FannyLo_KNN

April 10, 2023

0.0.1 Importing Libraries

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import sys
```

0.0.2 Reading In Data

[2]:
```
data=pd.read_csv('fetal_health.csv')
data
```

[2]:

	baseline value	accelerations	fetal_movement	uterine_contractions	\
0	120.0	0.000	0.000	0.000	
1	132.0	0.006	0.000	0.006	
2	133.0	0.003	0.000	0.008	
3	134.0	0.003	0.000	0.008	
4	132.0	0.007	0.000	0.008	
...	...	...	...	...	
2121	140.0	0.000	0.000	0.007	
2122	140.0	0.001	0.000	0.007	
2123	140.0	0.001	0.000	0.007	
2124	140.0	0.001	0.000	0.006	
2125	142.0	0.002	0.002	0.008	

	light_decelerations	severe_decelerations	prolongued_decelerations	\
0	0.000	0.0	0.0	
1	0.003	0.0	0.0	
2	0.003	0.0	0.0	
3	0.003	0.0	0.0	
4	0.000	0.0	0.0	
...	...	...	...	
2121	0.000	0.0	0.0	
2122	0.000	0.0	0.0	
2123	0.000	0.0	0.0	
2124	0.000	0.0	0.0	

```
2125                0.000                   0.0                         0.0

      abnormal_short_term_variability  mean_value_of_short_term_variability  \
0                                73.0                                   0.5
1                                17.0                                   2.1
2                                16.0                                   2.1
3                                16.0                                   2.4
4                                16.0                                   2.4
...                               ...                                   ...
2121                             79.0                                   0.2
2122                             78.0                                   0.4
2123                             79.0                                   0.4
2124                             78.0                                   0.4
2125                             74.0                                   0.4

      percentage_of_time_with_abnormal_long_term_variability  ...  \
0                                                    43.0      ...
1                                                     0.0      ...
2                                                     0.0      ...
3                                                     0.0      ...
4                                                     0.0      ...
...                                                   ...      ...
2121                                                 25.0      ...
2122                                                 22.0      ...
2123                                                 20.0      ...
2124                                                 27.0      ...
2125                                                 36.0      ...

      histogram_min  histogram_max  histogram_number_of_peaks  \
0              62.0          126.0                        2.0
1              68.0          198.0                        6.0
2              68.0          198.0                        5.0
3              53.0          170.0                       11.0
4              53.0          170.0                        9.0
...             ...            ...                        ...
2121          137.0          177.0                        4.0
2122          103.0          169.0                        6.0
2123          103.0          170.0                        5.0
2124          103.0          169.0                        6.0
2125          117.0          159.0                        2.0

      histogram_number_of_zeroes  histogram_mode  histogram_mean  \
0                            0.0           120.0           137.0
1                            1.0           141.0           136.0
2                            1.0           141.0           135.0
3                            0.0           137.0           134.0
4                            0.0           137.0           136.0
```

```
...                          ...             ...             ...
2121                         0.0           153.0           150.0
2122                         0.0           152.0           148.0
2123                         0.0           153.0           148.0
2124                         0.0           152.0           147.0
2125                         1.0           145.0           143.0

      histogram_median  histogram_variance  histogram_tendency  fetal_health
0                121.0                73.0                 1.0           2.0
1                140.0                12.0                 0.0           1.0
2                138.0                13.0                 0.0           1.0
3                137.0                13.0                 1.0           1.0
4                138.0                11.0                 1.0           1.0
...                ...                 ...                 ...           ...
2121             152.0                 2.0                 0.0           2.0
2122             151.0                 3.0                 1.0           2.0
2123             152.0                 4.0                 1.0           2.0
2124             151.0                 4.0                 1.0           2.0
2125             145.0                 1.0                 0.0           1.0

[2126 rows x 22 columns]
```

Converting 3 class problem into binary classification problem Originally: 1 = Normal ; 2 = Suspect; 3 = Pathological

Modification: 0 = non-pathological (previously 1 and 2); 1 = pathological

[3]:
```
data['fetal_health'] = data['fetal_health'].replace(1.0,0)
data['fetal_health'] = data['fetal_health'].replace(2.0,0)
data['fetal_health'] = data['fetal_health'].replace(3.0,1)
data
```

[3]:
```
      baseline value  accelerations  fetal_movement  uterine_contractions  \
0              120.0          0.000           0.000                 0.000
1              132.0          0.006           0.000                 0.006
2              133.0          0.003           0.000                 0.008
3              134.0          0.003           0.000                 0.008
4              132.0          0.007           0.000                 0.008
...              ...            ...             ...                   ...
2121           140.0          0.000           0.000                 0.007
2122           140.0          0.001           0.000                 0.007
2123           140.0          0.001           0.000                 0.007
2124           140.0          0.001           0.000                 0.006
2125           142.0          0.002           0.002                 0.008

      light_decelerations  severe_decelerations  prolongued_decelerations  \
0                   0.000                   0.0                       0.0
```

1	0.003	0.0	0.0
2	0.003	0.0	0.0
3	0.003	0.0	0.0
4	0.000	0.0	0.0
...	...	...	...
2121	0.000	0.0	0.0
2122	0.000	0.0	0.0
2123	0.000	0.0	0.0
2124	0.000	0.0	0.0
2125	0.000	0.0	0.0

	abnormal_short_term_variability	mean_value_of_short_term_variability	\
0	73.0	0.5	
1	17.0	2.1	
2	16.0	2.1	
3	16.0	2.4	
4	16.0	2.4	
...	...	...	
2121	79.0	0.2	
2122	78.0	0.4	
2123	79.0	0.4	
2124	78.0	0.4	
2125	74.0	0.4	

	percentage_of_time_with_abnormal_long_term_variability	...	\
0	43.0	...	
1	0.0	...	
2	0.0	...	
3	0.0	...	
4	0.0	...	
...	...	...	
2121	25.0	...	
2122	22.0	...	
2123	20.0	...	
2124	27.0	...	
2125	36.0	...	

	histogram_min	histogram_max	histogram_number_of_peaks	\
0	62.0	126.0	2.0	
1	68.0	198.0	6.0	
2	68.0	198.0	5.0	
3	53.0	170.0	11.0	
4	53.0	170.0	9.0	
...	...	...	...	
2121	137.0	177.0	4.0	
2122	103.0	169.0	6.0	
2123	103.0	170.0	5.0	

```
2124          103.0          169.0                       6.0
2125          117.0          159.0                       2.0

      histogram_number_of_zeroes  histogram_mode  histogram_mean  \
0                            0.0           120.0           137.0
1                            1.0           141.0           136.0
2                            1.0           141.0           135.0
3                            0.0           137.0           134.0
4                            0.0           137.0           136.0
...                          ...             ...             ...
2121                         0.0           153.0           150.0
2122                         0.0           152.0           148.0
2123                         0.0           153.0           148.0
2124                         0.0           152.0           147.0
2125                         1.0           145.0           143.0

      histogram_median  histogram_variance  histogram_tendency  fetal_health
0                121.0                73.0                 1.0           0.0
1                140.0                12.0                 0.0           0.0
2                138.0                13.0                 0.0           0.0
3                137.0                13.0                 1.0           0.0
4                138.0                11.0                 1.0           0.0
...                ...                 ...                 ...           ...
2121             152.0                 2.0                 0.0           0.0
2122             151.0                 3.0                 1.0           0.0
2123             152.0                 4.0                 1.0           0.0
2124             151.0                 4.0                 1.0           0.0
2125             145.0                 1.0                 0.0           0.0

[2126 rows x 22 columns]
```

[4]:
```
# to confirm that we have 0 and 1
data.iloc[0:10]
```

[4]:
```
   baseline value  accelerations  fetal_movement  uterine_contractions  \
0           120.0          0.000             0.0                 0.000
1           132.0          0.006             0.0                 0.006
2           133.0          0.003             0.0                 0.008
3           134.0          0.003             0.0                 0.008
4           132.0          0.007             0.0                 0.008
5           134.0          0.001             0.0                 0.010
6           134.0          0.001             0.0                 0.013
7           122.0          0.000             0.0                 0.000
8           122.0          0.000             0.0                 0.002
9           122.0          0.000             0.0                 0.003

   light_decelerations  severe_decelerations  prolongued_decelerations  \
```

0	0.000	0.0	0.000
1	0.003	0.0	0.000
2	0.003	0.0	0.000
3	0.003	0.0	0.000
4	0.000	0.0	0.000
5	0.009	0.0	0.002
6	0.008	0.0	0.003
7	0.000	0.0	0.000
8	0.000	0.0	0.000
9	0.000	0.0	0.000

	abnormal_short_term_variability	mean_value_of_short_term_variability	\
0	73.0	0.5	
1	17.0	2.1	
2	16.0	2.1	
3	16.0	2.4	
4	16.0	2.4	
5	26.0	5.9	
6	29.0	6.3	
7	83.0	0.5	
8	84.0	0.5	
9	86.0	0.3	

	percentage_of_time_with_abnormal_long_term_variability	...	histogram_min	\
0	43.0	...	62.0	
1	0.0	...	68.0	
2	0.0	...	68.0	
3	0.0	...	53.0	
4	0.0	...	53.0	
5	0.0	...	50.0	
6	0.0	...	50.0	
7	6.0	...	62.0	
8	5.0	...	62.0	
9	6.0	...	62.0	

	histogram_max	histogram_number_of_peaks	histogram_number_of_zeroes	\
0	126.0	2.0	0.0	
1	198.0	6.0	1.0	
2	198.0	5.0	1.0	
3	170.0	11.0	0.0	
4	170.0	9.0	0.0	
5	200.0	5.0	3.0	
6	200.0	6.0	3.0	
7	130.0	0.0	0.0	
8	130.0	0.0	0.0	
9	130.0	1.0	0.0	

```
   histogram_mode  histogram_mean  histogram_median  histogram_variance  \
0           120.0           137.0             121.0                73.0
1           141.0           136.0             140.0                12.0
2           141.0           135.0             138.0                13.0
3           137.0           134.0             137.0                13.0
4           137.0           136.0             138.0                11.0
5            76.0           107.0             107.0               170.0
6            71.0           107.0             106.0               215.0
7           122.0           122.0             123.0                 3.0
8           122.0           122.0             123.0                 3.0
9           122.0           122.0             123.0                 1.0

   histogram_tendency  fetal_health
0                 1.0           0.0
1                 0.0           0.0
2                 0.0           0.0
3                 1.0           0.0
4                 1.0           0.0
5                 0.0           1.0
6                 0.0           1.0
7                 1.0           1.0
8                 1.0           1.0
9                 1.0           1.0

[10 rows x 22 columns]
```

[5]:
```
X = data.drop('fetal_health', axis = 1)
X
```

[5]:
```
      baseline value  accelerations  fetal_movement  uterine_contractions  \
0              120.0          0.000           0.000                 0.000
1              132.0          0.006           0.000                 0.006
2              133.0          0.003           0.000                 0.008
3              134.0          0.003           0.000                 0.008
4              132.0          0.007           0.000                 0.008
...              ...            ...             ...                   ...
2121           140.0          0.000           0.000                 0.007
2122           140.0          0.001           0.000                 0.007
2123           140.0          0.001           0.000                 0.007
2124           140.0          0.001           0.000                 0.006
2125           142.0          0.002           0.002                 0.008

      light_decelerations  severe_decelerations  prolongued_decelerations  \
0                   0.000                   0.0                       0.0
1                   0.003                   0.0                       0.0
2                   0.003                   0.0                       0.0
3                   0.003                   0.0                       0.0
```

```
4                    0.000                    0.0                         0.0
...                    ...                    ...                         ...
2121                 0.000                    0.0                         0.0
2122                 0.000                    0.0                         0.0
2123                 0.000                    0.0                         0.0
2124                 0.000                    0.0                         0.0
2125                 0.000                    0.0                         0.0

      abnormal_short_term_variability  mean_value_of_short_term_variability  \
0                                73.0                                   0.5
1                                17.0                                   2.1
2                                16.0                                   2.1
3                                16.0                                   2.4
4                                16.0                                   2.4
...                               ...                                   ...
2121                             79.0                                   0.2
2122                             78.0                                   0.4
2123                             79.0                                   0.4
2124                             78.0                                   0.4
2125                             74.0                                   0.4

      percentage_of_time_with_abnormal_long_term_variability  ...  \
0                                                    43.0      ...
1                                                     0.0      ...
2                                                     0.0      ...
3                                                     0.0      ...
4                                                     0.0      ...
...                                                    ...     ...
2121                                                 25.0      ...
2122                                                 22.0      ...
2123                                                 20.0      ...
2124                                                 27.0      ...
2125                                                 36.0      ...

      histogram_width  histogram_min  histogram_max  \
0                64.0           62.0          126.0
1               130.0           68.0          198.0
2               130.0           68.0          198.0
3               117.0           53.0          170.0
4               117.0           53.0          170.0
...               ...            ...            ...
2121             40.0          137.0          177.0
2122             66.0          103.0          169.0
2123             67.0          103.0          170.0
2124             66.0          103.0          169.0
2125             42.0          117.0          159.0
```

```
      histogram_number_of_peaks  histogram_number_of_zeroes  histogram_mode  \
0                           2.0                         0.0           120.0
1                           6.0                         1.0           141.0
2                           5.0                         1.0           141.0
3                          11.0                         0.0           137.0
4                           9.0                         0.0           137.0
...                         ...                         ...             ...
2121                        4.0                         0.0           153.0
2122                        6.0                         0.0           152.0
2123                        5.0                         0.0           153.0
2124                        6.0                         0.0           152.0
2125                        2.0                         1.0           145.0

      histogram_mean  histogram_median  histogram_variance  histogram_tendency
0              137.0             121.0                73.0                 1.0
1              136.0             140.0                12.0                 0.0
2              135.0             138.0                13.0                 0.0
3              134.0             137.0                13.0                 1.0
4              136.0             138.0                11.0                 1.0
...              ...               ...                 ...                 ...
2121           150.0             152.0                 2.0                 0.0
2122           148.0             151.0                 3.0                 1.0
2123           148.0             152.0                 4.0                 1.0
2124           147.0             151.0                 4.0                 1.0
2125           143.0             145.0                 1.0                 0.0

[2126 rows x 21 columns]
```

[6]:
```
# Known output Y = classification result
y = data['fetal_health']
y
```

[6]:
```
0       0.0
1       0.0
2       0.0
3       0.0
4       0.0
       ...
2121    0.0
2122    0.0
2123    0.0
2124    0.0
2125    0.0
Name: fetal_health, Length: 2126, dtype: float64
```

[7]:
```
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,␣
↪random_state=142)
x_train
```

[7]:

	baseline value	accelerations	fetal_movement	uterine_contractions	\
199	120.0	0.000	0.013	0.000	
340	133.0	0.000	0.000	0.000	
440	142.0	0.001	0.003	0.001	
1427	144.0	0.006	0.000	0.004	
1015	139.0	0.007	0.000	0.005	
...	...	...	...	...	
1420	142.0	0.006	0.000	0.007	
1616	144.0	0.003	0.049	0.002	
1050	125.0	0.007	0.000	0.005	
511	154.0	0.007	0.001	0.002	
277	123.0	0.001	0.000	0.004	

	light_decelerations	severe_decelerations	prolongued_decelerations	\
199	0.001	0.0	0.000	
340	0.000	0.0	0.000	
440	0.002	0.0	0.000	
1427	0.000	0.0	0.000	
1015	0.000	0.0	0.000	
...	...	...	...	
1420	0.000	0.0	0.000	
1616	0.006	0.0	0.001	
1050	0.000	0.0	0.000	
511	0.000	0.0	0.000	
277	0.000	0.0	0.000	

	abnormal_short_term_variability	mean_value_of_short_term_variability	\
199	53.0	0.7	
340	75.0	0.2	
440	55.0	1.3	
1427	39.0	1.0	
1015	38.0	0.9	
...	...	...	
1420	42.0	1.0	
1616	66.0	3.4	
1050	26.0	1.3	
511	45.0	0.8	
277	54.0	0.5	

	percentage_of_time_with_abnormal_long_term_variability	...	\
199	7.0	...	
340	91.0	...	
440	10.0	...	

```
1427                                                5.0       ...
1015                                                0.0       ...
...                                                 ...       ...
1420                                                0.0       ...
1616                                                0.0       ...
1050                                                0.0       ...
511                                                 0.0       ...
277                                                 9.0       ...

      histogram_width  histogram_min  histogram_max  \
199              77.0           56.0          133.0
340               7.0          131.0          138.0
440             115.0           52.0          167.0
1427             43.0          136.0          179.0
1015             34.0          136.0          170.0
...               ...            ...            ...
1420             97.0           74.0          171.0
1616            113.0           67.0          180.0
1050             87.0           79.0          166.0
511              47.0          142.0          189.0
277              29.0          116.0          145.0

      histogram_number_of_peaks  histogram_number_of_zeroes  histogram_mode  \
199                         6.0                         0.0           123.0
340                         1.0                         0.0           133.0
440                        12.0                         3.0           148.0
1427                        1.0                         0.0           157.0
1015                        1.0                         0.0           144.0
...                         ...                         ...             ...
1420                        6.0                         0.0           148.0
1616                        7.0                         0.0           141.0
1050                        4.0                         0.0           131.0
511                         2.0                         1.0           161.0
277                         4.0                         0.0           126.0

      histogram_mean  histogram_median  histogram_variance  histogram_tendency
199            121.0             123.0                 3.0                 1.0
340            134.0             135.0                 0.0                 0.0
440            143.0             147.0                17.0                 1.0
1427           157.0             157.0                 4.0                 0.0
1015           146.0             146.0                 4.0                -1.0
...              ...               ...                 ...                 ...
1420           148.0             149.0                 5.0                 1.0
1616           110.0             147.0                89.0                 1.0
1050           133.0             132.0                11.0                 0.0
511            166.0             165.0                10.0                 0.0
277            128.0             129.0                 3.0                 0.0
```

```
[1700 rows x 21 columns]
```

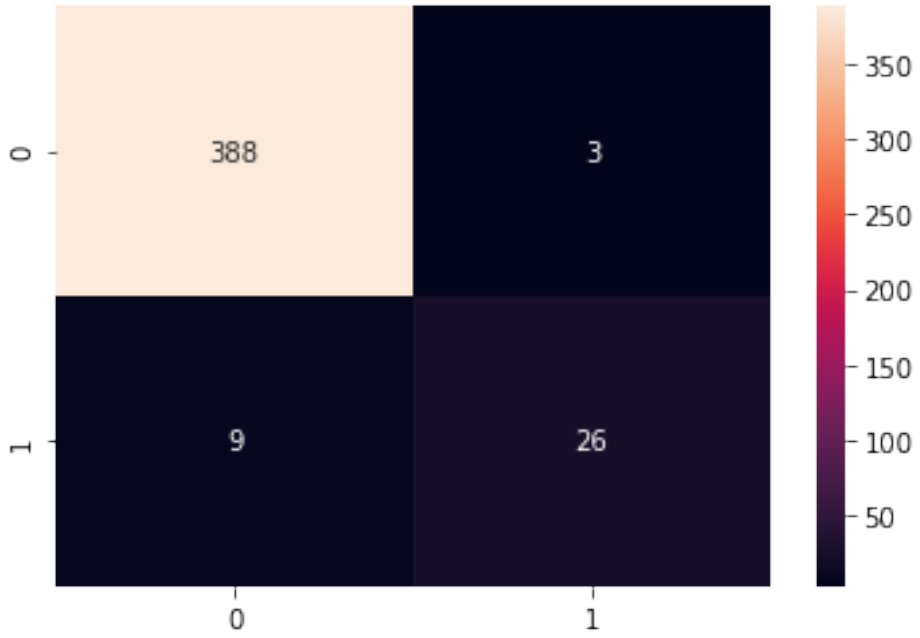
0.0.3 Training with KNN with 3 neighbors

[8]:
```
from sklearn.neighbors import KNeighborsClassifier
neigh = KNeighborsClassifier(n_neighbors=3)
```

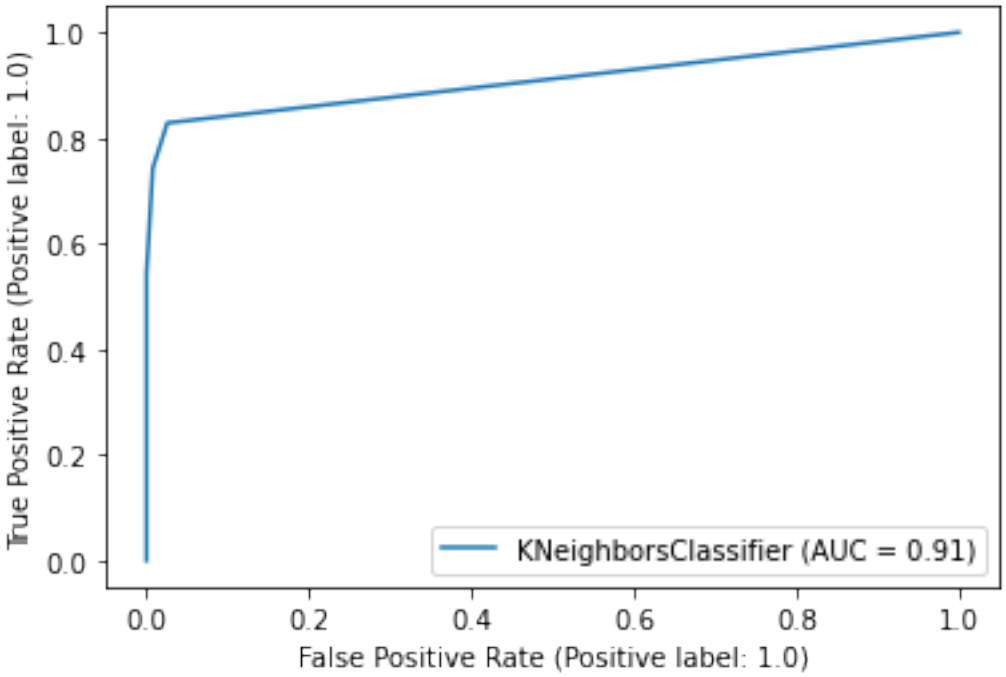
[9]:
```
neigh.fit(x_train,y_train)
y_pred_test = neigh.predict(x_test)
```

[10]:
```
#confusion matrix
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score

cm = confusion_matrix(y_test, y_pred_test)

sns.heatmap(cm, annot=True,annot_kws={"size": 10},fmt="d")
print(accuracy_score(y_test, y_pred_test))
```

```
0.971830985915493
```



[11]:
```
#ROC plot
from sklearn import  metrics
metrics.plot_roc_curve(neigh, x_test, y_test)
plt.show()
```

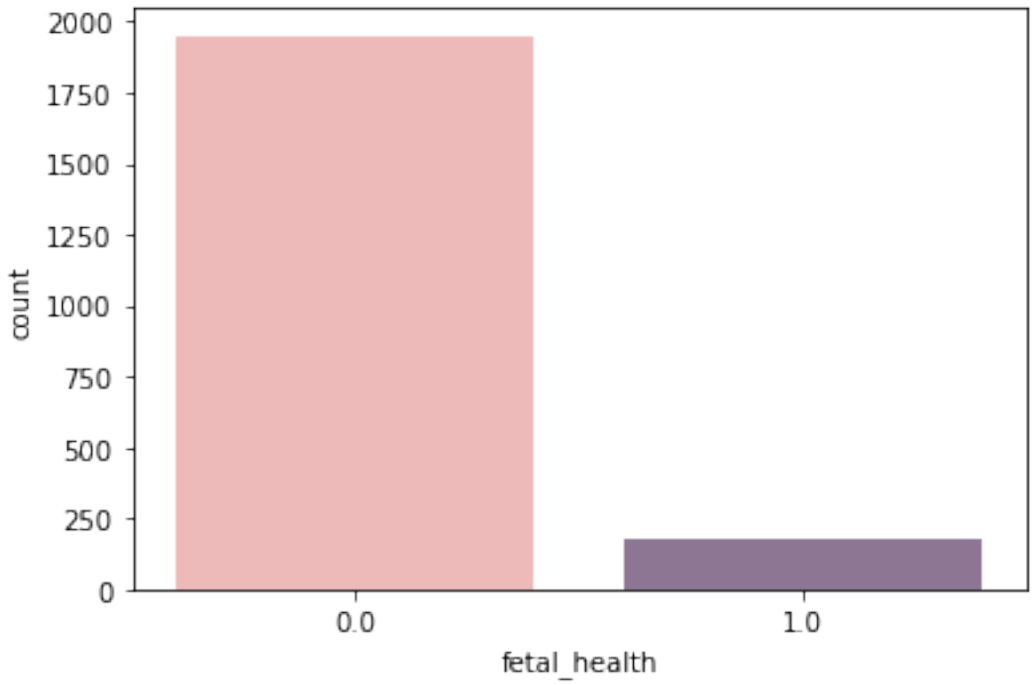


[12]:
```
from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred_test))
```

```
              precision    recall  f1-score   support

         0.0       0.98      0.99      0.98       391
         1.0       0.90      0.74      0.81        35

    accuracy                           0.97       426
   macro avg       0.94      0.87      0.90       426
weighted avg       0.97      0.97      0.97       426
```

[13]:
```
#Because the merging of two groups, our data is not balanced as a result. We␣
 ↪have way more class 0 than class 1
colours=["#f7b2b0","#8f7198", "#003f5c"]
sns.countplot(data= data, x="fetal_health",palette=colours)
```

[13]: <AxesSubplot:xlabel='fetal_health', ylabel='count'>



0.0.4 Class Imbalance

We have issue with class imabalance which can hinder our accuracy of model. Here are some methods that we can try - Undersampling of class 0 - Upsampling of class 1

Undersampling

[14]:
```
#undersampling
class_0 = data[data['fetal_health'] == 0]
class_0
```

[14]:

	baseline value	accelerations	fetal_movement	uterine_contractions	\
0	120.0	0.000	0.000	0.000	
1	132.0	0.006	0.000	0.006	
2	133.0	0.003	0.000	0.008	
3	134.0	0.003	0.000	0.008	
4	132.0	0.007	0.000	0.008	
...	...	...	...	...	
2121	140.0	0.000	0.000	0.007	
2122	140.0	0.001	0.000	0.007	
2123	140.0	0.001	0.000	0.007	
2124	140.0	0.001	0.000	0.006	
2125	142.0	0.002	0.002	0.008	

	light_decelerations	severe_decelerations	prolongued_decelerations	\

```
0                    0.000                    0.0                         0.0
1                    0.003                    0.0                         0.0
2                    0.003                    0.0                         0.0
3                    0.003                    0.0                         0.0
4                    0.000                    0.0                         0.0
...                    ...                    ...                         ...
2121                 0.000                    0.0                         0.0
2122                 0.000                    0.0                         0.0
2123                 0.000                    0.0                         0.0
2124                 0.000                    0.0                         0.0
2125                 0.000                    0.0                         0.0

      abnormal_short_term_variability  mean_value_of_short_term_variability  \
0                                73.0                                   0.5
1                                17.0                                   2.1
2                                16.0                                   2.1
3                                16.0                                   2.4
4                                16.0                                   2.4
...                               ...                                   ...
2121                             79.0                                   0.2
2122                             78.0                                   0.4
2123                             79.0                                   0.4
2124                             78.0                                   0.4
2125                             74.0                                   0.4

      percentage_of_time_with_abnormal_long_term_variability  ...  \
0                                                    43.0      ...
1                                                     0.0      ...
2                                                     0.0      ...
3                                                     0.0      ...
4                                                     0.0      ...
...                                                    ...     ...
2121                                                 25.0      ...
2122                                                 22.0      ...
2123                                                 20.0      ...
2124                                                 27.0      ...
2125                                                 36.0      ...

      histogram_min  histogram_max  histogram_number_of_peaks  \
0              62.0          126.0                        2.0
1              68.0          198.0                        6.0
2              68.0          198.0                        5.0
3              53.0          170.0                       11.0
4              53.0          170.0                        9.0
...             ...            ...                        ...
2121          137.0          177.0                        4.0
2122          103.0          169.0                        6.0
```

```
2123          103.0          170.0                       5.0
2124          103.0          169.0                       6.0
2125          117.0          159.0                       2.0

      histogram_number_of_zeroes  histogram_mode  histogram_mean  \
0                            0.0           120.0           137.0
1                            1.0           141.0           136.0
2                            1.0           141.0           135.0
3                            0.0           137.0           134.0
4                            0.0           137.0           136.0
...                          ...             ...             ...
2121                         0.0           153.0           150.0
2122                         0.0           152.0           148.0
2123                         0.0           153.0           148.0
2124                         0.0           152.0           147.0
2125                         1.0           145.0           143.0

      histogram_median  histogram_variance  histogram_tendency  fetal_health
0                121.0                73.0                 1.0           0.0
1                140.0                12.0                 0.0           0.0
2                138.0                13.0                 0.0           0.0
3                137.0                13.0                 1.0           0.0
4                138.0                11.0                 1.0           0.0
...                ...                 ...                 ...           ...
2121             152.0                 2.0                 0.0           0.0
2122             151.0                 3.0                 1.0           0.0
2123             152.0                 4.0                 1.0           0.0
2124             151.0                 4.0                 1.0           0.0
2125             145.0                 1.0                 0.0           0.0

[1950 rows x 22 columns]
```

[15]:
```
class_1 = data[data['fetal_health'] == 1]
class_1
```

[15]:
```
      baseline value  accelerations  fetal_movement  uterine_contractions  \
5              134.0          0.001           0.000                 0.010
6              134.0          0.001           0.000                 0.013
7              122.0          0.000           0.000                 0.000
8              122.0          0.000           0.000                 0.002
9              122.0          0.000           0.000                 0.003
...              ...            ...             ...                   ...
2036           128.0          0.000           0.002                 0.007
2037           128.0          0.000           0.002                 0.007
2038           128.0          0.000           0.002                 0.007
2039           128.0          0.000           0.003                 0.007
2048           128.0          0.000           0.000                 0.008
```

```
      light_decelerations  severe_decelerations  prolongued_decelerations  \
5                   0.009                   0.0                     0.002
6                   0.008                   0.0                     0.003
7                   0.000                   0.0                     0.000
8                   0.000                   0.0                     0.000
9                   0.000                   0.0                     0.000
...                   ...                   ...                       ...
2036                0.005                   0.0                     0.002
2037                0.005                   0.0                     0.002
2038                0.006                   0.0                     0.002
2039                0.006                   0.0                     0.002
2048                0.010                   0.0                     0.000

      abnormal_short_term_variability  mean_value_of_short_term_variability  \
5                                26.0                                   5.9
6                                29.0                                   6.3
7                                83.0                                   0.5
8                                84.0                                   0.5
9                                86.0                                   0.3
...                               ...                                   ...
2036                             65.0                                   2.6
2037                             66.0                                   2.7
2038                             65.0                                   2.5
2039                             65.0                                   2.5
2048                             63.0                                   4.2

      percentage_of_time_with_abnormal_long_term_variability  ...  \
5                                                        0.0  ...
6                                                        0.0  ...
7                                                        6.0  ...
8                                                        5.0  ...
9                                                        6.0  ...
...                                                      ...  ...
2036                                                     0.0  ...
2037                                                     0.0  ...
2038                                                     0.0  ...
2039                                                     0.0  ...
2048                                                     0.0  ...

      histogram_min  histogram_max  histogram_number_of_peaks  \
5              50.0          200.0                        5.0
6              50.0          200.0                        6.0
7              62.0          130.0                        0.0
8              62.0          130.0                        0.0
9              62.0          130.0                        1.0
...             ...            ...                        ...
```

```
2036           63.0          192.0                       6.0
2037           63.0          192.0                       4.0
2038           63.0          192.0                       6.0
2039           63.0          192.0                       5.0
2048           66.0          156.0                       5.0

      histogram_number_of_zeroes  histogram_mode  histogram_mean  \
5                            3.0            76.0           107.0
6                            3.0            71.0           107.0
7                            0.0           122.0           122.0
8                            0.0           122.0           122.0
9                            0.0           122.0           122.0
...                          ...             ...             ...
2036                         0.0           103.0            92.0
2037                         0.0           105.0            89.0
2038                         0.0           104.0            94.0
2039                         0.0           114.0            97.0
2048                         0.0            69.0            73.0

      histogram_median  histogram_variance  histogram_tendency  fetal_health
5                107.0               170.0                 0.0           1.0
6                106.0               215.0                 0.0           1.0
7                123.0                 3.0                 1.0           1.0
8                123.0                 3.0                 1.0           1.0
9                123.0                 1.0                 1.0           1.0
...                ...                 ...                 ...           ...
2036             114.0                28.0                 0.0           1.0
2037             112.0                21.0                 0.0           1.0
2038             115.0                28.0                 0.0           1.0
2039             116.0                25.0                 0.0           1.0
2048             118.0               128.0                 0.0           1.0

[176 rows x 22 columns]
```

[16]:
```
class_count_0, class_count_1 = data['fetal_health'].value_counts()
class_0_under = class_0.sample(class_count_1)
class_0_under #now class 0 and 1 have the same size
```

[16]:
```
      baseline value  accelerations  fetal_movement  uterine_contractions  \
102            125.0          0.008           0.013                 0.002
1482           132.0          0.000           0.000                 0.007
1257           118.0          0.000           0.000                 0.007
210            125.0          0.008           0.015                 0.001
1973           138.0          0.006           0.002                 0.003
...              ...            ...             ...                   ...
1561           144.0          0.003           0.000                 0.006
1516           130.0          0.000           0.000                 0.004
```

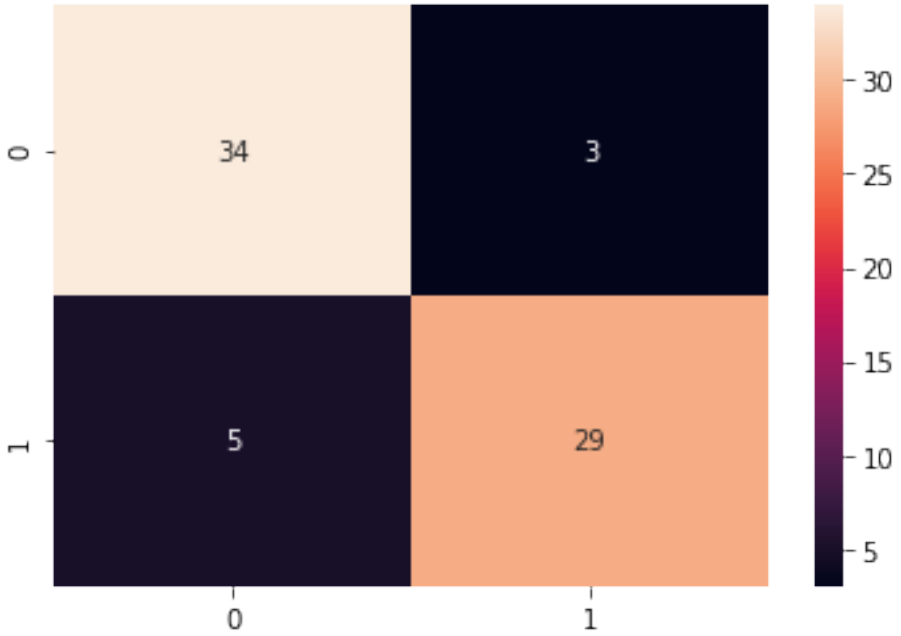
```
1389          121.0          0.007          0.000                0.010
1100          122.0          0.000          0.000                0.004
1233          125.0          0.000          0.000                0.008

      light_decelerations  severe_decelerations  prolongued_decelerations  \
102                 0.000                   0.0                       0.0
1482                0.001                   0.0                       0.0
1257                0.000                   0.0                       0.0
210                 0.000                   0.0                       0.0
1973                0.000                   0.0                       0.0
...                   ...                   ...                       ...
1561                0.000                   0.0                       0.0
1516                0.000                   0.0                       0.0
1389                0.000                   0.0                       0.0
1100                0.000                   0.0                       0.0
1233                0.000                   0.0                       0.0

      abnormal_short_term_variability  mean_value_of_short_term_variability  \
102                              24.0                                   1.6
1482                             47.0                                   0.7
1257                             26.0                                   1.2
210                              48.0                                   0.8
1973                             60.0                                   0.8
...                               ...                                   ...
1561                             40.0                                   0.8
1516                             28.0                                   1.2
1389                             41.0                                   0.8
1100                             22.0                                   1.8
1233                             32.0                                   1.1

      percentage_of_time_with_abnormal_long_term_variability  ...  \
102                                                      0.0  ...
1482                                                     1.0  ...
1257                                                     0.0  ...
210                                                      0.0  ...
1973                                                     0.0  ...
...                                                      ...  ...
1561                                                     5.0  ...
1516                                                     0.0  ...
1389                                                     0.0  ...
1100                                                     0.0  ...
1233                                                     3.0  ...

      histogram_min  histogram_max  histogram_number_of_peaks  \
102            68.0          152.0                        4.0
1482          114.0          145.0                        1.0
1257          100.0          144.0                        3.0
```

```
210            82.0          145.0                       9.0
1973          125.0          168.0                       1.0
...             ...            ...                       ...
1561          112.0          168.0                       2.0
1516          114.0          150.0                       2.0
1389          105.0          153.0                       3.0
1100          106.0          140.0                       2.0
1233           96.0          147.0                       4.0

      histogram_number_of_zeroes  histogram_mode  histogram_mean  \
102                          0.0           132.0           132.0
1482                         0.0           136.0           134.0
1257                         0.0           120.0           119.0
210                          0.0           133.0           132.0
1973                         0.0           148.0           144.0
...                          ...             ...             ...
1561                         0.0           157.0           153.0
1516                         2.0           131.0           132.0
1389                         0.0           150.0           135.0
1100                         1.0           126.0           125.0
1233                         0.0           126.0           125.0

      histogram_median  histogram_variance  histogram_tendency  fetal_health
102              133.0                 5.0                 1.0           0.0
1482             136.0                 2.0                 1.0           0.0
1257             120.0                 2.0                 0.0           0.0
210              133.0                 3.0                 1.0           0.0
1973             147.0                 6.0                 0.0           0.0
...                ...                 ...                 ...           ...
1561             155.0                 5.0                 1.0           0.0
1516             133.0                 2.0                 0.0           0.0
1389             137.0                32.0                 1.0           0.0
1100             127.0                 2.0                 0.0           0.0
1233             127.0                 2.0                 0.0           0.0

[176 rows x 22 columns]
```

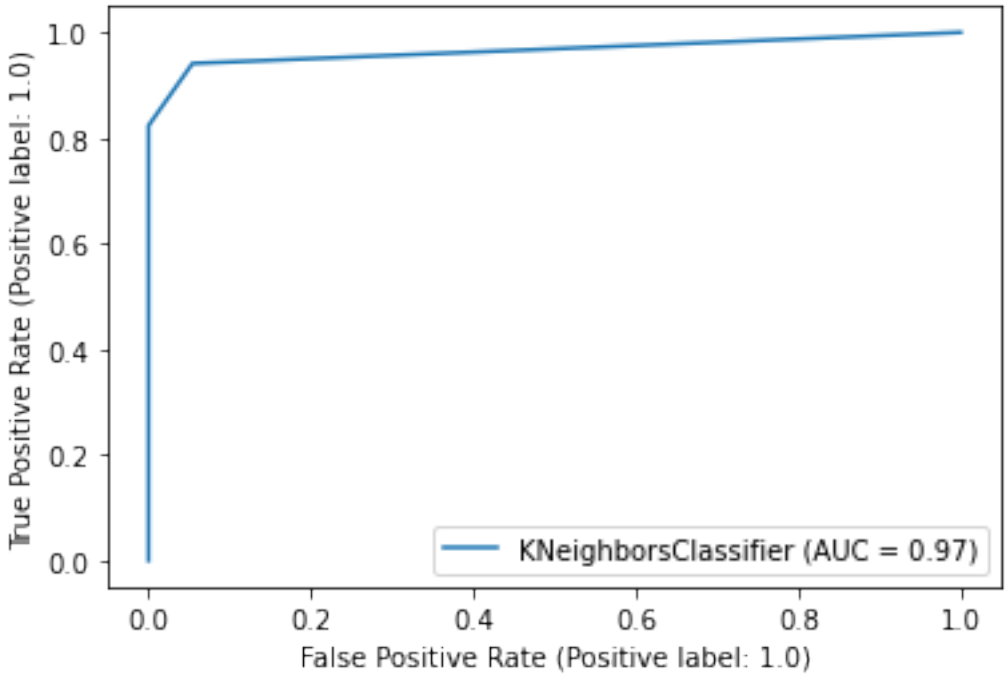
[17]:
```
data_under = pd.concat([class_0_under, class_1], axis=0)
X_under = data_under.drop('fetal_health', axis = 1)
y_under = data_under['fetal_health']
x_train_u,x_test_u,y_train_u,y_test_u =␣
 ↪train_test_split(X_under,y_under,test_size=0.2, random_state=142)
neigh_under = KNeighborsClassifier(n_neighbors=3)
neigh_under.fit(x_train_u,y_train_u)

y_pred_train_u = neigh_under.predict(x_train_u)
y_pred_test_u = neigh_under.predict(x_test_u)
```

[18]:
```
#confusion matrix
cm = confusion_matrix(y_test_u, y_pred_test_u)

sns.heatmap(cm, annot=True)
print(accuracy_score(y_test_u, y_pred_test_u))
```

```
0.8873239436619719
```



[19]:
```
metrics.plot_roc_curve(neigh, x_test_u, y_test_u)
plt.show()
```



[20]:
```
from sklearn.metrics import classification_report
print(classification_report(y_test_u, y_pred_test_u))
```

```
              precision    recall  f1-score   support

         0.0       0.87      0.92      0.89        37
         1.0       0.91      0.85      0.88        34

    accuracy                           0.89        71
   macro avg       0.89      0.89      0.89        71
weighted avg       0.89      0.89      0.89        71
```

Oversampling via SMOTE

[21]:
```
#oversampling
class_1_over = class_1.sample(class_count_0, replace=True)
class_1_over
```

[21]:

	baseline value	accelerations	fetal_movement	uterine_contractions	\
1491	132.0	0.000	0.000	0.008	
2025	129.0	0.000	0.000	0.006	
1752	134.0	0.004	0.001	0.001	
2027	129.0	0.000	0.001	0.006	
823	152.0	0.000	0.000	0.003	
...	...	...	...	...	

```
1755          134.0          0.003          0.000                0.000
240           125.0          0.000          0.009                0.000
328           146.0          0.000          0.000                0.000
769           134.0          0.000          0.000                0.000
1908          130.0          0.002          0.002                0.008

      light_decelerations  severe_decelerations  prolongued_decelerations  \
1491                0.002                   0.0                     0.002
2025                0.010                   0.0                     0.001
1752                0.003                   0.0                     0.003
2027                0.008                   0.0                     0.002
823                 0.000                   0.0                     0.000
...                   ...                   ...                       ...
1755                0.000                   0.0                     0.005
240                 0.000                   0.0                     0.000
328                 0.000                   0.0                     0.000
769                 0.000                   0.0                     0.000
1908                0.008                   0.0                     0.002

      abnormal_short_term_variability  mean_value_of_short_term_variability  \
1491                             30.0                                   1.7
2025                             65.0                                   2.6
1752                             61.0                                   1.8
2027                             65.0                                   2.8
823                              61.0                                   0.4
...                               ...                                   ...
1755                             65.0                                   1.8
240                              68.0                                   0.3
328                              84.0                                   0.2
769                              81.0                                   0.2
1908                             64.0                                   2.4

      percentage_of_time_with_abnormal_long_term_variability  ...  \
1491                                                  0.0      ...
2025                                                  0.0      ...
1752                                                  0.0      ...
2027                                                  0.0      ...
823                                                  71.0      ...
...                                                   ...      ...
1755                                                  0.0      ...
240                                                  75.0      ...
328                                                  81.0      ...
769                                                  91.0      ...
1908                                                  0.0      ...

      histogram_min  histogram_max  histogram_number_of_peaks  \
1491           60.0          150.0                        3.0
```

```
2025           50.0          151.0                       9.0
1752           66.0          184.0                       6.0
2027           50.0          151.0                       7.0
823           134.0          164.0                       4.0
...             ...            ...                       ...
1755           66.0          180.0                       8.0
240           120.0          128.0                       0.0
328           140.0          150.0                       1.0
769           133.0          136.0                       1.0
1908           50.0          211.0                       8.0

      histogram_number_of_zeroes  histogram_mode  histogram_mean  \
1491                         0.0            99.0           104.0
2025                         2.0           105.0            91.0
1752                         0.0            88.0           112.0
2027                         2.0           105.0            86.0
823                          0.0           160.0           158.0
...                          ...             ...             ...
1755                         1.0            88.0           100.0
240                          0.0           125.0           125.0
328                          0.0           146.0           145.0
769                          0.0           135.0           134.0
1908                         0.0            60.0            85.0

      histogram_median  histogram_variance  histogram_tendency  fetal_health
1491             102.0                33.0                 0.0           1.0
2025             113.0                13.0                 0.0           1.0
1752             111.0               182.0                -1.0           1.0
2027             112.0                13.0                 0.0           1.0
823              160.0                 1.0                 1.0           1.0
...                ...                 ...                 ...           ...
1755             101.0                83.0                -1.0           1.0
240              126.0                 0.0                 0.0           1.0
328              147.0                 0.0                 0.0           1.0
769              136.0                 0.0                 1.0           1.0
1908             112.0               243.0                -1.0           1.0

[1950 rows x 22 columns]
```

[22]:
```
data_over = pd.concat([class_1_over, class_0], axis=0)
data_over
```

[22]:
```
      baseline value  accelerations  fetal_movement  uterine_contractions  \
1491           132.0          0.000           0.000                 0.008
2025           129.0          0.000           0.000                 0.006
1752           134.0          0.004           0.001                 0.001
2027           129.0          0.000           0.001                 0.006
```

```
823           152.0          0.000           0.000                 0.003
...             ...            ...             ...                   ...
2121          140.0          0.000           0.000                 0.007
2122          140.0          0.001           0.000                 0.007
2123          140.0          0.001           0.000                 0.007
2124          140.0          0.001           0.000                 0.006
2125          142.0          0.002           0.002                 0.008

      light_decelerations  severe_decelerations  prolongued_decelerations  \
1491                0.002                   0.0                     0.002
2025                0.010                   0.0                     0.001
1752                0.003                   0.0                     0.003
2027                0.008                   0.0                     0.002
823                 0.000                   0.0                     0.000
...                   ...                   ...                       ...
2121                0.000                   0.0                     0.000
2122                0.000                   0.0                     0.000
2123                0.000                   0.0                     0.000
2124                0.000                   0.0                     0.000
2125                0.000                   0.0                     0.000

      abnormal_short_term_variability  mean_value_of_short_term_variability  \
1491                             30.0                                   1.7
2025                             65.0                                   2.6
1752                             61.0                                   1.8
2027                             65.0                                   2.8
823                              61.0                                   0.4
...                               ...                                   ...
2121                             79.0                                   0.2
2122                             78.0                                   0.4
2123                             79.0                                   0.4
2124                             78.0                                   0.4
2125                             74.0                                   0.4

      percentage_of_time_with_abnormal_long_term_variability  ...  \
1491                                                  0.0      ...
2025                                                  0.0      ...
1752                                                  0.0      ...
2027                                                  0.0      ...
823                                                  71.0      ...
...                                                   ...      ...
2121                                                 25.0      ...
2122                                                 22.0      ...
2123                                                 20.0      ...
2124                                                 27.0      ...
2125                                                 36.0      ...
```

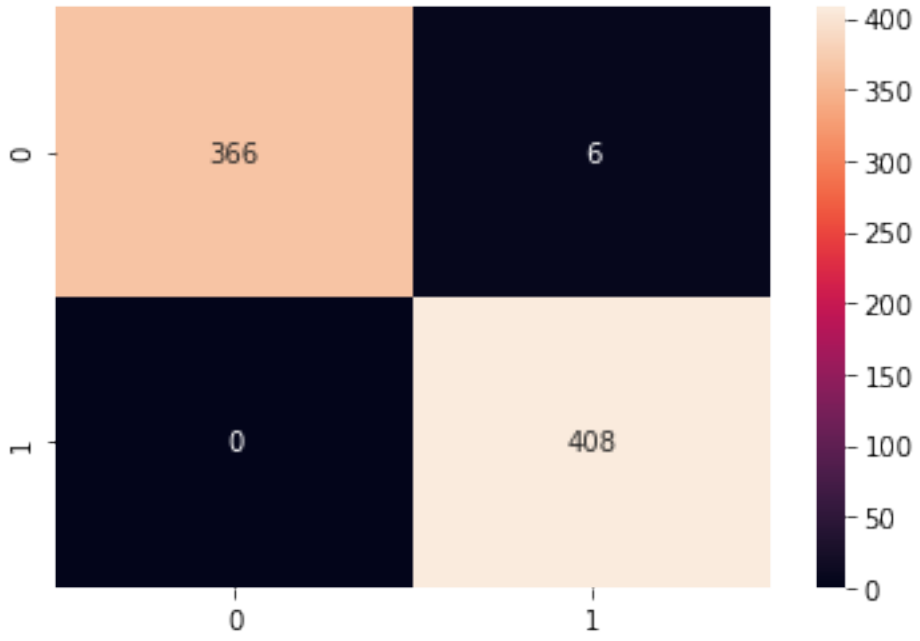
```
      histogram_min  histogram_max  histogram_number_of_peaks  \
1491           60.0          150.0                        3.0
2025           50.0          151.0                        9.0
1752           66.0          184.0                        6.0
2027           50.0          151.0                        7.0
823           134.0          164.0                        4.0
...             ...            ...                        ...
2121          137.0          177.0                        4.0
2122          103.0          169.0                        6.0
2123          103.0          170.0                        5.0
2124          103.0          169.0                        6.0
2125          117.0          159.0                        2.0

      histogram_number_of_zeroes  histogram_mode  histogram_mean  \
1491                         0.0            99.0           104.0
2025                         2.0           105.0            91.0
1752                         0.0            88.0           112.0
2027                         2.0           105.0            86.0
823                          0.0           160.0           158.0
...                          ...             ...             ...
2121                         0.0           153.0           150.0
2122                         0.0           152.0           148.0
2123                         0.0           153.0           148.0
2124                         0.0           152.0           147.0
2125                         1.0           145.0           143.0

      histogram_median  histogram_variance  histogram_tendency  fetal_health
1491             102.0                33.0                 0.0           1.0
2025             113.0                13.0                 0.0           1.0
1752             111.0               182.0                -1.0           1.0
2027             112.0                13.0                 0.0           1.0
823              160.0                 1.0                 1.0           1.0
...                ...                 ...                 ...           ...
2121             152.0                 2.0                 0.0           0.0
2122             151.0                 3.0                 1.0           0.0
2123             152.0                 4.0                 1.0           0.0
2124             151.0                 4.0                 1.0           0.0
2125             145.0                 1.0                 0.0           0.0

[3900 rows x 22 columns]
```

[23]:
```
X_over = data_over.drop('fetal_health', axis = 1)
y_over = data_over['fetal_health']
x_train_o,x_test_o,y_train_o,y_test_o =␣
 ↪train_test_split(X_over,y_over,test_size=0.2, random_state=142)
neigh_over = KNeighborsClassifier(n_neighbors=3)
neigh_over.fit(x_train_o,y_train_o)
```

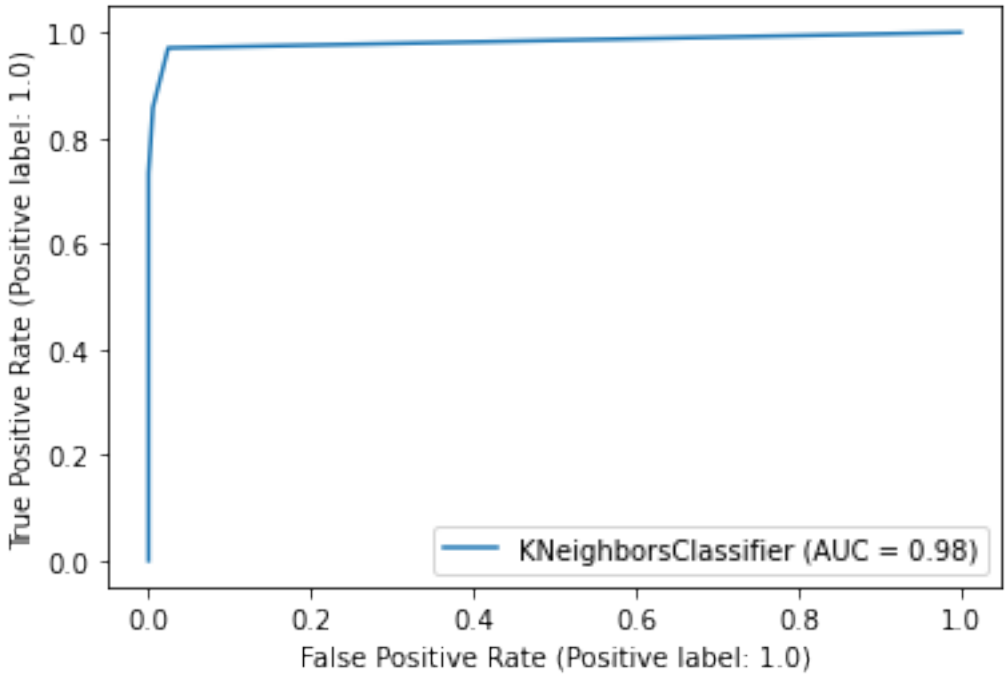
```
y_pred_train_o = neigh_over.predict(x_train_o)
y_pred_test_o = neigh_over.predict(x_test_o)
```

[24]:
```
#confusion matrix

cm = confusion_matrix(y_test_o, y_pred_test_o)

sns.heatmap(cm, annot=True,annot_kws={"size": 10},fmt="d")
print(accuracy_score(y_test_o, y_pred_test_o))
```

```
0.9923076923076923
```



[25]:
```
metrics.plot_roc_curve(neigh, x_test_o, y_test_o)
plt.show()
```



[26]:
```
from sklearn.metrics import classification_report
print(classification_report(y_test_o, y_pred_test_o))
```

```
              precision    recall  f1-score   support

         0.0       1.00      0.98      0.99       372
         1.0       0.99      1.00      0.99       408

    accuracy                           0.99       780
   macro avg       0.99      0.99      0.99       780
weighted avg       0.99      0.99      0.99       780
```