|  |  |
| --- | --- |
| **EX.NO: 2 Attribute Selection Model** | |
| IN [1]: | **import** pandas **as** pd  **import** numpy **as** np  **import** seaborn **as** sns  **import** matplotlib.pyplot **as** plt  **import** math |
| IN [2]: | df pd**.**read\_csv("Heart2.csv")  df**.**head(5) |
| IN [3] | df**.**head(14) |
| IN [4] | df['Thal'] **=** pd**.**factorize(df['Thal'])[0]**.**astype(np**.**int8)  df**.**head() |
| IN [5] | df['ChestPain'] **=** pd**.**factorize(df['ChestPain'])[0]**.**astype(np**.**int8)  df['AHD'] **=** pd**.**factorize(df['AHD'])[0]**.**astype(np**.**int8) |
| IN [6] | df**.**head() |
| IN [7] | cols**=**['Age','Sex','ChestPain','RestBP','Chol','Fbs','RestECG','MaxHR','ExAng','Oldpeak','Slope','Ca','Thal','AHD'] |
| IN [8] | plt**.**subplots(figsize**=**(20,15))  cm **=** np**.**corrcoef(df[cols]**.**values**.**T)  hm **=** sns**.**heatmap(cm,  cbar**=True**,  annot**=True**,  square**=True**,  fmt**=**'.2f',  annot\_kws**=**{'size': 10},  yticklabels**=**cols,  xticklabels**=**cols)  plt**.**show() |
| IN [9] | df**.**drop(['ChestPain','MaxHR'],axis**=**1,inplace**=True**) |
| IN [10] | df**.**head() |
| IN [11] | **def** inf(p,n):  diff **=** **-**p**/**14  diff1 **=** **-**n**/**14  ans **=** diff**\***(math**.**log(**-**1**\***diff)) **+** diff1**\***(math**.**log(diff1**\*-**1))  **return** ans |
| IN [12] | df['AHD']**.**value\_counts() |
| IN [13] | inf(8,6) |
| IN [14] | df**.**head() |
| IN [15] | df1 **=** df[(df['Sex'] **==** 1) **&** (df['AHD'] **==** 1)]  df1['AHD']**.**value\_counts() |
| IN [16] | df1 **=** df[(df['Sex'] **==** 1) **&** (df['AHD'] **==** 0)]  df1['AHD']**.**value\_counts() |
| IN [17] | df['Slope'] **=** [0 **if** x **==** 1 **else** 1 **if** x **==** 2 **else** 2 **for** x **in** df['Slope']]  df**.**head() |
| IN [18] | **def** info\_col(col\_name,n):  l **=** []  **for** i **in** range(0,n):  df1 **=** df[(df[col\_name] **==** i) **&** (df['AHD'] **==** 1)]  **if** df1**.**empty:  l**.**append(0)  **else**:  l**.**append(int(df1['AHD']**.**value\_counts()))  df1 **=** df[(df[col\_name] **==** i) **&** (df['AHD'] **==** 0)]  **if** df1**.**empty:  l**.**append(0)  **else** :  l**.**append(int(df1['AHD']**.**value\_counts()))  j **=** 0  s **=** 0  **while** j **<** len(l):  **if** l[j]**>**0 **and** l[j**+**1]**>**0:  s **=** s **+** ((l[j]**+**l[j**+**1])**/**14)**\***inf(int(l[j]),int(l[j**+**1]))  j **+=** 2  **return** s |
| IN [19] | info\_col('Sex',2) |
| IN [20] | info\_col('Fbs',2) |
| IN [21] | info\_col('ExAng',2) |
| IN [22] | info\_col('Thal',3) |
| IN [23] | info\_col('Ca',4) |
| IN [24] | info\_col('Slope',3) |

**OUTPUT:**

|  |  |
| --- | --- |
| OUT [8] | HEAT MAP |
| OUT [19] | Accuracy for Sex = 0.673487303713993 |
| OUT [20] | Accuracy for Fbs = 0.6535027367351511 |
| OUT [21] | Accuracy for ExAng = 0.5966591891645707 |
| OUT [22] | Accuracy for Thal = 0.5709878794955071 |
| OUT [23] | Accuracy for Ca = 0.3267543504994274 |
| OUT [24] | Accuracy for Slope = 0.4313867113134571 |