

Reading the Dataset

TIn [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import sklearn
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import MinMaxScaler

import joblib
import pickle
```

```
%matplotlib inline
```

Loading the Dataset

In [2]:

```
data=pd.read_csv("/content/chronickidneydisease.csv")
```

In [3]:

```
data.head()
```

Out[3]:

	i d	a g e	b p	sg	a l	s u	rb c	pc	pcc	ba	.	p c v	w c	rc	h t n	d m	c a d	ap pe t	p e	a n e	classif icatio n
0	0	48.0	8.0	1.02	1.0	0.0	NaN	normal	notpresent	notpresent	.	44	7800	5.2	yes	yes	no	good	no	no	ckd
1	1	70	50.0	1.02	4.0	0.0	NaN	normal	notpresent	notpresent	.	38	6000	NaN	no	no	no	good	no	no	ckd
2	2	62.0	8.0	1.01	2.0	3.0	normal	normal	notpresent	notpresent	.	31	7500	NaN	no	yes	no	poor	no	yes	ckd
3	3	48.0	7.0	1.005	4.0	0.0	normal	abnormal	present	notpresent	.	32	6700	3.9	yes	no	no	poor	yes	yes	ckd

	i	a	b	sg	a	s	rb	pc	pcc	ba	.	p	w	rc	h	d	c	ap	p	a	classif
	d	g	p		l	u	c				.	c	c		t	m	a	pe	e	n	ication
		e									.	v			n		d	t		e	n
4	4	5	8	1.	2	0	nor	nor	notp	notp	.	3	7	4.	n	n	n	go	n	n	ckd
		1.	0.	01	.	.	ma	mal	rese	rese	.	5	3	6	o	o	o	od	o	o	
		0	0	0	0	0	l		nt	nt	.		0								

5 rows × 26 columns

In [4]:
data.tail()

	i	a	b	sg	a	s	rb	pc	pcc	ba	.	p	w	rc	h	d	c	ap	p	a	classif
	d	g	p		l	u	c				.	c	c		t	m	a	pe	e	n	ication
		e									.	v			n		d	t		e	n
3	3	5	8	1.	0	0	nor	nor	notp	notp	.	4	6	4	n	n	n	go	n	n	notckd
9	9	5.	0.	02	.	.	ma	ma	rese	rese	.	7	7	.	o	o	o	od	o	o	
5	5	0	0	0	0	0	l	l	nt	nt	.		0	9							
											.		0								
3	3	4	7	1.	0	0	nor	nor	notp	notp	.	5	7	6	n	n	n	go	n	n	notckd
9	9	2.	0.	02	.	.	ma	ma	rese	rese	.	4	8	.	o	o	o	od	o	o	
6	6	0	0	5	0	0	l	l	nt	nt	.		0	2							
											.		0								
3	3	1	8	1.	0	0	nor	nor	notp	notp	.	4	6	5	n	n	n	go	n	n	notckd
9	9	2.	0.	02	.	.	ma	ma	rese	rese	.	9	6	.	o	o	o	od	o	o	
7	7	0	0	0	0	0	l	l	nt	nt	.		0	4							
											.		0								
3	3	1	6	1.	0	0	nor	nor	notp	notp	.	5	7	5	n	n	n	go	n	n	notckd
9	9	7.	0.	02	.	.	ma	ma	rese	rese	.	1	2	.	o	o	o	od	o	o	
8	8	0	0	5	0	0	l	l	nt	nt	.		0	9							
											.		0								
3	3	5	8	1.	0	0	nor	nor	notp	notp	.	5	6	6	n	n	n	go	n	n	notckd
9	9	8.	0.	02	.	.	ma	ma	rese	rese	.	3	8	.	o	o	o	od	o	o	
9	9	0	0	5	0	0	l	l	nt	nt	.		0	1							
											.		0								

5 rows × 26 columns

In [5]:
data.head(10)

Out[5]:

	id	age	bp	sg	al	su	Rbc	pc	pcc	ba	.	pcv	wc	rc	htn	dm	cad	appet	pe	ane	classification
	0	0	48.0	80.0	1.020	1.00	NaN	normal	notpresent	notpresent	.	44	7800	5.2	yes	yes	no	good	no	no	ckd
	1	1	70	50.0	1.020	4.00	NaN	normal	notpresent	notpresent	.	38	6000	NaN	no	no	no	good	no	no	ckd
	2	2	62.0	80.0	1.010	2.30	normal	normal	notpresent	notpresent	.	31	7500	NaN	no	yes	no	poor	no	yes	ckd
	3	3	48.0	70.0	1.005	4.00	normal	abnormal	present	notpresent	.	32	6700	3.9	yes	no	no	poor	yes	yes	ckd
	4	4	51.0	80.0	1.010	2.00	normal	normal	notpresent	notpresent	.	35	7300	4.6	no	no	no	good	no	no	ckd
	5	5	60.0	90.0	1.015	3.00	NaN	NaN	notpresent	notpresent	.	39	7800	4.4	yes	yes	no	good	yes	no	ckd
	6	6	68.0	70.0	1.010	0.00	NaN	normal	notpresent	notpresent	.	36	NaN	NaN	no	no	no	good	no	no	ckd
	7	7	24.0	NaN	1.015	2.40	normal	abnormal	notpresent	notpresent	.	44	6900	5	no	yes	no	good	yes	no	ckd
	8	8	52.0	100.0	1.015	3.00	normal	abnormal	present	notpresent	.	33	9600	4.0	yes	yes	no	good	no	yes	ckd
	9	9	53.0	90.0	1.020	2.00	abnormal	abnormal	present	notpresent	.	29	12100	3.7	yes	yes	no	poor	no	yes	ckd

10 rows × 26 columns

Drop id Column

In [6]:

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

Out[6]:

```
Index(['age', 'bp', 'sg', 'al', 'su', 'rbc', 'pc', 'pcc', 'ba', 'bgr', 'bu',
       'sc', 'sod', 'pot', 'hemo', 'pcv', 'wc', 'rc', 'htn', 'dm', 'cad',
       'appet', 'pe', 'ane', 'classification'],
      dtype='object')
```

Renaming the columns

In [7]:

```
data.columns=['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','appetite','pedal_edema','anemia','class']
data.columns
```

Out[7]:

```
Index(['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', 'appetite',
       'pedal_edema', 'anemia', 'class'],
      dtype='object')
```

In [8]:

```
data.info()
```

RangeIndex: 400 entries, 0 to 399

Data columns (total 25 columns):

#	Column	Non-Null Count	Dtype
0	age	391 non-null	float64
1	blood_pressure	388 non-null	float64
2	specific_gravity	353 non-null	float64
3	albumin	354 non-null	float64
4	sugar	351 non-null	float64
5	red_blood_cells	248 non-null	object
6	pus_cell	335 non-null	object
7	pus_cell_clumps	396 non-null	object
8	bacteria	396 non-null	object
9	blood glucose random	356 non-null	float64
10	blood_urea	381 non-null	float64
11	serum_creatinine	383 non-null	float64
12	sodium	313 non-null	float64
13	potassium	312 non-null	float64
14	hemoglobin	348 non-null	float64
15	packed_cell_volume	330 non-null	object
16	white_blood_cell_count	295 non-null	object
17	red_blood_cell_count	270 non-null	object

18	hypertension	398	non-null	object
19	diabetesmellitus	398	non-null	object
20	coronary_artery_disease	398	non-null	object
21	appetite	399	non-null	object
22	pedal_edema	399	non-null	object
23	anemia	399	non-null	object
24	class	400	non-null	object

dtypes: float64(11), object(14)
memory usage: 78.2+ KB