**MACHINE INTELLIGENCE LABORATORY**

**WEEK – 3**

**NAME :** SANMAT SANJAYAKUMAR PAYAGOUDAR

**SRN** **:** PES1UG20CS385

**CODE :**

**PES1UG20CS385.py**

import numpy as np

import pandas as pd

import random

def get\_entropy\_of\_dataset(df):

 entropy = 0

 target = df[[df.columns[-1]]].values

 \_, counts = np.unique(target, return\_counts=True)

 total\_count = np.sum(counts)

 for freq in counts:

    temp = freq / total\_count

    if temp != 0:

        entropy -= temp \* (np.log2(temp))

 return entropy

def get\_avg\_info\_of\_attribute(df, attribute):

 attribute\_values = df[attribute].values

 unique\_attribute\_values = np.unique(attribute\_values)

 rows = df.shape[0]

 entropy\_of\_attribute = 0

 for current\_value in unique\_attribute\_values:

     df\_slice = df[df[attribute] == current\_value]

     target = df\_slice[[df\_slice.columns[-1]]].values

     \_, counts = np.unique(target, return\_counts=True)

     total\_count = np.sum(counts)

     entropy = 0

     for freq in counts:

        temp = freq / total\_count

        if temp != 0:

             entropy -= temp \* np.log2(temp)

     entropy\_of\_attribute += entropy \* (np.sum(counts) / rows)

 return (abs(entropy\_of\_attribute))

def get\_information\_gain(df, attribute):

 information\_gain = 0

 entropy\_of\_attribute = get\_avg\_info\_of\_attribute(df, attribute)

 entropy\_of\_dataset = get\_entropy\_of\_dataset(df)

 information\_gain = entropy\_of\_dataset - entropy\_of\_attribute

 return information\_gain

def get\_selected\_attribute(df):

 information\_gains = {}

 selected\_column = ''

 max\_information\_gain = float("-inf")

 for attribute in df.columns[:-1]:

    information\_gain\_of\_attribute = get\_information\_gain(df, attribute)

    if information\_gain\_of\_attribute > max\_information\_gain:

        selected\_column = attribute

        max\_information\_gain = information\_gain\_of\_attribute

    information\_gains[attribute] = information\_gain\_of\_attribute

 return (information\_gains, selected\_column)

**OUTPUT :**

