# assignment 4

#import the package numpy

import numpy as np

#import the package pandas

import pandas as pd

#import imputer method

from sklearn.preprocessing import Imputer

# import apriori and association\_rules

from mlxtend.frequesntpatternsfrom mlxtend.frequent\_patterns import apriori

from mlxtend.frequent\_patterns import association\_rules

# read the house-votes-84 dataset using method read\_csv

data = pd.read\_csv('house-votes-84.data', header=None)

# print the dimension of the data set

print('rows\*columns', data.shape)

# converting the dataset to binary

data = data.replace('y', 1)

data = data.replace('n', 0)

data = data.replace('republican', 1)

data = data.replace('democrat', 0)

data = data.replace('?', np.NaN)

imp = Imputer(missing\_values='NaN', strategy='most\_frequent', axis=0)

imp.fit(data)

data\_clean = imp.transform(data)

#Columns list

columns = ['republican','handicapped-infants','water-project-cost-sharing','adoptionof-the-budget-resolution',

'physician-fee-freeze','el-salvador-aid','eligious-groups-inschools','anti-satellite-test-ban'

,'aid-to-nicaraguan-contras','mx-missile','immigration','synfuelscorporation-cutback',

'education-spending','superfund-right-to-sue','crime','dutyfree-exports','export-administration-act-south-africa']

data1 = pd.DataFrame(np.array(data\_clean), columns=columns)

data1['Democrat'] = data1.republican.apply(lambda x: 0 if x == 1 else 1)

#aplly apriori algorithm to data set with minsup = 0.3

freq\_itemsets=apriori(data1,min\_support=0.3,use\_colnames=True)

#print frequent datasets

print(freq\_itemsets)

data2 = pd.DataFrame(data=freq\_itemsets)

#aplly apriori algorithm to data set with minsup = 0.9

ass\_rules = association\_rules(data2,metric="confidence", min\_threshold=0.9)

print(ass\_rules)











