Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To make the clear division for making the decisions about the classes using Decision Tree and Random Forest**	
Experiment No: 08	**Date:19-10-2022**	**Enrolment No:92000133018**

Aim: To make the clear division for making the decisions about the classes using Decision Tree and Random Forest

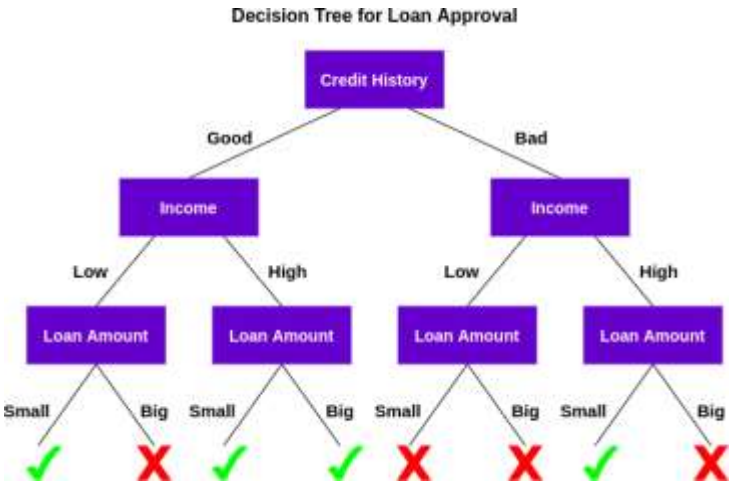
IDE: Google Colab

Theory:

Let's start with a thought experiment that will illustrate the difference between a decision tree and a random forest model. Suppose a bank has to approve a small loan amount for a customer and the bank needs to make a decision quickly. The bank checks the person's credit history and their financial condition and finds that they haven't re-paid the older loan yet. Hence, the bank rejects the application.

But here's the catch – the loan amount was very small for the bank's immense coffers and they could have easily approved it in a very low-risk move. Therefore, the bank lost the chance of making some money. Now, another loan application comes in a few days down the line but this time the bank comes up with a different strategy – multiple decision-making processes. Sometimes it checks for credit history first, and sometimes it checks for customer's financial condition and loan amount first. Then, the bank combines results from these multiple decision-making processes and decides to give the loan to the customer.

Even if this process took more time than the previous one, the bank profited using this method. This is a classic example where collective decision making outperformed a single decision-making process. A decision tree is a supervised machine learning algorithm that can be used for both classification and regression problems. A decision tree is simply a series of sequential decisions made to reach a specific result. Here's an illustration of a decision tree in action (using the above example):



Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To make the clear division for making the decisions about the classes using Decision Tree and Random Forest**	
Experiment No: 08	**Date:19-10-2022**	**Enrolment No:92000133018**

First, it checks if the customer has a good credit history. Based on that, it classifies the customer into two groups, i.e., customers with good credit history and customers with bad credit history. Then, it checks the income of the customer and again classifies him/her into two groups. Finally, it checks the loan amount requested by the customer. Based on the outcomes from checking these three features, the decision tree decides if the customer's loan should be approved or not.

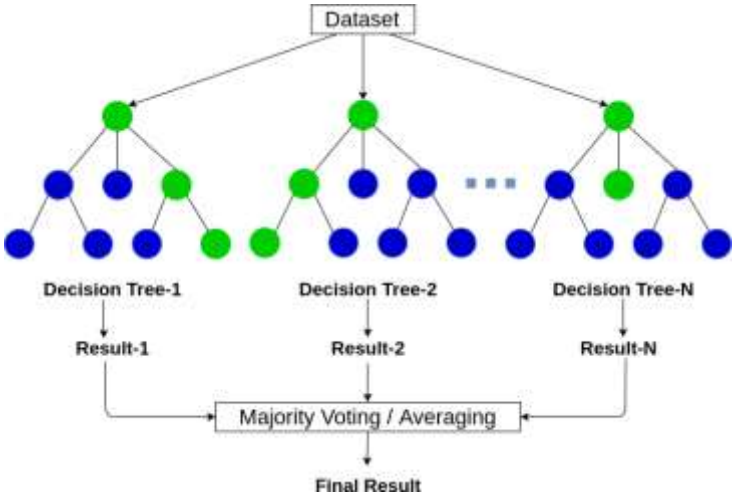
The features/attributes and conditions can change based on the data and complexity of the problem but the overall idea remains the same. So, a decision tree makes a series of decisions based on a set of features/attributes present in the data, which in this case were credit history, income, and loan amount.

An Overview of Random Forest

The decision tree algorithm is quite easy to understand and interpret. But often, a single tree is not sufficient for producing effective results. This is where the Random Forest algorithm comes into the picture. Random Forest is a tree-based machine learning algorithm that leverages the power of multiple decision trees for making decisions. As the name suggests, it is a "forest" of trees!

But why do we call it a "random" forest? That's because it is a forest of randomly created decision trees. Each node in the decision tree works on a random subset of features to calculate the output. The random forest then combines the output of individual decision trees to generate the final output.

The Random Forest Algorithm combines the output of multiple (randomly created) Decision Trees to generate the final output. T his process of combining the output of multiple individual models (also known as weak learners) is called **Ensemble Learning**.



Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To make the clear division for making the decisions about the classes using Decision Tree and Random Forest**	
Experiment No: 08	**Date:19-10-2022**	**Enrolment No:92000133018**

Program (Code):

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

from sklearn import datasets

iris = datasets.load_iris()

iris

class_name = iris.target_names

class_name

features=iris.feature_names

features

data = pd.DataFrame({
   'sepal length':iris.data[:,0],
   'sepal width':iris.data[:,1],
   'petal length':iris.data[:,2],
   'petal width':iris.data[:,3],
   'species':iris.target
})

data

#split the dataset to train and test data
from sklearn.model_selection import train_test_split

#features for training
x=data[['sepal length','sepal width','petal width','petal length']]
y = data['species'] #target label

x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.25) #25% in testing and 75% in training

from sklearn.ensemble import RandomForestClassifier
```

Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To make the clear division for making the decisions about the classes using Decision Tree and Random Forest**	
Experiment No: 08	**Date:19-10-2022**	**Enrolment No:92000133018**

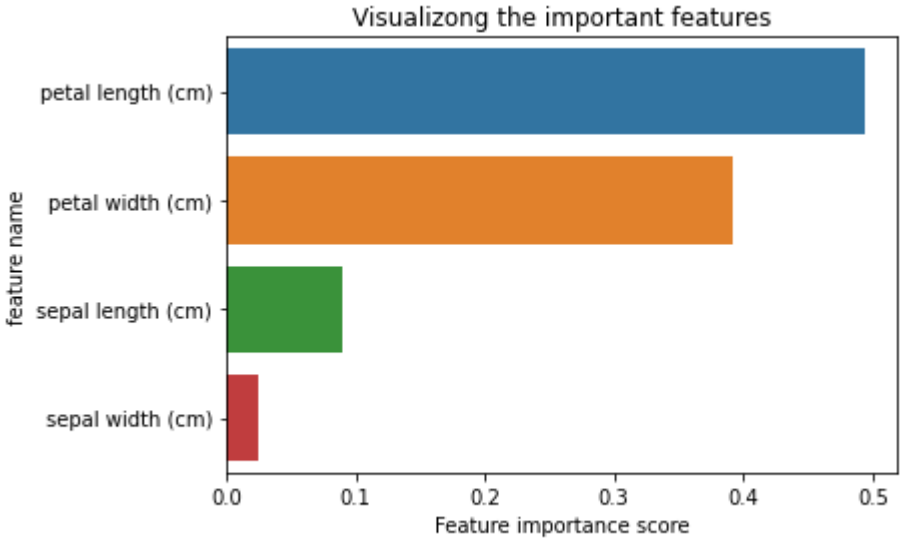
```
#creating the random forest classifier
forest = RandomForestClassifier(n_estimators=100)
forest.fit(x_train,y_train) # fit: training the RF model

y_pred = forest.predict(x_test)

#model accuracy
y_pred #predicted

y_test #testing dataset original

from sklearn import metrics
print('Accuracy score : ',metrics.accuracy_score(y_test,y_pred))

from sklearn.tree import export_graphviz

import pydot

tree = forest.estimators_[5]
export_graphviz(tree,out_file='/content/tree5.dot',feature_names=features)

(graph,)=pydot.graph_from_dot_file("/content/tree5.dot")
graph.write_png('/content/tree5.png')

iris.target_names[forest.predict([[15,5,8,4]])][0]

#finding import feature
feature_imp=pd.Series(forest.feature_importances_,index=features).sort_values(ascending=False)

feature_imp

# Commented out IPython magic to ensure Python compatibility.
import seaborn as sns
# %matplotlib inline

sns.barplot(x=feature_imp,y=feature_imp.index)
plt.xlabel("Feature importance score")
plt.ylabel("feature name")
plt.title("Visualizong the important features")
plt.show()
```

Results:

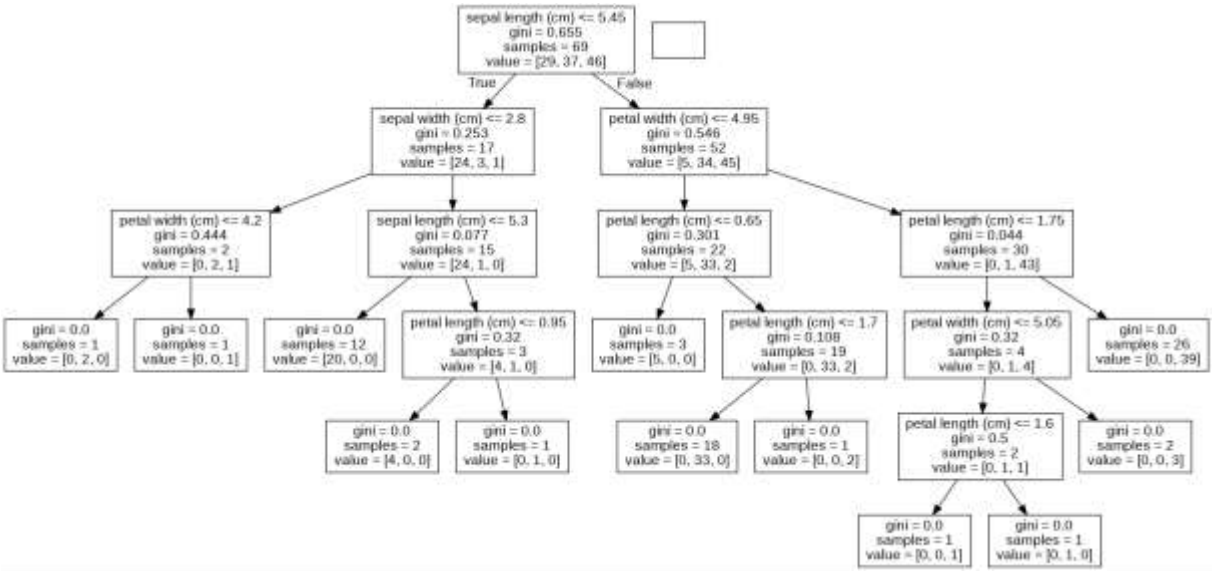
Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To make the clear division for making the decisions about the classes using Decision Tree and Random Forest**	
Experiment No: 08	**Date:19-10-2022**	**Enrolment No:92000133018**

To be attached with

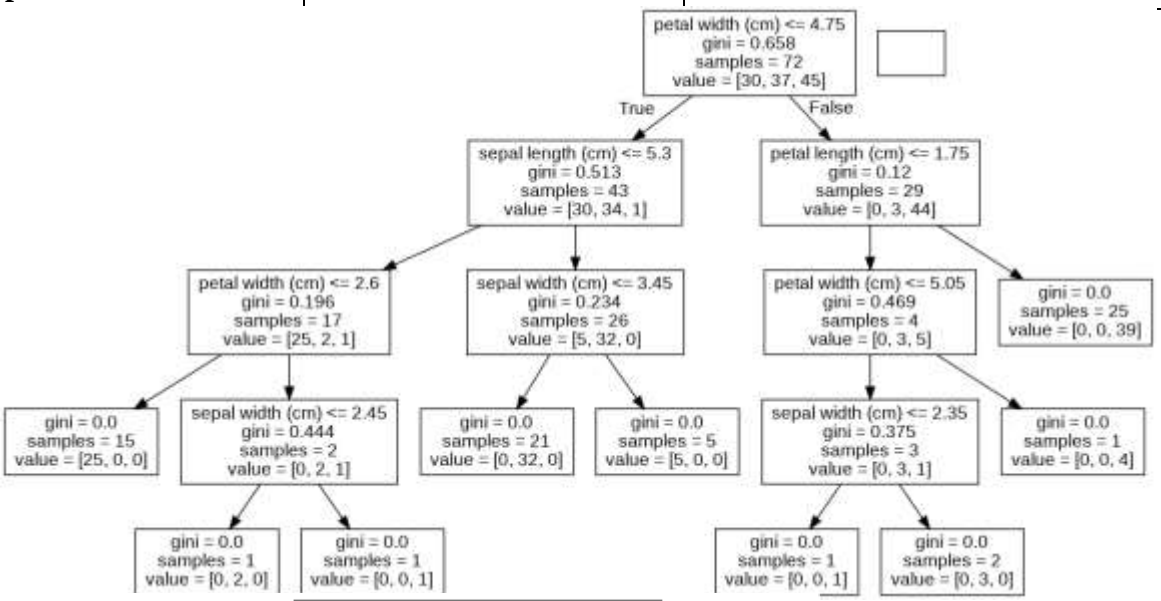
a. Decision Tree of classification

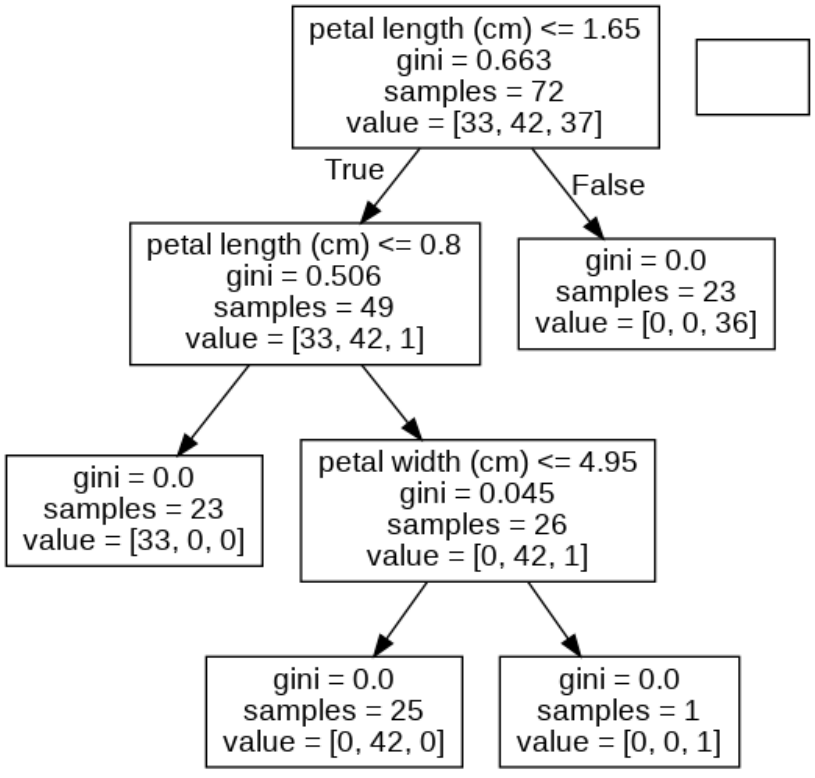


b. Mean error of decision tree with different number of estimators (consider atleast 10 values of estimators)

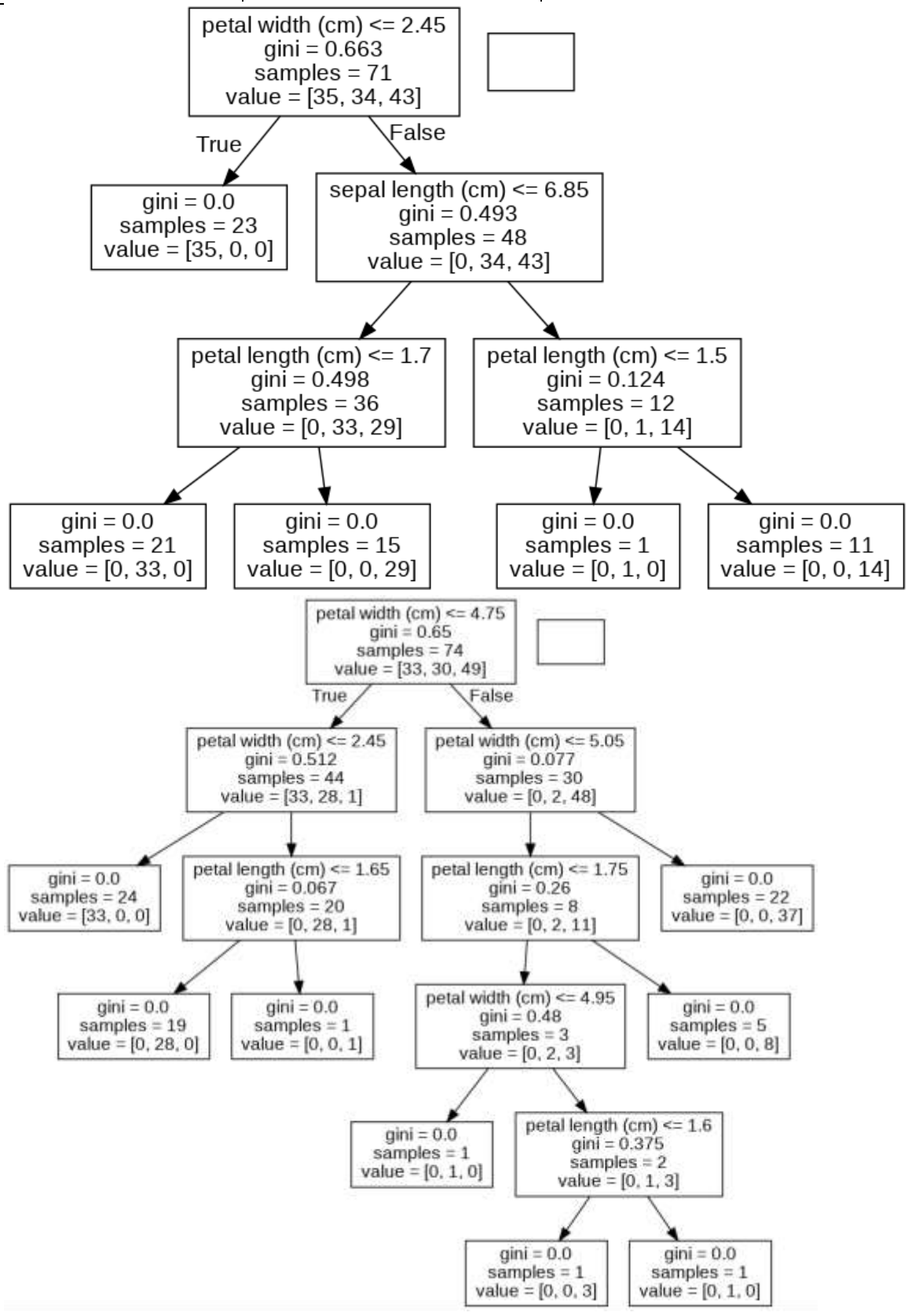


Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To make the clear division for making the decisions about the classes using Decision Tree and Random Forest**	
Experiment No: 08	**Date:19-10-2022**	**Enrolment No:92000133018**

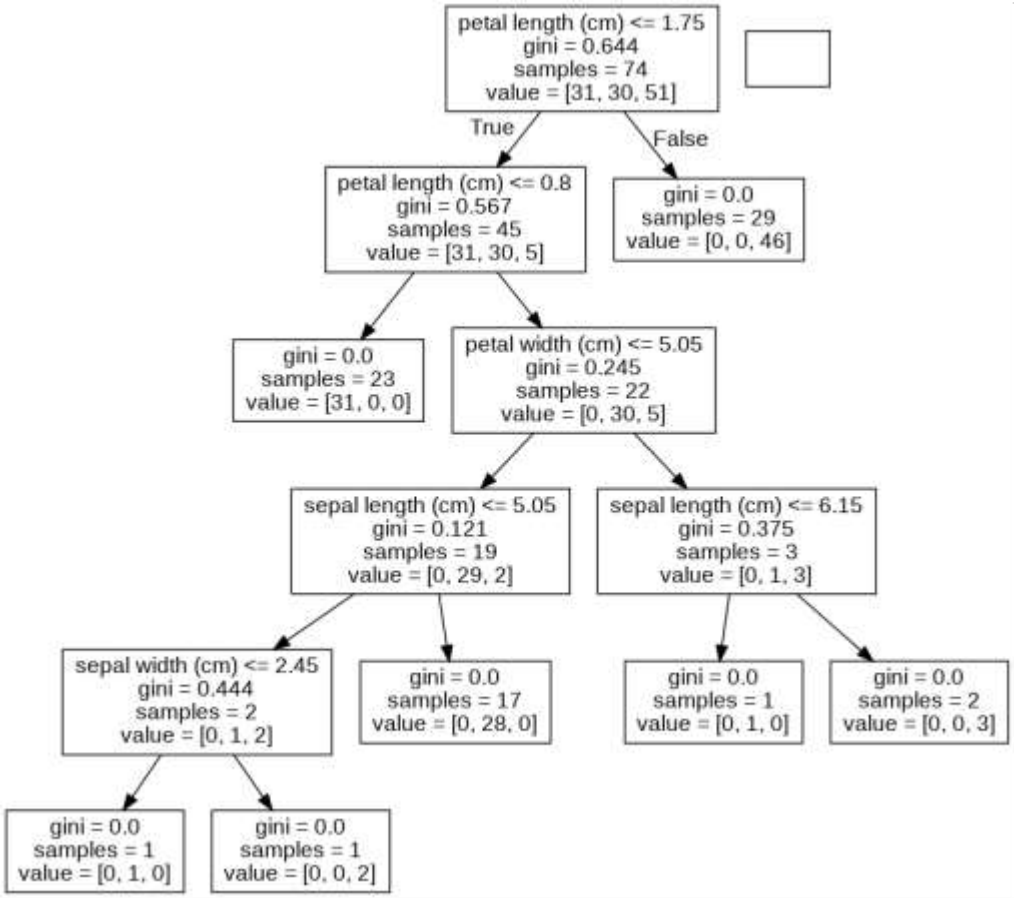




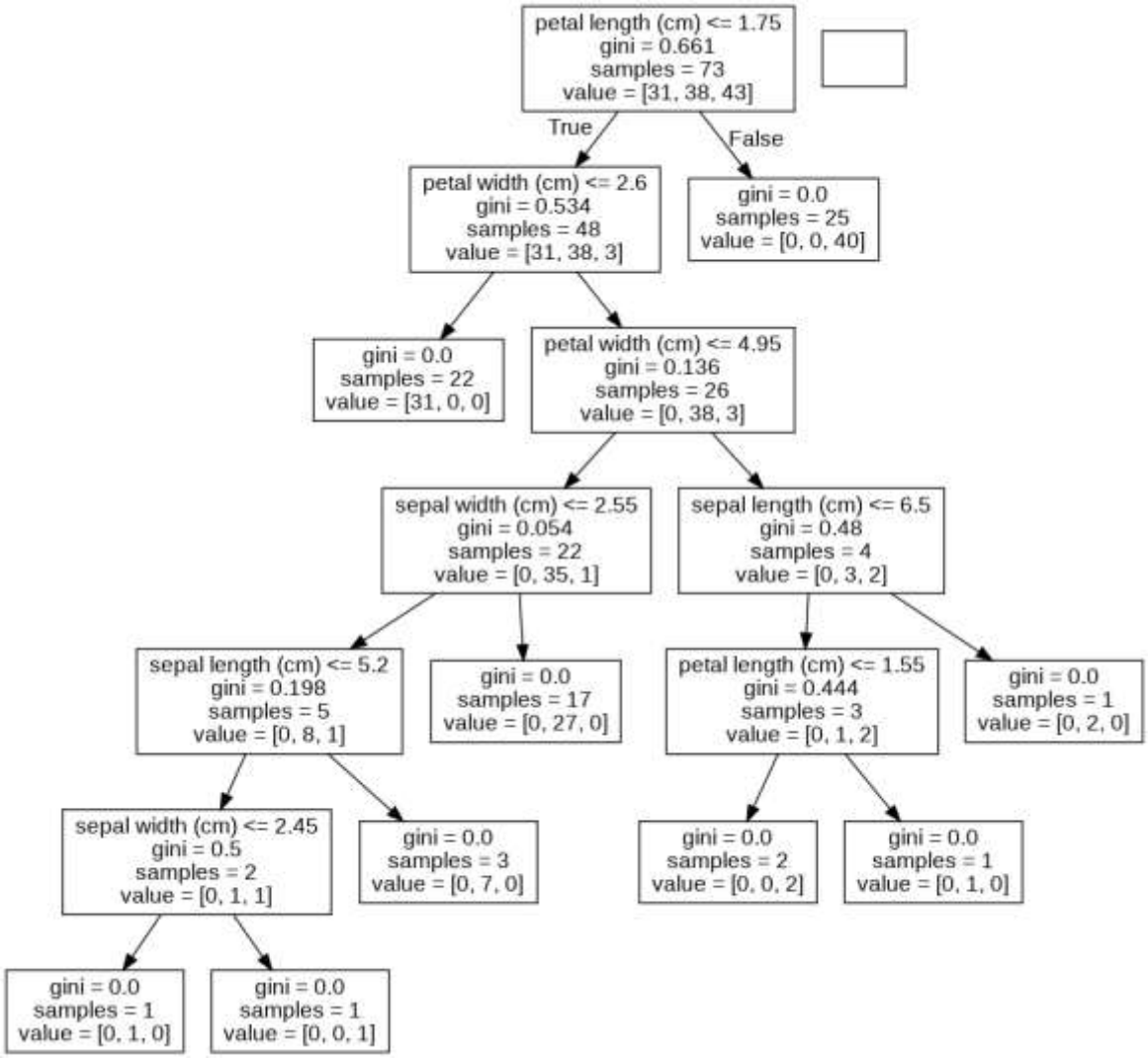
Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To make the clear division for making the decisions about the classes using Decision Tree and Random Forest**	
Experiment No: 08	**Date:19-10-2022**	**Enrolment No:92000133018**



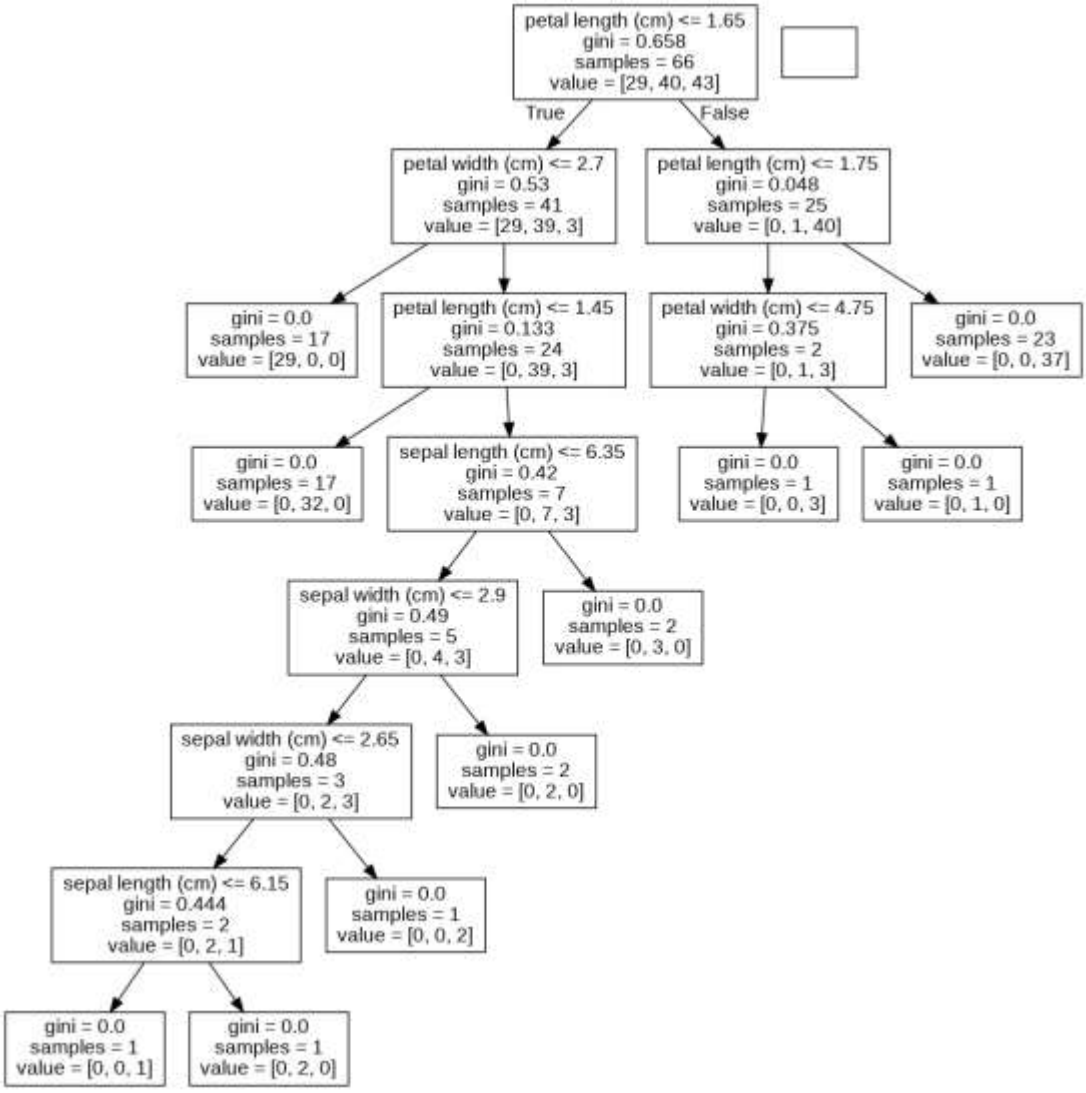
Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To make the clear division for making the decisions about the classes using Decision Tree and Random Forest**	
Experiment No: 08	**Date:19-10-2022**	**Enrolment No:92000133018**



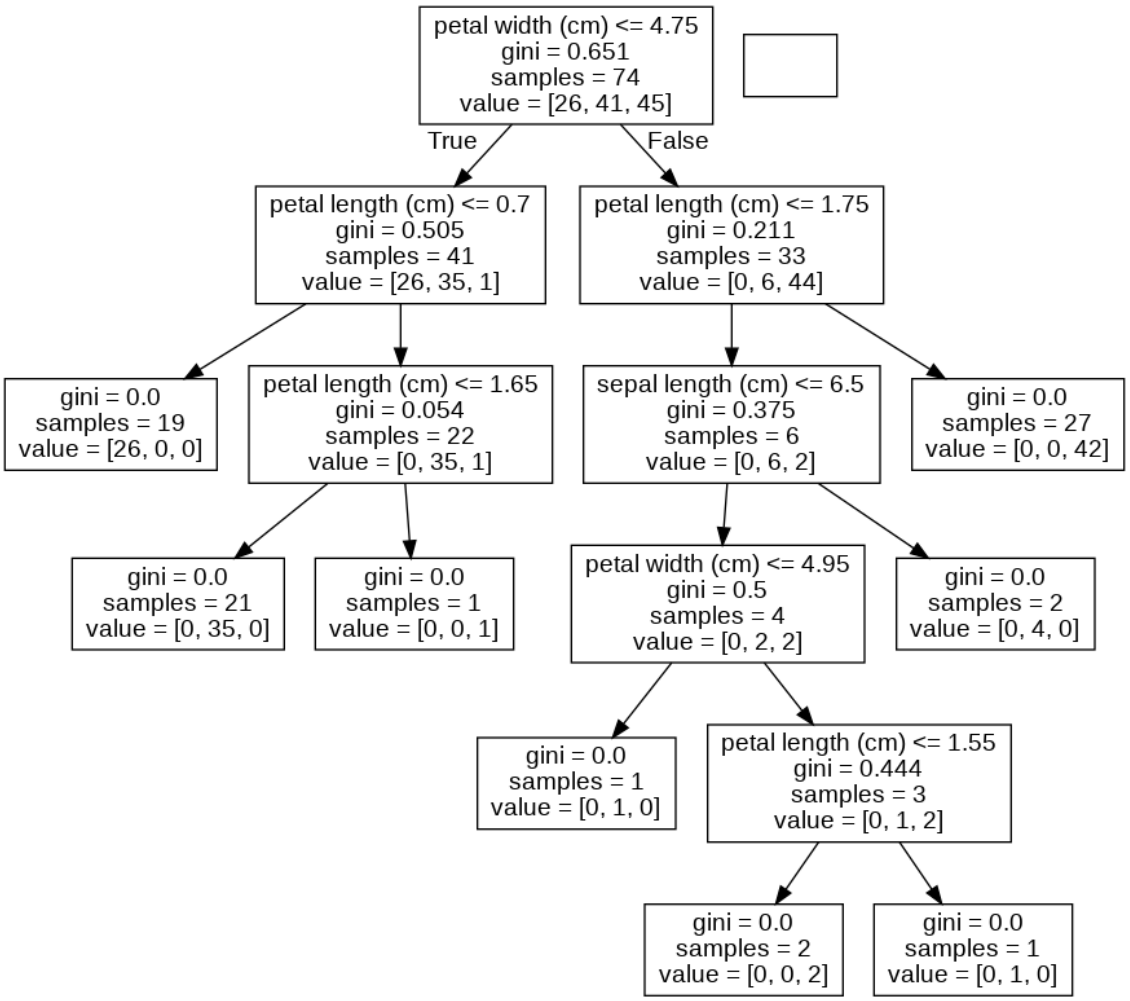
Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To make the clear division for making the decisions about the classes using Decision Tree and Random Forest**	
Experiment No: 08	**Date:19-10-2022**	**Enrolment No:92000133018**



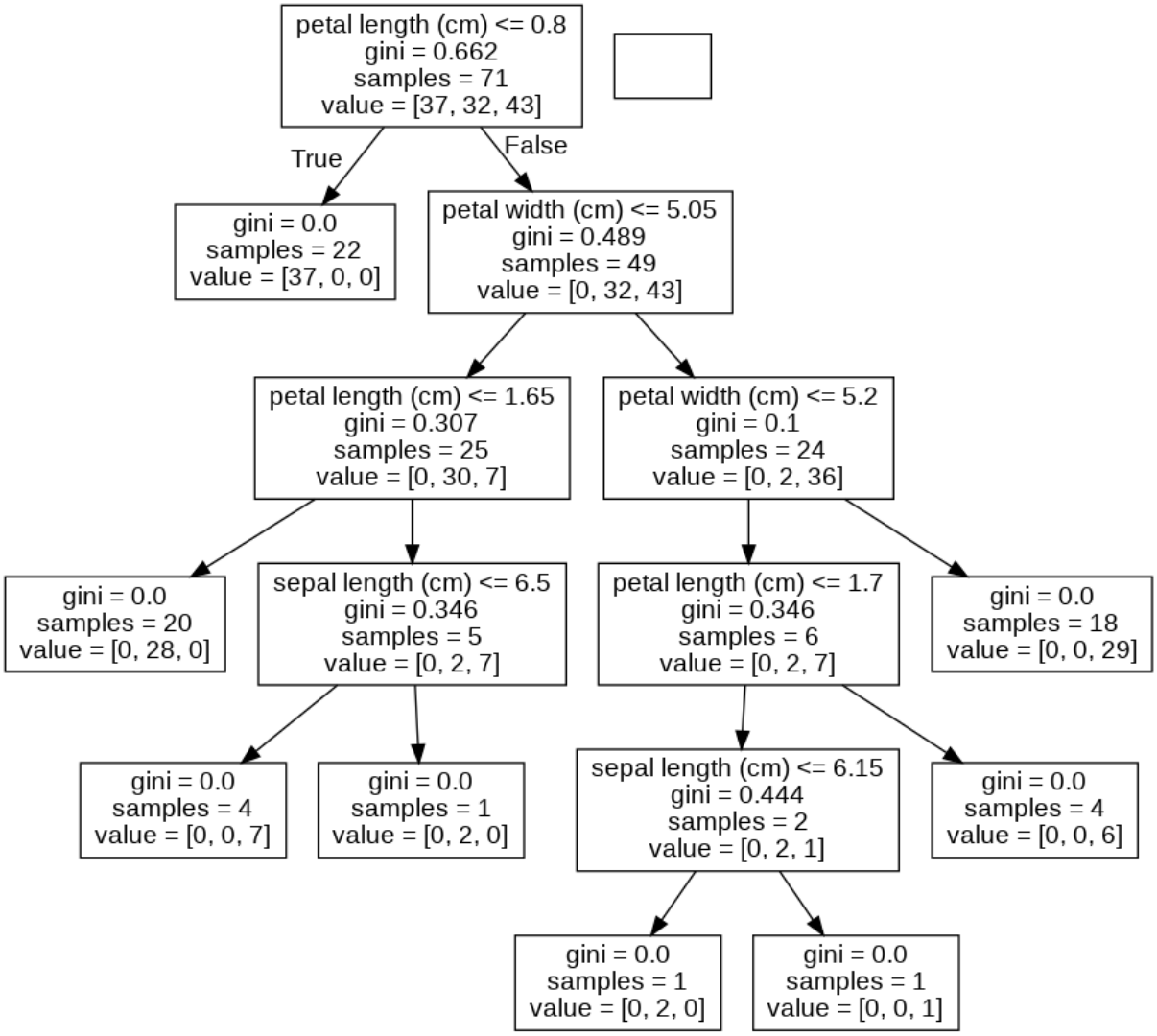
Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To make the clear division for making the decisions about the classes using Decision Tree and Random Forest**	
Experiment No: 08	**Date:19-10-2022**	**Enrolment No:92000133018**



Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To make the clear division for making the decisions about the classes using Decision Tree and Random Forest**	
Experiment No: 08	**Date:19-10-2022**	**Enrolment No:92000133018**



Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To make the clear division for making the decisions about the classes using Decision Tree and Random Forest**	
Experiment No: 08	**Date:19-10-2022**	**Enrolment No:92000133018**



Observation and Result Analysis:

a. Nature of the dataset

b. During Training Process

Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To make the clear division for making the decisions about the classes using Decision Tree and Random Forest**	
Experiment No: 08	**Date:19-10-2022**	**Enrolment No:92000133018**

c. After the training Process

d. Observation over the decision trees and random forest

Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To make the clear division for making the decisions about the classes using Decision Tree and Random Forest**	
Experiment No: 08	**Date:19-10-2022**	**Enrolment No:92000133018**

Post Lab Exercise:

a. Explain how does the decision tree works using statistical approach

b. What is the requirement of random forest over decision tree?

Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To make the clear division for making the decisions about the classes using Decision Tree and Random Forest**	
Experiment No: 08	**Date:19-10-2022**	**Enrolment No:92000133018**

c. What are the advantages of decision tree?

d. What are the limitations of decision tree?

e. Can Random Forest Algorithm be used both for Continuous and Categorical Target Variables?

f. How does a Random Forest Algorithm give predictions on an unseen dataset?

Post Lab Activity:

Consider any dataset from **https://archive.ics.uci.edu/ml/datasets.php** and perform the multiple variable linear regression analysis over the dataset and obtain the best fit line. Make sure that the dataset is not matching with your classmates. You can also select the dataset from other ML repositories with prior permission from your concerned subject faculty.