**PROJECT REPORT**

## Iris Flower Species Classification

Submitted by

**Mansoor Elahi Registration Number : 11902487**

**Kankan Bardhan Registration Number : 11902428**

**Course Code : INT-246**

Under the Guidance of

# Dr Sagar Pande

# School of Computer Science and Engineering

A picture containing text

Description automatically generated

**DECLARATION**

We hereby declare that the project work entitled (“Title of the project”) is an authentic record of our own work carried out as requirements of Project for the award of B.Tech degree in CSE from Lovely Professional University, Phagwara, under the guidance of Dr. Sagar Pande, during August to November 2020. All the information furnished in this project report is based on our own intensive work and is genuine.

Name of Student 1: Mansoor Elahi

Registration Number: 11902487

Name of Student 2: Kankan Bardhan

Registration Number: 11902428

**CERTIFICATE**

This is to certify that the declaration statement made by this group of students is correct to the best of my knowledge and belief. They have completed this Project under my guidance and supervision. The present work is the result of their original investigation, effort and study. No part of the work has ever been submitted for any other degree at any University. The Project is fit for the submission and partial fulfillment of the conditions for the award of B.Tech degree in CSE from Lovely Professional University, Phagwara.

**School of Computer Science and Engineering,**

Lovely Professional University,

Phagwara, Punjab.

Date : 18TH November 2021

**ACKNOWLEDGEMENT**

We would like to express our special thanks of gratitude to our mentor **‘Dr. Sagar Pande’** who game us the golden opportunity to do this wonderful project on the topic **‘IRIS flower species classification’** which also helped us in doing a lot of research and we come to know about so many new things.

Mansoor Elahi

Kankan Bardhan

**TABLE OF CONTENTS**

Title Page No.

1. **Introduction 6.**
   1. **Overview 6.**
   2. **Project Objective 6.**
2. **Problem Statement and Methodology 6.**
   1. **Problem Statement 6.**
   2. **Pattern Recognition 7.**

* 1. **Material 7.**
  2. **Methodology 7.**
     1. **Preprocessing 8.**
     2. **Artificial Neural Network 8,**
     3. **Training the Neural Network 9.**

**2.5. Summary 10.**

**3. Results 10.**

**4. Code 11.**

**5. Conclusion 14.**

**1. INTRODUCTION:**

**1.1. OVERVIEW:**

Bioinformatics is a promising and novel research area in the 21st century. This field is data-driven and aims at the understanding of relationships and gaining knowledge in biology. In order to extract this knowledge encoded in biological data, advanced computational technologies, algorithms, and tools need to be used. Basic problems in bioinformatics like protein structure prediction, multiple alignments of sequences, phylogenic inferences, etc. Are inherently non-deterministic polynomial-time hard in nature. To solve these kinds of problems artificial intelligence (AI) methods offer a powerful and efficient approach. Researchers have used AI techniques like Artificial Neural Networks (ANN), Fuzzy Logic, Genetic Algorithms, and Support Vector Machines to solve problems in bioinformatics. Artificial Neural Networks is one of the AI techniques commonly in use because of its ability to capture and represent complex input and output relationships among data. The purpose of this paper is to provide an overall understanding of ANN and its place in bioinformatics to a newcomer to the field.

**1.2 PROJECT OBJECTIVE**

The current study aims to identify the type of iris flowers by using the dataset that prepared in advance way by the expert biologists to study the flower types through some measurements and statistics for each type using data mining techniques and neural network classifiers.

**2. PROBLEM STATEMENT AND METHODOLOGY:**

**2.1. PROBLEM STATEMENT**

The real problem in this study is how to achieve a new way to classify Iris flowers and identify their type to study their behavior and help biologists with new ML techniques.

**2.2. PATTERN RECOGNITIION**

As machine learning, pattern recognition, can be treated as two different classification methods: supervised classification and unsupervised classification. They are quite similar to supervised learning and unsupervised learning. As supervised classification needs a teacher that gives the category of samples, the unsupervised classification is doing it the other way around. Pattern recognition is related to statistics, psychology, linguistics, computer science, biology and so on. It plays an important role in Artificial Intelligence and image processing.

**2.3. MATERIAL**

The IRIS data set includes three classes of 50 objects each, where each class refers to a type of IRIS plant. The attributed that already been predicted belongs to the class of IRIS plant. The list of attributes present in the IRIS can be described as categorical, nominal and continuous.

The data set is complete. This project makes use of the well-known IRIS dataset, which refers to three classes of 50 instances each, where each class refers to a type of IRIS plant. The first of the classes is linearly distinguishable from the remaining two, with the second two not being linearly separable from each other. The 150 instances, which are equally separated between the three classes, contain the following four numeric attributes:

* **sepal length**
* **sepal width**
* **petal length**
* **petal width**

the fifth attribute is the predictive attributes which is the class attribute that means each instance also includes an identifying class name, each of which is one of the following: IRIS **Setosa**, IRIS **Versicolour**, or IRIS **Virginica**.

**2.4 METHODOLOGY**

The proposed method of this work includes the following steps:

**2.4.1 PREPROCESSING**

Initially the dataset pre-processing was done by isolating the attribute part that have the features from the label part and cleaning the data by removing null or missing values.

The dataset is shuffled to ensure that all cases will be trained on the neural network and tested successfully.

In addition, the dataset has been configured to be an input to the neural network, by normalize it and make its values between zero and one to reduce the model over-fitting by the following equation.

**2.4.2 ARTIFICIAL NEURAL NETWORK**

Neural networks are composed of simple elements operating in parallel. The neuron model shown in Figure 2.1. is the one that widely used in artificial neural networks with some minor modifications on it.

**Shape

Description automatically generated**

Figure 2.2 Neural Network general diagram

The artificial neuron given in this figure has *N* input, denoted as *u1*, *u2, …, u N.* Each line connecting these inputs to the neuron is assigned a weight, which is denoted as *w1, w2, …, w N* respectively.

The threshold in artificial neuron is usually represented by Ф and the activation is given by this formula:

The inputs and weight are real values. A negative value for a weight indicates an inhibitory connection while a positive value indicating excitatory one. If Ф is positive, it is usually referred as bias. For its mathematical convenience (+) sign is used in the activation formula. Sometimes, the threshold is combined for simplicity into the summation part by assuming an imaginary input u0 = +1 and a connection weight w0 = Ф. Hence the activation formula becomes:

The neuron output function f(a) can be:

Linear:

Sigmoid:

**2.4.3 TRAINING THE NEURAL NETWORK**

Neural networks have been trained to perform complex functions in various fields of application including pattern recognition, identification, classification, speech, vision and control systems.

After loading and cleaning the dataset, the feature vectors are calculated for each follower type in the dataset. These feature vectors are used as inputs to train the networks. In training algorithm, the IRIS flower feature vectors that belong to same class are used as input to the network, the target will be binary number for every class in dataset example for class one that have id 1:

Target= [ 1,0,0], class two 2: target = [0,1,0], and so on.

**2.5. SUMMARY**

These steps show the summary of IRIS flower and neural network

* Load the dataset and clean it to extract the feature vector.
* Normalize the extracted features.
* Split the dataset into training set and test set to evaluate the neural network.
* Create neural network and train it on the training set.
* For each new IRIS flower image to be identified, calculate its feature vector.
* Use these feature vectors as network inputs and simulate network with these inputs.

**3. RESULTS**

The proposed method is tested on IRIS flower dataset, we use 105/150 of dataset for training process and 45/150 for testing process to ensure our network is working properly.

|  |  |  |  |
| --- | --- | --- | --- |
| **Si. No.** | **Error rate** | **Accuracy (%)** | **Training Time (s/step)** |
| 4 | **0.8889** | **66.67** | **0** |
| 6 | 0.8180 | 67.62 | 0 |
| 8 | 0.7700 | 66.67 | 4ms |
| 10 | 0.7166 | 70.48 | 0 |
| 11 | **0.6948** | **67.62** | **0** |
| 13 | 0.6566 | 70.48 | 0 |
| 16 | 0.6169 | 67.62 | 4ms |
| 19 | 0.5840 | 92.38 | 4ms |

Our system yields a good result when use IRIS flower features and back propagation neural network, this project has been implemented using Python programming language.

**4. CODE**

# Importing Packages

import numpy as np

import keras

from keras.models import Sequential

from keras.layers import Dense

from sklearn import datasets

# Loading Dataset

data = datasets.load\_iris()

print(data)

x = data.data

y = data.target

z = data.target\_names

print(data.feature\_names)

print(x[0])

x = data.data

y = data.target

z = data.target\_names

print(data.feature\_names)

print(x[0])

# Split Dataset

from sklearn.model\_selection import train\_test\_split

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size=0.3)

# Data Shape

print(x\_train.shape)

print(x\_test.shape)

# Building the Model

model= Sequential()

model.add(Dense(100,input\_shape=(4,), activation="relu"))

model.add(Dense(3, activation='softmax'))

# Compile the Model

model.compile(optimizer="adam", loss="sparse\_categorical\_crossentropy", metrics=["accuracy"])

# Fit the Model

model.fit(x\_train,y\_train, epochs=20)

# Evaluate the Model

model.evaluate(x\_test, y\_test)

# Predict for the first 10 Observations

pred=model.predict(x\_test[:10])

print(pred)

p=np.argmax(pred, axis=1)

print(p)

print(y\_test[:10])

# Prediction Result

for i in p:

print("Predicted-Class: {}, Name: {}".format(i,z[i]))

**5. CONCLUSION**

This part will outline the most critical conclusions reached after studying the problem and the proposed solution, as summarized in the following point

* IRIS flowers dataset is obtained and pre-processed to extract the features.
* The feature vector is constructed and shuffled to prepare it for training process.
* Create neural network with back-propagation algorithm with different hidden neurons size to yield the optimal error rate.
* Train the network using the training set until the error reach the required error-threshold.
* Save the neural network with the adjusted weight for later identification.
* Write the experimental results and discuss it.
* Finally, suggest a future work.