International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

IRIS SPECEIS CLASSIFIER

K.Venkata sai
B.Tech AI&ML
Malla Reddy University
Hyderabad

A.Shiva kumari
B.Tech AI&ML
Malla Reddy University
Hyderabad

Challa Nanda kishor
B.Tech-AI&ML
Malla Reddy University
Hyderabad

D.Venkatesh
B.Tech-AI&ML
Malla Reddy University
Hyderabad

K.Venkatesham
B.Tech –AI&ML
Malla Reddy University
Hyderabad

A.Venkata sai
B.Tech-AIM&ML
Malla ReddyUniversity
Hyderabad

Dr.Ravinder
Professor
Artificial Intellignce and Machine Learning
Malla ReddyUniversity
Hyderabad

Abstract : This Python project aims to classify the species of Iris flowers based on the measurements of petal and sepal lengths and breadths using machine learning techniques. The Iris dataset, which is commonly used in machine learning applications, will be utilized for this project. The project will involve data preprocessing, feature engineering, and model building using various classification algorithms such as logistic regression, decision trees, random forests, and support vector machines. The performance of these models will be evaluated using accuracy, precision, recall, F1-score, and confusion matrix. The outcome of this project will be a Python application that can predict the species of an Iris flower given its petal and sepal measurements with high accuracy. This application can be used in various fields, such as agriculture and botany, to quickly identify the species of Iris flowers without the need for manual identification.

Keywords:-iris-species-classifier,irisdataset,python.

I. INTRODUCTION

The Iris dataset is a classic dataset in machine learning and statistics that is often used as a benchmark for classification algorithms. It was first introduced by the British statistician and biologist Ronald Fisher in 1936. The dataset consists of 150 observations of iris flowers from three different species: Iris setosa, Iris versicolor, and Iris virginica. For each observation, there are four measurements: sepal length, sepal width, petal length, and petal width, all measured in centimeters. The goal of the dataset is to use these four features to classify the flowers into one of the three species. The dataset is often used for supervised learning, where the goal is to train a machine learning model to predict the species of a new observation based on its four features. The Iris dataset is a well-known and popular dataset in machine learning, and it has been used in many different applications, from classification algorithms to exploratory data analysis and visualization. It is often used as a benchmark for new machine learning algorithms and techniques, and it has become a standard reference.

II. MACHINE LEARNING ALGORITHM

In this iris project we used used random forest machine learning algorithm.

The Random Forest classification algorithm is a popular machine learning algorithm used for classification tasks such as the classification of iris flowers. It is an ensemble learning method that

constructs a large number of decision trees at training time and outputs the class that is the mode of the classes predicted by individual trees.

Here's how the Random Forest algorithm works:

1. First, the algorithm randomly selects a subset of features from the dataset.
2. It then constructs a decision tree using the selected subset of features and a subset of the training data.
3. The algorithm repeats steps 1 and 2 to construct multiple decision trees, each using a different random subset of features and training data.
4. Once all decision trees are constructed, the algorithm combines their predictions by selecting the class that appears most frequently among the individual trees.

In the case of iris flower classification, the Random Forest algorithm can be trained on the four measurements of the flowers (sepal length, sepal width, petal length, and petal width) to predict the species of the iris flower. By constructing multiple decision trees, the algorithm can learn complex relationships between the features and the target variable and achieve high accuracy in classifying the flowers.

III. REQUIRED TOOLS

- MS Word
- Jupyter Notebook
- Python3
- Data set
- Machine Learning Algorithm
- Pickle

IV. DATA ANALYSIS

The Iris dataset is a classic dataset in machine learning and statistics that is often used as a benchmark for classification algorithms. It was first introduced by the British statistician and biologist Ronald Fisher in 1936.

The dataset consists of 150 observations of iris flowers from three different species: Iris setosa, Iris versicolor, and Iris virginica. For each observation, there are four measurements: sepal length, sepal width, petal length, and petal width, all measured in centimeters.

The goal of the dataset is to use these four features to classify the flowers into one of the three species. The dataset is often used for supervised learning, where the goal is to train a machine learning model to predict the species of a new observation based on its four features.

The Iris dataset is a well-known and popular dataset in machine learning, and it has been used in many different applications, from classification algorithms to exploratory data analysis and visualization. It is often used as a benchmark for new machine learning algorithms and techniques, and it has become a standard reference dataset in the field.

The Iris dataset is often distributed in CSV (Comma Separated Values) file format, which is a plain text file that contains data in a tabular form. Each row in the CSV file represents a single observation of an iris flower, while each column contains a specific feature of that observation.

The CSV file contains a header row that specifies the name of each column, and subsequent rows contain the actual data. The four columns in the CSV file correspond to the four measurements of the iris flowers: sepal length, sepal width, petal length, and petal width. The last column in the file corresponds to the species of the iris flower, which can take one of three values: "Iris-setosa", "Irisversicolor", or "Iris-virginica".

TABLE 1 CSV file

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...	...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

V.METHODOLOGY

The iris classifier uses the Random Forest classification algorithm to train a model to classify iris flowers based on their sepal length, sepal width, petal length, and petal width measurements.

Here's a brief summary of the steps involved in the training of data:

1. Load the Iris dataset from a CSV file using the pandas library.
2. Split the dataset into independent variables (X) and dependent variables (y), where X consists of the four measurement columns, and y consists of the class labels.
3. Split the data into training and testing sets using the `train_test_split` function from the `sklearn.model_selection` module.
4. Apply feature scaling to standardize the independent variables using the `StandardScaler` function from the `sklearn.preprocessing` module.
5. Instantiate a `RandomForestClassifier` object to create a decision tree-based model.
6. Fit the model to the training data using the `fit` method of the `RandomForestClassifier` object.
7. Save the trained model as a pickle file using the `pickle.dump` function from the `pickle` module.

The Random Forest algorithm used in this code is an ensemble learning algorithm that trains multiple decision trees on random subsets of the training data and combines their predictions to make a final classification. The algorithm is often used for classification and regression tasks in machine learning because it tends to perform well on a wide range of datasets and is resistant to overfitting.

The web application built using the Flask framework in Python. The application loads a trained machine learning model from a pickle file and uses it to predict the species of an iris flower based on user input.

Here's a brief summary of the steps involved in the loading and predicting the trained data: 1. Import the necessary libraries, including Flask, NumPy, and pickle.

2. Load the trained model from the pickle file using the `pickle.load` function and store it in a variable named `model`.
3. Create a Flask application using the `Flask` class from the `flask` module.
4. Define two routes for the application: a root route that renders an HTML template, and a predict route that receives user input and returns a prediction.
5. In the predict route, extract the user input from the request using the `request.form.values()` method, convert it to a list of floats, and store it in a variable named `float_features`.
6. Convert the `float_features` list to a NumPy array and store it in a variable named `features`. 7. Use the `model.predict` method to make a prediction based on the `features` array and store the result in a variable named `prediction`.
8. Render the same HTML template as in the root route but with an additional variable named `prediction_text` that displays the predicted species of the iris flower.
9. Run the Flask application using the `run` method of the `Flask` object.

Overall, this web application allows users to input measurements of an iris flower and get a prediction of the flower species using a pretrained Random Forest classification model.

VI.MERITS OF PROPOSED SYSTEM

1.Improved model performance

While the Random Forest classifier used in this project performed well, there is always room for improvement. We can explore other machine learning algorithms, such as Support Vector Machines (SVMs) or Gradient Boosting, to see if they can produce better results.

2.Additional features

The iris dataset used in this project only contains information on the sepal and petal lengths and widths. We can explore adding other features, such as flower color or texture, to improve the model's accuracy.

3.Real-time data

Currently, the model only makes predictions based on user input through a Flask web application. We can explore integrating the model with real-time data sources, such as sensors, to make predictions on the fly.

4. Deployment to a cloud platform The current deployment of the model is on a local machine. We can explore deploying the model to a cloud platform, such as Amazon Web Services or Google Cloud Platform, to make it accessible to a wider audience.

5.Mobile application

We can explore creating a mobile application that uses the trained model to make predictions on the go. This would allow users to take pictures of Iris flowers and get instant predictions on the species.

VII.ARCHITECTURE

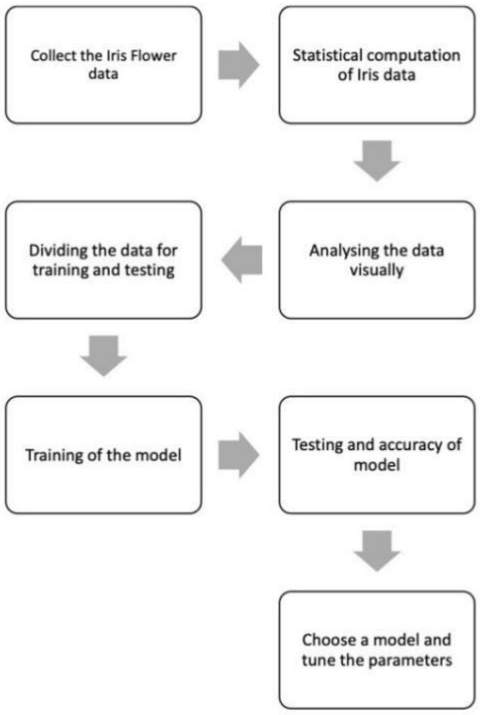

Figure 1 : architecture

VIII. EXPERIMENTAL RESULTS

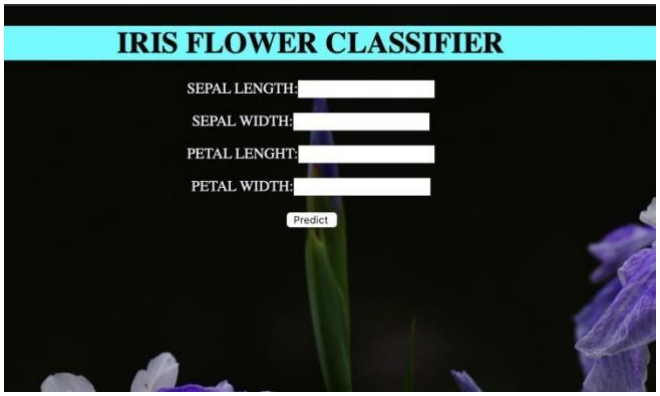

Figure 2 :interface of taking input from user

Figure 3 : flower prediction

IX. CONCLUSION

In summary, , the Python project on iris flower classification has successfully demonstrated the use of machine learning algorithms for predicting the species of Iris flowers based on their sepal and petal lengths and widths. The project involved data preprocessing, exploratory data analysis, feature scaling, and model training using a Random Forest classifier. The trained model was then used to predict the species of an Iris flower based on user input through a Flask web application.

Overall, this project serves as a great introduction to machine learning and demonstrates the practical applications of these techniques in real-world scenarios.

We sincerely thank all faculty members and non- . teaching staff of informatics department for . cooperation. Special thanks to our colleagues and friends for providing us with useful information Comments, suggestions.

REFERENCES

• International Journal on Soft Computing (IJSC) Vol.3, No.1, February 2012

•http://lab.fs.unilj.si/lasin/wp/IMIT_files/neural/doc/seminar8.pdf

• https://archive.ics.uci.edu/ml/datasets/iris

•https://www.neuraldesigner.com/learning/examples

ACKNOWLEDGEMENT

It gives us great pleasure and a moment of immense satisfaction to express our deep gratitude to our respected Prof.Ravinder and our respected mentor Prof.P.Bhavani whose constant encouragement made it possible
to work with enthusiasm. Their eternal motivation, patience and excellent expertise in discussion during the progress of the project the works have benefited us to an extent that cannot be expressed and the breadth of knowledge of the field of Computer Engineering allowed us to realize that theoretical knowledge always it helps develop effective operating software that is a blend of all the key subjects of the field. We are
we are greatly indebted to them for their invaluable guidance and ever-ready support for successful completion
of this project over time. The work under their guidance was fruitful and unforgettable.
We express our gratitude to Dr. Thayyaba
Khatoon, Head of CSE (Artificial Intelligence and Machine Learning), for her encouragement and for providing an excellent academic environment for work in the department for providing adequate background.