

In [1]:

from matplotlib import pyplot as plt

In [2]:

import pandas as pd

In [4]:

df=pd.read_csv("C:\\Users\\User\\Desktop\\IRIS.csv")

In [5]:

df

Out[5]:

	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
...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 5 columns

In [6]:

features=df.iloc[:, :-1].values
label=df.iloc[:, -1].values

In [7]:

df["species"].value_counts()

Out[7]:

Iris-setosa 50
Iris-versicolor 50
Iris-virginica 50
Name: species, dtype: int64

In [8]:

from sklearn.model_selection import train_test_split

In [9]:

x_train,x_test,y_train,y_test =train_test_split(features,label)

In [10]:

x_train.shape,x_test.shape,y_train.shape,y_test.shape

Out[10]:

((112, 4), (38, 4), (112,), (38,))

In [12]:

from sklearn.tree import DecisionTreeClassifier

In [15]:

from sklearn.metrics import classification_report

In [16]:

dt=DecisionTreeClassifier().fit(features,label)

In [17]:

y_pred=dt.predict(x_test)

In [19]:

print(classification_report(y_pred,y_test))

	precision	recall	f1-score	support
Iris-setosa	1.00	1.00	1.00	18
Iris-versicolor	1.00	1.00	1.00	7
Iris-virginica	1.00	1.00	1.00	13
accuracy			1.00	38
macro avg	1.00	1.00	1.00	38
weighted avg	1.00	1.00	1.00	38

In [20]:

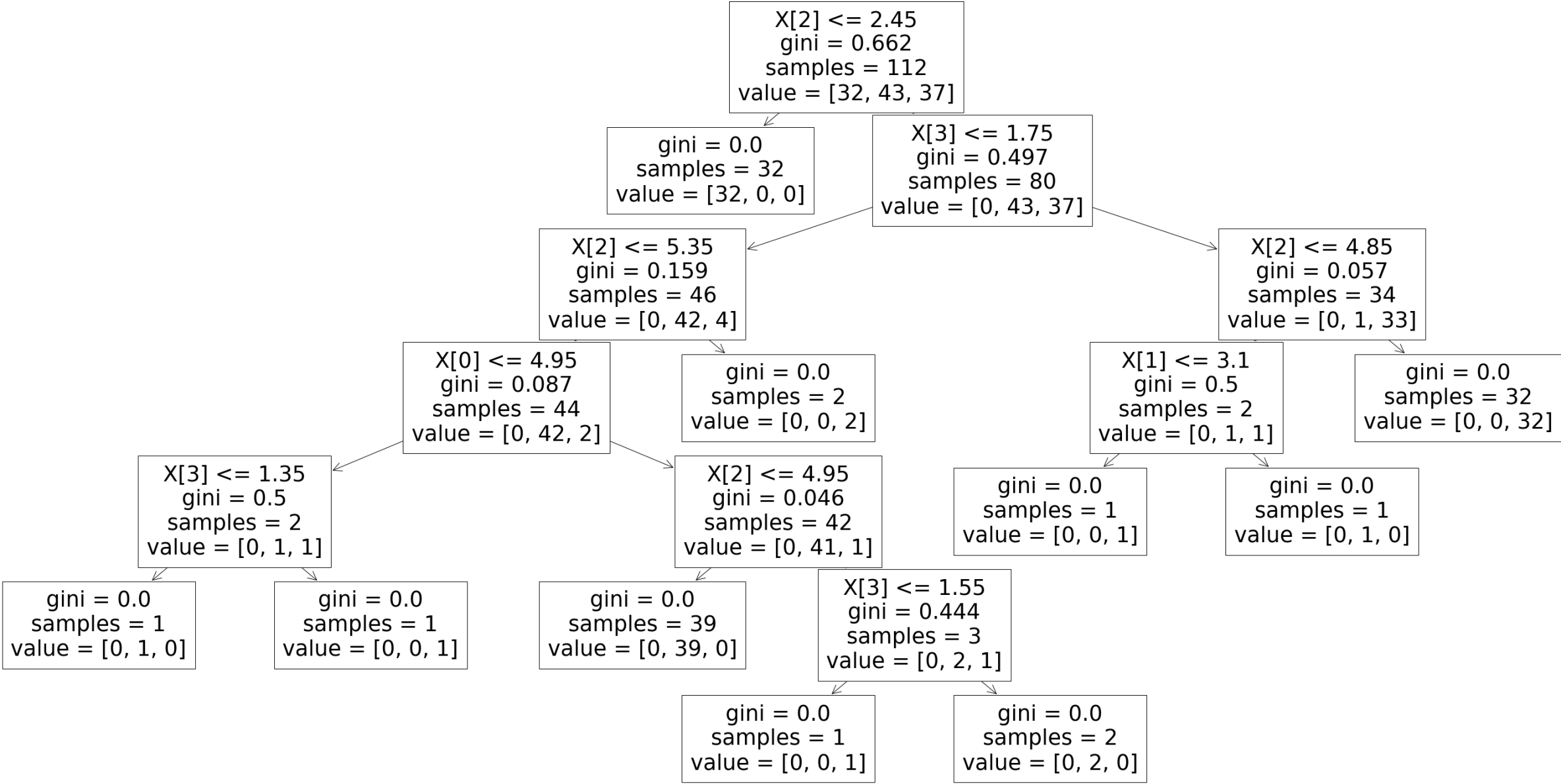
from sklearn import tree

In [27]:

plt.figure(figsize=(50,25))
tree.plot_tree(dt.fit(x_train, y_train))

Out[27]:

[Text(1511.25, 1261.9285714285713, 'X[2] <= 2.45\ngini = 0.662\nsamples = 112\nvalue = [32, 43, 37]'),
Text(1278.75, 1067.7857142857142, 'gini = 0.0\nsamples = 32\nvalue = [32, 0, 0]'),
Text(1743.75, 1067.7857142857142, 'X[3] <= 1.75\ngini = 0.497\nsamples = 80\nvalue = [0, 43, 37]'),
Text(1162.5, 873.6428571428571, 'X[2] <= 5.35\ngini = 0.159\nsamples = 46\nvalue = [0, 42, 4]'),
Text(930.0, 679.5, 'X[0] <= 4.95\ngini = 0.087\nsamples = 44\nvalue = [0, 42, 2]'),
Text(465.0, 485.3571428571429, 'X[3] <= 1.35\ngini = 0.5\nsamples = 2\nvalue = [0, 1, 1]'),
Text(232.5, 291.2142857142858, 'gini = 0.0\nsamples = 1\nvalue = [0, 1, 0]'),
Text(697.5, 291.2142857142858, 'gini = 0.0\nsamples = 1\nvalue = [0, 0, 1]'),
Text(1395.0, 485.3571428571429, 'X[2] <= 4.95\ngini = 0.046\nsamples = 42\nvalue = [0, 41, 1]'),
Text(1162.5, 291.2142857142858, 'gini = 0.0\nsamples = 39\nvalue = [0, 39, 0]'),
Text(1627.5, 291.2142857142858, 'X[3] <= 1.55\ngini = 0.444\nsamples = 3\nvalue = [0, 2, 1]'),
Text(1395.0, 97.07142857142867, 'gini = 0.0\nsamples = 1\nvalue = [0, 0, 1]'),
Text(1860.0, 97.07142857142867, 'gini = 0.0\nsamples = 2\nvalue = [0, 2, 0]'),
Text(1395.0, 679.5, 'gini = 0.0\nsamples = 2\nvalue = [0, 0, 2]'),
Text(2325.0, 873.6428571428571, 'X[2] <= 4.85\ngini = 0.057\nsamples = 34\nvalue = [0, 1, 33]'),
Text(2092.5, 679.5, 'X[1] <= 3.1\ngini = 0.5\nsamples = 2\nvalue = [0, 1, 1]'),
Text(1860.0, 485.3571428571429, 'gini = 0.0\nsamples = 1\nvalue = [0, 0, 1]'),
Text(2325.0, 485.3571428571429, 'gini = 0.0\nsamples = 1\nvalue = [0, 1, 0]'),
Text(2557.5, 679.5, 'gini = 0.0\nsamples = 32\nvalue = [0, 0, 32]')]



In []: