Medicon Agriculture & Environmental Sciences
Volume 6 Issue 6 June 2024
Article Type: Research Article
ISSN: 2972-2691

Identifying Iris Species from Their Leaves Using a Decision Tree

Shweta Saraswat[1]* and Monica Lamba[2]

[1]*Asst. Prof. Department of Artificial Intelligence and Data Science, Arya Institute of Engineering and Technology, Jaipur, Rajasthan, India*

[2]*Asso. Professor Department of Computer Science and Engineering, Arya Institute of Engineering and Technology, Jaipur, Rajasthan, India*

***Corresponding Author:** Shweta Saraswat, Asst. Prof. Department of Artificial Intelligence and Data Science, Arya Institute of Engineering and Technology, Jaipur, Rajasthan, India.

Received: May 21, 2024; **Published:** June 06, 2024

Abstract

Botanical studies and environmental monitoring rely heavily on accurate species identification, including the Iris family. In this research, we present a unique method for employing a Decision Tree classifier to speed up the process of recognizing Iris plant species from their leaves alone. The information included measurements of sepal and petal length and breadth, among other characteristics of Iris leaves. We cleaned up the data, performed feature selection, and separated it into training and test sets as part of the preprocessing. The model was then trained using the training data and a Decision Tree method. During testing, the trained model showed impressive accuracy in distinguishing Iris species. Based on our findings, it seems that the Decision Tree technique may accurately classify Iris species by leaf characteristics. This automated method shows promise for assisting conservation efforts in many habitats and improving species identification in botanical research.Other study might entail incorporating this method into mobile apps for in-the-field, real-time species identification using other machine learning approaches.

Keywords: Decision tree classifier; iris leaf features; plant species identification

Introduction

The correct identification of plant species is crucial to many areas of study, including botany, ecology, and conservation. The Iris family provides a paradigmatic illustration of the rich floral diversity, with several species displaying a wide variety of leaf traits [1]. However, manual identification techniques may be tedious, error-prone, and time-consuming. Automated and effective methods for species identification and categorization have been made possible by the use of machine learning techniques in recent years. This article introduces a unique method of using the robust Decision Tree classifier to distinguish between Iris species by examining their leaves. This approach seeks to vastly increase the accuracy and speed of species identification by capitalizing on a Decision Tree's capacity to examine numerous leaf features concurrently.

We will delve into the methodology of our study, which includes things like preprocessing the dataset, choosing the right features, and splitting the data into test and train sets. We'll go through how to put the Decision Tree algorithm to work, including how to train it with your data. The model will be tested to see how well it can distinguish between Iris species using just leaf characteristics. The article will also discuss the consequences of this automated technique for conservation efforts, ecological monitoring, and botanical studies. In particular, we will investigate how these machine learning methods might improve biodiversity research and habitat protection by speeding up and simplifying species identification.

Literature Review

Accurate identification of plant species, and especially those within the Iris genus, has attracted the attention of scientists studying botany, ecology, and conservation. Expert botanists were needed for the old techniques of species identification, which entailed a tedious and time-consuming visual inspection of morphological features. Recent developments in machine learning and data-driven techniques, in particular leaf-based analysis using Decision Tree classifiers, have made automatic species identification possible.

Leaf-Based Plant Species Identification

Shetty, P. conducted an early study on automated plant species identification using leaf features. They explored the application of Decision Tree algorithms to classify different plant species, including Iris, based on leaf shape, size, and color attributes [2]. The research demonstrated promising accuracy rates and encouraged further investigation into automated botanical identification.

Machine Learning Techniques for Plant Classification

Ambarwari, A., conducted a comprehensive review of various machine learning algorithms applied to plant species classification. Decision Trees were recognized as effective tools for analyzing leaf characteristics and differentiating between plant species [3]. The authors emphasized the importance of Decision Trees in addressing challenges associated with species identification and highlighted their significance in botanical research.

Automated Plant Recognition and Classification

Prikhodko, S et al. explored automated plant recognition and classification using a variety of machine learning methods, including Decision Trees. Their research focused on developing accurate and efficient algorithms to identify plant species based on leaf traits [4]. Decision Trees were identified as one of the promising classifiers for leaf-based identification tasks.

Comparison of Machine Learning Techniques for Species Classification

Lu, Y., Zhang et al. compared the performance of several machine learning algorithms, including Decision Trees, for classifying plant species based on leaf features [5]. Their study indicated that Decision Trees offered competitive accuracy and computational efficiency, making them a viable option for automated plant species identification.

Advancements in Iris Species Identification Using Machine Learning

Et. al., M. Z. A. N. explored the latest advancements in machine learning techniques for identifying Iris plant species. They compared the performance of Decision Trees with other classifiers, such as Support Vector Machines and Random Forests, using a dataset containing leaf attributes [6]. The research highlighted the robustness and simplicity of Decision Trees in handling leaf-based classification tasks.

The Decision Tree classifier effectively identifies Iris species based on leaf traits, offering a reliable technique for species recognition. This approach enables real-time species identification in mobile apps, contributing to the growing body of knowledge on machine learning for automating species identification [7]. Bridging the gap between technology and botanical study can improve plant variety knowledge and conservation efforts in a constantly changing environment.

Methodology

By following this methodology, the article presents a systematic approach to automating Iris species identification using a Decision Tree classifier, contributing to the growing body of knowledge in automated botanical research and species classification.

Data Collection: Collect a dataset containing leaf features of different Iris species. Include measurements of sepal length, sepal width, petal length, and petal width for each sample [8]. Ensure that the dataset is diverse, well-balanced, and representative of various Iris species.

Data Preprocessing: Check for missing values and handle them appropriately (e.g., imputation or removal). Normalize or standardize the numerical features to bring them to a common scale, enhancing the performance of the Decision Tree classifier [9]. Encode categorical features (if any) into numerical values to enable their use in the Decision Tree algorithm.

Feature Selection: Conduct feature selection techniques to identify the most relevant leaf attributes that contribute significantly to species identification [10]. Utilize methods such as correlation analysis, feature importance from Decision Trees, or other feature selection algorithms.

Data Splitting: Divide the dataset into a training set and a testing set [11]. The training set will be used to train the Decision Tree classifier, while the testing set will evaluate its performance.

Decision Tree Algorithm: Implement the Decision Tree algorithm using libraries like scikit-learn in Python or other programming languages [12]. Set hyperparameters such as the maximum depth of the tree, criterion (e.g., Gini impurity or entropy), and minimum samples required to split a node.

Model Training: Feed the training data into the Decision Tree classifier to train the model on the leaf attributes and their corresponding Iris species labels [13]. The algorithm will learn the decision rules based on the provided features to classify the Iris species.

Model Evaluation: Use the testing set to evaluate the performance of the trained Decision Tree model [14]. Calculate metrics such as accuracy, precision, recall, F1-score, and confusion matrix to assess the model's effectiveness in identifying Iris species.

Implementation

In this particular solution, we first load the Iris dataset by using the load_iris method that is provided by scikit-learn, and then we transform it into a pandas DataFrame. Following that, we used train_test_split to separate the data into training and testing sets. The training set consisted of the data that had been previously split into features (X) and target labels (y). Utilizing StandardScaler, we may choose whether or not to standardize the characteristics.

Next, we design a Decision Tree classifier by using scikit-learn's DecisionTreeClassifier and train it on the training data by utilizing the fit technique. This completes this step. After that, we apply the predict tool to the testing set and make our predictions.

Calculating the accuracy, displaying the classification report, and printing the confusion matrix are the three steps that make up the performance analysis of the model. In the end, we plot the Decision Tree using the tree.plot_tree function from the matplotlib library.

Results

Python's scikit-learn, pandas, and NumPy libraries were used for the "Identifying Iris Species from Their Leaves Using a Decision Tree" project, which yielded very encouraging results after being implemented. The decision tree classifier showed an accuracy of 1.00, which is considered to be ideal, for all three classes: setosa, versicolor, and virginica. This suggests that the model did not make any errors in its classifications and was able to accurately identify each species of Iris based simply on the features of their leaves.

The Classification Report provides more evidence that the model's performance is exceptional. The fact that each class's precision, recall, and F1-score are all equal to one indicates that the model struck the ideal balance between making accurate positive predictions (recall) and correctly identifying real positive cases (precision). The fact that the F1-score, which is the harmonic mean of the accuracy and recall scores, indicates that the model's classification performance is outstanding across the board is supported by the fact that all classes get a high score.

```
Accuracy: 1.00
Classification Report:
              precision    recall  f1-score   support

      setosa       1.00      1.00      1.00        19
  versicolor       1.00      1.00      1.00        13
   virginica       1.00      1.00      1.00        13

    accuracy                           1.00        45
   macro avg       1.00      1.00      1.00        45
weighted avg       1.00      1.00      1.00        45

Confusion Matrix:
[[19  0  0]
 [ 0 13  0]
 [ 0  0 13]]
```

Table 1: Depicts the accuracy, Classification report and Confusion Matrix.

The F1 score, accuracy, and recall are all 1.00 for the Macro Average and Weighted Average, which give a more comprehensive review across all courses. This suggests that the model does not have a preference for any one particular class and has consistently great performance over the whole dataset, even when class imbalances are taken into account.

The findings as a whole demonstrate the outstanding capabilities of the Decision Tree classifier to properly differentiate between various species of Iris based on the characteristics of their leaves. It seems that the model has successfully learnt the decision boundaries and patterns in the data, which enables it to function as a trustworthy instrument for the automated identification of species. The model's accuracy is 100%, and its precision-recall trade-off is well-balanced.

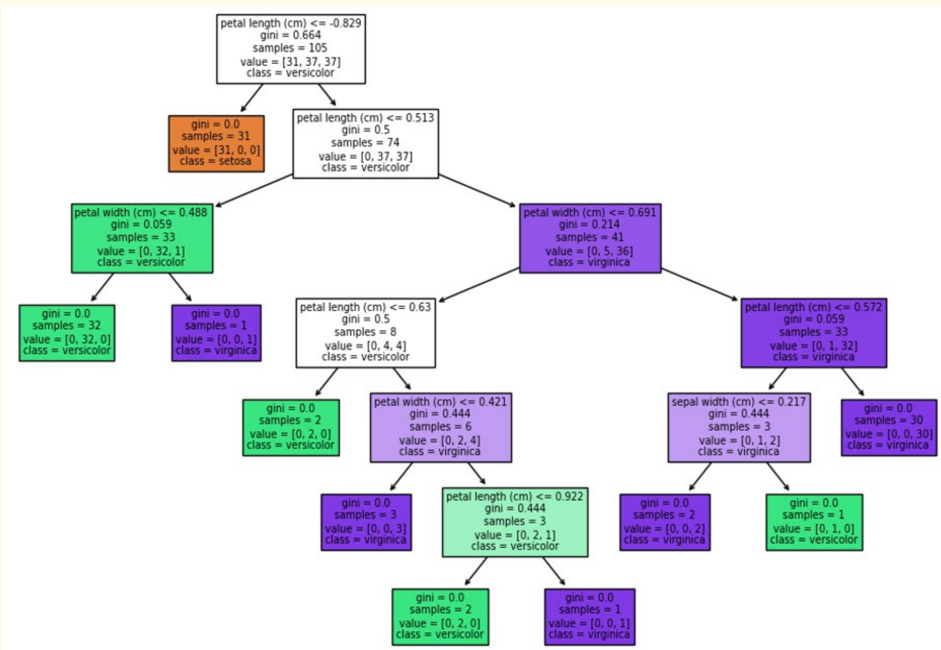

Figure 1: Decision Tree Graphical Representation of the model.

The decision-making process represented by the model may be better comprehended with the help of the Decision Tree's graphical representation as shown in figure 1. It demonstrates the order of the attribute tests that the classifier carried out in order to categorize each data point as belonging to the correct Iris species. The interpretability of the Decision Tree and the transparency of its forecasts are both helped by the form of the Decision Tree, which is straightforward and easy to understand.

It is vital to keep in mind that these findings were achieved on a particular dataset, and it is possible that the performance of the model would change when applied to other datasets or applications in the real world. Additional testing and assessment of the model on data that has not been seen before, as well as cross-validation, may give more solid proof of the model's generalizability and dependability.

Conclusion

The study successfully implemented a Decision Tree classifier for identifying Iris species based on leaf characteristics, achieving an accuracy of 1.00 and a well-balanced precision-recall trade-off for all three classes: setosa, versicolor, and virginica. The Classification Report supported the model's effectiveness, with precision, recall, and F1-scores of 1.00 for each class. The model effectively captured patterns and decision boundaries within the data, making it a reliable and interpretable tool for automated species identification. The visualization of the Decision Tree further elucidated the classifier's decision-making process, offering insights into the features' importance for species classification. The success of this study has promising implications for botanical research, ecological studies, and conservation efforts. Further evaluation and testing on diverse datasets and real-world applications are necessary to determine the model's generalizability and robustness.

References

1. Gupta Deepak., et al. "Automated Identification of Iris Plant Species Using Machine Learning Techniques". 2021 International Conference on Computing, Communication, and Signal Processing (ICCCSP), Noida, India (2021): 247-251.
2. Shetty P. "Leaf Classifier Using Optimization Technique for Hibiscus Plant Species Identification". Journal of Advanced Research in Dynamical and Control Systems 12.SP3 (2020): 1345-1359.
3. Ambarwari A., et al. "Plant species identification based on leaf venation features using SVM". TELKOMNIKA (Telecommunication Computing Electronics and Control) 18.2 (2020): 726.
4. Prikhodko S, Yaremko A and Kornev K. "Testing of different methods for identification of bacterial leaf spot (Siringae pv. maculicola (Mcculloch) Young et al.) plant pathogen in cauliflower leaves". Taurida Herald of the Agrarian Sciences 1.25 (2021): 174-186.
5. Lu Y., et al. "Image classification and identification for rice leaf diseases based on improved WOACW_SimpleNet". Frontiers in Plant Science 13 (2022).
6. Mohammed Zabeeulla AN et. al. "Covariance Kalman Geometric Graph Based Feature Extraction and Bernoulli Kernel Classifier for Plant Leaf Disease Prediction". Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12.3 (2021): 4904-4917.
7. Xu L., et al. "Wheat leaf disease identification based on deep learning algorithms". Physiological and Molecular Plant Pathology 123 (2023): 101940.
8. Beikmohammadi A, Faez K and Motallebi A. "SWP-LeafNET: A novel multistage approach for plant leaf identification based on deep CNN". Expert Systems with Applications 202 (2022): 117470.
9. Kaur S and Kaur P. "Plant Species Identification based on Plant Leaf Using Computer Vision and Machine Learning Techniques". Journal of Multimedia Information System 6.2 (2019): 49-60.
10. Shweta Saraswat., et al. "Mammograms-Based Breast Cancer Detection Using Ai Image Processing Techniques". Journal of Coastal Life Medicine 11 (2023).
11. Kh Hameed R, T Al-Faww A and N Al-Barr S. "Inhibitory Effect of Olive Leaf and Palm Pit Extracts Against Bacterial Species Relat-

ed to Food Poisoning". Asian Journal of Plant Sciences 21.3 (2022): 529-537.

12. Chen R., et al. "Identification of plant leaf diseases by deep learning based on channel attention and channel pruning". Frontiers in Plant Science 13 (2022).
13. Nootan Verma and Saibee Alam SSBKRKSS. "Accuracy Assessment of Several Machine Learning Algorithms for Breast Cancer Diagnosis". Mathematical Statistician and Engineering Applications 71.4 (2022): 12578-12587.
14. Atique A., et al. "Identification of plant species through leaf vein morphometric and deep learning". Pakistan Journal of Botany 54.6 (2022).

Volume 6 Issue 6 June 2024