In [115...

import pandas as pd
import pickle
import warnings

warnings.filterwarnings('ignore')

In [141...

data=pd.read\_csv("C:\\Users\\reshma\_koduri\\OneDrive\\Documents\\archive (3)\\medica
data

Out[141...

	Student ID	Age	Gender	Height	Weight	Blood Type	ВМІ	Temperature	Heart Rate	Bl Press
0	1.0	18.0	Female	161.777924	72.354947	0	27.645835	NaN	95.0	1
1	2.0	NaN	Male	152.069157	47.630941	В	NaN	98.714977	93.0	1
2	3.0	32.0	Female	182.537664	55.741083	А	16.729017	98.260293	76.0	1
3	NaN	30.0	Male	182.112867	63.332207	В	19.096042	98.839605	99.0	1
4	5.0	23.0	Female	NaN	46.234173	0	NaN	98.480008	95.0	1
•••										
199995	NaN	24.0	Male	176.503260	95.756997	В	30.737254	99.170685	65.0	1
199996	99997.0	29.0	Female	163.917675	45.225194	NaN	16.831734	97.865785	62.0	1
199997	99998.0	34.0	Female	NaN	99.648914	NaN	33.189303	98.768210	60.0	
199998	99999.0	30.0	Female	156.446944	50.142824	Α	20.486823	98.994212	61.0	1
199999	100000.0	20.0	Female	153.927409	99.928405	0	42.175189	98.595817	95.0	1

200000 rows × 13 columns

a=data.tail(1800)

In [142...

Out[142...

	Student ID	Age	Gender	Height	Weight	Blood Type	ВМІ	Temperature	Heart Rate	Blo Press
198200	98201.0	18.0	Male	NaN	NaN	А	15.165705	99.296993	94.0	1(
198201	98202.0	18.0	Female	193.034462	71.347347	В	NaN	97.878775	77.0	1(
198202	98203.0	28.0	Male	184.871975	79.299260	0	23.202083	98.959156	63.0	1(
198203	NaN	28.0	NaN	198.473003	61.309481	В	15.564127	99.679238	71.0	12
198204	98205.0	18.0	Female	159.772722	43.143915	А	16.901073	99.305172	85.0	1.
•••			•••			•••				
199995	NaN	24.0	Male	176.503260	95.756997	В	30.737254	99.170685	65.0	12
199996	99997.0	29.0	Female	163.917675	45.225194	NaN	16.831734	97.865785	62.0	12
199997	99998.0	34.0	Female	NaN	99.648914	NaN	33.189303	98.768210	60.0	ć
199998	99999.0	30.0	Female	156.446944	50.142824	А	20.486823	98.994212	61.0	1(
199999	100000.0	20.0	Female	153.927409	99.928405	0	42.175189	98.595817	95.0	13

1800 rows × 13 columns

In [143...

data.tail(10)

Out[143...

	Student ID	Age	Gender	Height	Weight	Blood Type	ВМІ	Temperature	Heart Rate	Blo Press
199990	99991.0	21.0	Female	183.735110	51.172076	AB	15.158238	97.998790	67.0	ć
199991	99992.0	28.0	Male	183.499177	NaN	А	26.527962	97.321680	70.0	1.
199992	99993.0	34.0	Male	161.590030	90.877589	В	34.803881	98.728836	70.0	ć
199993	99994.0	22.0	Male	NaN	46.155224	А	NaN	98.331019	93.0	1(
199994	99995.0	22.0	Male	159.486907	NaN	А	27.631082	98.971976	86.0	13
199995	NaN	24.0	Male	176.503260	95.756997	В	30.737254	99.170685	65.0	12
199996	99997.0	29.0	Female	163.917675	45.225194	NaN	16.831734	97.865785	62.0	12
199997	99998.0	34.0	Female	NaN	99.648914	NaN	33.189303	98.768210	60.0	ć
199998	99999.0	30.0	Female	156.446944	50.142824	А	20.486823	98.994212	61.0	1(
199999	100000.0	20.0	Female	153.927409	99.928405	0	42.175189	98.595817	95.0	1;

In [144...

a.describe()

Out[144...

	Student ID	Age	Height	Weight	ВМІ	Temperature	Heart Rate
count	1633.000000	1627.000000	1604.000000	1614.000000	1626.000000	1618.000000	1619.000000
mean	99098.347214	26.249539	174.607063	69.947989	23.390458	98.600257	79.924027
std	518.285621	4.976123	14.791578	17.338499	7.259284	0.508875	11.491052
min	98201.000000	18.000000	150.000041	40.009476	10.458955	96.843803	60.000000
25%	98654.000000	22.000000	161.589501	54.486845	17.756278	98.254190	70.000000
50%	99097.000000	27.000000	174.619626	69.434012	22.626262	98.596611	80.000000
75%	99542.000000	31.000000	187.507729	84.790252	28.099608	98.950415	90.000000
max	100000.000000	34.000000	199.980696	99.980893	43.694294	100.587479	99.000000

In [145...

a.info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 1800 entries, 198200 to 199999

Data columns (total 13 columns):

#	Column	Non-Null Count	Dtype
0	Student ID	1633 non-null	float64
1	Age	1627 non-null	float64
2	Gender	1611 non-null	object
3	Height	1604 non-null	float64

```
4
                                1614 non-null
               Weight
                                                float64
           5
                                1650 non-null
               Blood Type
                                                object
                                1626 non-null
                                                float64
           6
           7
               Temperature
                               1618 non-null
                                                float64
               Heart Rate
                              1619 non-null
           8
                                                float64
           9
               Blood Pressure 1614 non-null
                                                float64
           10 Cholesterol
                              1603 non-null
                                                float64
           11 Diabetes
                                1625 non-null
                                                object
           12 Smoking
                                1637 non-null
                                                object
          dtypes: float64(9), object(4)
          memory usage: 182.9+ KB
In [146...
           print('Gender:',a['Gender'].unique())
           print('Blood Type:',a['Blood Type'].unique())
           print('Diabetes:',a['Diabetes'].unique())
           print('Smoking:',a['Smoking'].unique())
          Gender: ['Male' 'Female' nan]
          Blood Type: ['A' 'B' 'O' 'AB' nan]
          Diabetes: ['No' nan 'Yes']
          Smoking: [nan 'No' 'Yes']
In [147...
           a.isna().sum()
          Student ID
                             167
Out[147...
                             173
          Age
          Gender
                             189
          Height
                             196
          Weight
                             186
          Blood Type
                             150
          BMI
                             174
          Temperature
                             182
          Heart Rate
                             181
          Blood Pressure
                             186
          Cholesterol
                             197
          Diabetes
                             175
          Smoking
                             163
          dtype: int64
In [148...
           a['Gender'].unique()
          array(['Male', 'Female', nan], dtype=object)
Out[148...
In [149...
           a['Student ID']=a['Student ID'].fillna(a['Student ID'].mean())
           a['Age']=a['Age'].fillna(a['Age'].mean())
           a['Gender']=a['Gender'].fillna(a['Gender'].mode()[0])
           a['Height']=a['Height'].fillna(a['Height'].mean())
           a['Weight']=a['Weight'].fillna(a['Weight'].mean())
           a['Blood Type']=a['Blood Type'].fillna(a['Blood Type'].mode()[0])
           a['BMI']=a['BMI'].fillna(a['BMI'].mean())
           a['Temperature']=a['Temperature'].fillna(a['Temperature'].mean())
           a['Heart Rate']=a['Heart Rate'].fillna(a['Heart Rate'].mean())
           a['Blood Pressure']=a['Blood Pressure'].fillna(a['Blood Pressure'].median())
           a['Cholesterol']=a['Cholesterol'].fillna(a['Cholesterol'].mean())
           a['Diabetes']=a['Diabetes'].fillna(a['Diabetes'].mode()[0])
           a['Smoking']=a['Smoking'].fillna(a['Smoking'].mode()[0])
In [150...
           #a.fillna('unknown',inplace=True)
```

```
In [151...
             a.isna().sum()
            Student ID
                                 0
Out[151...
            Age
                                 0
            Gender
                                 0
            Height
                                 0
            Weight
                                 0
            Blood Type
                                 0
            Temperature
                                 0
            Heart Rate
                                 0
            Blood Pressure
                                 0
            Cholesterol
                                 0
            Diabetes
                                 0
            Smoking
                                 0
            dtype: int64
In [152...
             a['Gender']=a['Gender'].map({'Male':1, 'Female':0})
Out[152...
                                                                          Blood
                                                                                                          Heart
                        Student ID Age Gender
                                                      Height
                                                                 Weight
                                                                                       BMI Temperature
                                                                           Type
                                                                                                           Rate
                                                1 174.607063
            198200
                      98201.000000
                                    18.0
                                                                                                            94.0
                                                               69.947989
                                                                                15.165705
                                                                                               99.296993
            198201
                      98202.000000
                                    18.0
                                                   193.034462
                                                               71.347347
                                                                                 23.390458
                                                                                               97.878775
                                                                                                            77.0
                                    28.0
            198202
                      98203.000000
                                                   184.871975
                                                              79.299260
                                                                                 23.202083
                                                                                               98.959156
                                                                                                            63.0
                                                                              0
            198203
                      99098.347214
                                    28.0
                                                   198.473003
                                                               61.309481
                                                                                 15.564127
                                                                                               99.679238
                                                                                                            71.0
            198204
                      98205.000000
                                    18.0
                                                                                 16.901073
                                                                                                            85.0
                                                   159.772722
                                                              43.143915
                                                                                               99.305172
            199995
                      99098.347214
                                                   176.503260
                                                               95.756997
                                                                                 30.737254
                                                                                               99.170685
                                                                                                            65.0
                                    24.0
                                                                              В
            199996
                      99997.000000
                                    29.0
                                                   163.917675
                                                               45.225194
                                                                                 16.831734
                                                                                               97.865785
                                                                                                            62.0
            199997
                      99998.000000
                                    34.0
                                                   174.607063
                                                               99.648914
                                                                                 33.189303
                                                                                               98.768210
                                                                                                            60.0
            199998
                      99999.000000
                                    30.0
                                                   156.446944
                                                               50.142824
                                                                                 20.486823
                                                                                               98.994212
                                                                                                            61.0
            199999 100000.000000 20.0
                                                   153.927409
                                                               99.928405
                                                                                 42.175189
                                                                                                            95.0
                                                                                               98.595817
           1800 rows × 13 columns
In [153...
             a['Diabetes']=a['Diabetes'].map({'Yes':1,'No':0})
Out[153...
                                                                          Blood
                                                                                                          Heart
                        Student ID
                                    Age Gender
                                                      Height
                                                                 Weight
                                                                                       BMI
                                                                                           Temperature
                                                                           Type
                                                                                                           Rate
            198200
                      98201.000000
                                    18.0
                                                   174.607063
                                                               69.947989
                                                                                 15.165705
                                                                                               99.296993
                                                                                                            94.0
            198201
                      98202.000000
                                    18.0
                                                   193.034462
                                                               71.347347
                                                                                 23.390458
                                                                                               97.878775
                                                                                                            77.0
            198202
                      98203.000000
                                    28.0
                                                   184.871975
                                                              79.299260
                                                                                 23.202083
                                                                                               98.959156
                                                                                                            63.0
            198203
                      99098.347214
                                    28.0
                                                   198.473003
                                                               61.309481
                                                                                 15.564127
                                                                                               99.679238
                                                                                                            71.0
            198204
                      98205.000000
                                                 159.772722 43.143915
                                                                                16.901073
                                                                                               99.305172
                                                                                                            85.0
                                   18.0
```

	Student ID	Age	Gender	Height	Weight	Blood Type	ВМІ	Temperature	Heart Rate
•••						•••			•••
199995	99098.347214	24.0	1	176.503260	95.756997	В	30.737254	99.170685	65.0
199996	99997.000000	29.0	0	163.917675	45.225194	В	16.831734	97.865785	62.0
199997	99998.000000	34.0	0	174.607063	99.648914	В	33.189303	98.768210	60.0
199998	99999.000000	30.0	0	156.446944	50.142824	А	20.486823	98.994212	61.0
199999	100000.000000	20.0	0	153.927409	99.928405	0	42.175189	98.595817	95.0

1800 rows × 13 columns

```
In [154... a['Smoking']=a['Smoking'].map({'Yes':1,'No':0})
a
```

Out[154...

	Student ID	Age	Gender	Height	Weight	Blood Type	ВМІ	Temperature	Heart Rate
198200	98201.000000	18.0	1	174.607063	69.947989	А	15.165705	99.296993	94.0
198201	98202.000000	18.0	0	193.034462	71.347347	В	23.390458	97.878775	77.0
198202	98203.000000	28.0	1	184.871975	79.299260	0	23.202083	98.959156	63.0
198203	99098.347214	28.0	1	198.473003	61.309481	В	15.564127	99.679238	71.0
198204	98205.000000	18.0	0	159.772722	43.143915	Α	16.901073	99.305172	85.0
•••									
199995	99098.347214	24.0	1	176.503260	95.756997	В	30.737254	99.170685	65.0
199996	99997.000000	29.0	0	163.917675	45.225194	В	16.831734	97.865785	62.0
199997	99998.000000	34.0	0	174.607063	99.648914	В	33.189303	98.768210	60.0
199998	99999.000000	30.0	0	156.446944	50.142824	Α	20.486823	98.994212	61.0
199999	100000.000000	20.0	0	153.927409	99.928405	0	42.175189	98.595817	95.0

1800 rows × 13 columns

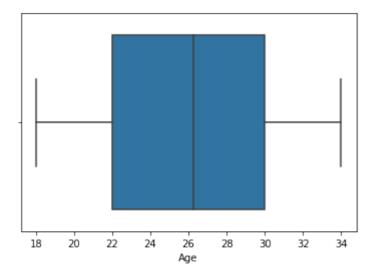
```
In [155...
#dummies = pd.get_dummies(a[['Gender', 'Blood Type' , 'Smoking']], drop_first = True
#data_dummies = pd.concat([a, dummies], axis = 1)
#data_dummies.head()
#data_dummies = data_dummies.drop(['Gender', 'Blood Type', 'Smoking'], axis=1)
#data_dummies.head()

In [156...
data.groupby(['Diabetes']).count()
```

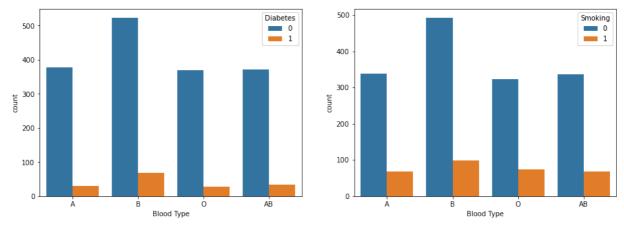
	1	D 1	Age Gende		Weight	Type	BMI	Tempe	rature	Hear Rate	
Diabetes	5										
No	14618	34 146	123 14611	4 145984	146166	146136	146118	1	146146	14607	1 1460
Yes	1622	23 16	227 1621	4 16268	16197	16220	16192		16239	16240	) 162
4											
a['Gen	ıder'].ı	ınique(	()								
array([	[1, 0],	dtype:	=int64)								
a.grou	ıpby(['(	Gender'	']).count(	()							
	Student	Ade	Height \	Neight	ood ype BMI	Tempe	erature	Heart Rate	Blood Pressure	Chol	esterol
Gender				·	•						
0	786	786	786	786	786 786		786	786	786		786
1	1014	1014	1014	1014 1	014 1014		1014	1014	1014		1014
4											
b=a.dr	op(['St	udent	ID'],axis	5=1)							
			ID'],axis								
c=pd.g	get_dumn	nies(b,		<b>:</b> )	t BN	11 Tem	perature	Heart Rate		i Ch	nolestero
c=pd.g c	get_dumn	nies(b	dtype=int	Weight			<b>perature</b> 9.296993	Rate	Pressu	ire Ch	nolestero
c=pd.g c	et_dumm	nies(b, ender	dtype=int Height	Weight	9 15.16570	5 9		<b>Rate</b> 94.0	Pressu	ire Ch	
c=pd.g c	Age G 18.0	ender	Height	Weight 69.947989 71.347347	) 15.16570 7 23.39045	5 9 8 9	9.296993	94.0 77.0	109 109	5.0 6.0	196.
c=pd.g c	Age G 18.0 18.0 28.0	ender	Height  174.607063  193.034462  184.871975	Weight 69.947989 71.347347	<ul><li>15.16570</li><li>23.39045</li><li>23.20208</li></ul>	5 9 8 9 3 9	9.296993 7.878775	94.0 77.0 63.0	109 109	5.0 6.0 9.0	196. 157.
c=pd.g c	Age G 18.0 18.0 28.0 28.0	ender  1  0  1	Height  174.607063  193.034462  184.871975	Weight 69.947989 71.347347 79.299260 61.309481	23.39045 23.20208 1 15.56412	5 9 8 9 3 9 7 9	9.296993 7.878775 8.959156	94.0 77.0 63.0 71.0	109 109 109 122	5.0 6.0 9.0	196. 157. 204.
c=pd·g c 198200 198201 198202 198203 198204	Age G  18.0  18.0  28.0  28.0  18.0	ender  1  0  1	Height 174.607063 193.034462 184.871975 198.473003 159.772722	Weight 69.947989 71.347347 79.299260 61.309481 43.143915	15.16570 7 23.39045 0 23.20208 1 15.56412 5 16.90107	5 9 8 9 3 9 7 9 3 9	9.296993 7.878775 8.959156 9.679238 9.305172	94.0 77.0 63.0 71.0 85.0	109 109 109 123 118	5.0 5.0 9.0 2.0 3.0	196. 157. 204. 224. 166.
c=pd·g c 198200 198201 198202 198203 198204 	Age G  18.0  18.0  28.0  18.0  28.0  18.0  24.0	nies(b) ender  1 0 1 0 1	Height 174.607063 193.034462 184.871975 198.473003 159.772722 176.503260	Weight 69.947989 71.347347 79.299260 61.309481 43.143915  95.756997	15.16570 23.39045 23.20208 15.56412 16.90107 30.73725	5 9 8 9 3 9 7 9 3 9 	9.296993 7.878775 8.959156 9.679238 9.305172 	94.0 77.0 63.0 71.0 85.0 	109 109 109 123 118	5.0 5.0 9.0 9.0 3.0	196. 157. 204. 224. 166.
c=pd·g c  198200 198201 198202 198203 198204 199995	Age G  18.0  18.0  28.0  18.0  24.0  29.0	nies(b) ender  1 0 1 0 1 0	Height 174.607063 193.034462 184.871975 198.473003 159.772722 176.503260 163.917675	Weight 69.947989 71.347347 79.299260 61.309481 43.143915 95.756997 45.225194	15.16570 23.39045 23.20208 15.56412 16.90107 30.73725 16.83173	5 9 8 9 3 9 7 9 3 9  4 9	9.296993 7.878775 8.959156 9.679238 9.305172  9.170685	94.0 77.0 63.0 71.0 85.0  65.0	109 109 109 123 118 129	5.0 5.0 9.0 2.0 3.0 	196. 157. 204. 224. 166. 130. 198.
c=pd·g c  198200 198201 198202 198203 198204 199995 199996 199997	Age G  18.0  18.0  28.0  28.0  18.0  29.0  34.0	1 0 1 1 0 1 0 0	Height 174.607063 193.034462 184.871975 198.473003 159.772722 176.503260 163.917675 174.607063	Weight 69.947989 71.347347 79.299260 61.309481 43.143915 95.756997 45.225194 99.648914	15.16570 23.39045 23.20208 15.56412 16.90107 30.73725 16.83173 133.18930	5 9 8 9 3 9 7 9 3 9  4 9 4 9	9.296993 7.878775 8.959156 9.679238 9.305172  9.170685 7.865785	94.0 77.0 63.0 71.0 85.0  65.0 62.0	109 109 123 118 129 90	5.0 5.0 6.0 9.0 2.0 3.0 	196. 157. 204. 224. 166. 130. 198.
c=pd·g c  198200 198201 198202 198203 198204 199995	Age G  18.0  18.0  28.0  28.0  18.0  24.0  29.0  34.0  30.0	1 0 1 1 0 1 0 0 0 0	Height 174.607063 193.034462 184.871975 198.473003 159.772722 176.503260 163.917675	Weight 69.947989 71.347347 79.299260 61.309481 43.143915 95.756997 45.225194 99.648914 50.142824	15.16570 23.39045 23.20208 15.56412 16.90107 30.73725 16.83173 133.18930 20.48682	5 9 8 9 3 9 7 9 3 9 4 9 4 9 3 9	9.296993 7.878775 8.959156 9.679238 9.305172  9.170685	94.0 77.0 63.0 71.0 85.0 65.0 62.0 60.0 61.0	109 109 123 118 129 100	5.0 5.0 6.0 9.0 2.0 3.0  1.0 5.0 0.0	196. 157. 204. 224. 166. 130. 198.

```
import seaborn as sb
import matplotlib.pyplot as plt
sb.boxplot(b.Age)
```

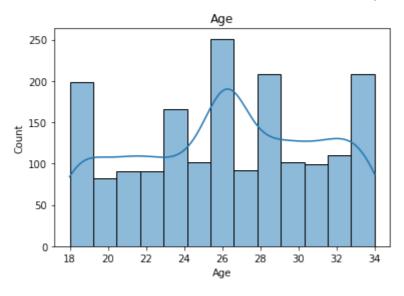
Out[166... <AxesSubplot:xlabel='Age'>



```
fig, ax = plt.subplots(1,2, figsize=(15,5))
sb.countplot(x='Blood Type', hue='Diabetes', data=b, ax=ax[0])
sb.countplot(x='Blood Type', hue='Smoking', data=b, ax=ax[1])
fig.show()
```



```
sb.histplot(a, x='Age', kde=True)
plt.title('Age')
plt.show()
```



```
In [171...
           y=c['Diabetes']
           198200
                      0
Out[171...
           198201
                      0
           198202
                     0
           198203
                     0
           198204
                      0
           199995
                     0
           199996
           199997
                      0
           199998
                      0
           199999
           Name: Diabetes, Length: 1800, dtype: int64
```

In [172...

x=c.drop(['Diabetes'],axis=1)
x

Out[172...

	Age	Gender	Height	Weight	ВМІ	Temperature	Heart Rate	Blood Pressure	Cholesterol
198200	18.0	1	174.607063	69.947989	15.165705	99.296993	94.0	105.0	196.0
198201	18.0	0	193.034462	71.347347	23.390458	97.878775	77.0	106.0	157.0
198202	28.0	1	184.871975	79.299260	23.202083	98.959156	63.0	109.0	204.0
198203	28.0	1	198.473003	61.309481	15.564127	99.679238	71.0	122.0	224.0
198204	18.0	0	159.772722	43.143915	16.901073	99.305172	85.0	118.0	166.0
•••									
199995	24.0	1	176.503260	95.756997	30.737254	99.170685	65.0	121.0	130.0
199996	29.0	0	163.917675	45.225194	16.831734	97.865785	62.0	125.0	198.0
199997	34.0	0	174.607063	99.648914	33.189303	98.768210	60.0	90.0	154.0
199998	30.0	0	156.446944	50.142824	20.486823	98.994212	61.0	106.0	225.0
199999	20.0	0	153.927409	99.928405	42.175189	98.595817	95.0	133.0	132.0

```
In [173...
  from sklearn.model_selection import train_test_split
  x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.33,random_state=42)
In [174...
  print(x_train.shape, y_train.shape)
  (1206, 14) (1206,)
In [175...
  from sklearn.linear_model import LogisticRegression
  classifier=LogisticRegression()
  classifier.fit(x_train, y_train)
Out[175...
  ▼ LogisticRegression
  LogisticRegression()
In [176...
  ypred=classifier.predict(x_test)
  ypred
  Out[176...
    dtype=int64)
In [177...
  from sklearn.metrics import confusion matrix
  confusion matrix(y test,ypred)
  array([[545,
      0],
Out[177...
      0]], dtype=int64)
    [ 49,
```

```
In [178...
   from sklearn.metrics import accuracy_score
   accuracy_score(y_test,ypred)
   0.9175084175084175
Out[178...
In [179...
   ypred
   Out[179...
     dtype=int64)
In [181...
   from sklearn.model selection import GridSearchCV
   from sklearn.ensemble import RandomForestClassifier
   reg=RandomForestClassifier()
   n estimators=[25,50,75,100,125,150,175,200]
   criterion=['gini','entropy']
   max_depth=[3,5,10]
   parameters={'n estimators': n estimators,'criterion':criterion,'max depth':max depth
   rfc reg = GridSearchCV(reg, parameters)
   rfc_reg.fit(x_train,y_train)
       GridSearchCV
Out[181...
   ▶ estimator: RandomForestClassifier
     RandomForestClassifier
In [182...
   rfc_reg.best_params_
   {'criterion': 'gini', 'max depth': 3, 'n estimators': 25}
Out[182...
```

```
In [184...
  reg=RandomForestClassifier(n_estimators=25,criterion='gini',max_depth=10)
  reg.fit(x_train,y_train)
Out[184...
      RandomForestClassifier
  RandomForestClassifier(max_depth=10, n_estimators=25)
In [185...
  ypred=reg.predict(x_test)
  ypred
  Out[185...
    0,
      0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
    0, 0,
      0.
      0,
      0,
      0,
       0,
        0,
         0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
    dtype=int64)
In [186...
  from sklearn.metrics import accuracy_score
  accuracy_score(y_test,ypred)
  0.9158249158249159
Out[186...
In [ ]:
```