```python
#importing libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import os

os.getcwd()

'C:\\Users\\pc'

path='C:\\Users\\pc\\downloads\\'
data=pd.read_csv(path+'chronickidneydisease.csv')
data.head(10)
```

```
   id   age     bp     sg   al   su       rbc        pc         pcc  \
0   0  48.0   80.0  1.020  1.0  0.0       NaN    normal  notpresent
1   1   7.0   50.0  1.020  4.0  0.0       NaN    normal  notpresent
2   2  62.0   80.0  1.010  2.0  3.0    normal    normal  notpresent
3   3  48.0   70.0  1.005  4.0  0.0    normal  abnormal     present
4   4  51.0   80.0  1.010  2.0  0.0    normal    normal  notpresent
5   5  60.0   90.0  1.015  3.0  0.0       NaN       NaN  notpresent
6   6  68.0   70.0  1.010  0.0  0.0       NaN    normal  notpresent
7   7  24.0    NaN  1.015  2.0  4.0    normal  abnormal  notpresent
8   8  52.0  100.0  1.015  3.0  0.0    normal  abnormal     present
9   9  53.0   90.0  1.020  2.0  0.0  abnormal  abnormal     present

           ba  ...  pcv     wc   rc  htn   dm  cad appet   pe  ane  \
0  notpresent  ...   44   7800  5.2  yes  yes   no  good   no   no
1  notpresent  ...   38   6000  NaN   no   no   no  good   no   no
2  notpresent  ...   31   7500  NaN   no  yes   no  poor   no  yes
3  notpresent  ...   32   6700  3.9  yes   no   no  poor  yes  yes
4  notpresent  ...   35   7300  4.6   no   no   no  good   no   no
5  notpresent  ...   39   7800  4.4  yes  yes   no  good  yes   no
6  notpresent  ...   36    NaN  NaN   no   no   no  good   no   no
7  notpresent  ...   44   6900    5   no  yes   no  good  yes   no
8  notpresent  ...   33   9600  4.0  yes  yes   no  good   no  yes
9  notpresent  ...   29  12100  3.7  yes  yes   no  poor   no  yes

   classification
0             ckd
1             ckd
2             ckd
3             ckd
4             ckd
5             ckd
6             ckd
7             ckd
8             ckd
9             ckd
```

```
[10 rows x 26 columns]

data.tail(10)

      id   age    bp     sg   al   su      rbc       pc         pcc
ba  \
390  390  52.0  80.0  1.025  0.0  0.0   normal   normal  notpresent
notpresent
391  391  36.0  80.0  1.025  0.0  0.0   normal   normal  notpresent
notpresent
392  392  57.0  80.0  1.020  0.0  0.0   normal   normal  notpresent
notpresent
393  393  43.0  60.0  1.025  0.0  0.0   normal   normal  notpresent
notpresent
394  394  50.0  80.0  1.020  0.0  0.0   normal   normal  notpresent
notpresent
395  395  55.0  80.0  1.020  0.0  0.0   normal   normal  notpresent
notpresent
396  396  42.0  70.0  1.025  0.0  0.0   normal   normal  notpresent
notpresent
397  397  12.0  80.0  1.020  0.0  0.0   normal   normal  notpresent
notpresent
398  398  17.0  60.0  1.025  0.0  0.0   normal   normal  notpresent
notpresent
399  399  58.0  80.0  1.025  0.0  0.0   normal   normal  notpresent
notpresent

     ...  pcv    wc   rc  htn  dm  cad  appet  pe ane classification
390  ...   52  6300  5.3   no  no   no   good  no  no         notckd
391  ...   44  5800  6.3   no  no   no   good  no  no         notckd
392  ...   46  6600  5.5   no  no   no   good  no  no         notckd
393  ...   54  7400  5.4   no  no   no   good  no  no         notckd
394  ...   45  9500  4.6   no  no   no   good  no  no         notckd
395  ...   47  6700  4.9   no  no   no   good  no  no         notckd
396  ...   54  7800  6.2   no  no   no   good  no  no         notckd
397  ...   49  6600  5.4   no  no   no   good  no  no         notckd
398  ...   51  7200  5.9   no  no   no   good  no  no         notckd
399  ...   53  6800  6.1   no  no   no   good  no  no         notckd

[10 rows x 26 columns]

data.shape

(400, 26)

data.columns=['id','age','blood_pressure','specific_gravity','albumin'
,'sugar','red_blood_cells',
          'pus_cell','pus_cell_clumps','bacteria','blood glucose
random',
```

```
'blood_urea','serum_creatinine','sodium','potassium','hemoglobin','pac
ked_cell_volume',
               'white_blood_cell_count',
'red_blood_cell_count' ,'hypertension','diabetesmellitus','coronary_ar
tery_disease',
               'apettite','pedal_edema','anemia','class']
```

data.columns

```
Index(['id', 'age', 'blood_pressure', 'specific_gravity', 'albumin',
'sugar',
       'red_blood_cells', 'pus_cell', 'pus_cell_clumps', 'bacteria',
       'blood glucose random', 'blood_urea', 'serum_creatinine',
'sodium',
       'potassium', 'hemoglobin', 'packed_cell_volume',
       'white_blood_cell_count', 'red_blood_cell_count',
'hypertension',
       'diabetesmellitus', 'coronary_artery_disease', 'apettite',
       'pedal_edema', 'anemia', 'class'],
      dtype='object')
```

data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 400 entries, 0 to 399
Data columns (total 26 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   id                       400 non-null    int64
 1   age                      391 non-null    float64
 2   blood_pressure           388 non-null    float64
 3   specific_gravity         353 non-null    float64
 4   albumin                  354 non-null    float64
 5   sugar                    351 non-null    float64
 6   red_blood_cells          248 non-null    object
 7   pus_cell                 335 non-null    object
 8   pus_cell_clumps          396 non-null    object
 9   bacteria                 396 non-null    object
 10  blood glucose random     356 non-null    float64
 11  blood_urea               381 non-null    float64
 12  serum_creatinine         383 non-null    float64
 13  sodium                   313 non-null    float64
 14  potassium                312 non-null    float64
 15  hemoglobin               348 non-null    float64
 16  packed_cell_volume       330 non-null    object
 17  white_blood_cell_count   295 non-null    object
 18  red_blood_cell_count     270 non-null    object
 19  hypertension             398 non-null    object
 20  diabetesmellitus         398 non-null    object
 21  coronary_artery_disease  398 non-null    object
 22  apettite                 399 non-null    object
```

```
 23  pedal_edema               399 non-null    object
 24  anemia                    399 non-null    object
 25  class                     400 non-null    object
dtypes: float64(11), int64(1), object(14)
memory usage: 81.4+ KB

data.drop(['id'],axis=1,inplace=True)
data
```

	age	blood_pressure	specific_gravity	albumin	sugar
					red_blood_cells
0	48.0	80.0	1.020	1.0	0.0 NaN
1	7.0	50.0	1.020	4.0	0.0 NaN
2	62.0	80.0	1.010	2.0	3.0 normal
3	48.0	70.0	1.005	4.0	0.0 normal
4	51.0	80.0	1.010	2.0	0.0 normal
..	...	...	...	...	... ...
395	55.0	80.0	1.020	0.0	0.0 normal
396	42.0	70.0	1.025	0.0	0.0 normal
397	12.0	80.0	1.020	0.0	0.0 normal
398	17.0	60.0	1.025	0.0	0.0 normal
399	58.0	80.0	1.025	0.0	0.0 normal

	pus_cell	pus_cell_clumps	bacteria	blood glucose random	...
0	normal	notpresent	notpresent	121.0	...
1	normal	notpresent	notpresent	NaN	...
2	normal	notpresent	notpresent	423.0	...
3	abnormal	present	notpresent	117.0	...
4	normal	notpresent	notpresent	106.0	...
..	...	...	...	...	...
395	normal	notpresent	notpresent	140.0	...

				...
396	normal	notpresent	notpresent	75.0 ...
397	normal	notpresent	notpresent	100.0 ...
398	normal	notpresent	notpresent	114.0 ...
399	normal	notpresent	notpresent	131.0 ...

	packed_cell_volume	white_blood_cell_count	red_blood_cell_count \
0	44	7800	5.2
1	38	6000	NaN
2	31	7500	NaN
3	32	6700	3.9
4	35	7300	4.6
..	...	...	...
395	47	6700	4.9
396	54	7800	6.2
397	49	6600	5.4
398	51	7200	5.9
399	53	6800	6.1

	hypertension	diabetesmellitus	coronary_artery_disease	apettite \
0	yes	yes	no	good
1	no	no	no	good
2	no	yes	no	poor
3	yes	no	no	poor
4	no	no	no	good
..	...	...	...	...

395	no	no	no	good
396	no	no	no	good
397	no	no	no	good
398	no	no	no	good
399	no	no	no	good

```
     pedal_edema anemia   class
0            no     no     ckd
1            no     no     ckd
2            no    yes     ckd
3           yes    yes     ckd
4            no     no     ckd
..          ...    ...     ...
395          no     no  notckd
396          no     no  notckd
397          no     no  notckd
398          no     no  notckd
399          no     no  notckd

[400 rows x 25 columns]
```

```python
#target column
data['class'].unique()
```

```
array(['ckd', 'ckd\t', 'notckd'], dtype=object)
```

```python
#rectify target column
data['class']=data['class'].replace('ckd\t','ckd')
data['class'].unique()
```

```
array(['ckd', 'notckd'], dtype=object)
```

```python
#fetching categorical column
cat=data.select_dtypes(include=['object']).columns.tolist()
cat
```

```
['red_blood_cells',
 'pus_cell',
 'pus_cell_clumps',
 'bacteria',
 'packed_cell_volume',
 'white_blood_cell_count',
 'red_blood_cell_count',
 'hypertension',
```

```python
 'diabetesmellitus',
 'coronary_artery_disease',
 'apettite',
 'pedal_edema',
 'anemia',
 'class']

#removing column which are not categorical
cat.remove('red_blood_cell_count')
cat.remove('packed_cell_volume')
cat.remove( 'white_blood_cell_count')
cat

['red_blood_cells',
 'pus_cell',
 'pus_cell_clumps',
 'bacteria',
 'hypertension',
 'diabetesmellitus',
 'coronary_artery_disease',
 'apettite',
 'pedal_edema',
 'anemia',
 'class']

num=data.select_dtypes(include=['float64']).columns.tolist()#fetch
numerical column

num.remove('specific_gravity')#remove which are not numerical
num.remove('albumin')
num.remove('sugar')
num

['age',
 'blood_pressure',
 'blood glucose random',
 'blood_urea',
 'serum_creatinine',
 'sodium',
 'potassium',
 'hemoglobin']

#adding column which is numerical
num.append('red_blood_cell_count')
num.append('packed_cell_volume')
num.append( 'white_blood_cell_count')
num

['age',
 'blood_pressure',
 'blood glucose random',
 'blood_urea',
```

```
 'serum_creatinine',
 'sodium',
 'potassium',
 'hemoglobin',
 'red_blood_cell_count',
 'packed_cell_volume',
 'white_blood_cell_count']

sns.pairplot(data,hue='class')
fig=plt.figure(figsize=(10,5))
fig
```

```
C:\Users\pc\anaconda3\lib\site-packages\seaborn\distributions.py:305:
UserWarning: Dataset has 0 variance; skipping density estimate.
  warnings.warn(msg, UserWarning)
C:\Users\pc\anaconda3\lib\site-packages\seaborn\distributions.py:305:
UserWarning: Dataset has 0 variance; skipping density estimate.
  warnings.warn(msg, UserWarning)

<Figure size 720x360 with 0 Axes>
```
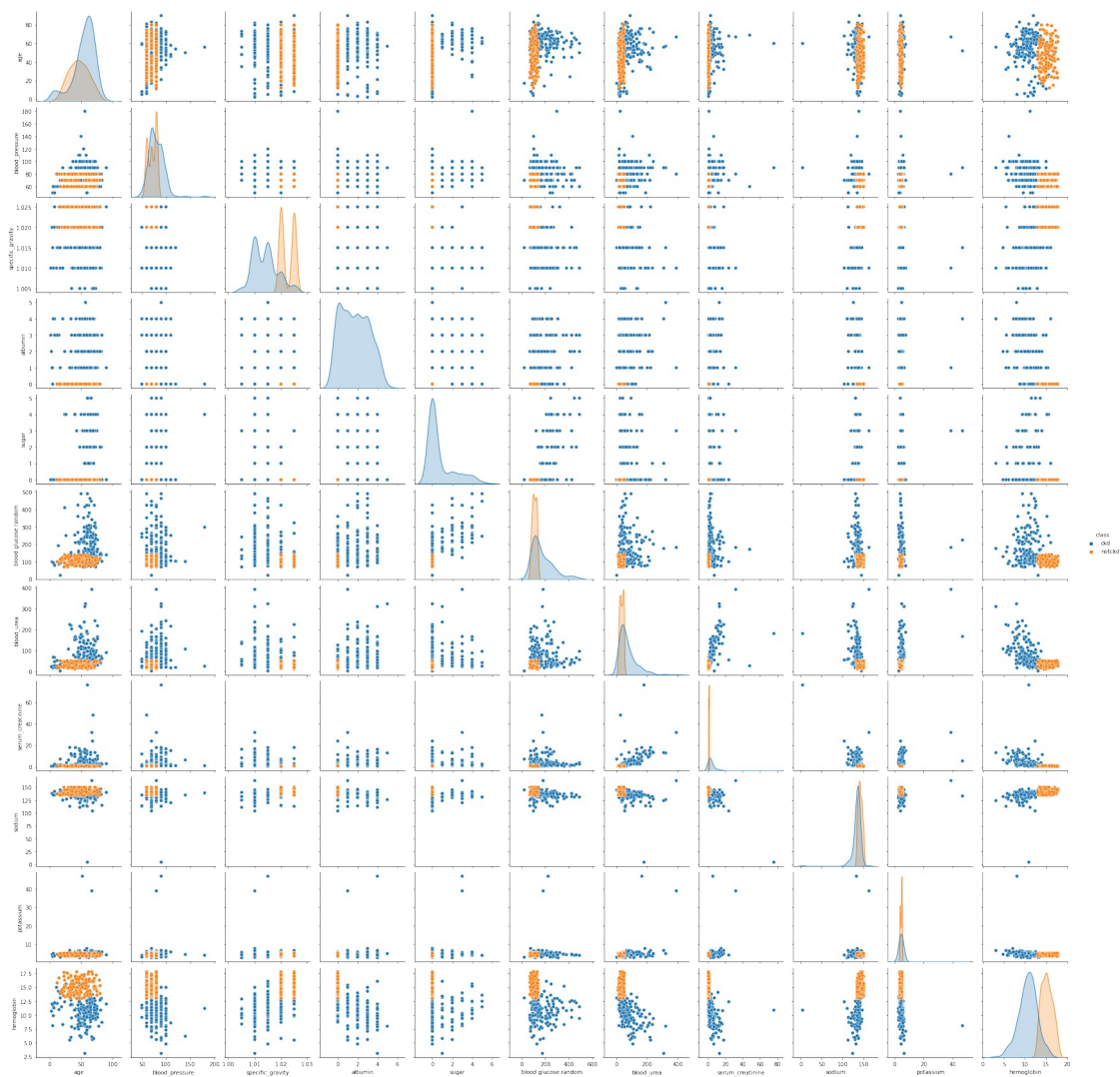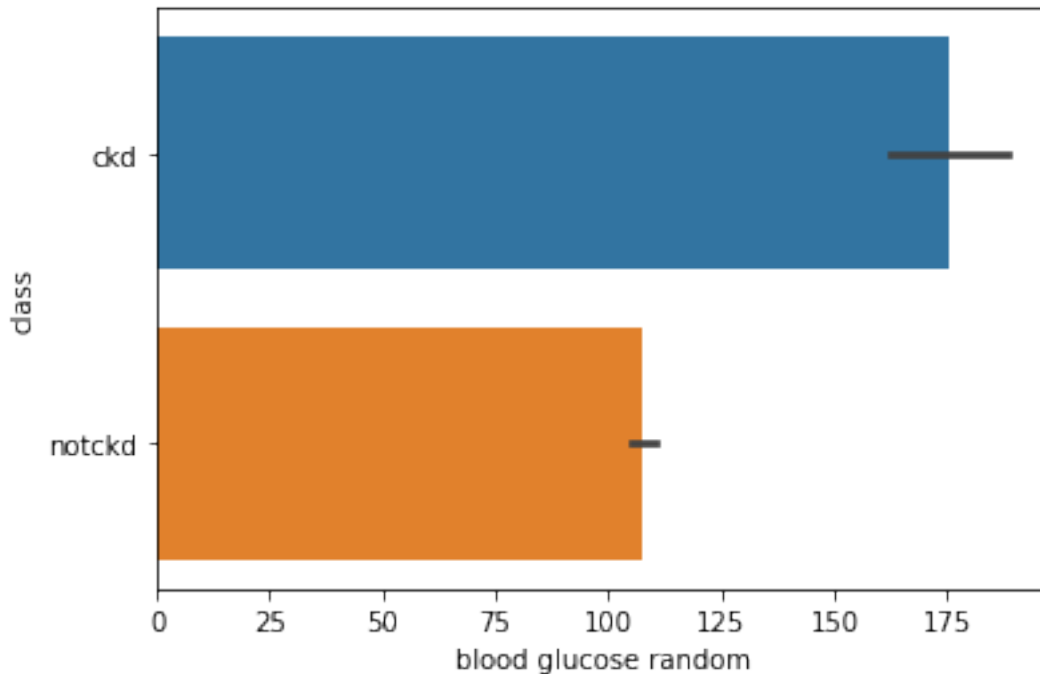
```
<Figure size 720x360 with 0 Axes>

sns.barplot(x='blood glucose random',y='class',data=data)

<AxesSubplot:xlabel='blood glucose random', ylabel='class'>
```

```
cat.append('specific_gravity')#adding column which is categorical
cat.append('albumin')
cat.append('sugar')
cat

['red_blood_cells',
 'pus_cell',
 'pus_cell_clumps',
 'bacteria',
 'hypertension',
 'diabetesmellitus',
 'coronary_artery_disease',
 'apettite',
 'pedal_edema',
 'anemia',
 'class',
 'specific_gravity',
 'albumin',
 'sugar']

a=data['coronary_artery_disease'].unique()
#b=data['sugar'].unique()
#c=data['albumin'].unique()
#d=data['specific_gravity'].unique()
#e=data['anemia'].unique()
#f=data['pedal_edema'].unique()#
#g=data['apettite'].unique()
h=data['diabetesmellitus'].unique()
```

```python
#i=data['bacteria'].unique()
#j=data['hypertension'].unique()#
#k=data['red_blood_cell_count'].unique()
#l=data['pus_cell'].unique()
#m=data['pus_cell_clumps'].unique()
```

```python
a,h
```

```
(array(['no', 'yes', '\tno', nan], dtype=object),
 array(['yes', 'no', ' yes', '\tno', '\tyes', nan], dtype=object))
```

```python
#rectifying the categorical column classes
data['coronary_artery_disease']=data['coronary_artery_disease'].replac
e('\tno','no')
data['coronary_artery_disease'].unique()
```

```
array(['no', 'yes', nan], dtype=object)
```

```python
data['diabetesmellitus']=data.diabetesmellitus.replace(to_replace={'ye
s':'yes','\tyes':'yes','\tno':'no'})
data['diabetesmellitus'].unique()
```

```
array(['yes', 'no', ' yes', nan], dtype=object)
```

```python
#handling missing value
data.isna().sum()
```

```
age                        9
blood_pressure            12
specific_gravity          47
albumin                   46
sugar                     49
red_blood_cells          152
pus_cell                  65
pus_cell_clumps            4
bacteria                   4
blood glucose random      44
blood_urea                19
serum_creatinine          17
sodium                    87
potassium                 88
hemoglobin                52
packed_cell_volume        70
white_blood_cell_count   105
red_blood_cell_count     130
hypertension               2
diabetesmellitus           2
coronary_artery_disease    2
apettite                   1
```

```
pedal_edema                    1
anemia                         1
class                          0
dtype: int64
```

```python
#before handling the numeric variable which is considered as string
should be convert to numerical
data.red_blood_cell_count=pd.to_numeric(data.red_blood_cell_count,erro
rs='coerce')
data.packed_cell_volume=pd.to_numeric(data.packed_cell_volume,errors='
coerce')
data.white_blood_cell_count=pd.to_numeric(data.white_blood_cell_count,
errors='coerce')
```

```python
#handle numerical column null values
data['blood_pressure'].fillna(data['blood_pressure'].mean(),inplace=Tr
ue)
data['blood_urea'].fillna(data['blood_urea'].mean(),inplace=True)
data['blood glucose random'].fillna(data['blood glucose
random'].mean(),inplace=True)
data['serum_creatinine'].fillna(data['serum_creatinine'].mean(),inplac
e=True)
data['sodium'].fillna(data['sodium'].mean(),inplace=True)
data['potassium'].fillna(data['potassium'].mean(),inplace=True)
data['hemoglobin'].fillna(data['hemoglobin'].mean(),inplace=True)

data['pus_cell'].fillna(data['pus_cell'].mode()[0],inplace=True)
data['age'].fillna(data['age'].mode()[0],inplace=True)
data['pus_cell_clumps'].fillna(data['pus_cell_clumps'].mode()
[0],inplace=True)
data['bacteria'].fillna(data['bacteria'].mode()[0],inplace=True)
data['red_blood_cell_count'].fillna(data['red_blood_cell_count'].mode(
)[0],inplace=True)
data['red_blood_cells'].fillna(data['red_blood_cells'].mode()
[0],inplace=True)
data['white_blood_cell_count'].fillna(data['white_blood_cell_count'].m
ode()[0],inplace=True)
data['packed_cell_volume'].fillna(data['packed_cell_volume'].mode()
[0],inplace=True)
data['hypertension'].fillna(data['hypertension'].mode()
[0],inplace=True)
data['diabetesmellitus'].fillna(data['diabetesmellitus'].mode()
[0],inplace=True)
data['coronary_artery_disease'].fillna(data['coronary_artery_disease']
.mode()[0],inplace=True)
data['apettite'].fillna(data['apettite'].mode()[0],inplace=True)
data['pedal_edema'].fillna(data['pedal_edema'].mode()[0],inplace=True)
data['anemia'].fillna(data['anemia'].mode()[0],inplace=True)
data['specific_gravity'].fillna(data['specific_gravity'].mode()
[0],inplace=True)
```

```python
data['albumin'].fillna(data['albumin'].mode()[0],inplace=True)
data['sugar'].fillna(data['sugar'].mode()[0],inplace=True)

data.isnull().sum()
```

```
age                          0
blood_pressure               0
specific_gravity             0
albumin                      0
sugar                        0
red_blood_cells              0
pus_cell                     0
pus_cell_clumps              0
bacteria                     0
blood glucose random         0
blood_urea                   0
serum_creatinine             0
sodium                       0
potassium                    0
hemoglobin                   0
packed_cell_volume           0
white_blood_cell_count       0
red_blood_cell_count         0
hypertension                 0
diabetesmellitus             0
coronary_artery_disease      0
apettite                     0
pedal_edema                  0
anemia                       0
class                        0
dtype: int64
```

```python
from sklearn.preprocessing import LabelEncoder
for i in cat:
    print('label of encoder= ',i)
    lei=LabelEncoder()
    print(data[i])
    data[i]=lei.fit_transform(data[i])
    print(data[i])
    print('*'*100)
```

```
label of encoder=  red_blood_cells
0       normal
1       normal
2       normal
3       normal
4       normal
```

```
         ...
395     normal
396     normal
397     normal
398     normal
399     normal
Name: red_blood_cells, Length: 400, dtype: object
0       1
1       1
2       1
3       1
4       1
        ..
395     1
396     1
397     1
398     1
399     1
Name: red_blood_cells, Length: 400, dtype: int32
****************************************************************
****************************
label of encoder=  pus_cell
0         normal
1         normal
2         normal
3       abnormal
4         normal
          ...
395       normal
396       normal
397       normal
398       normal
399       normal
Name: pus_cell, Length: 400, dtype: object
0       1
1       1
2       1
3       0
4       1
        ..
395     1
396     1
397     1
398     1
399     1
Name: pus_cell, Length: 400, dtype: int32
****************************************************************
****************************
label of encoder=  pus_cell_clumps
0       notpresent
```

```
1       notpresent
2       notpresent
3          present
4       notpresent
           ...
395     notpresent
396     notpresent
397     notpresent
398     notpresent
399     notpresent
Name: pus_cell_clumps, Length: 400, dtype: object
0       0
1       0
2       0
3       1
4       0
       ..
395     0
396     0
397     0
398     0
399     0
Name: pus_cell_clumps, Length: 400, dtype: int32
*************************************************************************
*****************************
label of encoder=  bacteria
0       notpresent
1       notpresent
2       notpresent
3       notpresent
4       notpresent
           ...
395     notpresent
396     notpresent
397     notpresent
398     notpresent
399     notpresent
Name: bacteria, Length: 400, dtype: object
0       0
1       0
2       0
3       0
4       0
       ..
395     0
396     0
397     0
398     0
399     0
Name: bacteria, Length: 400, dtype: int32
```

```
*******************************************************************
****************************
label of encoder=  hypertension
0       yes
1        no
2        no
3       yes
4        no
        ...
395      no
396      no
397      no
398      no
399      no
Name: hypertension, Length: 400, dtype: object
0        1
1        0
2        0
3        1
4        0
        ..
395      0
396      0
397      0
398      0
399      0
Name: hypertension, Length: 400, dtype: int32
*******************************************************************
****************************
label of encoder=  diabetesmellitus
0       yes
1        no
2       yes
3        no
4        no
        ...
395      no
396      no
397      no
398      no
399      no
Name: diabetesmellitus, Length: 400, dtype: object
0        2
1        1
2        2
3        1
4        1
        ..
395      1
396      1
```

```
397     1
398     1
399     1
Name: diabetesmellitus, Length: 400, dtype: int32
*******************************************************************
****************************
label of encoder=  coronary_artery_disease
0       no
1       no
2       no
3       no
4       no
        ..
395     no
396     no
397     no
398     no
399     no
Name: coronary_artery_disease, Length: 400, dtype: object
0       0
1       0
2       0
3       0
4       0
        ..
395     0
396     0
397     0
398     0
399     0
Name: coronary_artery_disease, Length: 400, dtype: int32
*******************************************************************
****************************
label of encoder=  apettite
0       good
1       good
2       poor
3       poor
4       good
        ...
395     good
396     good
397     good
398     good
399     good
Name: apettite, Length: 400, dtype: object
0       0
1       0
2       1
3       1
```

```
4        0
        ..
395      0
396      0
397      0
398      0
399      0
Name: apettite, Length: 400, dtype: int32
***********************************************************************
***************************
label of encoder=  pedal_edema
0        no
1        no
2        no
3       yes
4        no
        ...
395      no
396      no
397      no
398      no
399      no
Name: pedal_edema, Length: 400, dtype: object
0        0
1        0
2        0
3        1
4        0
        ..
395      0
396      0
397      0
398      0
399      0
Name: pedal_edema, Length: 400, dtype: int32
***********************************************************************
***************************
label of encoder=  anemia
0        no
1        no
2       yes
3       yes
4        no
        ...
395      no
396      no
397      no
398      no
399      no
Name: anemia, Length: 400, dtype: object
```

```
0       0
1       0
2       1
3       1
4       0
       ..
395     0
396     0
397     0
398     0
399     0
Name: anemia, Length: 400, dtype: int32
***********************************************************************
****************************
label of encoder=   class
0         ckd
1         ckd
2         ckd
3         ckd
4         ckd
         ...
395     notckd
396     notckd
397     notckd
398     notckd
399     notckd
Name: class, Length: 400, dtype: object
0       0
1       0
2       0
3       0
4       0
       ..
395     1
396     1
397     1
398     1
399     1
Name: class, Length: 400, dtype: int32
***********************************************************************
****************************
label of encoder=   specific_gravity
0       1.020
1       1.020
2       1.010
3       1.005
4       1.010
         ...
395     1.020
396     1.025
```

```
397    1.020
398    1.025
399    1.025
Name: specific_gravity, Length: 400, dtype: float64
0       3
1       3
2       1
3       0
4       1
       ..
395     3
396     4
397     3
398     4
399     4
Name: specific_gravity, Length: 400, dtype: int64
*******************************************************************
*****************************
label of encoder=  albumin
0       1.0
1       4.0
2       2.0
3       4.0
4       2.0
       ...
395     0.0
396     0.0
397     0.0
398     0.0
399     0.0
Name: albumin, Length: 400, dtype: float64
0       1
1       4
2       2
3       4
4       2
       ..
395     0
396     0
397     0
398     0
399     0
Name: albumin, Length: 400, dtype: int64
*******************************************************************
*****************************
label of encoder=  sugar
0       0.0
1       0.0
2       3.0
3       0.0
```

```
4       0.0
        ...
395     0.0
396     0.0
397     0.0
398     0.0
399     0.0
Name: sugar, Length: 400, dtype: float64
0       0
1       0
2       3
3       0
4       0
        ..
395     0
396     0
397     0
398     0
399     0
Name: sugar, Length: 400, dtype: int64
*****************************************************************
*****************************

data.corr().T

                        age   blood_pressure   specific_gravity
albumin  \
age                   1.000000      0.110293          -0.205802
0.171106
blood_pressure        0.110293      1.000000          -0.157111
0.119031
specific_gravity     -0.205802     -0.157111           1.000000 -
0.479962
albumin               0.171106      0.119031          -0.479962
1.000000
sugar                 0.177136      0.126987          -0.292053
0.287751
red_blood_cells      -0.017158     -0.131418           0.253894 -
0.394844
pus_cell             -0.117133     -0.147637           0.365353 -
0.561713
pus_cell_clumps       0.146163      0.047775          -0.306426
0.417868
bacteria              0.059716      0.104172          -0.231704
0.377935
blood glucose random  0.238305      0.136297          -0.329284
0.297452
blood_urea            0.282333      0.148290          -0.261826
0.282849
serum_creatinine      0.312846      0.213072          -0.421925
0.443853
```

sodium	-0.115827	-0.148548	0.330574	-0.365727
potassium	0.065573	0.068564	0.061206	0.052186
hemoglobin	0.001869	0.002287	-0.005032	-0.000883
packed_cell_volume	-0.254880	-0.209832	0.491157	-0.470819
white_blood_cell_count	0.176520	0.012344	-0.238735	0.214295
red_blood_cell_count	-0.199655	-0.164157	0.371683	-0.368210
hypertension	0.383346	0.162290	-0.323643	0.406057
diabetesmellitus	0.343494	0.188286	-0.352804	0.312978
coronary_artery_disease	0.231246	0.045068	-0.135814	0.200957
apettite	0.169656	0.113066	-0.230975	0.303145
pedal_edema	0.126072	0.061431	-0.253803	0.411080
anemia	0.052191	0.173970	-0.184155	0.229556
class	-0.314369	-0.261914	0.659504	-0.531562

	sugar	red_blood_cells	pus_cell	pus_cell_clumps
age	0.177136	-0.017158	-0.117133	0.146163
blood_pressure	0.126987	-0.131418	-0.147637	0.047775
specific_gravity	-0.292053	0.253894	0.365353	-0.306426
albumin	0.287751	-0.394844	-0.561713	0.417868
sugar	1.000000	-0.092940	-0.190062	0.168091
red_blood_cells	-0.092940	1.000000	0.377394	-0.102948
pus_cell	-0.190062	0.377394	1.000000	-0.520118
pus_cell_clumps	0.168091	-0.102948	-0.520118	1.000000
bacteria	0.119399	-0.184402	-0.330401	0.275082
blood glucose random	0.430407	-0.148463	-0.243161	0.156971
blood_urea	0.092368	-0.171132	-0.244872	

```
                              0.195091
serum_creatinine             0.197113          -0.226404 -0.345946
0.286231
sodium                      -0.182241           0.185515  0.270137          -
0.162873
potassium                    0.034984          -0.024983 -0.023251
0.030641
hemoglobin                   0.000183           0.000141  0.000928          -
0.003318
packed_cell_volume          -0.180658           0.278263  0.417339          -
0.297980
white_blood_cell_count       0.161132          -0.021104 -0.107753
0.158224
red_blood_cell_count        -0.149470           0.166161  0.365479          -
0.238772
hypertension                 0.254268          -0.140538 -0.291719
0.195623
diabetesmellitus             0.431277          -0.148374 -0.204596
0.167585
coronary_artery_disease      0.229301          -0.111493 -0.172295
0.188029
apettite                     0.069216          -0.160868 -0.274985
0.189688
pedal_edema                  0.116442          -0.199285 -0.350227
0.104356
anemia                       0.042464          -0.107625 -0.260566
0.175861
class                       -0.294555           0.282642  0.375154          -
0.265313
```

```
                         bacteria  blood glucose random  ...  \
age                      0.059716              0.238305  ...
blood_pressure           0.104172              0.136297  ...
specific_gravity        -0.231704             -0.329284  ...
albumin                  0.377935              0.297452  ...
sugar                    0.119399              0.430407  ...
red_blood_cells         -0.184402             -0.148463  ...
pus_cell                -0.330401             -0.243161  ...
pus_cell_clumps          0.275082              0.156971  ...
bacteria                 1.000000              0.122735  ...
blood glucose random     0.122735              1.000000  ...
blood_urea               0.190039              0.162540  ...
serum_creatinine         0.252775              0.322716  ...
sodium                  -0.173304             -0.186014  ...
potassium                0.069461              0.155887  ...
hemoglobin               0.000254              0.004052  ...
packed_cell_volume      -0.188624             -0.282095  ...
white_blood_cell_count   0.095702              0.176186  ...
red_blood_cell_count    -0.186593             -0.216739  ...
hypertension             0.089046              0.350230  ...
```

```
diabetesmellitus        0.081995              0.481638   ...
coronary_artery_disease 0.162395              0.135306   ...
apettite                0.149126              0.208179   ...
pedal_edema             0.134732              0.190206   ...
anemia                  0.052208              0.129143   ...
class                  -0.186871             -0.402788   ...

                        packed_cell_volume  white_blood_cell_count  \
age                            -0.254880                  0.176520
blood_pressure                 -0.209832                  0.012344
specific_gravity                0.491157                 -0.238735
albumin                        -0.470819                  0.214295
sugar                          -0.180658                  0.161132
red_blood_cells                 0.278263                 -0.021104
pus_cell                        0.417339                 -0.107753
pus_cell_clumps                -0.297980                  0.158224
bacteria                       -0.188624                  0.095702
blood glucose random           -0.282095                  0.176186
blood_urea                     -0.503641                  0.115036
serum_creatinine               -0.604186                  0.169639
sodium                          0.424787                 -0.089453
potassium                      -0.182086                 -0.072308
hemoglobin                      0.004826                  0.008046
packed_cell_volume              1.000000                 -0.193722
white_blood_cell_count         -0.193722                  1.000000
red_blood_cell_count            0.642322                 -0.094497
hypertension                   -0.565524                  0.135532
diabetesmellitus               -0.443637                  0.170342
coronary_artery_disease        -0.297873                  0.012900
apettite                       -0.374184                  0.165936
pedal_edema                    -0.382286                  0.172608
anemia                         -0.513150                  0.043373
class                           0.656471                 -0.290293

                        red_blood_cell_count  hypertension
diabetesmellitus  \
age                            -0.199655       0.383346
0.343494
blood_pressure                 -0.164157       0.162290
0.188286
specific_gravity                0.371683      -0.323643          -
0.352804
albumin                        -0.368210       0.406057
0.312978
sugar                          -0.149470       0.254268
0.431277
red_blood_cells                 0.166161      -0.140538          -
0.148374
pus_cell                        0.365479      -0.291719          -
0.204596
```

	red_blood_cell_count	hypertension	diabetesmellitus
pus_cell_clumps	-0.238772	0.195623	0.167585
bacteria	-0.186593	0.089046	0.081995
blood glucose random	-0.216739	0.350230	0.481638
blood_urea	-0.359463	0.498302	0.348893
serum_creatinine	-0.457078	0.598725	0.442709
sodium	0.260523	-0.396040	-0.317605
potassium	-0.126877	0.105932	0.123861
hemoglobin	0.008229	-0.001810	-0.000476
packed_cell_volume	0.642322	-0.565524	-0.443637
white_blood_cell_count	-0.094497	0.135532	0.170342
red_blood_cell_count	1.000000	-0.500994	-0.383711
hypertension	-0.500994	1.000000	0.589859
diabetesmellitus	-0.383711	0.589859	1.000000
coronary_artery_disease	-0.285634	0.325479	0.272335
apettite	-0.358803	0.345070	0.328017
pedal_edema	-0.286417	0.371026	0.311226
anemia	-0.366149	0.347802	0.186708
class	0.446612	-0.590438	-0.546786

	coronary_artery_disease	apettite	pedal_edema
age	0.231246	0.169656	0.126072
blood_pressure	0.045068	0.113066	0.061431
specific_gravity	-0.135814	-0.230975	-0.253803
albumin	0.200957	0.303145	0.411080
sugar	0.229301	0.069216	0.116442
red_blood_cells	-0.111493	-0.160868	-

0.199285
pus_cell   -0.172295 -0.274985   -
0.350227
pus_cell_clumps   0.188029 0.189688
0.104356
bacteria   0.162395 0.149126
0.134732
blood glucose random   0.135306 0.208179
0.190206
blood_urea   0.221482 0.241566
0.218813
serum_creatinine   0.333104 0.390556
0.297344
sodium   -0.182805 -0.225565   -
0.268777
potassium   0.132716 0.044508
0.053371
hemoglobin   -0.005071 0.005470
0.004343
packed_cell_volume   -0.297873 -0.374184   -
0.382286
white_blood_cell_count   0.012900 0.165936
0.172608
red_blood_cell_count   -0.285634 -0.358803   -
0.286417
hypertension   0.325479 0.345070
0.371026
diabetesmellitus   0.272335 0.328017
0.311226
coronary_artery_disease   1.000000 0.156104
0.172295
apettite   0.156104 1.000000
0.417055
pedal_edema   0.172295 0.417055
1.000000
anemia   0.047700 0.254942
0.207025
class   -0.236088 -0.393341   -
0.375154

	anemia	class
age	0.052191	-0.314369
blood_pressure	0.173970	-0.261914
specific_gravity	-0.184155	0.659504
albumin	0.229556	-0.531562
sugar	0.042464	-0.294555
red_blood_cells	-0.107625	0.282642
pus_cell	-0.260566	0.375154
pus_cell_clumps	0.175861	-0.265313
bacteria	0.052208	-0.186871

```
blood glucose random       0.129143 -0.402788
blood_urea                 0.342801 -0.404937
serum_creatinine           0.351231 -0.623903
sodium                    -0.217086  0.473949
potassium                  0.081940 -0.032281
hemoglobin                -0.003242 -0.011530
packed_cell_volume        -0.513150  0.656471
white_blood_cell_count     0.043373 -0.290293
red_blood_cell_count      -0.366149  0.446612
hypertension               0.347802 -0.590438
diabetesmellitus           0.186708 -0.546786
coronary_artery_disease    0.047700 -0.236088
apettite                   0.254942 -0.393341
pedal_edema                0.207025 -0.375154
anemia                     1.000000 -0.325396
class                     -0.325396  1.000000

[25 rows x 25 columns]

selcols=['red_blood_cells','pus_cell',
 'diabetesmellitus',
'coronary_artery_disease','blood_urea','pedal_edema','anemia',
 'blood glucose random']

x=pd.DataFrame(data,columns=selcols)
y=pd.DataFrame(data,columns=['class'])
print(x.shape)
print(y.shape)

(400, 8)
(400, 1)

from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,rando
m_state=2)
print(x_train.shape)
print(y_train.shape)
print(x_test.shape)
print(y_test.shape)

(320, 8)
(320, 1)
(80, 8)
(80, 1)

from sklearn.linear_model import LogisticRegression
lgr=LogisticRegression()
lgr.fit(x_train,y_train)

C:\Users\pc\anaconda3\lib\site-packages\sklearn\utils\
validation.py:72: DataConversionWarning: A column-vector y was passed
when a 1d array was expected. Please change the shape of y to
```

```
(n_samples, ), for example using ravel().
  return f(**kwargs)
C:\Users\pc\anaconda3\lib\site-packages\sklearn\linear_model\
_logistic.py:762: ConvergenceWarning: lbfgs failed to converge
(status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as
shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
  n_iter_i = _check_optimize_result(
```

LogisticRegression()

```
y_pred=lgr.predict(x_test)
y_pred1=lgr.predict([[140,45,0,0,0,0,0,0]])
print(y_pred1)
```

[1]

```
from sklearn.metrics import accuracy_score
acc=accuracy_score(y_test,y_pred)
acc
```

0.9125

```
from sklearn.metrics import confusion_matrix
conf_mat=confusion_matrix(y_test,y_pred)
conf_mat
```

```
array([[47,  7],
       [ 0, 26]], dtype=int64)
```

```
import pickle
pickle.dump(lgr,open('CKD.pkl','wb'))
```