

VIETNAM NATIONAL UNIVERSITY, HO CHI MINH CITY
UNIVERSITY OF TECHNOLOGY
FACULTY OF COMPUTER SCIENCE AND ENGINEERING



Machine Learning - CO3001

Assignment

Iris Flower Classification using several Machine Learning models

Instructor: Truong Thi Thai Minh
Class: CC01 - Group 1
Student(s): Tran Anh Kiet - 2152147
Le Duc Hoang Nam - 2111795
Vo Thien Nam - 2111817

HO CHI MINH CITY, OCTOBER 2023



Contents

1	Introduction	2
2	Literature Review	2
2.1	K-nearest neighbors	2
2.2	Random Forest	2
2.3	Logistic Regression	2
3	Methodology	2
3.1	Data Collection	2
3.2	Data Preprocessing	3
3.3	Model Training	3
4	Results and Discussion	3
4.1	Model Performance	3
4.2	Analysis	4
5	Conclusion	4
6	References	4

1 Introduction

Irises influenced the design of the French fleur-de-lis, are commonly used in the Japanese art of flower arrangement known as Ikebana, and underlie the floral scents of the “essence of violet” perfume. They’re also the subject of this well-known machine learning project, in which you must create an ML model capable of sorting irises based on five factors into one of three classes, Iris Setosa, Iris Versicolour, and Iris Virginica.

To get started, the data set below includes 50 instances of each of the three iris classes for a total of 150 instances. While one of the classes is linearly separable, the other two are not. Our goal is to create a model capable of classifying each iris instance into the appropriate class based on four attributes: sepal length, sepal width, petal length, and petal width.



Figure 1: Iris Setosa

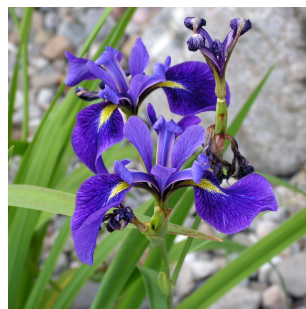


Figure 2: Iris Versicolour



Figure 3: Iris Virginica

2 Literature Review

Discuss previous work and findings related to the Iris classification problem. For classifying, i use 3 popular model in machine learning

2.1 K-nearest neighbors

2.2 Random Forest

2.3 Logistic Regression

3 Methodology

3.1 Data Collection

The dataset was obtained from the UCI Machine Learning Repository. For more information, visit the [UCI Iris Dataset](#) page.

This is one of the earliest datasets used in the literature on classification methods and widely used in statistics and machine learning. The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are not linearly separable from each other.

Predicted attribute: class of iris plant.

This is an exceedingly simple domain.

3.2 Data Preprocessing

1. **Loading the Dataset:** The dataset is loaded into a pandas DataFrame. This dataset includes features like sepal length, sepal width, petal length, and petal width, along with the class labels.
2. **Feature Selection:** The features (sepal length, sepal width, petal length, and petal width) are separated from the class labels. This is done by splitting the DataFrame into X (features) and y (labels).
3. **Splitting the Dataset:** The dataset is split into training and test sets. *train_test_split()*. This is a common practice in machine learning to evaluate the model on data that it hasn't seen during training. The test size is set to 20% of the total dataset.
4. **Feature Scaling:** Standardization of features using *StandardScaler()*.

The Python code for these steps is as follows:

```
1 import pandas as pd
2 from sklearn.model_selection import train_test_split
3 from sklearn.preprocessing import StandardScaler
4
5 # Load dataset
6 url = "https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"
7 names = ['sepal-length', 'sepal-width', 'petal-length', 'petal-width', 'class']
8 dataset = pd.read_csv(url, names=names)
9
10 # Feature Selection
11 X = dataset.iloc[:, :-1].values
12 y = dataset.iloc[:, -1].values
13
14 # Splitting the Dataset
15 X_train, X_test, y_train
16 y_test = train_test_split(X, y, test_size=0.2, random_state=1)
17
18 # Feature Scaling
19 scaler = StandardScaler()
20 X_train = scaler.fit_transform(X_train)
21 X_test = scaler.transform(X_test)
```

3.3 Model Training

Explain how you trained your model.

1. K-nearest neighbors
2. Linear Regression
3. Logistic Regression

4 Results and Discussion

4.1 Model Performance

Present the results of your model.



4.2 Analysis

Analyze and interpret the results.

5 Conclusion

Summarize the main findings, the implications of your work, and any future work.

6 References

<https://machinelearningcoban.com/2017/01/08/knn/>