IARE

INSTITUTE OF AERONAUTICAL ENGINEERING

(An Autonomous Institute affiliated to JNTUH, Hyderabad)
Dundigal, Hyderabad - 500 043

LABORATORY WORK BOOK

Name of the Student: HIMAKAR C

Class: CSE-B Semester: VI

Roll Number									
2	1	9	5	1	A	0	5	6	5

Course Code: ACIC08 Course Name: DMKD Laboratory

Name of the Course Faculty: Dr. D DURGA BHAVANI Faculty ID: IARE10927

Exercise Number: Week Number: 05 Date: 23/4/24

S. No.	Exercise Number	EXERCISE NAME	MARKS AWARDED					
			Aim/ Preparation	Algorithm / Procedure / Performance in the Lab	Source Code / Calculations and Graphs	Program Execution / Results and Error Analysis	Viva - Voce	Total
			4	4	4	4	4	20
1	5.1	Remove Rows/Attributes	4	4	4	4	4	20
2	5.2	Replace with mean or mode						
3	5.3	Discretion and normalization						
4								
5								
6								
7								
8								
9								
10								
11								
12								

Signature of the Student

Signature of the Fac

5.1 Remove rows/attributes

```
import pandas as pd
col = ['preg', 'plas', 'pres', 'skin', 'test', 'mass', 'pedi',
       'age', 'class']
d = pd.read_csv('pima-indians-diabetes.csv', names=col)
co = ['plas', 'pres', 'skin', 'test', 'mass']
d[co] = d[co].replace(0, None)
d = d.dropna()
print('Data after removing rows with missing
      values:')
print(d.head())
```

INPUT/OUTPUT:

Data after removing rows with missing values:

	preg	plas	pres	skin	test	mass	pedi	age	class
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
6	3	78	50	32	88	31.0	0.248	26	1
8	2	197	70	45	543	30.5	0.158	53	1
13	1	189	60	23	846	30.1	0.398	59	1

ROLL NUMBER :

5.2 Replace with mean or mode

Program that replaces missing values with mean or mode for numerical & categorical colums

```
import pandas as pd
cols = ['preg', 'plas', 'pres', 'skin', 'test', 'mass',
        'pedi', 'age', 'class']
d = pd.read_csv('pima-diabetes.csv', names=cols)
co = ['plas', 'pres', 'skin', 'test', 'mass']
d[co] = d[co].replace(0, None)
for c in d.columns:
    if c in co:
        d[c] = d[c].fillna(d[c].mean())
    else:
        d[c] = d[c].fillna(d[c].mode())
print('Data after replacing missing values:')
print(d.head())
```

Data after replacing missing values:

	Preg	Plas	Pres	skin	test	mass	Pedi	age	class
0	6	148.0	72.0	35.00	155.54	33.6	0.62	50	1
1	1	85.0	66.0	29.00	155.54	26.6	0.35	31	0
2	8	183.0	64.0	29.15	155.54	23.3	0.67	32	1
3	1	89.0	66.0	23.00	94.00	28.1	0.167	21	0
4	0	137.0	40.0	35.00	168.00	43.1	2.28	33	1

ROLL NUMBER :

5.3 Program to perform transformation of data using Discretion and normalization on given dataset.

```
import pandas as pd
from sklearn.preprocessing import KBinsDiscretizer, MinMaxScaler
cols = ['preg', 'plas', 'pres', 'skin', 'test', 'mass',
        'pedi', 'age', 'class']
d = pd.read_csv('pima-indians-diabetes.csv', names = cols)
dis = KBinsDiscretizer(n_bins=5, encode = 'ordinal',
                       strategy = 'quantile')
db = dis.fit_transform(d)
s = MinMaxScaler()
dn = s.fit_transform(db)
dn = pd.DataFrame(dn, columns = cols)
print('First few rows of transformed dataset: ')
print(dn.head())
```

First few rows of transformed dataset:

	Preg	Plas	Pres	Skin	test	mass	Pedi	age	class
0	0.75	1.00	0.50	1.00	0.0	0.50	0.75	1.00	0.0
1	0.25	0.00	0.25	0.66	0.0	0.25	0.50	0.50	0.0
2	1.00	1.00	0.25	0.00	0.0	0.00	0.75	0.50	0.0
3	0.25	0.00	0.25	0.33	0.5	0.25	0.00	0.00	0.0
4	0.00	0.75	0.00	1.00	1.0	1.00	1.00	0.75	0.0