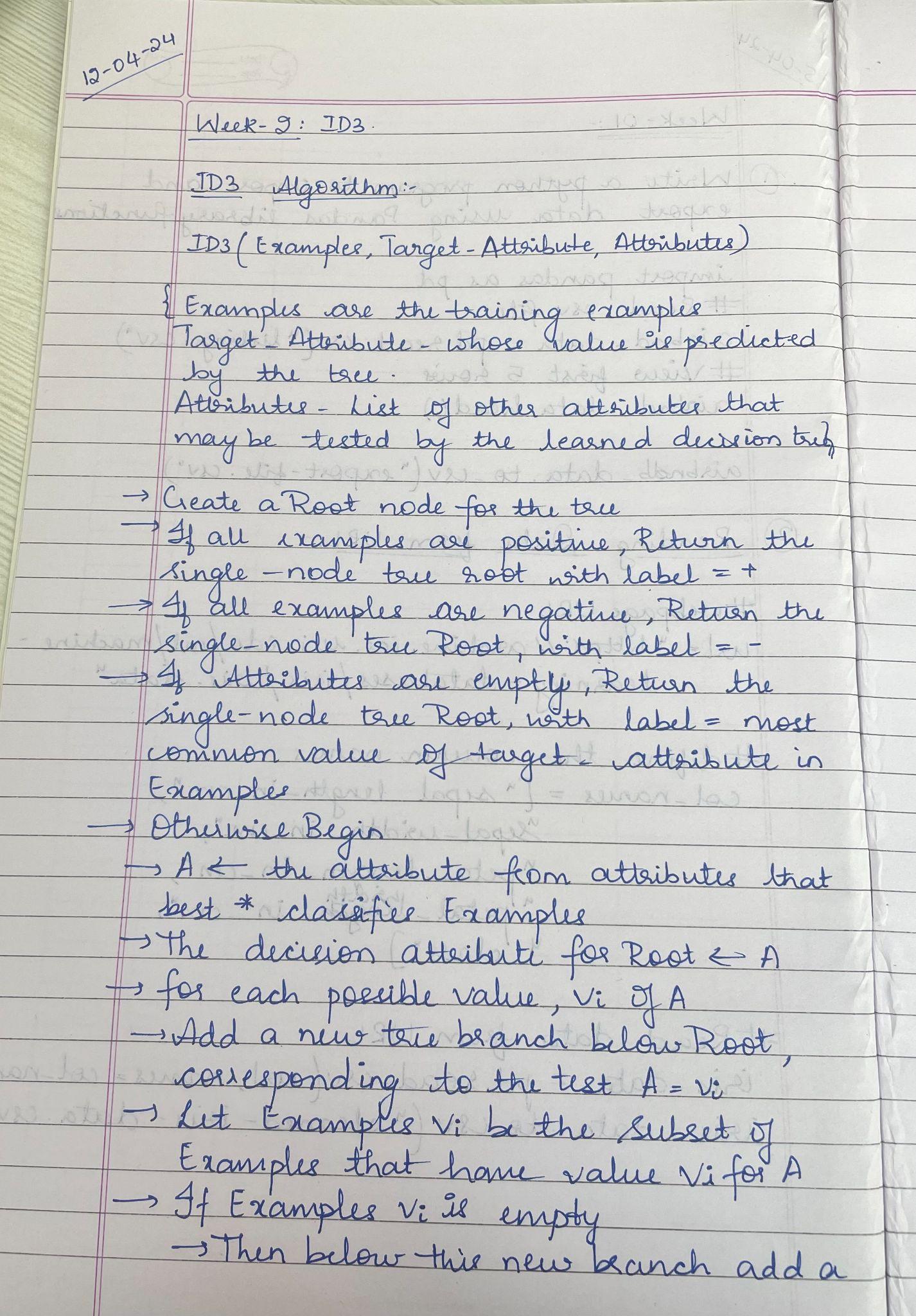
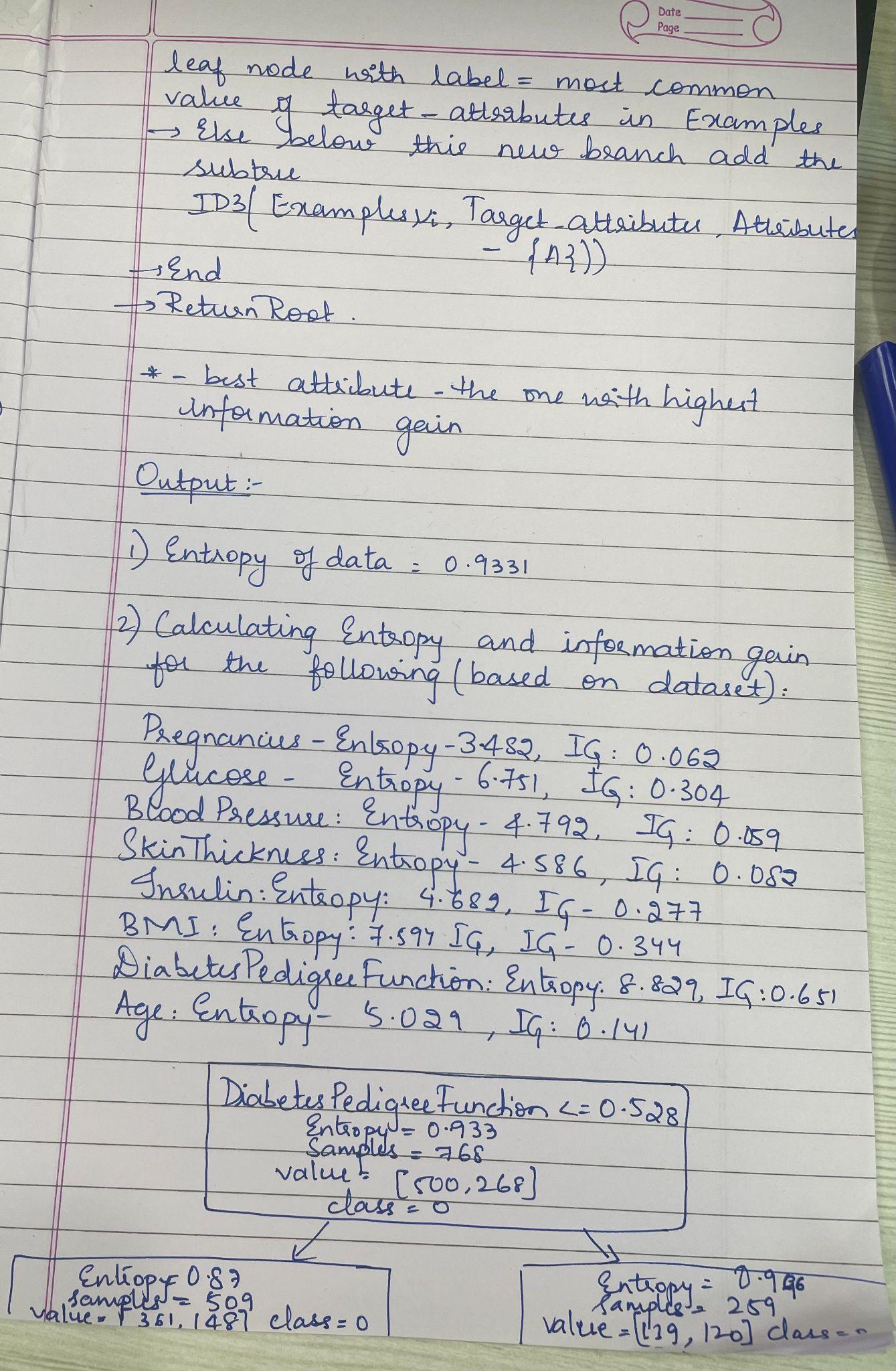
WEEK-2

**Use an appropriate data set for building the decision tree (ID3) and apply this knowledge to classify a new sample.**

Observation Screenshot:





**Code and Output:**

import pandas as pd

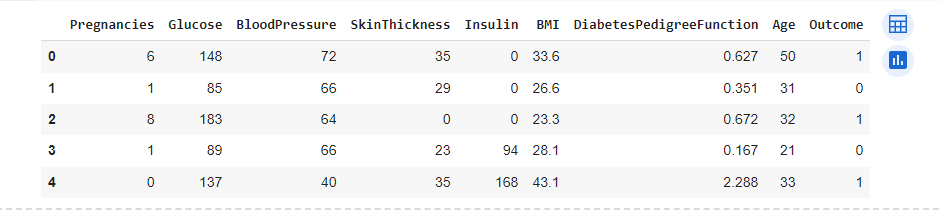
from sklearn.tree import DecisionTreeClassifier, plot\_tree

import matplotlib.pyplot as plt

import math

df = pd.read\_csv('/content/diabetes.csv')

df.head()

****

def calculate\_entropy(data, target\_column):

total\_rows = len(data)

target\_values = data[target\_column].unique()

entropy = 0

for value in target\_values:

# Calculate the proportion of instances with the current value

value\_count = len(data[data[target\_column] == value])

proportion = value\_count / total\_rows

entropy -= proportion \* math.log2(proportion)

return entropy

entropy\_outcome = calculate\_entropy(df, 'Outcome')

print(f"Entropy of the dataset: {entropy\_outcome}")



def calculate\_entropy(data, target\_column): # for each categorical variable

total\_rows = len(data)

target\_values = data[target\_column].unique()

entropy = 0

for value in target\_values:

# Calculate the proportion of instances with the current value

value\_count = len(data[data[target\_column] == value])

proportion = value\_count / total\_rows

entropy -= proportion \* math.log2(proportion) if proportion != 0 else 0

return entropy

def calculate\_information\_gain(data, feature, target\_column):

# Calculate weighted average entropy for the feature

unique\_values = data[feature].unique()

weighted\_entropy = 0

for value in unique\_values:

subset = data[data[feature] == value]

proportion = len(subset) / len(data)

weighted\_entropy += proportion \* calculate\_entropy(subset, target\_column)

# Calculate information gain

information\_gain = entropy\_outcome - weighted\_entropy

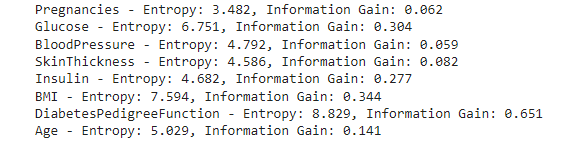
return information\_gain

for column in df.columns[:-1]:

entropy = calculate\_entropy(df, column)

information\_gain = calculate\_information\_gain(df, column, 'Outcome')

print(f"{column} - Entropy: {entropy:.3f}, Information Gain: {information\_gain:.3f}")



# Feature selection for the first step in making decision tree

selected\_feature = 'DiabetesPedigreeFunction'

# Create a decision tree

clf = DecisionTreeClassifier(criterion='entropy', max\_depth=1)

X = df[[selected\_feature]]

y = df['Outcome']

clf.fit(X, y)

plt.figure(figsize=(8, 6))

plot\_tree(clf, feature\_names=[selected\_feature], class\_names=['0', '1'], filled=True, rounded=True)

plt.show()

