BioGecko

Vol 12 Issue 03 2023

ISSN NO: 2230-5807

A Comparative Analysis of Machine Learning Based Algorithms for Iris Flower Classification Application

Hemlata Ohal[1], Payal Dhindale[2],Eshwari Bhandkar[3],Tanusri Bokka[4],Shravani Shinde[5]

[1]Computer Science and Engineering, DVK MIT World Peace University, India,hemaohal@gmail.com,

[2]Diploma in Computer Engineering, MIT Polytechnic, Pune, payaldhindale@gmail.com

[3]Diploma in Computer Engineering, MIT Polytechnic, Pune,eshwaribhandkar@gmail.com

[4]Diploma in Computer Engineering, MIT Polytechnic, Pune,tanusribokka@gmail.com,

Abstract: In order to identify IRIS flower species, we are using machine learning to extract knowledge from data in a semi-automated fashion. The response is categorical in supervised learning classification, meaning that its values are confined to a finite unordered set. The classification task has been streamlined using Scikit-Learn tools. This work's main objective is to categorize

IRIS flowers using sci-kit methods and machine learning. The question at hand is how to determine the species of IRIS flowers based on measurements of the blooms' characteristics. Patterns must be identified by analysing the petal and sepal sizes of the IRIS flower and how predictions are made by deriving the pattern from the IRIS flower class in order to categorize the IRIS data set. To train the machine learning model, we utilize data.

Keywords: *Semi-automated, Supervised, Scikit-Learn tools, Scikit methods.*

1 Introduction

It is becoming more typical to classify iris blossoms using machine learning. Sepal width, sepal length and petal width, petal length are the four characteristics of iris flower. These form the special features to iris flower. These features identify iris flower in to three classes in the iris dataset: Virginica, Setosa, and Versicolor. We will request user input on things like petal and sepal length and breadth. The system will analyse the user-provided variables and forecast the kind of Iris flower based on the inputs.

In this system, we employed SVM, Logistic Regression, and Decision Tree techniques to categorize the Iris blooms, and we based our classification on the accuracy that these algorithms provided. SVM demonstrated the maximum accuracy, at 99%. The algorithm has then undergone Bagging, Boosting, and Voting.

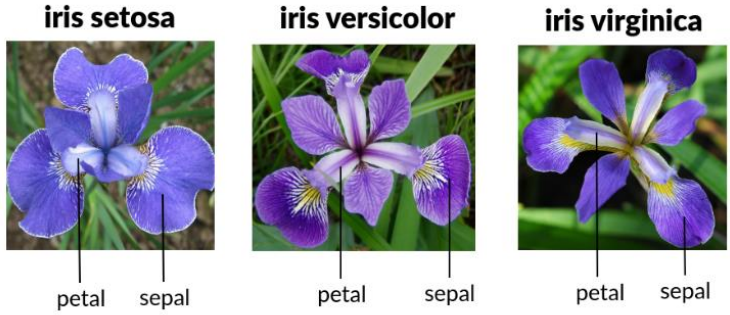

Fig 1 Iris flower Species

Learning to anticipate something or glean information from data is what machine learning is all about. Artificial intelligence includes machine learning. ML algorithms create a model using training data, sometimes referred to as sample data, and then use that model to make predictions about fresh data.
Three categories can be used to categorise machine learning:

Supervised machine learning: This form of machine learning is learned using training data that has been accurately labelled. Labeled data refers to training data that has already been associated with the desired result.

Unsupervised machine learning: Unsupervised machine learning, does not utilise labelled data. Using untagged data, it discovered patterns. In essence, it generates a set of objects depending on the supplied traits and data.

Semi supervised machine learning: It is in the middle between supervised and unsupervised learning. It consists primarily of untagged data with a tiny quantity of tagged data.

So, in this system, we are utilising supervised machine learning method based on the preceding information.

2. Literature Survey

Kholerdi, Hedyeh A., Nima TaheriNejad, and Axel Jantsch used neural network approaches, and they generated 17% accuracy. (Hedyeh A. et al *2018)*
K Nearest and Logistic regression were utilised by Rao, T Srinivas, and etla. (Rao et al *2021)*
Using Random Forest in AWS, Pachipala and Yellamma categorised the iris blossoms according to species. (Pachipala et al *2022*)
Rivero, Daniel, et al. used genetic programming (GP) to resolve a classification problem from a database and demonstrated how we may change this tool in two different ways: to improve performance and to make mistake detection possible. (Daniel, et al. *2003*)
In their study, Mithy, S. A., et al. showed how to handle the classification problem using algorithms including SVM, Logistic Regression, KNN, Random Forest decision, K-means clustering and K-medoids. Additionally, they worked on four advanced features for the scikit implementation tool. Scientists then applied classification and regression algorithms to the iris dataset after recognising and studying the patterns. (Mithy, et al. *2022*)

Shivam and Vatshayan focused on categorising IRIS flowers using scikit and machine learning methods. In this work, the Iris Dataset was utilised to train the machine learning model, and performance and accuracy were assessed using supervised learning techniques. (Shivam et al *2019*).
In order to categorise the floral design, Pawar, Lokesh, and associates focused on developing a distinctive classification approach to identify the plant's iris. It is recommended that the pattern be recognised and classified using an ideal ensemble model. An ensemble model is advised to improve performance over the initial Decision Tree, OneR, Adaboost, Random Forest, and Bayesnet models. (Lokesh et al *2022*)
In the approach proposed by Gupta, Tina, and colleagues, three classification models— support vector machine (SVM), logistic regression, and K-nearest neighbours (KNN)—are used to assess and pre-process the data using exploratory data analysis (EDA). All recommended models showed maximum accuracy of 96.43, 98.21, and 94.64 percent, respectively, when tested on the Iris dataset. (Gupta, et al *2022*)
Singh, Anshuman, and Rohan Akash developed a machine learning model for flower extraction utilising the IRIS database and the flask web framework. There are 50 samples Iris virginica, Iris versicolor and Iris setosa, included in the data set. Every sample's sepal and petal length and width, measured in inches, were recorded. (Anshuman et al *2022)*
The goal of Prathima, Pala, and T. Ranjith Kumar was to increase public knowledge about the creation of a machine learning model. We make use of an iris dataset made accessible via Scikit tools. An problem with supervised learning classification has been identified for this dataset. The dataset is then split into training and test sets, and processed using the Scikit tool. Using new samples, a model is developed and evaluated using the K-Nearest Neighbours method. (Pala et al *2021)*
Pavel, Jirava, and Kupka Ji are examples of a recent trend in problem-solving that is built on the use of different conventional methodologies and techniques. A categorisation model employs two computational intelligence components. It refers to the rough and hazy sets that serve as the foundation for the hybrid model for data categorisation. Even using data with uncertainty is permitted. This model is implemented in MATLAB, tested on more data files, and contrasted with other, well-established classification techniques. (Jirava et al *2007*)

3. Proposed Methodology

Dataset:
There are several Iris Flower datasets available. The Setosa, Versicolor, and Virginca are three different varieties of iris blooms that are covered in a total of 150 datasets. The gathered Iris Datasets are used to build the model for machine learning. Scikit-learn includes a few common datasets, including the order-specific Iris dataset. The load irirs work is imported from Scikit-Learn.
PRE-PROCESSING: Data preprocessing is the process of altering or encoding the data so that the computer can quickly parse it. Alternatively, the algorithm can now easily comprehend the data's characteristics.
Label encoding:
Labels must be encoded as integers in order to be machine-readable. Labeling is the procedure involved in this. The employment of those labels by machine learning algorithms might subsequently be decided upon more effectively. It is an essential step for the structured dataset in supervised learning.
Normalization: You may need to preprocess in machine learning models. Normalization is a preprocessing technique. The values are rescaled into a [0, 1] range during normalisation. In situations when all parameters must have the same positive scale, this may thus be helpful.

Xchanged=(X−Xmin)/(Xmax−Xmin). Hence, "Normalization" in the business sector often refers to "normalising the range of numbers to be from 0.0 to 1.0."

As a result of standardisation, data are rescaled. This standardisation have a standard deviation () of 1 and mean () of 0.Xchanged=(X−μ)/σ

Our system architecture is depicted in the system Architecture figure 2.

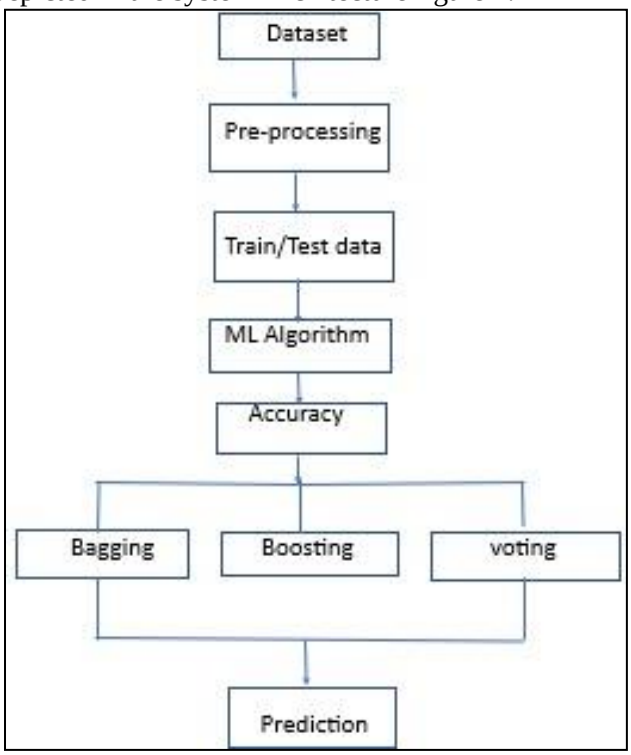

Fig 2. System Architecture

In contrast, "Standardization" refers to the process of standardising a range of data in order to calculate how far off from the mean each number is. In addition, it has DATA VISUALIZATION.

Model Building

We will separate our data into training and testing sets. The model gets trained from this data so as to subsequently generalise it to additional data. As the training set has a known outcome. We have the test dataset to see how well our model predicts on this subset.

ML ALGORITH:

In order to decide which strategy is best for our model, we will train our model using some of the algorithms.This is a list of commonly used algorithms: DecisionTreeClassifier, SVC, and Logistic Regression

ACCURACY: By refining our model, we were able to achieve 99% SVM accuracy.

(ALGO)BAGGING: It just compiles all the estimates from various estimators to provide final estimates. It use a variety of sampling methods, including pasting, bagging/bootstrap aggregation, random subspaces, and random patches, to train several estimators. This classifier is typically applied to decision trees and other high variance classifiers.

BOOSTING, or gradient classifier, is a sequential model that combines several different decision trees, with the results of one decision tree being used to train the next, and so on.
Voting classifiers are a particular kind of machine learning estimator that build a number of base models or estimators and then make predictions based on average their output. The aggregating criteria can be used with voting for each estimator output.
PREDICTIONS: Species belonging to their respective groups are predicted by using user inputs like length and width of sepal and petal length and width.

4. Results

The following Table 1 shows the accuracy achieved by every algorithm.

Table 1: Accuracy by three algorithms

Sr. No.	Algorithm	Accuracy
1.	Decision Tree	1.0%
2.	Logistic Regression	98%
3.	SVM	99%

The below graph in figure 3 and 4 shows the comparison of sepal length and sepal width among the Iris Setosa, Iris Verginica and Iris Versicolor.

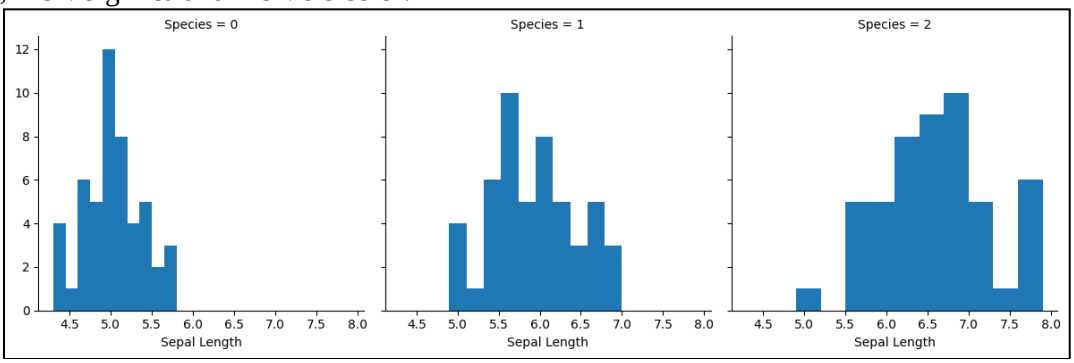

Fig 3. Comparison between Sepal length of the Species

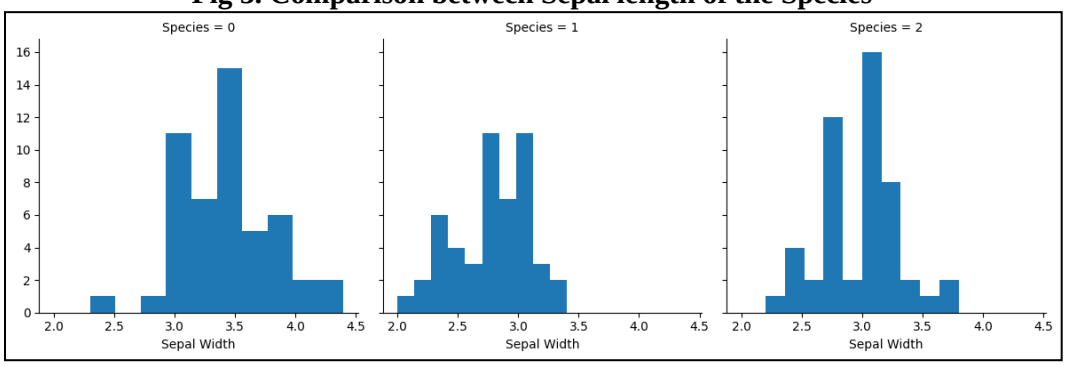

Fig 4. Comparison between Sepal width of the Species

The below graph in figure 5 and 6shows the comparison of petal length and petal width among the three classes of Iris flower.

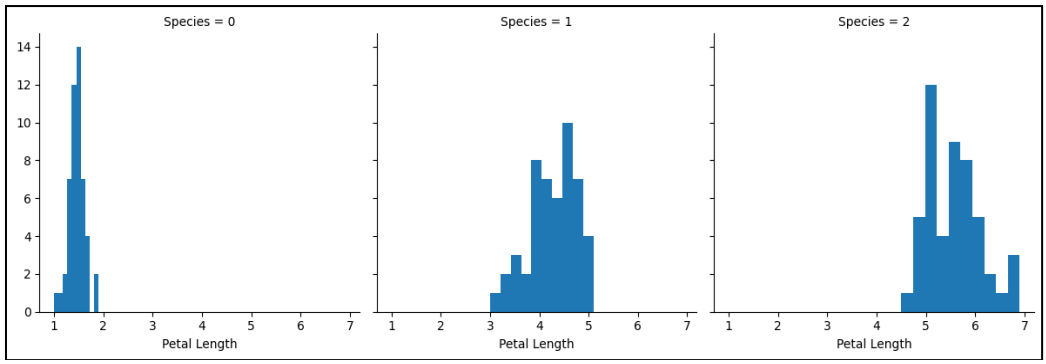

Fig 5. Comparison between Petal length of the Species

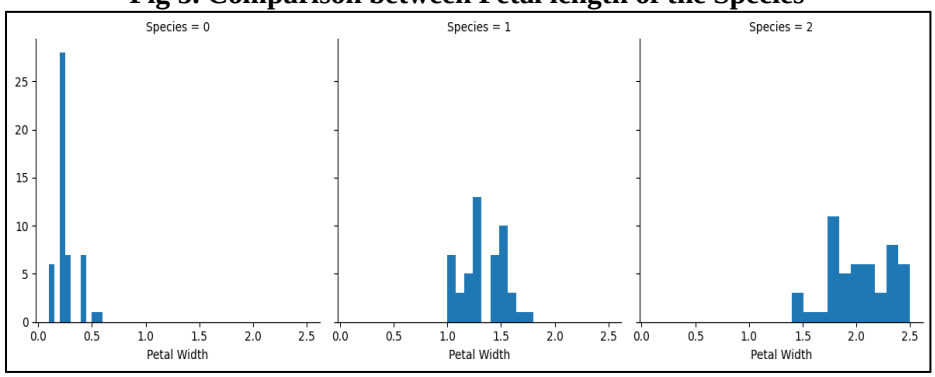

Fig 6. Comparison between Petal Width of the Species

Comparison

Three common machine learning classification methods are decision trees, logistic regression, and support vector machines (SVM).The three algorithms are contrasted in the following discussion:

1. SVM: The hyperplane that optimises the margin between the two classes is found using SVM, a binary classification technique.

   It is a powerful method that handles both linear and non-linear datasets with ease.SVM is efficient at managing high-dimensional data and small to medium-sized datasets.

   In our work, SVM's accuracy is 99%.

2. Logistic Regression: A function called as logistic function is used to describe the likelihood of a binary result in the binary classification process k. It is a simple, efficient method that works well for datasets that can be linearly separated. Logistic regression also makes it easy to understand how the model generates predictions since it is interpretable.

Logistic regression has a 97% accuracy rate.

3. Decision Trees: Decision trees are a categorization technique that uses a tree-like representation of alternatives and their results, can handle categorical and numerical data, is understandable, and can handle missing values in the dataset.1.0% is the accuracy of logistic regression.

SVM is, in short, the most effective method for this assignment.

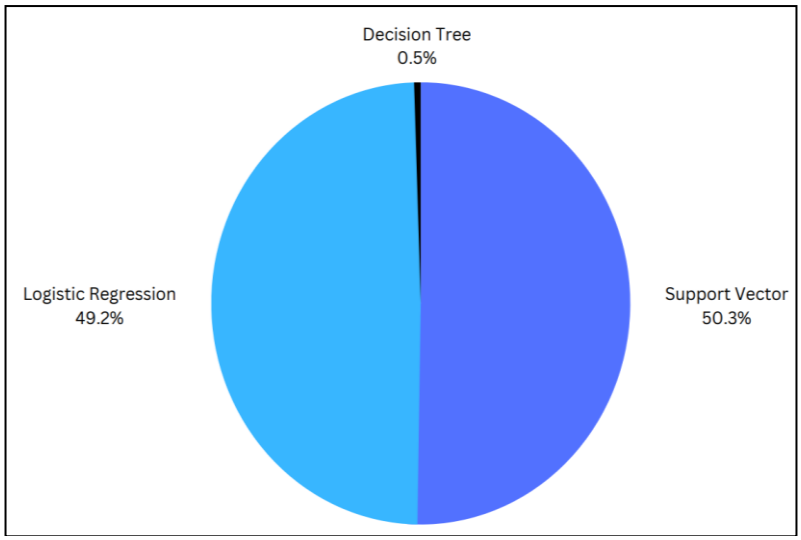

Fig 7. Comparison between algorithms

Future Scope

As there hasn't been any study on the Iris Flower Disease dataset to far, there hasn't been any actual implementation of the Iris Flower Disease. In light of this, the works's advancement may be achieved by identifying numerous Iris flower illnesses, as classifying the flower is the first stage in the process of identifying diseases.

Same work can be extended for leaf disease detection as well.

Conclusion

Flower classification is a vital, simple, and fundamental exercise for every student studying machine learning. Any machine learning student should extensively research the iris blossoms dataset. In this work, we utilised Logistic Regression whose accuracy came to be 0.97, Support Vector Machine's accuracy is achieved 0.99, and Decision Tree's accuracy is 0.1, three machine learning approaches that may be used to categorise data.

In this work, we have shared the knowledge of how to build our own supervised machine learning model of Iris Flower Classification. Through this work, we also shared the knowledge about machine learning, data analysis, data visualization, model construction, etc.

References

Anshuman, and Rohan Akash, Singh. "Flower Classifier Web App Using Ml & Flask Web Framework." *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. IEEE, 2022.

Daniel, Rivero, et al. "Extracting knowledge from databases with Genetic Programming: Iris flower classification problem." *Proceedings of CIMCA and IAWTIC* (2003).

Gupta, Tina, et al. "Classification of Flower Dataset using Machine Learning Models." *2022 4th International Conference on Artificial Intelligence and Speech Technology (AIST)*. IEEE, 2022.

Hedyeh A. Kholerdi, Nima TaheriNejad, and Axel Jantsch. "Enhancement of classification of small data sets using self-awareness—An iris flower case-study." *2018 IEEE international symposium on circuits and systems (ISCAS)*. IEEE, 2018.
Jirava, Křupka Jiří and Pala. "Classification model based on rough and fuzzy sets theory." *6th WSEAS international conference on Computational intelligence, man-machine systems and cybernetics*. 2007.
Lokesh, Pawar, et al. "Optimised ensembled machine learning model for iris plant classification." *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2022.
Mithy, S. A., et al. "Classification of Iris Flower Dataset using Different Algorithms." *Int. J. Sci. Res. in Mathematical and Statistical Sciences Vol* 9.6 (2022).Organisation
Pala, Prathima, and T. Ranjith Kumar. "A Model Identifying Iris Species using Machine Learning." (2021).
Pachipala, Yellamma, et al. "Iris Flower Classification by using Random Forest in AWS." *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE, 2022.
Rao, T. Srinivas, et al. "Iris Flower Classification Using Machine Learning." *Network* 9.6 (2021).
Shivam, Vatshayan. "Performance evaluation of supervised learning for iris flower species." (2019).