brain-project-1

September 25, 2024

0.0.1 NAME: HRISHIKESH SHIVPUTRA KAMBLE

0.0.2 PROJECT:-

SOLVING CLASSIFICATION PREDICTION FOR “BRAIN STROKE” DATASET USING “LOGISTIC REGRESSION, NAIVES BAYES CLASSIFICATION,SUPPORT VECTOR CLASSIFIER,K NEAREST NEIGHBOUR, DESICION TREE CLASSIFIER”.

0.0.3 DATA:-

1) GENDER: “MALE”, “FEMALE” OR “OTHER”
2) AGE: AGE OF THE PATIENT
3) HYPERTENSION: 0 IF THE PATIENT DOESN’T HAVE HYPERTENSION, 1 IF THE PATIENT HAS HYPERTENSION
4) HEART DISEASE: 0 IF THE PATIENT DOESN’T HAVE ANY HEART DISEASES, 1 IF THE PATIENT HAS A HEART DISEASE
5) EVER-MARRIED: “NO” OR “YES”
6) WORK TYPE: “CHILDREN”, “GOVTJOV”, “NEVER WORKED”, “PRIVATE” OR “SELF-EMPLOYED”
7) RESIDENCETYPE: “RURAL” OR “URBAN”
8) AVG GLUCOSE LEVEL: AVERAGE GLUCOSE LEVEL IN BLOOD
9) BMI: BODY MASS INDEX
10) SMOKING_STATUS: “FORMERLY SMOKED”, “NEVER SMOKED”, “SMOKES” OR “UNKNOWN”
11) STROKE: 1 IF THE PATIENT HAD A STROKE OR 0 IF NOT

0.0.4 APPROACH:

1.LOAD THE REQUIRED LIBRARIES SUCH AS PANDAS,MATPLOTLIB,SEABORN ALONG WITH GIVEN DATASET.

2.PERFORM EDA ON THE GIVEN DATASET.

3.CONVERT ALL THE REQUIRED COLUMNS INTO NUMERIAL COLUMNS USING GET DUMMIES FUNCTION FROM PANDAS LIBRARY.

4.CONVERTING ALL REQUIRED FEATURES IN NUMERIAL , CHECK FOR CORRELATION BETWEEN FEATURES AND TARGET AND CONSIDER THE ONLY FEATURES WITH HIGHER CORRELATION.

5.IMPORT "LOGISTIC REGRESSION, NAIVES BAYES CLASSIFICATION,SUPPORT VECTOR CLASSIFIER,K NEAREST NEIGHBOUR", AND SPLIT THE GIVEN DATASET INTO TRAINING AND TESTING DATA USING TRAIN_TEST_SPLIT FUNCTION.THEN CALUCLATE ACCURACY SCORE USING SKLEARN LIBRARY BY IMPORTING METRICS.

6.ONCE WE GET ACCURACY SCORE OF ALL MODELS FOR BOTH TRAING AND TESTING DATA, CREATE A DATAFRAME AND LOAD ALL THE ACCURACY OF ALL MODEL.

7.VISUALIZATION: ONCE THE DATASET IS CREATED PLOT THE ACCURACIES OF ALL THE MODELS USING BARPLOT USING MATPLOTLIB.

[1]:
```
import pandas as pd                              # LOADING ALL THE REQUIRED␣
 ↪LIBRARIES.
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
import warnings
warnings.filterwarnings("ignore")
```

[2]:
```
data=pd.read_csv(r"C:\Users\Hrishikesh\Desktop\DATA SCIENCE\brain_stroke.csv")
```

[3]:
```
data.columns
```

[3]:
```
Index(['gender', 'age', 'hypertension', 'heart_disease', 'ever_married',
       'work_type', 'Residence_type', 'avg_glucose_level', 'bmi',
       'smoking_status', 'stroke'],
      dtype='object')
```

[4]:
```
data.isna().sum()      # CHECK NULL VALUES
```

[4]:
```
gender               0
age                  0
hypertension         0
heart_disease        0
ever_married         0
work_type            0
Residence_type       0
avg_glucose_level    0
bmi                  0
smoking_status       0
stroke               0
dtype: int64
```

[5]:
```
data.info()                          # SHOWS ALL INFORMATION REGARDING THE DATA␣
 ↪SUCH AS NULL VALUE, COLUMNS,DATATYPES
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4981 entries, 0 to 4980
Data columns (total 11 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   gender             4981 non-null   object 
 1   age                4981 non-null   float64
 2   hypertension       4981 non-null   int64  
 3   heart_disease      4981 non-null   int64  
 4   ever_married       4981 non-null   object 
 5   work_type          4981 non-null   object 
 6   Residence_type     4981 non-null   object 
 7   avg_glucose_level  4981 non-null   float64
 8   bmi                4981 non-null   float64
 9   smoking_status     4981 non-null   object 
 10  stroke             4981 non-null   int64  
dtypes: float64(3), int64(3), object(5)
memory usage: 428.2+ KB
```

[6]:
```
data.describe()            # SHOWS THE ALL DETAILS REGARDING ALL NUMERICAL␣
↪COLUMNS
```

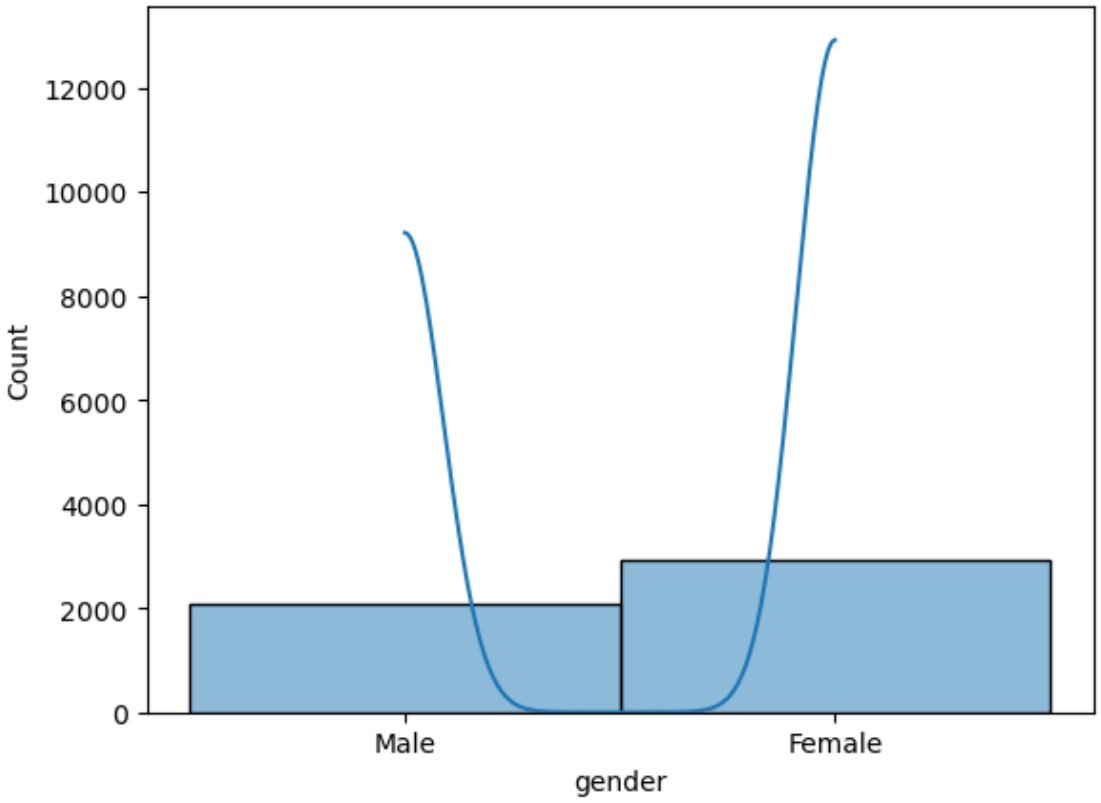
[6]:

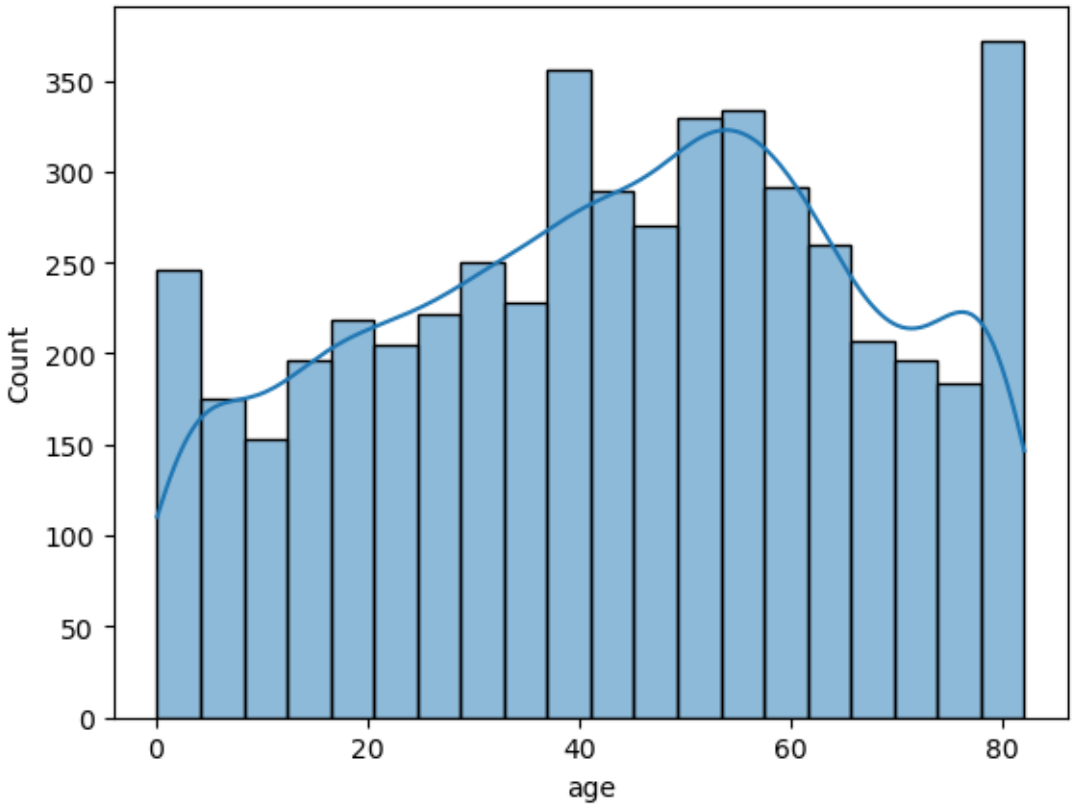
	age	hypertension	heart_disease	avg_glucose_level	bmi	stroke
count	4981.000000	4981.000000	4981.000000	4981.000000	4981.000000	4981.000000
mean	43.419859	0.096165	0.055210	105.943562	28.498173	0.049789
std	22.662755	0.294848	0.228412	45.075373	6.790464	0.217531
min	0.080000	0.000000	0.000000	55.120000	14.000000	0.000000
25%	25.000000	0.000000	0.000000	77.230000	23.700000	0.000000
50%	45.000000	0.000000	0.000000	91.850000	28.100000	0.000000
75%	61.000000	0.000000	0.000000	113.860000	32.600000	0.000000
max	82.000000	1.000000	1.000000	271.740000	48.900000	1.000000

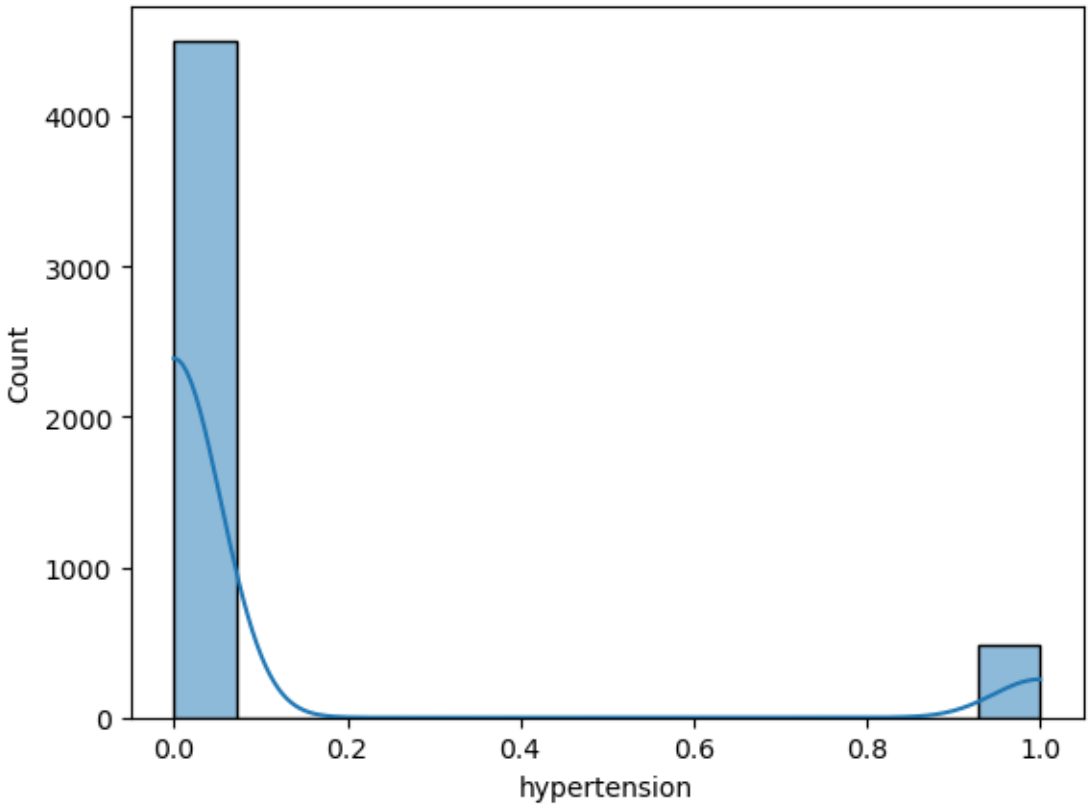
[7]:
```
data.shape                      # SHOWS THE NUMBER OF ROWS AND COLUMNS
```

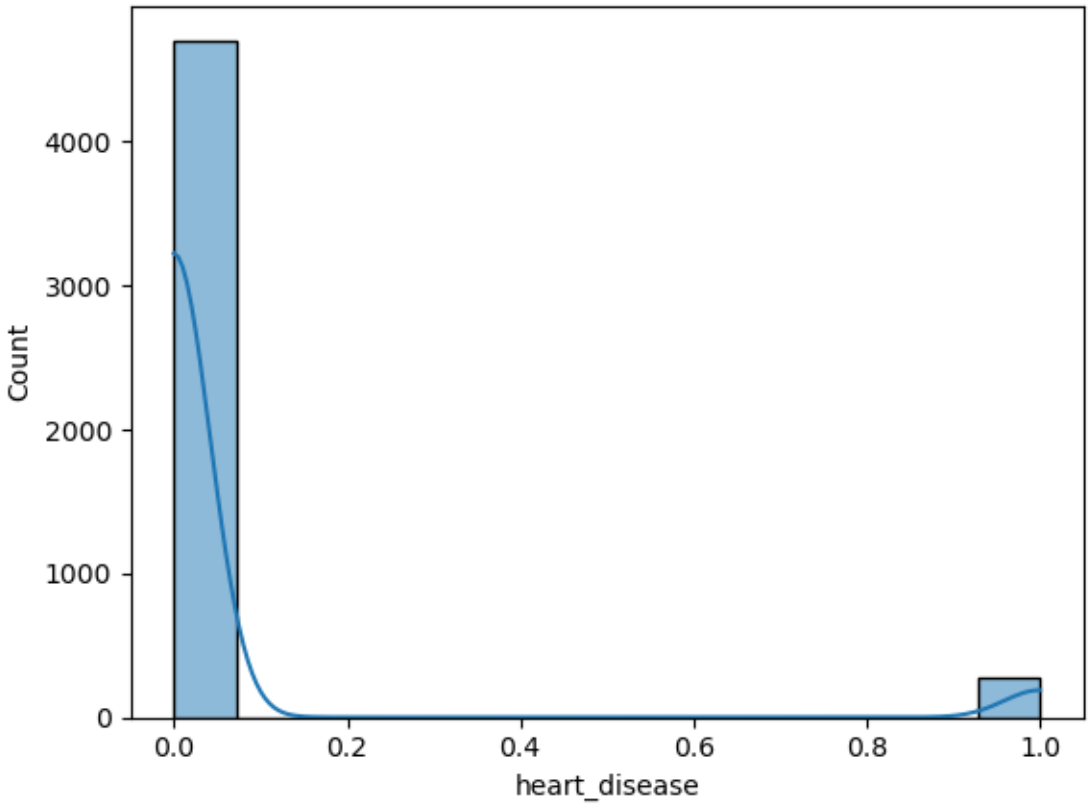
[7]:
```
(4981, 11)
```

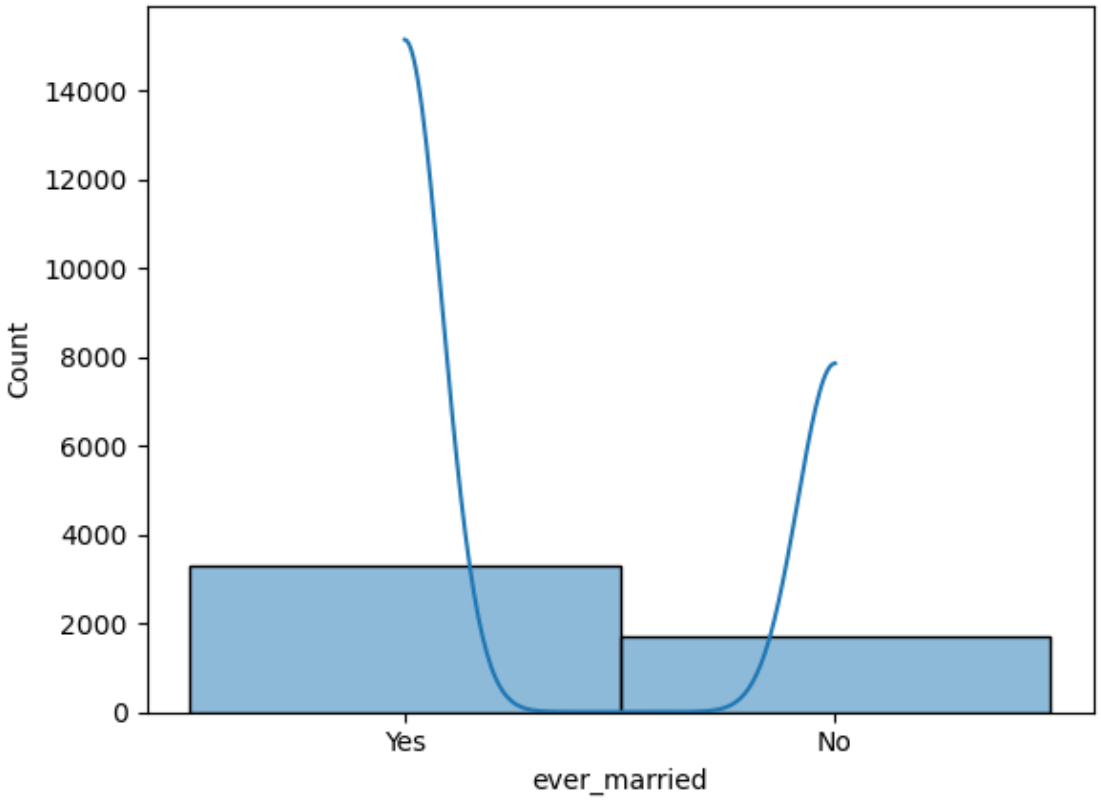
[8]:
```
for i in data.columns:
    sns.histplot(data[i],kde=True)     # PLOT HISTPLOT TO SEE DATA DISTRIBUTION
    plt.show()
```

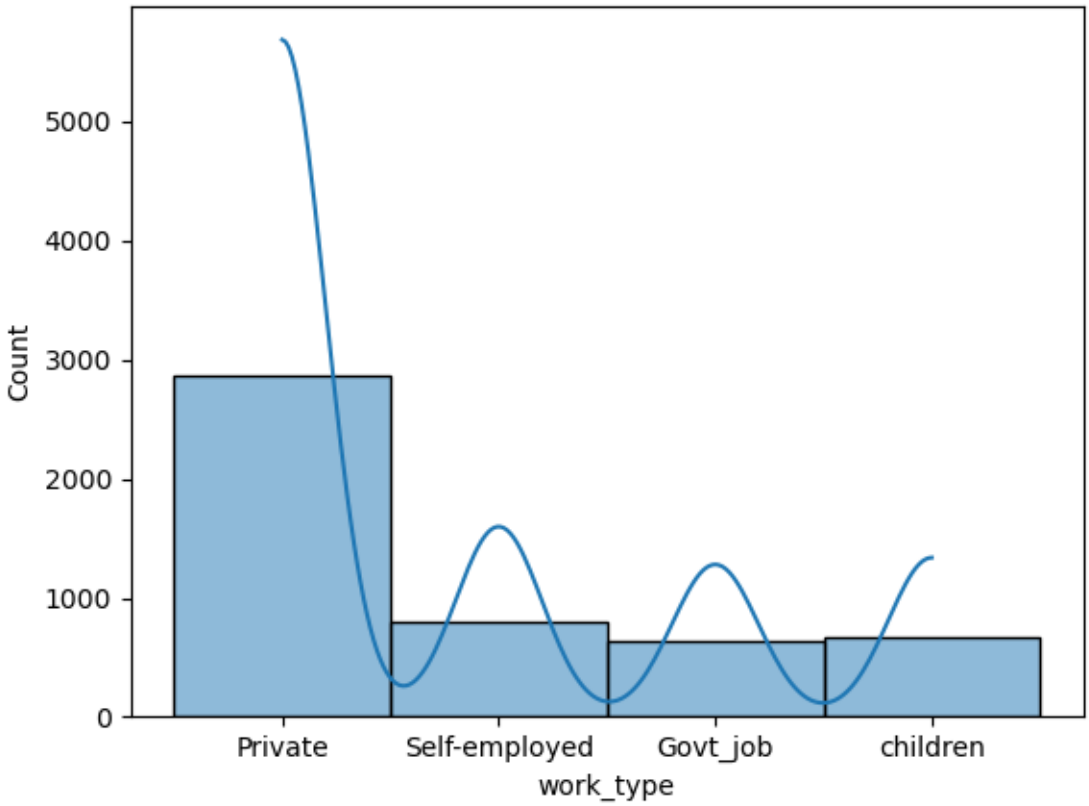


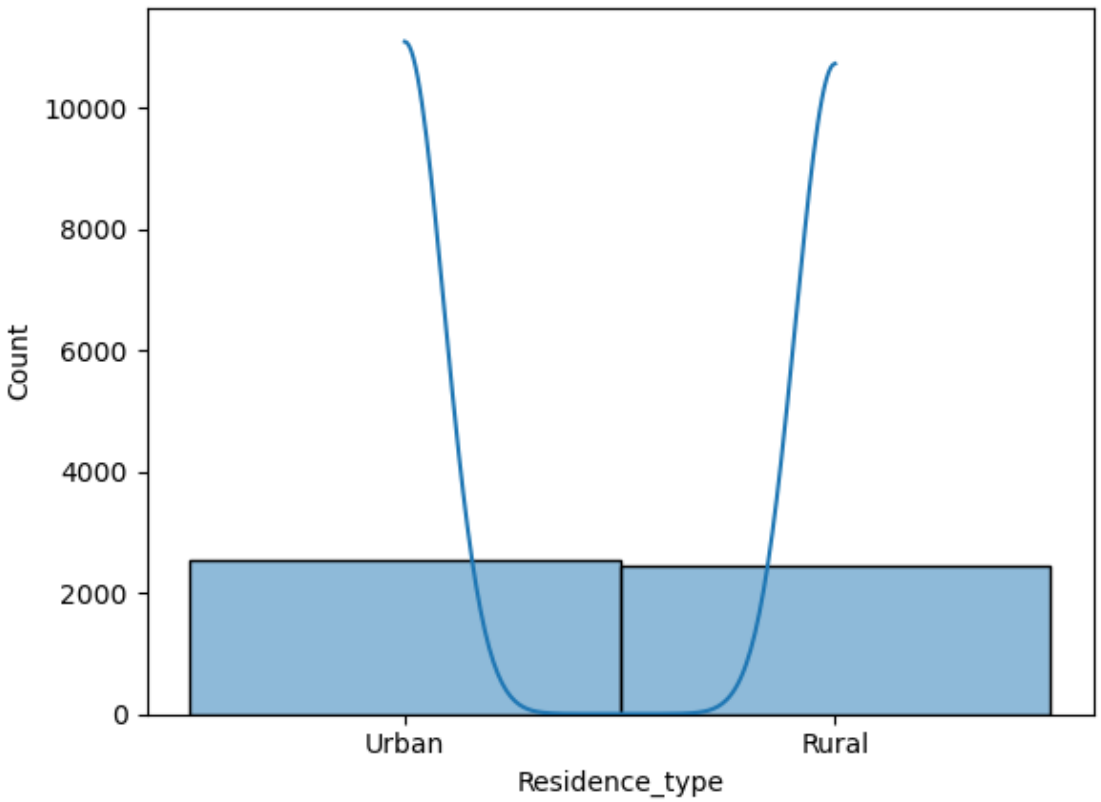

Count
350
300
250
200
150
100
50
0
0
20
40
60
80
age

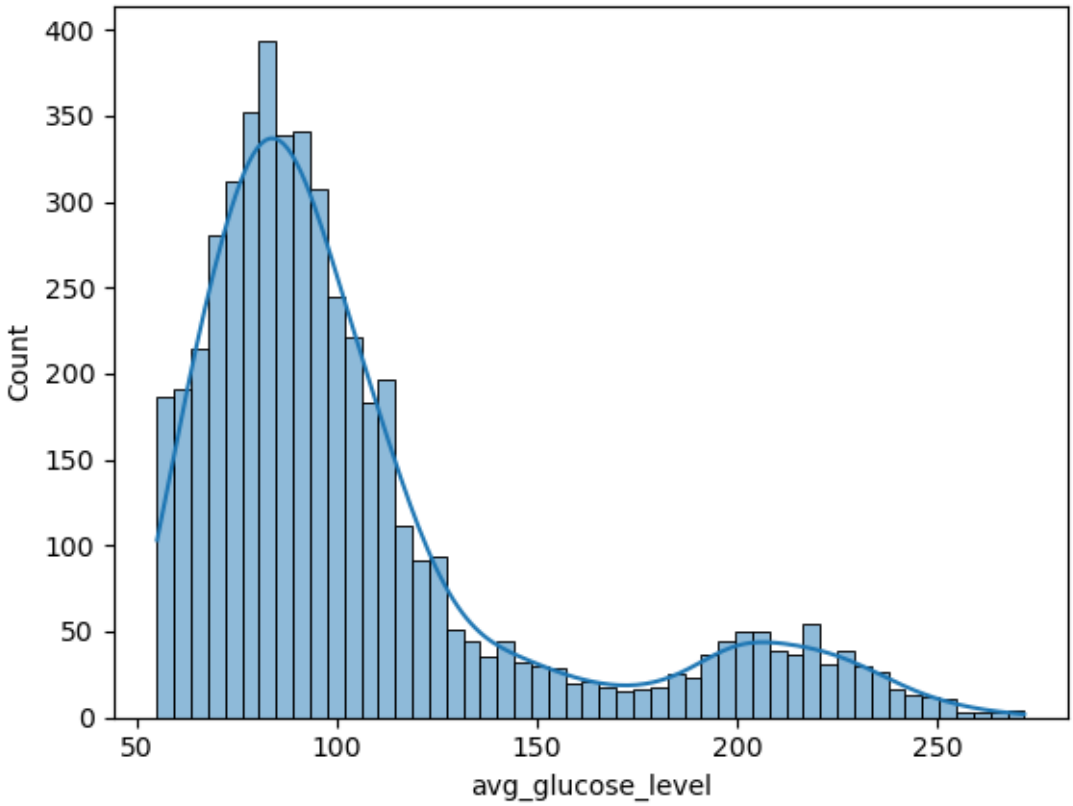

4000
3000
2000
1000
0
Count
0.0
0.2
0.4
0.6
0.8
1.0
hypertension

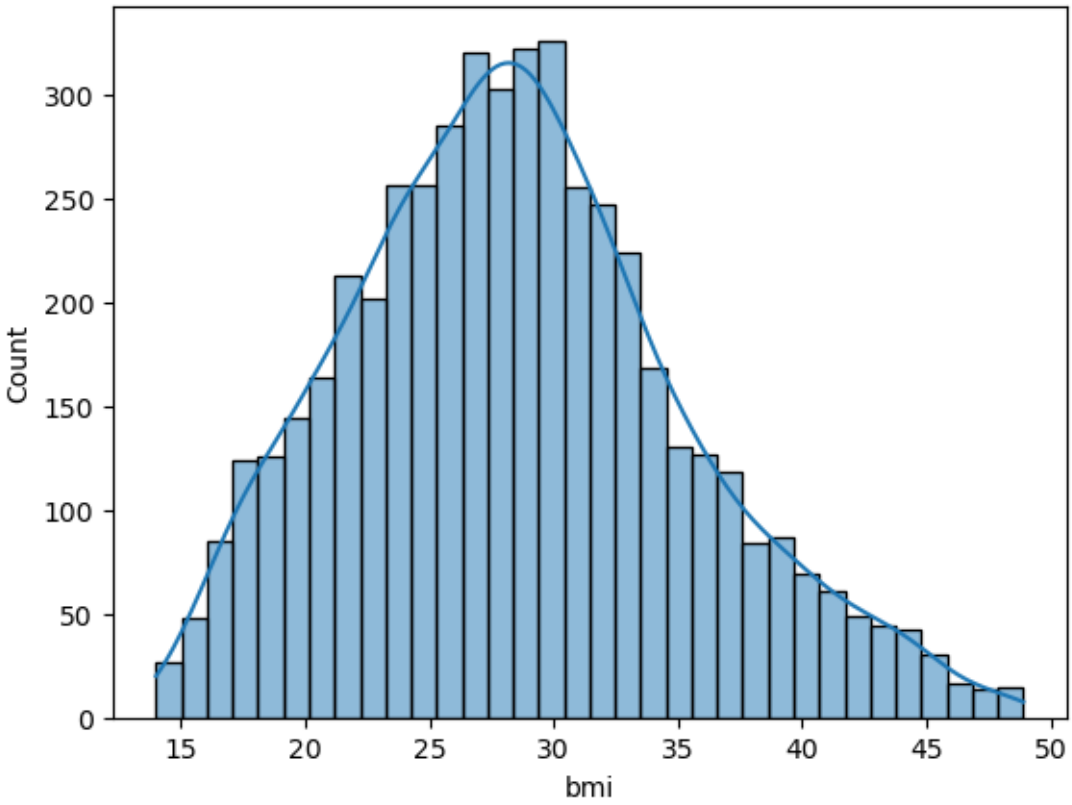

4000
3000
2000
1000
0
Count
0.0
0.2
0.4
0.6
0.8
1.0
heart_disease

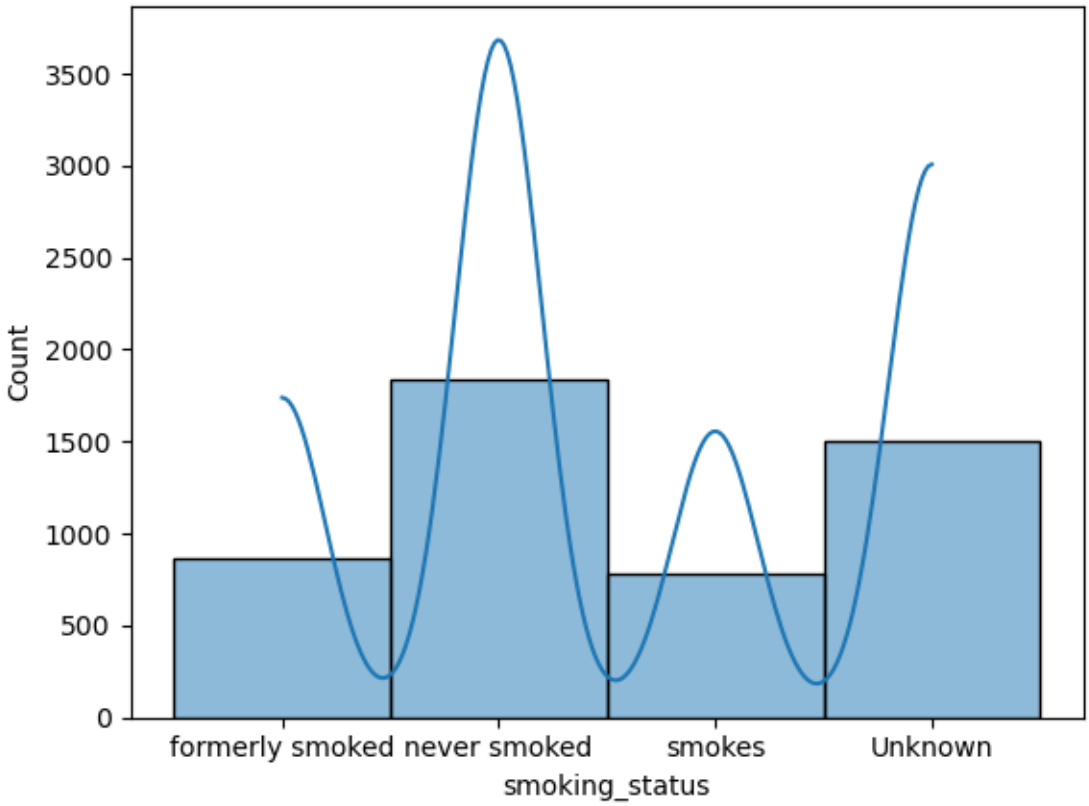

14000
12000
10000
8000
6000
4000
2000
0
Count
Yes
No
ever_married

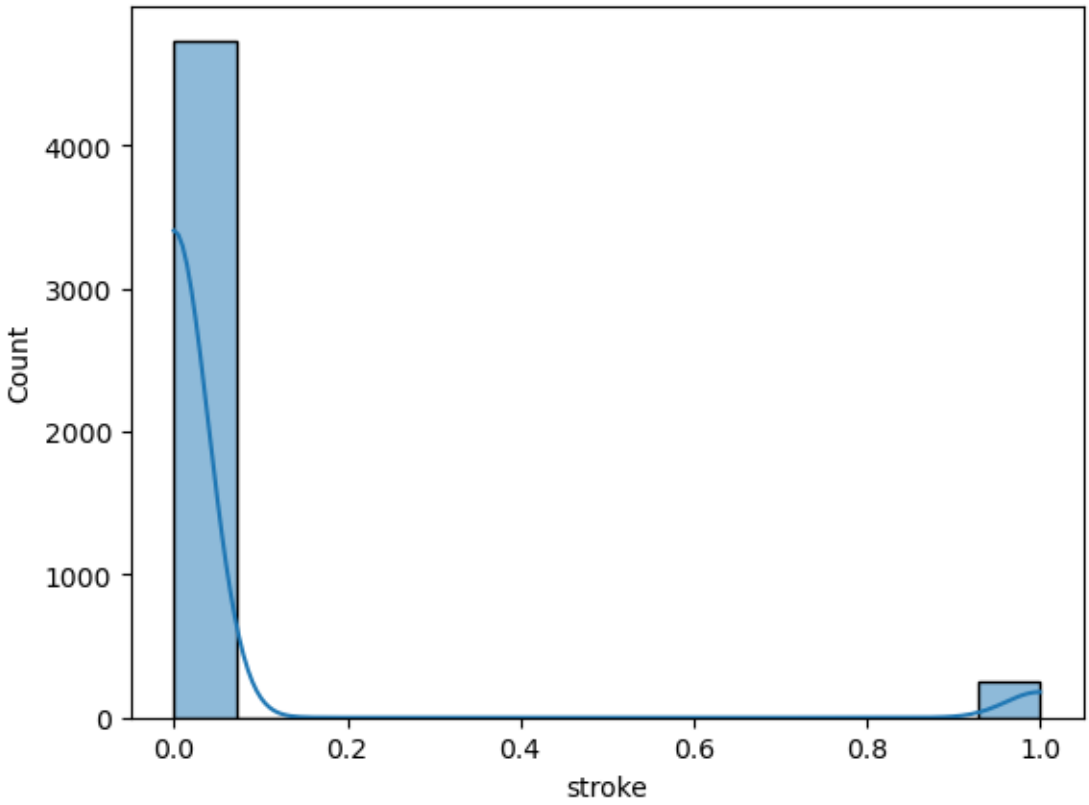

5000
4000
3000
2000
1000
0
Count
Private
Self-employed
Govt_job
children
work_type


10000
8000
6000
4000
2000
0
Count
Urban
Rural
Residence_type


400
350
300
250
200
150
100
50
0
Count
50
100
150
200
250
avg_glucose_level


Count
300
250
200
150
100
50
0
15
20
25
30
35
40
45
50
bmi


3500
3000
2500
2000
1500
1000
500
0
Count
formerly smoked
never smoked
smokes
Unknown
smoking_status



[9]:
```
data
```

[9]:
```
      gender   age  hypertension  heart_disease ever_married      work_type  \
0       Male  67.0             0              1          Yes        Private
1       Male  80.0             0              1          Yes        Private
2     Female  49.0             0              0          Yes        Private
3     Female  79.0             1              0          Yes  Self-employed
4       Male  81.0             0              0          Yes        Private
…      …   …            …              …           …              …
4976    Male  41.0             0              0           No        Private
4977    Male  40.0             0              0          Yes        Private
4978  Female  45.0             1              0          Yes       Govt_job
4979    Male  40.0             0              0          Yes        Private
4980  Female  80.0             1              0          Yes        Private

     Residence_type  avg_glucose_level   bmi   smoking_status  stroke
0             Urban             228.69  36.6  formerly smoked       1
1             Rural             105.92  32.5     never smoked       1
2             Urban             171.23  34.4           smokes       1
3             Rural             174.12  24.0     never smoked       1
4             Urban             186.21  29.0  formerly smoked       1
```

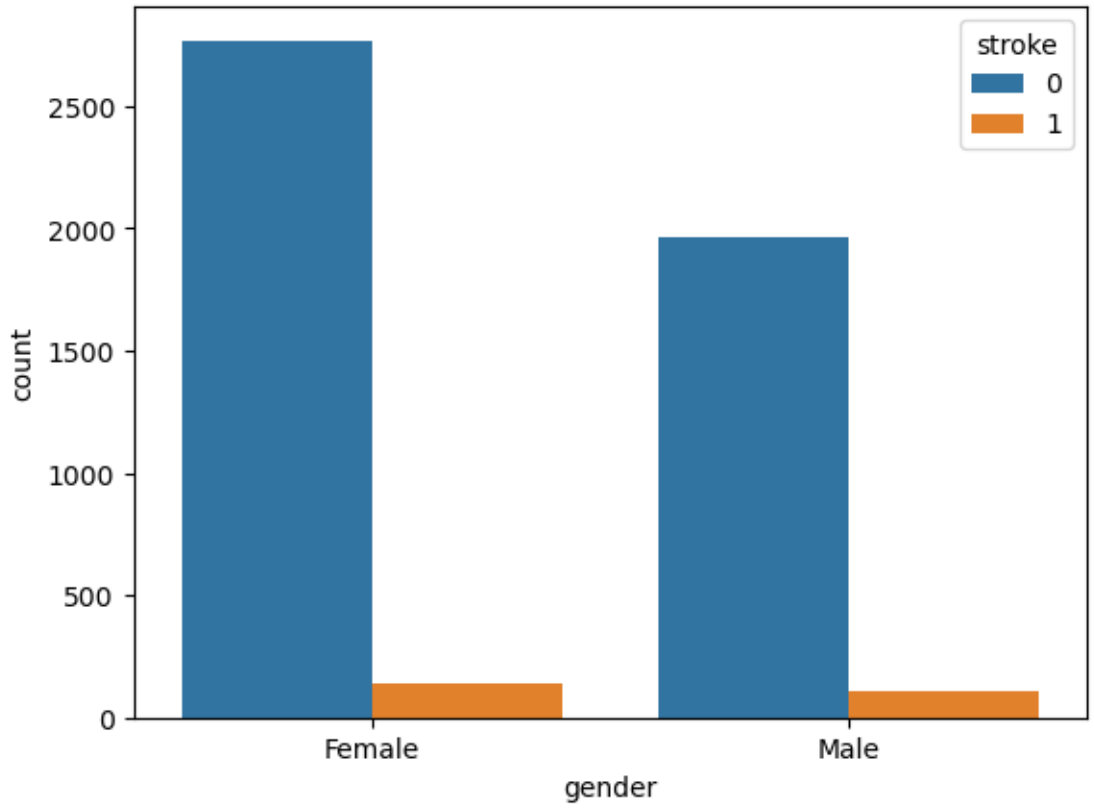
```
…              …                …   …                …      …
4976           Rural            70.15  29.8  formerly smoked       0
4977           Urban           191.15  31.1           smokes       0
4978           Rural            95.02  31.8           smokes       0
4979           Rural            83.94  30.0           smokes       0
4980           Urban            83.75  29.1     never smoked       0

[4981 rows x 11 columns]
```

[10]:
```
x=data.groupby("gender")["stroke"].value_counts().reset_index()
x
```

[10]:
```
   gender  stroke  count
0  Female       0   2767
1  Female       1    140
2    Male       0   1966
3    Male       1    108
```

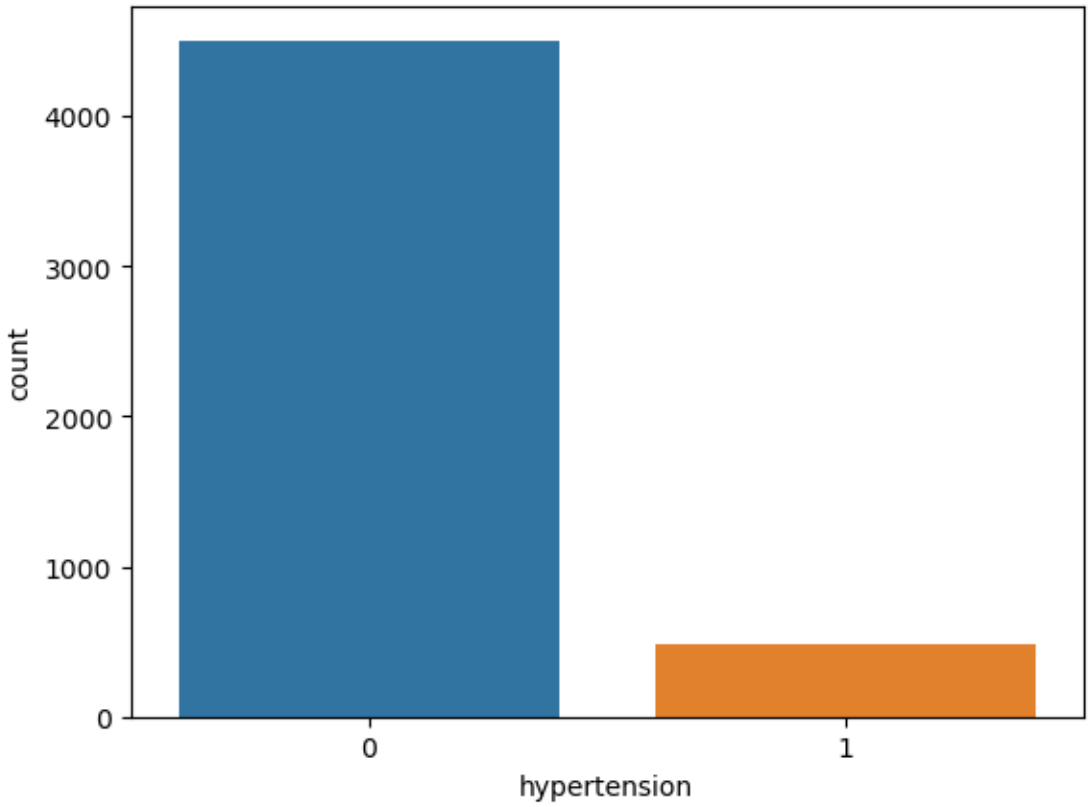
[11]:
```
sns.barplot(data=x,x="gender",y="count",hue="stroke")
```

[11]: <Axes: xlabel='gender', ylabel='count'>



[12]:
```
sns.countplot(data=data,x="hypertension")
```

[12]:
```
<Axes: xlabel='hypertension', ylabel='count'>
```



[13]:
```
data.groupby("stroke")["hypertension"].value_counts().reset_index()
```

[13]:

	stroke	hypertension	count
0	0	0	4320
1	0	1	413
2	1	0	182
3	1	1	66

[14]:
```
stroke_y=data.loc[data["stroke"]==1]
stroke_y
```

[14]:

	gender	age	hypertension	heart_disease	ever_married	work_type	\
0	Male	67.0	0	1	Yes	Private	
1	Male	80.0	0	1	Yes	Private	
2	Female	49.0	0	0	Yes	Private	
3	Female	79.0	1	0	Yes	Self-employed	
4	Male	81.0	0	0	Yes	Private	

```
…      …   …           …             …           …            …
4815    Male  79.0             0             0          Yes        Private
4816    Male  74.0             0             0          Yes        Private
4817  Female  76.0             1             1          Yes  Self-employed
4818    Male  74.0             0             0          Yes  Self-employed
4819    Male  71.0             1             0          Yes  Self-employed

     Residence_type  avg_glucose_level   bmi   smoking_status  stroke
0             Urban             228.69  36.6  formerly smoked       1
1             Rural             105.92  32.5     never smoked       1
2             Urban             171.23  34.4           smokes       1
3             Rural             174.12  24.0     never smoked       1
4             Urban             186.21  29.0  formerly smoked       1
…               …                  …   …              …     …
4815          Rural             114.77  27.2  formerly smoked       1
4816          Urban             167.13  28.7          Unknown       1
4817          Urban             199.86  31.7           smokes       1
4818          Rural              60.98  28.1     never smoked       1
4819          Rural              87.80  30.8          Unknown       1

[248 rows x 11 columns]
```

[16]:
```
x=pd.
 ↪get_dummies(data[["smoking_status","gender","ever_married","work_type","Residence_type"]],d
 ↪replace({True:1,False:0})
x            # CONVERT CATEGORICAL COLUMNS INTO NUMERICAL USING DUMMIES
```

[16]:
```
      smoking_status_formerly smoked  smoking_status_never smoked  \
0                                  1                            0
1                                  0                            1
2                                  0                            0
3                                  0                            1
4                                  1                            0
…                                  …                            …
4976                               1                            0
4977                               0                            0
4978                               0                            0
4979                               0                            0
4980                               0                            1

      smoking_status_smokes  gender_Male  ever_married_Yes  work_type_Private  \
0                         0            1                 1                  1
1                         0            1                 1                  1
2                         1            0                 1                  1
3                         0            0                 1                  0
4                         0            1                 1                  1
…                       …            …                 …                  …
```

```
4976                        0            1                 0                   1
4977                        1            1                 1                   1
4978                        1            0                 1                   0
4979                        1            1                 1                   1
4980                        0            0                 1                   1

      work_type_Self-employed  work_type_children  Residence_type_Urban
0                           0                   0                     1
1                           0                   0                     0
2                           0                   0                     1
3                           1                   0                     0
4                           0                   0                     1
…                         …                   …                     …
4976                        0                   0                     0
4977                        0                   0                     1
4978                        0                   0                     0
4979                        0                   0                     0
4980                        0                   0                     1

[4981 rows x 9 columns]
```

[17]:
```
data
```

[17]:
```
      gender   age  hypertension  heart_disease ever_married      work_type  \
0       Male  67.0             0              1          Yes        Private
1       Male  80.0             0              1          Yes        Private
2     Female  49.0             0              0          Yes        Private
3     Female  79.0             1              0          Yes  Self-employed
4       Male  81.0             0              0          Yes        Private
…      …   …           …             …           …            …
4976    Male  41.0             0              0           No        Private
4977    Male  40.0             0              0          Yes        Private
4978  Female  45.0             1              0          Yes       Govt_job
4979    Male  40.0             0              0          Yes        Private
4980  Female  80.0             1              0          Yes        Private

     Residence_type  avg_glucose_level   bmi   smoking_status  stroke
0             Urban             228.69  36.6  formerly smoked       1
1             Rural             105.92  32.5     never smoked       1
2             Urban             171.23  34.4           smokes       1
3             Rural             174.12  24.0     never smoked       1
4             Urban             186.21  29.0  formerly smoked       1
…             …                 …   …              …     …
4976          Rural              70.15  29.8  formerly smoked       0
4977          Urban             191.15  31.1           smokes       0
4978          Rural              95.02  31.8           smokes       0
4979          Rural              83.94  30.0           smokes       0
```

```
4980          Urban              83.75  29.1     never smoked       0

[4981 rows x 11 columns]
```

[18]:
```
data=pd.concat([data,x],axis=1).
 ↪drop(columns=["smoking_status","ever_married","work_type","Residence_type","gender"])
data                    # CONCATE DUMMIES DATA WITH ORIGINAL DATASET
```

[18]:
```
       age  hypertension  heart_disease  avg_glucose_level   bmi  stroke  \
0     67.0             0              1             228.69  36.6       1
1     80.0             0              1             105.92  32.5       1
2     49.0             0              0             171.23  34.4       1
3     79.0             1              0             174.12  24.0       1
4     81.0             0              0             186.21  29.0       1
…     …              …              …                 …    …      …
4976  41.0             0              0              70.15  29.8       0
4977  40.0             0              0             191.15  31.1       0
4978  45.0             1              0              95.02  31.8       0
4979  40.0             0              0              83.94  30.0       0
4980  80.0             1              0              83.75  29.1       0

      smoking_status_formerly smoked  smoking_status_never smoked  \
0                                  1                            0
1                                  0                            1
2                                  0                            0
3                                  0                            1
4                                  1                            0
…                                …                            …
4976                               1                            0
4977                               0                            0
4978                               0                            0
4979                               0                            0
4980                               0                            1

      smoking_status_smokes  gender_Male  ever_married_Yes  work_type_Private  \
0                         0            1                 1                  1
1                         0            1                 1                  1
2                         1            0                 1                  1
3                         0            0                 1                  0
4                         0            1                 1                  1
…                       …            …                 …                  …
4976                      0            1                 0                  1
4977                      1            1                 1                  1
4978                      1            0                 1                  0
4979                      1            1                 1                  1
4980                      0            0                 1                  1
```

```
      work_type_Self-employed  work_type_children  Residence_type_Urban
0                           0                   0                     1
1                           0                   0                     0
2                           0                   0                     1
3                           1                   0                     0
4                           0                   0                     1
…                         …                   …                     …
4976                        0                   0                     0
4977                        0                   0                     1
4978                        0                   0                     0
4979                        0                   0                     0
4980                        0                   0                     1

[4981 rows x 15 columns]
```

[19]:
```
data=data[['age', 'hypertension', 'heart_disease', 'avg_glucose_level', 'bmi',␣
 ↪'smoking_status_formerly smoked',
       'smoking_status_never smoked', 'smoking_status_smokes', 'gender_Male',
       'ever_married_Yes', 'work_type_Private', 'work_type_Self-employed',
       'work_type_children', 'Residence_type_Urban',
       'stroke']]
```

[20]:
```
data.corr()*100           #CHECKING FOR CORRELATION BETWEEN FEATURES AND TARGET
```

[20]:
```
                                    age  hypertension  heart_disease  \
age                          100.000000     27.811956      26.485169
hypertension                  27.811956    100.000000      11.197364
heart_disease                 26.485169     11.197364     100.000000
avg_glucose_level             23.676268     17.002767      16.684657
bmi                           37.370310     15.876244       6.092647
smoking_status_formerly smoked  23.550811      5.679747       6.754129
smoking_status_never smoked   12.261730      6.526698      -2.272694
smoking_status_smokes          7.089943      3.074894       4.401079
gender_Male                   -2.653843      2.148476       8.647553
ever_married_Yes              67.713657     16.453409      11.476489
work_type_Private             11.102048     -0.417718      -0.160001
work_type_Self-employed       32.683483     11.046797       8.747408
work_type_children           -63.686595    -12.892427      -9.297401
Residence_type_Urban           1.715450     -0.475503       0.212545
stroke                        24.647787     13.196524      13.461031

                              avg_glucose_level         bmi  \
age                                   23.676268   37.370310
hypertension                          17.002767   15.876244
heart_disease                         16.684657    6.092647
avg_glucose_level                    100.000000   18.634817
bmi                                   18.634817  100.000000
```

```
smoking_status_formerly smoked           6.698903   12.015566
smoking_status_never smoked              2.472661   10.932215
smoking_status_smokes                    1.787274   10.071028
gender_Male                              5.579594   -1.209292
ever_married_Yes                        15.072374   37.169006
work_type_Private                        2.076356   21.182000
work_type_Self-employed                  5.841942    8.558153
work_type_children                     -10.196024  -48.425704
Residence_type_Urban                     0.134561    1.318494
stroke                                  13.322733    5.692566

                                 smoking_status_formerly smoked  \
age                                                   23.550811
hypertension                                           5.679747
heart_disease                                          6.754129
avg_glucose_level                                      6.698903
bmi                                                   12.015566
smoking_status_formerly smoked                       100.000000
smoking_status_never smoked                          -35.105727
smoking_status_smokes                                -19.720833
gender_Male                                            4.510887
ever_married_Yes                                      17.203936
work_type_Private                                      2.268483
work_type_Self-employed                                9.218600
work_type_children                                   -16.130989
Residence_type_Urban                                   0.982495
stroke                                                 6.532000

                                 smoking_status_never smoked  \
age                                                12.261730
hypertension                                        6.526698
heart_disease                                      -2.272694
avg_glucose_level                                   2.472661
bmi                                                10.932215
smoking_status_formerly smoked                   -35.105727
smoking_status_never smoked                      100.000000
smoking_status_smokes                            -32.850987
gender_Male                                      -10.238666
ever_married_Yes                                  10.411985
work_type_Private                                 10.993599
work_type_Self-employed                            3.089779
work_type_children                               -23.652935
Residence_type_Urban                              -2.689246
stroke                                            -0.480609

                                 smoking_status_smokes  gender_Male  \
age                                           7.089943    -2.653843
```

hypertension	3.074894	2.148476
heart_disease	4.401079	8.647553
avg_glucose_level	1.787274	5.579594
bmi	10.071028	-1.209292
smoking_status_formerly smoked	-19.720833	4.510887
smoking_status_never smoked	-32.850987	-10.238666
smoking_status_smokes	100.000000	1.334940
gender_Male	1.334940	100.000000
ever_married_Yes	10.623449	-2.897115
work_type_Private	9.676944	-2.870645
work_type_Self-employed	-0.339571	-2.963462
work_type_children	-16.655330	9.027526
Residence_type_Urban	3.048988	-0.430116
stroke	0.856073	0.886987

	ever_married_Yes	work_type_Private	\
age	67.713657	11.102048	
hypertension	16.453409	-0.417718	
heart_disease	11.476489	-0.160001	
avg_glucose_level	15.072374	2.076356	
bmi	37.169006	21.182000	
smoking_status_formerly smoked	17.203936	2.268483	
smoking_status_never smoked	10.411985	10.993599	
smoking_status_smokes	10.623449	9.676944	
gender_Male	-2.897115	-2.870645	
ever_married_Yes	100.000000	14.613875	
work_type_Private	14.613875	100.000000	
work_type_Self-employed	19.166761	-50.945810	
work_type_children	-54.885099	-45.896796	
Residence_type_Urban	0.819076	-1.610378	
stroke	10.839811	1.045879	

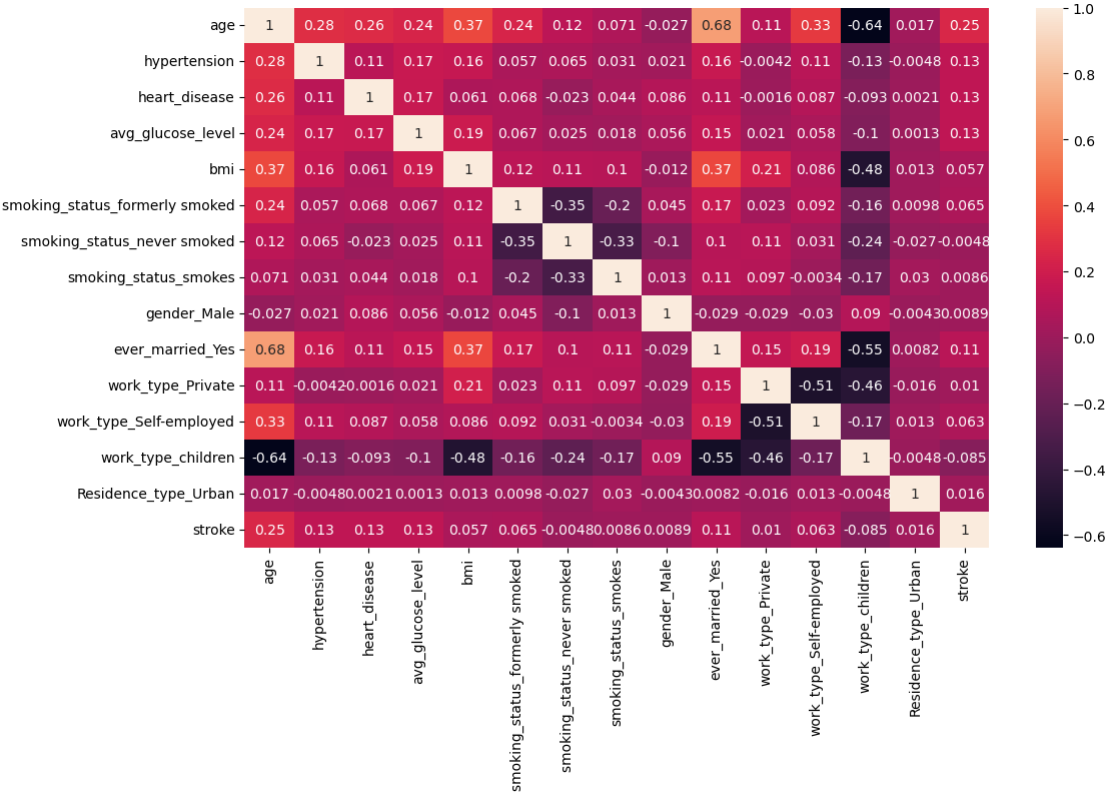
	work_type_Self-employed	work_type_children	\
age	32.683483	-63.686595	
hypertension	11.046797	-12.892427	
heart_disease	8.747408	-9.297401	
avg_glucose_level	5.841942	-10.196024	
bmi	8.558153	-48.425704	
smoking_status_formerly smoked	9.218600	-16.130989	
smoking_status_never smoked	3.089779	-23.652935	
smoking_status_smokes	-0.339571	-16.655330	
gender_Male	-2.963462	9.027526	
ever_married_Yes	19.166761	-54.885099	
work_type_Private	-50.945810	-45.896796	
work_type_Self-employed	100.000000	-17.340654	
work_type_children	-17.340654	100.000000	
Residence_type_Urban	1.342711	-0.482494	

```
stroke                                 6.264289          -8.507456
```

	Residence_type_Urban	stroke
age	1.715450	24.647787
hypertension	-0.475503	13.196524
heart_disease	0.212545	13.461031
avg_glucose_level	0.134561	13.322733
bmi	1.318494	5.692566
smoking_status_formerly smoked	0.982495	6.532000
smoking_status_never smoked	-2.689246	-0.480609
smoking_status_smokes	3.048988	0.856073
gender_Male	-0.430116	0.886987
ever_married_Yes	0.819076	10.839811
work_type_Private	-1.610378	1.045879
work_type_Self-employed	1.342711	6.264289
work_type_children	-0.482494	-8.507456
Residence_type_Urban	100.000000	1.649415
stroke	1.649415	100.000000

[21]:
```
plt.figure(figsize=(12,7))
sns.heatmap(data.corr(),annot=True)
```

[21]: <Axes: >



[22]:
```
F=data[[ 'age', 'hypertension', 'heart_disease', 'avg_glucose_level', 'bmi',
       'smoking_status_formerly smoked', 'smoking_status_never smoked',
       'smoking_status_smokes', 'gender_Male', 'ever_married_Yes',
       'work_type_Private', 'work_type_Self-employed', 'work_type_children',
       'Residence_type_Urban']]
T=data["stroke"]                                    # STORE DATA INTO FEATURES␣
 ↪AND TARGET ACCORDINGLY
```

[23]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(F,T)
                                                    # SPLIT THE DATASET INTO␣
 ↪TRAIN AND TESTING DATA
```

[24]:
```
from sklearn.preprocessing import MinMaxScaler
M=MinMaxScaler()            # IMPORT MINMAX SCALER FOR STANDARDIZATION
```

[25]:
```
x_train
```

[25]:
```
       age  hypertension  heart_disease  avg_glucose_level   bmi  \
4370  66.0             0              0              74.88  32.6
2625  70.0             1              0             118.81  26.0
1729  63.0             0              0             104.70  21.0
678    2.0             0              0              73.62  20.8
3190  71.0             1              1              67.06  26.7
…      …              …              …                …    …
2548  65.0             0              0              95.88  28.5
1297  12.0             0              0              90.42  28.9
3966  25.0             0              0             229.94  23.5
4547  13.0             0              0              73.48  22.9
1252  59.0             0              0              89.96  28.1

      smoking_status_formerly smoked  smoking_status_never smoked  \
4370                               0                            1
2625                               0                            0
1729                               1                            0
678                                0                            0
3190                               0                            0
…                                …                            …
2548                               0                            1
1297                               0                            0
3966                               0                            1
4547                               0                            0
1252                               0                            0

      smoking_status_smokes  gender_Male  ever_married_Yes  work_type_Private  \
4370                      0            0                 1                  0
```

```
2625                          1                1                    1                    0
1729                          0                1                    1                    0
678                           0                1                    0                    0
3190                          1                1                    1                    0
…                           …                …                    …                    …
2548                          0                1                    1                    0
1297                          0                1                    0                    0
3966                          0                1                    0                    1
4547                          0                0                    0                    0
1252                          0                0                    1                    1

      work_type_Self-employed  work_type_children  Residence_type_Urban
4370                        1                   0                     0
2625                        1                   0                     0
1729                        1                   0                     0
678                         0                   1                     0
3190                        1                   0                     0
…                         …                   …                     …
2548                        1                   0                     1
1297                        0                   1                     0
3966                        0                   0                     0
4547                        0                   1                     0
1252                        0                   0                     1

[3735 rows x 14 columns]
```

[26]:
```
x_train[['avg_glucose_level','bmi']]=M.
 ↪fit_transform(x_train[['avg_glucose_level','bmi']])
x_test[['avg_glucose_level','bmi']]=M.
 ↪transform(x_test[['avg_glucose_level','bmi']])
    # FIT THE DATA INTO MODEL AND DO THE STANDARDIZATION
```

[27]:
```
x_train
```

[27]:
```
        age  hypertension  heart_disease  avg_glucose_level       bmi  \
4370  66.0             0              0           0.091220  0.532951
2625  70.0             1              0           0.294017  0.343840
1729  63.0             0              0           0.228880  0.200573
678    2.0             0              0           0.085403  0.194842
3190  71.0             1              1           0.055120  0.363897
…     …              …              …                   …       …
2548  65.0             0              0           0.188164  0.415473
1297  12.0             0              0           0.162958  0.426934
3966  25.0             0              0           0.807035  0.272206
4547  13.0             0              0           0.084757  0.255014
1252  59.0             0              0           0.160835  0.404011
```

```
      smoking_status_formerly smoked  smoking_status_never smoked  \
4370                                0                            1
2625                                0                            0
1729                                1                            0
678                                 0                            0
3190                                0                            0
…                                 …                            …
2548                                0                            1
1297                                0                            0
3966                                0                            1
4547                                0                            0
1252                                0                            0

      smoking_status_smokes  gender_Male  ever_married_Yes  work_type_Private  \
4370                      0            0                 1                  0
2625                      1            1                 1                  0
1729                      0            1                 1                  0
678                       0            1                 0                  0
3190                      1            1                 1                  0
…                       …            …                 …                  …
2548                      0            1                 1                  0
1297                      0            1                 0                  0
3966                      0            1                 0                  1
4547                      0            0                 0                  0
1252                      0            0                 1                  1

      work_type_Self-employed  work_type_children  Residence_type_Urban
4370                        1                   0                    0
2625                        1                   0                    0
1729                        1                   0                    0
678                         0                   1                    0
3190                        1                   0                    0
…                         …                   …                    …
2548                        1                   0                    1
1297                        0                   1                    0
3966                        0                   0                    0
4547                        0                   1                    0
1252                        0                   0                    1

[3735 rows x 14 columns]
```

[28]:
```
x_test
```

[28]:
```
       age  hypertension  heart_disease  avg_glucose_level       bmi  \
3512  19.0             0              0           0.297572  0.309456
4717  38.0             0              0           0.264380  0.352436
2453  42.0             0              0           0.068646  0.916905
```

```
3784   2.0             0              0           0.157880  0.091691
1974  29.0             0              0           0.033930  0.584527
…      …              …              …                …       …
563   40.0             0              0           0.074231  0.375358
3088  47.0             0              0           0.078848  0.544413
1102  14.0             0              0           0.135445  0.223496
1005  18.0             0              0           0.225372  0.266476
96    61.0             1              0           0.096898  0.381089

      smoking_status_formerly smoked  smoking_status_never smoked  \
3512                                0                            0
4717                                0                            0
2453                                0                            0
3784                                0                            0
1974                                1                            0
…                                   …                            …
563                                 0                            1
3088                                0                            0
1102                                0                            0
1005                                0                            1
96                                  0                            0

      smoking_status_smokes  gender_Male  ever_married_Yes  work_type_Private  \
3512                      0            1                 0                  1
4717                      0            1                 0                  1
2453                      1            0                 1                  1
3784                      0            1                 0                  0
1974                      0            0                 0                  0
…                         …            …                 …                  …
563                       0            0                 1                  1
3088                      0            1                 1                  1
1102                      0            0                 0                  1
1005                      0            1                 0                  1
96                        1            1                 1                  0

      work_type_Self-employed  work_type_children  Residence_type_Urban
3512                        0                   0                    1
4717                        0                   0                    0
2453                        0                   0                    0
3784                        0                   1                    1
1974                        1                   0                    0
…                           …                   …                    …
563                         0                   0                    1
3088                        0                   0                    0
1102                        0                   0                    0
1005                        0                   0                    0
96                          1                   0                    0
```

```
[1246 rows x 14 columns]
```

0.1 LOGISTIC REGRESSION:-

[29]:
```
from sklearn.linear_model import LogisticRegression
L=LogisticRegression()                                    #IMPORT LOGISTIC␣
 ↪REGRESSION AND FIT
L.fit(x_train,y_train)
```

[29]:
```
LogisticRegression()
```

[30]:
```
L1=L.score(x_train,y_train)*100
L1                                  #TRAINING ACCURACY
```

[30]:
```
95.1004016064257
```

[31]:
```
L2=L.score(x_test,y_test)*100
L2                                 #TESTING ACCURACY
```

[31]:
```
94.86356340288926
```

0.2 SVC:-

[32]:
```
from sklearn.svm import SVC
S=SVC()                                        #IMPORT SVC AND FIT
S.fit(x_train,y_train)
```

[32]:
```
SVC()
```

[33]:
```
S1=S.score(x_train,y_train)*100
S1                                  #TRAINING ACCURACY
```

[33]:
```
95.07362784471218
```

[34]:
```
S2=S.score(x_test,y_test)*100
S2                                  #TESTING ACCURACY
```

[34]:
```
94.86356340288926
```

0.3 NAIVES BAYES:-

[35]:
```
from sklearn.naive_bayes import␣
 ↪GaussianNB,ComplementNB,MultinomialNB,BernoulliNB
G=GaussianNB()
C=ComplementNB()
```

```
M=MultinomialNB()
B=BernoulliNB()                                    #IMPORT NAIVES BAYES AND␣
 ↪FIT
```

0.3.1 GaussianNB:

[36]:
```
G.fit(x_train,y_train)
```

[36]:
```
GaussianNB()
```

[37]:
```
G1=G.score(x_train,y_train)*100
G1                                #TRAINING ACCURACY
```

[37]:
```
83.02543507362785
```

[38]:
```
G2=G.score(x_test,y_test)*100     #TESTING ACCURACY
G2
```

[38]:
```
81.46067415730337
```

0.3.2 BernoulliNB:

[39]:
```
B.fit(x_train,y_train)
```

[39]:
```
BernoulliNB()
```

[40]:
```
B1=B.score(x_train,y_train)*100      #TRAINING ACCURACY
B1
```

[40]:
```
94.80589022757697
```

[41]:
```
B2=B.score(x_test,y_test)*100        #TESTING ACCURACY
B2
```

[41]:
```
94.30176565008026
```

[ ]:
```

```

0.3.3 ComplementNB:-

[42]:
```
C.fit(x_train,y_train)
```

[42]:
```
ComplementNB()
```

[43]:
```
C1=C.score(x_train,y_train)*100       #TRAINING ACCURACY
C1
```

[43]: 64.28380187416332

[44]:
```
C2=C.score(x_test,y_test)*100        #TESTING ACCURACY
C2
```

[44]: 60.75441412520064

[ ]:

0.3.4 MultinomialNB:-

[45]:
```
M.fit(x_train,y_train)
```

[45]: MultinomialNB()

[46]:
```
M1=M.score(x_train,y_train)*100       #TRAINING ACCURACY
M1
```

[46]: 95.07362784471218

[47]:
```
M2=M.score(x_test,y_test)*100       #TESTING ACCURACY
M2
```

[47]: 94.86356340288926

0.4 K NEAREST NEIGHBOUR:-

[48]:
```
from sklearn.neighbors import KNeighborsClassifier
K=KNeighborsClassifier()                                    #IMPORT K NEAREST␣
 ↪NEIGHBOUR AND FIT
```

[49]:
```
K.fit(x_train,y_train)
```

[49]: KNeighborsClassifier()

[50]:
```
K1=K.score(x_train,y_train)*100      #TRAINING ACCURACY
K1
```

[50]: 95.18072289156626

[51]:
```
K2=K.score(x_test,y_test)*100
K2                                  #TESTING ACCURACY
```

[51]: 94.70304975922953

[ ]:

0.5 DECISION TREE CLASSIFIER:-

[52]:
```
from sklearn.tree import DecisionTreeClassifier
D=DecisionTreeClassifier()                          #IMPORT DECISION TREE␣
 ↪CLASSIFIER AND FIT
```

[53]:
```
D.fit(x_train,y_train)
```

[53]:
```
DecisionTreeClassifier()
```

[54]:
```
DT1=D.score(x_train,y_train)*100          #TRAINING ACCURACY
DT1
```

[54]:
```
100.0
```

[55]:
```
DT2=D.score(x_test,y_test)*100
DT2                                       #TESTING ACCURACY
```

[55]:
```
90.36918138041734
```

0.6 ACCURACY GRAPH:-

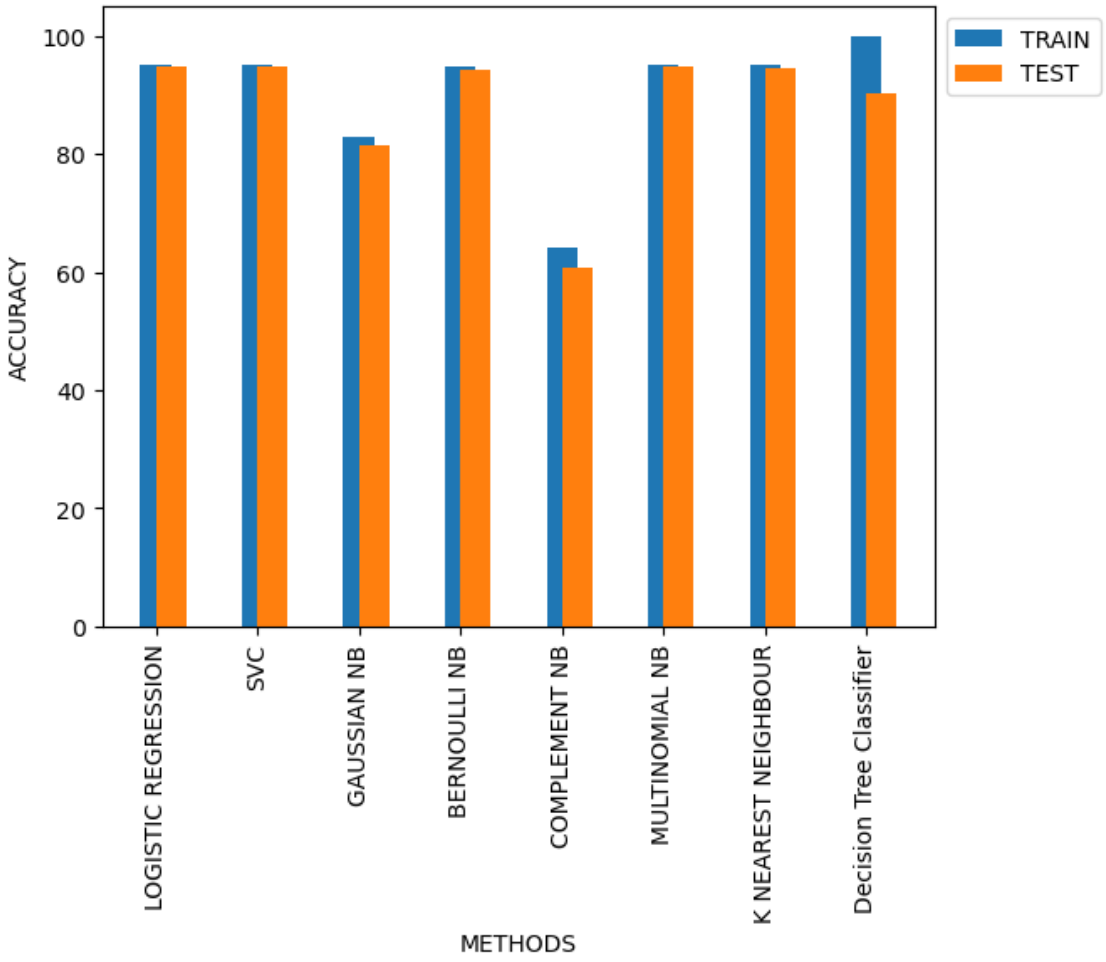
[58]:
```
A={"METHODS":["LOGISTIC REGRESSION","SVC","GAUSSIAN NB","BERNOULLI␣
 ↪NB","COMPLEMENT NB","MULTINOMIAL NB","K NEAREST NEIGHBOUR","DECISION TREE␣
 ↪CLASSIFIER"],"TRAIN ACCURACY":[L1,S1,G1,B1,C1,M1,K1,DT1],"TEST ACCURACY":
 ↪[L2,S2,G2,B2,C2,M2,K2,DT2]}
A=pd.DataFrame(A)
A=np.around(A,2)
A
```

[58]:

	METHODS	TRAIN ACCURACY	TEST ACCURACY
0	LOGISTIC REGRESSION	95.10	94.86
1	SVC	95.07	94.86
2	GAUSSIAN NB	83.03	81.46
3	BERNOULLI NB	94.81	94.30
4	COMPLEMENT NB	64.28	60.75
5	MULTINOMIAL NB	95.07	94.86
6	K NEAREST NEIGHBOUR	95.18	94.70
7	DECISION TREE CLASSIFIER	100.00	90.37

[57]:
```
plt.bar(A["METHODS"],A["TRAIN ACCURACY"],width=0.3,label="TRAIN")
plt.bar(A["METHODS"],A["TEST ACCURACY"],align="edge",width=0.3,label="TEST")
plt.legend(bbox_to_anchor=[1,0,0,1])
plt.xlabel("METHODS")
plt.ylabel("ACCURACY")
plt.xticks(rotation=90)
```

```
plt.show()                          # PLOT THE ACCURACY CHART BETWEEN ALL MODELS␣
 ↪ACCURACIES
```



0.7 CONCLUSION:

0.7.1 FROM THE ABOVE BAR CHART IT IS CLEAR THAT SVC IS BEST FOR CLASSIFICATION FOR THIS DATASET WITH 95% ACCURACY.

[ ]:

[ ]: