|---|---|---|
| K-means | 0.8116 | 0.1883 | 45 |
| Decision trees | 0.9924 | 0.76 | 15 |

## 4 Conclusion

We have shown that the classification of plant species can be done automatically. We used automatic learning, which is an area of artificial intelligence. It aims to use algorithms to reproduce all activities that are repetitive and require large amounts of data. We are based on the stem and leaf characteristics of the plant; because these characteristics are common to all plants, whatever their stage of evolution. We used K-Means algorithms and decision trees to classify plants by three types, trees, shrubs and grasses. We found that decision trees classify plants with better accuracy than any K-means algorithms. We hope to further our research by using other classification methods.

## References

1. Picouet, D., et al.: les règles de la taxonimie: nommer les espèces (2018)
2. Kaya, A., et al.: Analysis of transfer learning for deep neural network based plant classification models. Comput. Electron. Agric. (2019)
3. Saleem, G., et al.: Automated analysis of visual leaf shape features for plant classification. Comput. Electron. Agric. (2019)
4. Coulon, A.: supplément à l'intelligence artificielle, la Lettre 111, Printemps (2018)
5. Paquette, D., et al.: Clés des 16 genres de Cypéraceaes, Flora Quebeca, 19 Septembre 2016
6. Paquette, D., et al.: Clés des verges d'or, Flora Québéca (2016)
7. Sabourin, A., et al.: clé des crucifères., flora Québéca, mars 2018
8. Guillaume, R., de La Clergerie, É.V.: Analyse automatique de documents botaniques: le projet Biotim. In: Proceedings of TIA 2005: Journées Terminologie; Intelligence Artificielle, Rouen, France, April 2005
9. Boyd, R., et al.: Une base de données informatisée transdisciplinaire de la flore chez les sémé du burkina faso: un outil pour l'étude du lien nature-société (2014)
10. Pegliasco, G.: Classer une fleur selon des critères observables: Initiation au Machine Learning avec Python - La pratique
11. Piernot, T., et al.: Flora Bellissima, un nouvel outil pour découvrir la flore, mars 2014
12. Dyrmann, M., et al.: Plants species classification using deep convolutional neural network (2016). https://doi.org/10.1016/j.biosystemseng.2016.08.024
13. Zhao, Z.-Q., et al.: ApLeaf: an efficient android-based plant leaf identification system. Neurocomputing (2014)
14. Tippannavar, S., et al.: A machine learning system for recognition of vegetable plant and classification of abnormality using leaf texture analysis. Int. J. Sci. Eng. Res. **8**(6), 1558–1563 (2017)
15. Hutchinson, J., Dalziel, J.M., Keay, R.W.J., Hepper, N.: Flora of West Tropical Africa (2014)
16. Forum des Marais Atlantiques 2017, Herbier Numérique, Flore en zonz humide, Région nouvelles d'Aquitaine, 93 p. http://www.forum-zones-humides.org
17. Schoonderwoerd, K.M., et al.: Zygotic dormancy underlies prolonged seed development in Franklinia alatamaha (Theaceae): a most unusual case of reproductive phenology in angiosperms. Bot. J. Linn. Soc. **181**(1), 70–83 (2016)

18. Dundar, M., Kou, Q., Zhang, B., He, Y., Rajwa, B.: Simplicity of Kmeans versus deepness of deep learning: a case of unsupervised feature learning with limited data. In: 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), pp. 883–888 (2015)
19. Bholowalia, P., Kumar, A.: EBK-means: a clustering technique based on elbow method and k-means in WSN. Int. J. Comput. Appl. **105** (2014)
20. Beraud, P.: MSFT, 5 August 2014
21. Biernat, E., et al.: Data science: fondamentaux et études de cas, EYROLLES (2015)