In [5]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import GridSearchCV, train_test_split, cross_val_score
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import r2_score
import scipy.stats as stats
import seaborn as sns

%matplotlib inline
```

In [6]:
```
df = pd.read_csv('CKD dataset.csv')
data = df
data.head()
```

Out[6]:

	age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	...	wbcc	rbcc	htn	dm	cad	appet	pe	ane	class	Unname 2
0	NaN	48.0	80.0	1.020	1.0	0.0	NaN	normal	notpresent	notpresent	...	44.0	7800.0	5.2	yes	yes	no	good	no	no	ck
1	NaN	7.0	50.0	1.020	4.0	0.0	NaN	normal	notpresent	notpresent	...	38.0	6000.0	NaN	no	no	no	good	no	no	ck
2	NaN	62.0	80.0	1.010	2.0	3.0	normal	normal	notpresent	notpresent	...	31.0	7500.0	NaN	no	yes	no	poor	no	yes	ck
3	NaN	48.0	70.0	1.005	4.0	0.0	normal	abnormal	present	notpresent	...	32.0	6700.0	3.9	yes	no	no	poor	yes	yes	ck
4	NaN	51.0	80.0	1.010	2.0	0.0	normal	normal	notpresent	notpresent	...	35.0	7300.0	4.6	no	no	no	good	no	no	ck

5 rows x 26 columns

In [7]:
```
data.shape
```

Out[7]:
```
(400, 26)
```

In [7]:
```
data.shape
```

Out[7]:
```
(400, 26)
```

In [8]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 400 entries, 0 to 399
Data columns (total 26 columns):
age             0 non-null float64
bp              391 non-null float64
sg              388 non-null float64
al              353 non-null float64
su              354 non-null float64
rbc             351 non-null float64
pc              248 non-null object
pcc             335 non-null object
ba              396 non-null object
bgr             396 non-null object
bu              356 non-null float64
sc              381 non-null float64
sod             383 non-null float64
pot             313 non-null float64
hemo            312 non-null float64
pcv             348 non-null float64
wbcc            329 non-null float64
rbcc            294 non-null float64
htn             269 non-null float64
dm              398 non-null object
cad             398 non-null object
appet           398 non-null object
pe              399 non-null object
ane             399 non-null object
class           399 non-null object
Unnamed: 25     400 non-null object
dtypes: float64(15), object(11)
memory usage: 81.3+ KB
```

In [9]:
```
df.describe()
```

Out[9]:

	age	bp	sg	al	su	rbc	bu	sc	sod	pot	hemo	
count	0.0	391.000000	388.000000	353.000000	354.000000	351.000000	356.000000	381.000000	383.000000	313.000000	312.000000	348
mean	NaN	51.483376	76.469072	1.017408	1.016949	0.450142	148.036517	57.425722	3.072454	137.528754	4.627244	12
std	NaN	17.169714	13.683637	0.005717	1.352679	1.099191	79.281714	50.503006	5.741126	10.408752	3.193904	.
min	NaN	2.000000	50.000000	1.005000	0.000000	0.000000	22.000000	1.500000	0.400000	4.500000	2.500000	3
25%	NaN	42.000000	70.000000	1.010000	0.000000	0.000000	99.000000	27.000000	0.900000	135.000000	3.800000	10
50%	NaN	55.000000	80.000000	1.020000	0.000000	0.000000	121.000000	42.000000	1.300000	138.000000	4.400000	12
75%	NaN	64.500000	80.000000	1.020000	2.000000	0.000000	163.000000	66.000000	2.800000	142.000000	4.900000	15
max	NaN	90.000000	180.000000	1.025000	5.000000	5.000000	490.000000	391.000000	76.000000	163.000000	47.000000	17

In [7]:
```
data.shape
```

Out[7]:
```
(400, 26)
```

In [8]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 400 entries, 0 to 399
Data columns (total 26 columns):
age             0 non-null float64
bp              391 non-null float64
sg              388 non-null float64
al              353 non-null float64
su              354 non-null float64
rbc             351 non-null float64
pc              248 non-null object
pcc             335 non-null object
ba              396 non-null object
bgr             396 non-null object
bu              356 non-null float64
sc              381 non-null float64
sod             383 non-null float64
pot             313 non-null float64
hemo            312 non-null float64
pcv             348 non-null float64
wbcc            329 non-null float64
rbcc            294 non-null float64
htn             269 non-null float64
dm              398 non-null object
cad             398 non-null object
appet           398 non-null object
pe              399 non-null object
ane             399 non-null object
class           399 non-null object
Unnamed: 25     400 non-null object
dtypes: float64(15), object(11)
memory usage: 81.3+ KB
```

In [9]:
```
df.describe()
```

Out[9]:

	age	bp	sg	al	su	rbc	bu	sc	sod	pot	hemo	
count	0.0	391.000000	388.000000	353.000000	354.000000	351.000000	356.000000	381.000000	383.000000	313.000000	312.000000	348
mean	NaN	51.483376	76.469072	1.017408	1.016949	0.450142	148.036517	57.425722	3.072454	137.528754	4.627244	12
std	NaN	17.169714	13.683637	0.005717	1.352679	1.099191	79.281714	50.503006	5.741126	10.408752	3.193904	.
min	NaN	2.000000	50.000000	1.005000	0.000000	0.000000	22.000000	1.500000	0.400000	4.500000	2.500000	3
25%	NaN	42.000000	70.000000	1.010000	0.000000	0.000000	99.000000	27.000000	0.900000	135.000000	3.800000	10
50%	NaN	55.000000	80.000000	1.020000	0.000000	0.000000	121.000000	42.000000	1.300000	138.000000	4.400000	12
75%	NaN	64.500000	80.000000	1.020000	2.000000	0.000000	163.000000	66.000000	2.800000	142.000000	4.900000	15
max	NaN	90.000000	180.000000	1.025000	5.000000	5.000000	490.000000	391.000000	76.000000	163.000000	47.000000	17

In [10]:
```
df.isna().sum()
```

Out[10]:
```
age             400
bp                9
sg               12
al               47
su               46
rbc              49
pc              152
pcc              65
ba                4
bgr               4
bu               44
sc               19
sod              17
pot              87
hemo             88
pcv              52
wbcc             71
rbcc            106
htn             131
dm                2
cad               2
appet             2
pe                1
ane               1
class             1
Unnamed: 25       0
dtype: int64
```

In [ ]: