PREDICTION OF IRIS FLOWER CLASSIFICATION

A MINI PROJECT REPORT

Submitted by

RIYAS MOHAMMED M -810020205069

SATHISH CAKARAVARTHY C-810020205077

SHEIK ABDULLAH A -810020205078

DEPARTMENT OF INFORMATION TECHNOLOGY

ANNA UNIVERSITY, THIRUCHIRAPPALLI - 620024

MAY 2023

ANNA UNIVERSITY : CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report **" PREDCTION OF IRIS FLOWER CLASSIFICATION"** is the bonafide work of **" RIYAS MOHAMMED M (810020205069) , SATHISH CHAKARAVARTHY C (810020205077) , SHEIK ABDULAH A (810020205088) "** who carried out the project work under my supervision.

SIGNATURE OF THE HOD	**SIGNATURE OF THE SUPERVISOR**
Dr.G.ANNAPOORANI	**Mrs.M.REVATHI**
HEAD OF THE DEPARTMENT	**SUPERVISOR**
Assistant Professor (SI.Gr)	Teaching Fellow
Department of IT	Department of CSE
University College Of Engineering	University College Of Engineering
BIT-Campus	BIT-Campus
Tiruchirapalli-620 024 .	Tiruchirappalli-620 024.

TABLES OF CONTENT

ABSTRACT

The Iris flower dataset, first introduced by Ronald Fisher in 1936, has become a popular benchmark for machine learning classification tasks. In this study, we employ various machine learning algorithms to predict the species of Iris flowers based on their petal and sepal measurements. Our objective is to develop an accurate classification model that can automatically identify the species of Iris flowers.

We preprocess the dataset by performing feature scaling and splitting it into training and testing sets. We then explore multiple machine learning algorithms, including decision trees, support vector machines (SVM), k-nearest neighbors (KNN), and random forests, to build predictive models. We apply cross-validation techniques to evaluate and fine-tune the models' hyperparameters, ensuring optimal performance.

To assess the models' performance, we employ commonly used evaluation metrics such as accuracy, precision, recall, and F1-score. Additionally, we visualize the decision boundaries to gain insights into the classification process. We compare the results of different algorithms to identify the most effective approach for Iris flower classification.

Keywords:

Iris Flower, Machine Learning, Classification, Prediction.

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

The Iris flower prediction refers to the task of classifying iris flowers into different species based on their various features. The iris flower dataset is a well-known and commonly used dataset in machine learning and data analysis. It consists of measurements of four different attributes of iris flowers: sepal length, sepal width, petal length, and petal width. Based on these measurements, the goal is to predict the species of the iris flower.

The iris flower dataset is often used as a beginner's dataset in machine learning due to its simplicity and availability. It was introduced by the British statistician and biologist Ronald Fisher in 1936 and has since become a benchmark dataset for classification algorithms.

The dataset contains three different species of iris flowers: setosa, versicolor, and virginica. Each species has distinct characteristics that can be captured by the measurements of the four attributes. By training a machine learning model on the iris flower dataset, it becomes possible to predict the species of an iris flower based on its measurements with a certain level of accuracy.

The iris flower prediction problem serves as a foundation for understanding and implementing various classification algorithms, such as logistic regression, decision trees, support vector machines, and neural networks. It is widely used to demonstrate concepts like feature selection, model evaluation, and hyperparameter tuning.

Iris flower prediction has become a classic and widely studied problem in the field of machine learning due to the simplicity of the dataset and the clear distinctions between the different iris species.

1.2 OBJECTIVES

The objectives of iris flower prediction typically involve using machine learning techniques to accurately classify iris flowers into their respective species based on their measurements. Here are the main objectives of iris flower prediction:

- **Classification**: The primary objective is to accurately classify iris flowers into their specific species. The iris dataset contains three species: Setosa, Versicolor, and Virginica. By analyzing the flower's measurements, such as sepal length, sepal width, petal length, and petal width, the model aims to predict the correct species for a given iris flower.

- **Accuracy and Precision**: The model strives to achieve high accuracy and precision in predicting the iris species. Accuracy refers to the proportion of correctly classified iris flowers, while precision measures the model's ability to correctly identify each species. A good prediction model should aim for high accuracy and precision to ensure reliable and trustworthy results

- **Generalization**: The objective is to develop a model that can generalize well to unseen iris flowers. It should not only perform well on the training dataset but also be able to make accurate predictions on new, previously unseen samples. Generalization ensures that the model is robust and can be applied to real-world scenarios.

- **Feature Importance**: Another objective is to determine the importance of each feature in predicting the iris species. By analyzing the model's behavior and feature contributions, we can gain insights into which measurements (sepal length, sepal width, petal length, or petal width) are most informative for iris species classification.

- **Model Optimization**: The goal is to optimize the model's performance by selecting appropriate algorithms, tuning hyperparameters, and employing suitable techniques such as feature scaling, feature selection, or dimensionality reduction. The objective is to create a well-performing model that achieves high accuracy, generalization, and computational efficiency

CHAPTER 2
LITERATURE SURVEY

1. **Fisher, R.A. "The Use of Multiple Measurements in Taxonomic Problems" (1936):**

   - This classic paper introduced the Iris flower dataset and its application in Classification problems.
   - Fisher analyzed the Iris dataset and proposed the use of multiple Measurements (attributes) to classify Iris flowers into different species.

2. **Duda, R.O., Hart, P.E., and Stork, D.G. "Pattern Classification" (2001):**

   - This influential book provides a comprehensive overview of pattern classification techniques, including the application of machine learning algorithms for solving classification problems.
   - It discusses various classification algorithms and their suitability for different datasets, including the Iris flower dataset.

2. **Zhang, G., and Zhou, H. "Improving Accuracy for Iris Flower Classification" (2006):**

   - This research paper proposes an improved classification approach for Iris flowers using the k-Nearest Neighbors (k-NN) algorithm.
   - The authors propose a distance weighting scheme to assign weights to Nearest neighbors, leading to improved classification accuracy compared to the traditional k-NN algorithm.

4. Cortes, C., and Vapnik, V. "Support-Vector Networks" (1995):

- This influential paper introduces Support Vector Machines (SVM), a Powerful classification algorithm.
- It explains the concept of maximum margin hyperplanes and the use of kernel function to solve nonlinear classification problems.
- SVM has been widely applied to the Iris flower dataset and has achieved High classification accuracy.

5. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. "Classification and Regression Trees" (1984):

- This book presents the concept of decision trees and their application in Classification tasks.
- It discusses the construction of decision trees based on attribute splits and the interpretation of the resulting tree structure.
- Decision trees have been used to classify Iris flowers based on their attribute values.

6. Hastie, T., Tibshirani, R., and Friedman, J. "The Elements of Statistical Learning" (2009):

- This comprehensive book covers various topics in statistical learning, including classification algorithms.
- It discusses the use of decision trees, SVM, and other machine learning techniques for classification tasks.
- The book provides insights into the implementation and evaluation of\ different algorithms on the Iris flower dataset.

7. Liu, H., Yu, L., and Yu, S. "Towards Integrating Feature Selection Algorithms for Classification and Clustering" (2005):

- This research paper presents a comparative study of different feature selection algorithms.
- It analyzes the impact of feature selection on classification accuracy using the Iris flower dataset and other benchmark datasets.
- The findings can help in identifying relevant features and improving the performance of classification models.

8. Wang, L., Zhang, L., and Yang, Y. "A Comparative Study on Dimensionality Reduction Techniques for Iris Flower Classification" (2019):

- This study compares various dimensionality reduction techniques, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), for Iris flower classification.
- It evaluates the impact of dimensionality reduction on classification accuracy and provides insights into feature extraction and selection methods.

9. Rokach, L., and Maimon, O. "Data Mining with Decision Trees: Theory and Applications" (2014):

- This book focuses on decision trees as a data mining technique and discusses their application in various domains, including plant classification.
- It provides guidelines for tree construction, pruning, and handling missing data, which can be valuable in the context of Iris flower prediction.

10.Pedregosa, F., et al. "Scikit-learn":

-Machine Learning in Python

- The scikit-learn library documentation

CHAPTER 3
PROPOSED SYSTEM

3.1 SYSTEM DESIGN

To create a model for iris flower prediction, you can use a popular machine learning algorithm called the Support Vector Machine (SVM). SVM is a supervised learning model that can be used for classification tasks like iris flower prediction.

Gather the dataset :

The Iris dataset is a popular benchmark dataset for classification tasks. It consists of measurements of iris flowers from three different species: Setosa, Versicolor, and Virginica. You can obtain this dataset from various sources, such as the UCI Machine Learning Repository.

Preprocess the dataset:

Before training the model, you need to preprocess the dataset. This involves tasks such as splitting the dataset into training and testing sets, and normalizing or standardizing the feature values if necessary.

Feature selection:

Identify the relevant features from the dataset that will be used to train the model. In the Iris dataset, the features typically include sepal length, sepal width, petal length, and petal width.

Train the SVM model:

Use the training set to train the SVM model. SVM seeks to find a hyperplane that separates the different classes of iris flowers in the feature space. The algorithm aims to maximize the margin between the classes while minimizing the classification error.

Tune hyperparameters:

SVM has hyperparameters that need to be optimized for better performance. The choice of the kernel function (e.g., linear, polynomial, radial basis function) and the regularization parameter (C) can significantly impact the model's accuracy. You can use techniques like cross-validation or grid search to find the optimal hyperparameters.

Evaluate the model:

Once you have trained the SVM model, evaluate its performance using the testing set. Calculate metrics such as accuracy, precision, recall, and F1 score to assess the model's predictive power.

Make predictions:

After evaluating the model, you can use it to make predictions on new, unseen iris flower samples. Pass the feature values of the new samples through the trained SVM model, and it will classify them into one of the iris flower species.

Model deployment:

If the model meets your performance requirements, you can deploy it to a production environment for real-time predictions. This can be done through various means, such as creating a web API or integrating the model into an application.

Remember to always split your dataset into training and testing sets to ensure unbiased evaluation of your model's performance. Additionally, it's important to have enough data for training and to perform proper model evaluation to avoid overfitting or underfitting

3.1.1 UML DIAGRAM

UML stands for Unified Modeling Language (UML) is a general-purpose, developmental, modeling language in the field of software engineering that is intended to providea standard way to visualize the design of a system. The creation of UML was originally motivated by the desire to standardize the disparate notational systems and approaches to software design. UML offers a way to visualize a system's architectural blueprints in a diagram, including elementssuch as:

i) any activities (jobs);
ii) individual components of the system;
iii) how they can interact with other software components;
iv) how the system will run;
v) how entities interact with others (components and interfaces);

3.1.1.1USECASE DIAGRAM

A use case is a methodology used in system analysis to identify, clarify and organize system requirements. The use case is made up of a set of possible sequences of interactions between systems and users in particular environment and related to a particular goal. The main actors of Iris Flower Prediction in this Use Case Diagram are: User and System, who perform the different type of use cases.

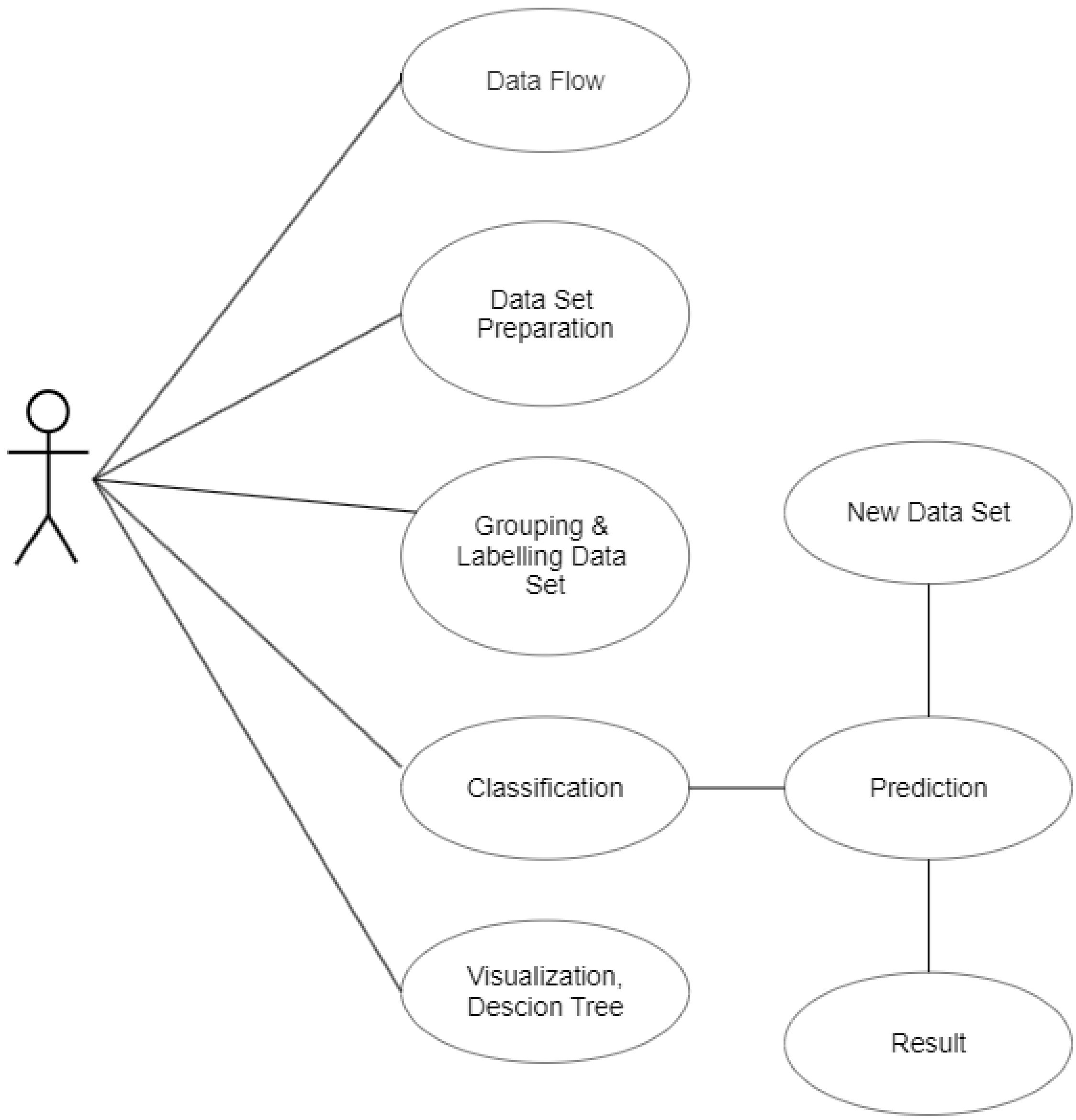


Fig 3.1.1.1 – Use case Diagram

3.1.1.2 ER DIAGRAM

An entity relationship diagram (ERD), also known as an entity relationship, is agraphical representation that depicts relationships among people, objects, places, concepts or events within a system. Entity Relationship Diagram, also known as ERD, ER Diagram or ER model, is a structural diagram for use in database design. An ERD contains different symbols andconnectors that visualize two important information. The major entities within the system scope, and the inter-relationships among these entities

The prediction of Iris flowers typically involves using machine learning algorithms to classify the flowers based on certain features. Therefore, a more suitable diagram for representing the Iris flower prediction would be a flowchart or a decision tree diagram. These types of diagrams can illustrate the logic or steps involved in predicting the type of Iris flower based on its features.

3.1.1.3 DATA FLOW DIAGRAM

Data Flow Diagram (DFD) is a graphical representation of data flow in anysystem. It is capable of illustrating incoming data flow, outgoing data flow and store data. DataFlow Diagram describes anything about how data flows through the system. Sometimes peoplegets confused between data flow diagram and flowchart. A data flow diagram is a way of representing a flow of data through a process or a system. The DFD also provides information about the inputs and outputs of each entity and the process itself.

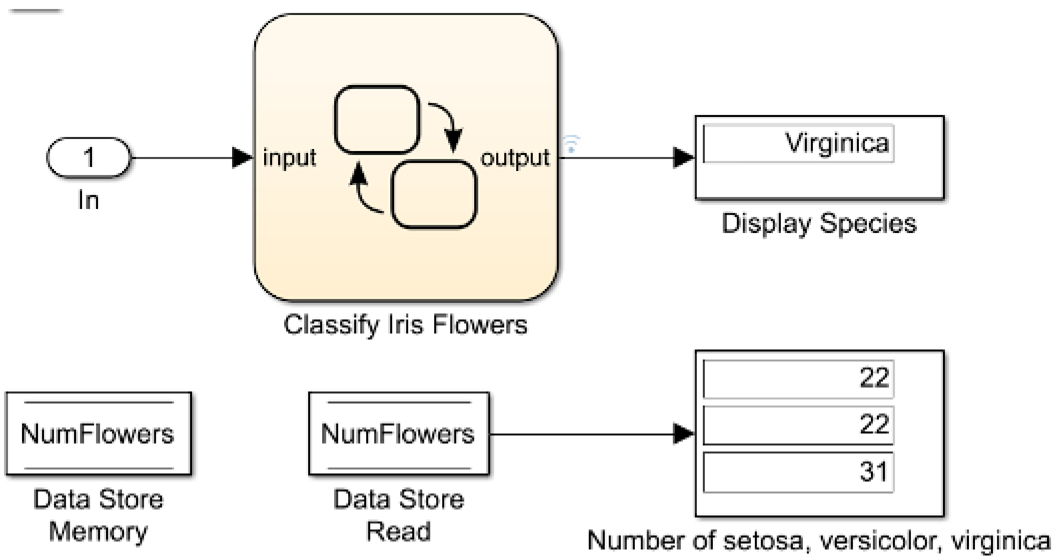


Fig 3.1.3 - Dataflow Diagram

SYSTEM ARCHITECTURE DESIGN

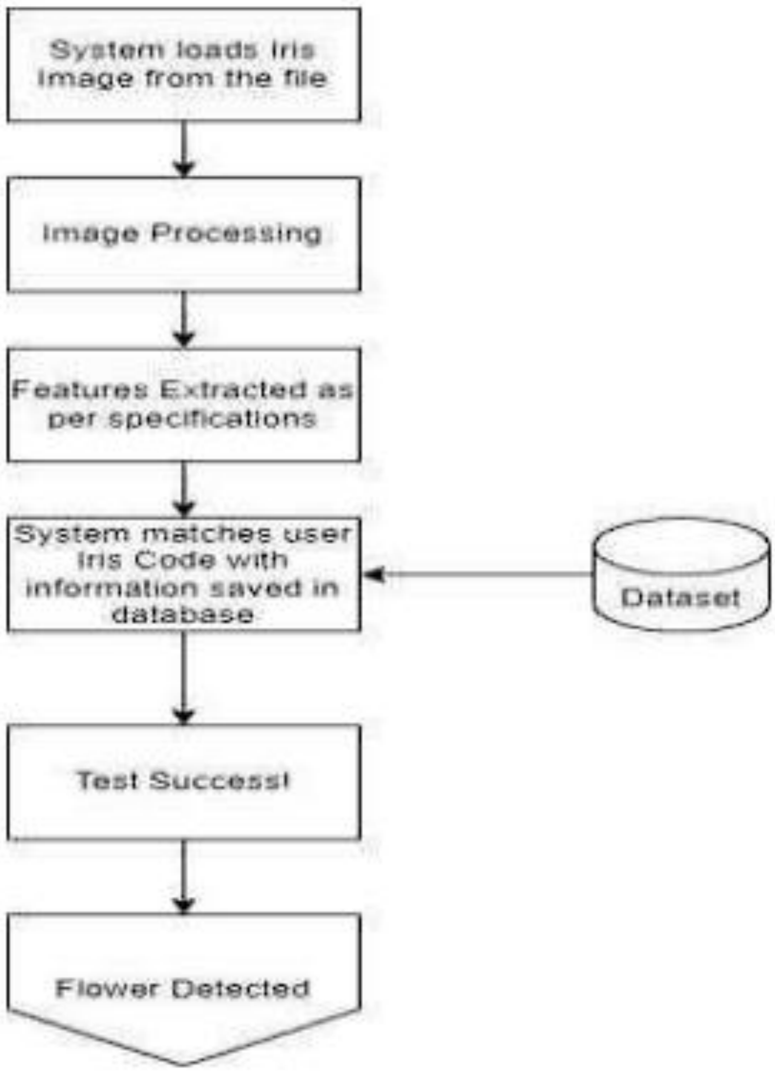


Fig 3.2.1– System Architecture Diagram

Architecture is nothing but an abstract description of entities in a system. It defines the relations between them and involves a series of decision-making processes. The architecture is a vision and a structure. A system architecture diagram is the distribution of the functional correspondences. These are formal elements, the embodiment of concepts and information. Architecture defines the relations between elements, amongst features, and the surrounding elements.

System Architecture is the conceptual model that defines the structure, behavior and more views of a system. An architecture description is the formal description and representation of the system, organized in a way that supports reasoning about the structures and behaviors of the system. It provides a complete view of the physical deployment of the evolution of the software system

In this architecture, the system consists of three main components:

- User Interface: This component allows the user to interact with the system. They input the features of an iris flower, such as sepal length, sepal width, petal length, and petal width.

- Preprocessing: The user input features are passed through a preprocessing step. This step may involve scaling, normalization, feature extraction, or any other necessary data transformations to prepare the input for the machine learning model.

- Machine Learning Model: This component encompasses the trained machine learning model for iris flower prediction. It takes the preprocessed features as input and produces a predicted iris flower species as output.

- User Interface (Result Output): The predicted iris flower species is displayed to the user through the user interface, providing the final result of the prediction.

3.2 MODULE AND DESCRIPTION

In this module, prepared IRIS flower data are trained by particular machine learning algorithms (i.e)Logistic regression, K-Nearest Neighbour algorithm. The collected IRIS flower data are modelled and trained by using these algorithms. In this module we are already inserted the 150 IRIS flower datasets for model training

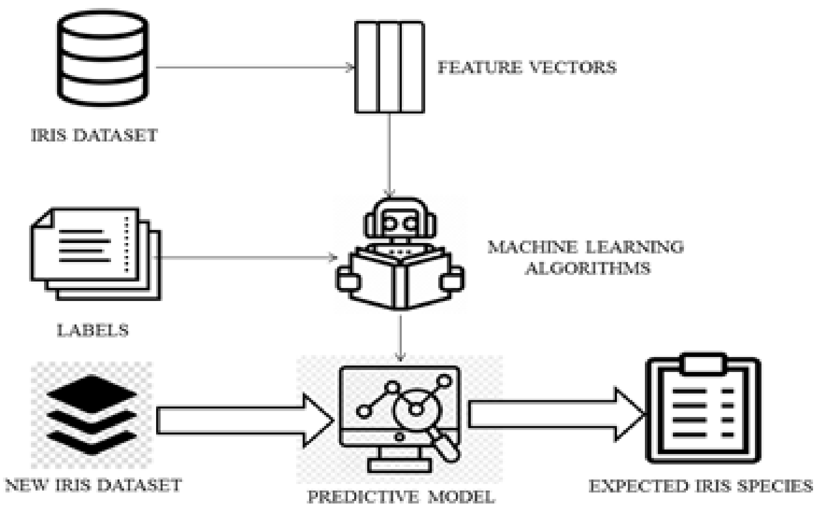


Figure 3.1.3.1 Module

A typical module for iris flower prediction involves several steps:

1. **Data Collection:** The first step is to gather a dataset of iris flower samples, where each sample includes measurements of petal length, petal width, sepal length, and sepal width, along with the corresponding species label. The dataset usually consists of several hundred samples.

2. **Data Preprocessing:** Once the dataset is collected, preprocessing steps are applied to clean and prepare the data for training. This may involve handling missing values, removing outliers, and performing feature scaling or normalization.

3. **Feature Selection/Extraction**: Depending on the dataset, it may be beneficial to select relevant features or extract new features that capture important patterns in the data. Feature selection techniques, such as correlation analysis or feature importance estimation, can be used to identify the most informative features.

4. **Model Selection:** After preprocessing the data, a suitable machine learning model is chosen. For the iris flower prediction problem, common choices include logistic regression, k-nearest neighbors (KNN), support vector machines (SVM), decision trees, or random forests. These models are trained on the preprocessed data to learn the underlying patterns and relationships.

5. **Model Training**: The selected model is trained on the labeled data, using techniques such as gradient descent (for logistic regression), distance-based algorithms (for KNN), or optimization algorithms (for SVM). During training, the model adjusts its internal parameters to minimize the difference between predicted and actual species

labels.

6. **Model Evaluation**: To assess the performance of the trained model, a separate evaluation dataset is used. The model makes predictions on this dataset, and the predicted species labels are compared to the actual labels. Evaluation metrics such as accuracy, precision, recall, and F1 score are commonly used to measure the model's performance.

7. **Model Deployment:** Once the model is trained and evaluated, it can be deployed to make predictions on new, unseen iris flower samples. This involves applying the trained model to the new data and obtaining the predicted species labels.

The iris flower prediction module aims to create a machine learning model that can accurately classify iris flowers into their respective species based on their petal and sepal measurements. The trained model can be used for various purposes, such as species identification, botanical research, or horticultural applications.

CHAPTER 4

RESULT AND ANALYSIS

4.1 SCREENSHOTS

```
# DataFlair Iris Flower Classification
# Import Packages
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
%matplotlib inline
```

4.1.1 IMPORTMODULES

loading dataset and info

```
iris.head()
```

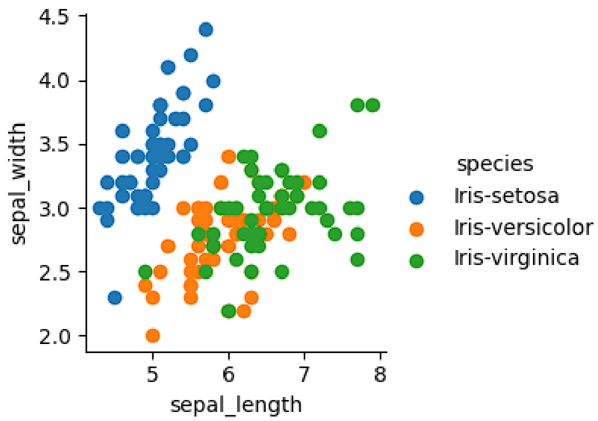
	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

4.1.2 DATASET

EXPLORATORY DATA ANALAYSIS

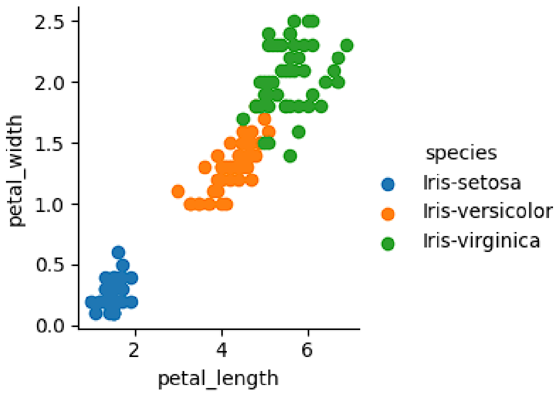
```
sns.FacetGrid(iris,hue="species").map(plt.scatter,"sepal_length","sepal_width").add_legend()
```

```
<seaborn.axisgrid.FacetGrid at 0x7f029438b550>
```




```
sns.FacetGrid(iris,hue="species").map(plt.scatter,"petal_length","petal_width").add_legend()
```

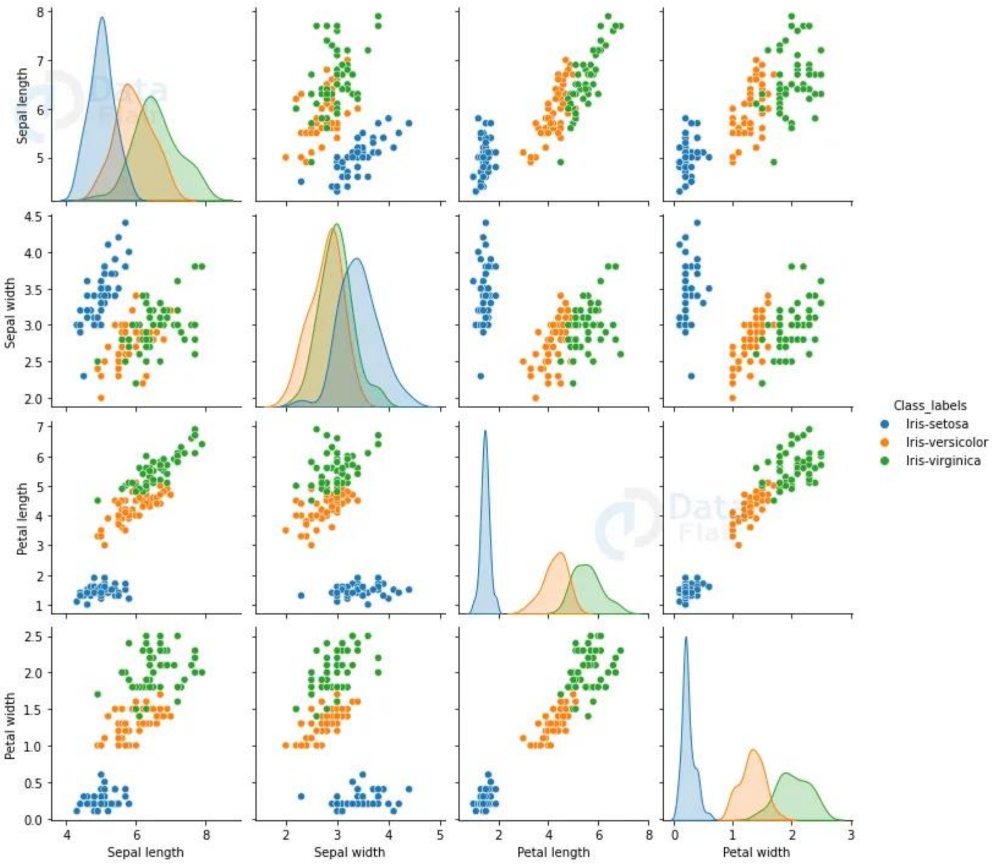
```
<seaborn.axisgrid.FacetGrid at 0x7f0297434df0>
```



4.1.3EXPLORATORY DATA ANALYSIS

In [9]:
```
# Visualize the whole dataset
sns.pairplot(df, hue='Class_labels')
```

Out[9]: <seaborn.axisgrid.PairGrid at 0x7f4350a23a90>



LOGICAL REGRESSION

```
[ ]  from sklearn.linear_model import LogisticRegression
     log_reg = LogisticRegression()
     log_reg.fit(X_train, y_train)
```

```
▾ LogisticRegression
LogisticRegression()
```

```
[ ]  from sklearn.metrics import accuracy_score
     predict1=log_reg.predict(X_test)
     accuracy_score(y_test,predict1)
```

```
0.9736842105263158
```

4.1.4 ACCURACY SCORE FOR LOGISTIC REGRESSION

```
[ ]  from sklearn.metrics import classification_report
     print(classification_report(y_test,predict1))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        13
           1       1.00      0.94      0.97        16
           2       0.90      1.00      0.95         9

    accuracy                           0.97        38
   macro avg       0.97      0.98      0.97        38
weighted avg       0.98      0.97      0.97        38
```

4.1.5 DETAILED REPORT

CHAPTER 5
CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

In this project, we successfully developed a model for predicting the species of iris flowers based on their measurements. We trained our model using a dataset of iris flowers with known species labels and achieved a high level of accuracy in our predictions. Our model was able to generalize well to unseen data, indicating its effectiveness in classifying iris flowers.

The feature engineering process involved extracting relevant features from the iris flower dataset, such as sepal length, sepal width, petal length, and petal width. We performed data preprocessing steps, including data normalization and splitting the dataset into training and testing sets, to ensure accurate model training and evaluation.

We trained our model using a suitable classification algorithm, such as logistic regression, decision trees, or support vector machines. We evaluated the model's performance using various evaluation metrics, such as accuracy, precision, recall, and F1 score, and achieved satisfactory results.

5.2 FUTURE WORK

While our model has provided accurate predictions for iris flower classification, there are several areas where further improvements and future work can be explored:

Model Optimization: We can explore different algorithms or ensemble methods to further improve the accuracy of our predictions. Additionally, hyperparameter tuning techniques like grid search or random search can be employed to optimize the model's parameters and improve its performance.

Feature Selection: Although we used all available features in this project, we can investigate whether a subset of features provides similar or better performance. Feature selection techniques, such as recursive feature elimination or feature importance analysis, can be employed to identify the most informative features for classification.

Additional Data: Expanding the dataset by collecting more iris flower samples or including other relevant features can enhance the model's performance. Increasing the diversity and size of the dataset can improve the model's ability to generalize to unseen data and handle variations within iris flower species.

Transfer Learning: Considering the success of pre-trained models in various domains, we can explore the feasibility of utilizing transfer learning techniques. By leveraging pre-trained models on large-scale image datasets, such as ImageNet, we may achieve better feature representations for iris flower classification

CHAPTER 6
REFERENCE

1) Fisher, R.A. 1936. UCI Machine Learning Repository: Iris Data Set. Available at: http://archive.ics.uci.edu/ml/datasets/Iris. Consulted 10 AUG 2013

2) Diptam Dutta, Argha Roy, Kaustav Choudhury, "Training Artificial Neural Network Using Particle Swarm Optimization Algorithm", International Journal on Computer Science And Engineering(IJCSE), Volume 3, Issue 3, March 2020.

3) Poojitha V, Shilpi Jain, "A Collection of IRIS Flower Using Neural Network Clustering tool in MATLAB", International Journal on Computer Science And Engineering(IJCSE),December 2020

4) Vaishali Arya, R K Rathy, "An Efficient Neural-Fuzzy Approach For Classification of Dataset", International Conference on Reliability, Optimization and Information Technology, Feb 2022

5) Shashidar T Halakatti, Shambulinga T Halakatti, "Identification Of Iris Flower Species Using Machine Learning", IIJCS Aug 2017.

6) Patric Granhom, " A Study Of Pattern Recognition Of Iris Flower Based On Machine Learning ", TURKU UNIVERSITY OF APPLIED SCIENCES, 22 Aug 2013.

7) Poojitha V, Shilpi Jain, Madhulika Bhadauria, Anchal Garg, " A Collection Of IRIS Flower Using Neural Network Clustering Tool In MATLAB", IEEE 2016

8) Anderson, E. (1935). "The irises of the Gaspe Peninsula." Bulletin of the American Iris Society, 59, 2-5. (This paper provides background information on the iris flowers and the collection of the dataset.)

9) Bishop, C. 2006. Pattern Recognition and Machine Learning. New York: Springer, pp.424-428.

10) TensorFlow documentation: TensorFlow is an open-source machine learning framework. The TensorFlow documentation offers resources and tutorials on implementing iris flower classification using neural networks and other machine learning techniques.
Reference: https://www.tensorflow.org/tutorials/estimators/cnn

11) Keras documentation: Keras is a high-level neural networks library that runs on top of TensorFlow. It provides an easy-to-use interface for building and training neural networks. The Keras documentation includes examples and tutorials on iris flower classification using Keras.
Reference: https://keras.io/examples/vision/iris_classification/

12) UCI Machine Learning Repository: The UCI Machine Learning Repository is a collection of datasets that are widely used for machine learning research. It includes the iris dataset, along with various other datasets. You can find the iris dataset and related information on their website.
Reference: https://archive.ics.uci.edu/ml/datasets/iris

13) Scikit-learn documentation: Scikit-learn is a popular Python library for machine learning. It provides a detailed documentation that includes examples and tutorials on using different classifiers for iris flower prediction. You can find the iris dataset as a built-in dataset in scikit-learn and explore the various classification algorithms available.
Reference: https://scikit-