Stroke Disease Detection and Prediction

Using Robust learning approaches

Team members -

1) Sreenidhi Ganachari - 19BCE7230

2) Sandeep Chowdary - 19BCE7042

3) Lokesh Bandaru - 19BCD7111

4) Gopikrishna Kamepalli - 19BCE7204

Guided by - Dr . Srinivas Rao Battula

Objective

Stroke is a medical disorder in which blood arteries in brain are ruptured causing damage to brain .Stroke is the greatest cause of death or disability.Early detection and appropriate measures are best to prevent further damage to affected area of brain or other parts in body . The dataset contains 5110 observations and 12 attributes .

The 12 attributes used in prediction are :-

1)id

2)gender

3)age

4)hypertension

4)heart_disease

6) ever _ married

7)work_type

8)Residence_type-rural, urban

9)avg_glucose_level

10)BMI

11)smoking_status

12)stroke _ 1 or 0

Machine Learning Algorithms executed

1)K means Algorithm

2)Naive Bayes Algorithm

3)KNN Algorithm

4)SVM (sigmoid) Algorithm

5)Random forest Algorithm

6)Logistic Regression Algorithm

7)Voting classifier Algorithm

8)SVM (linear) Algorithm

9)Decision Tree Algorithm

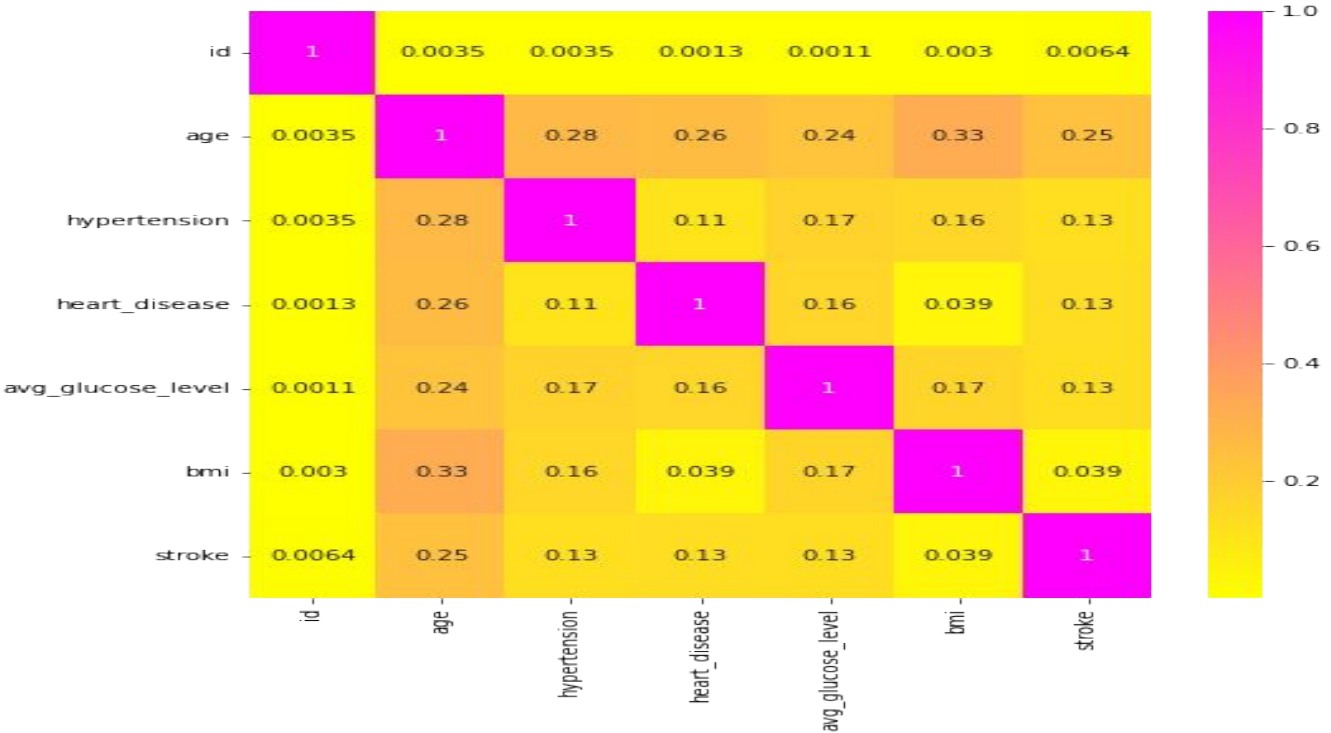
10)AdaBoost Algorithm

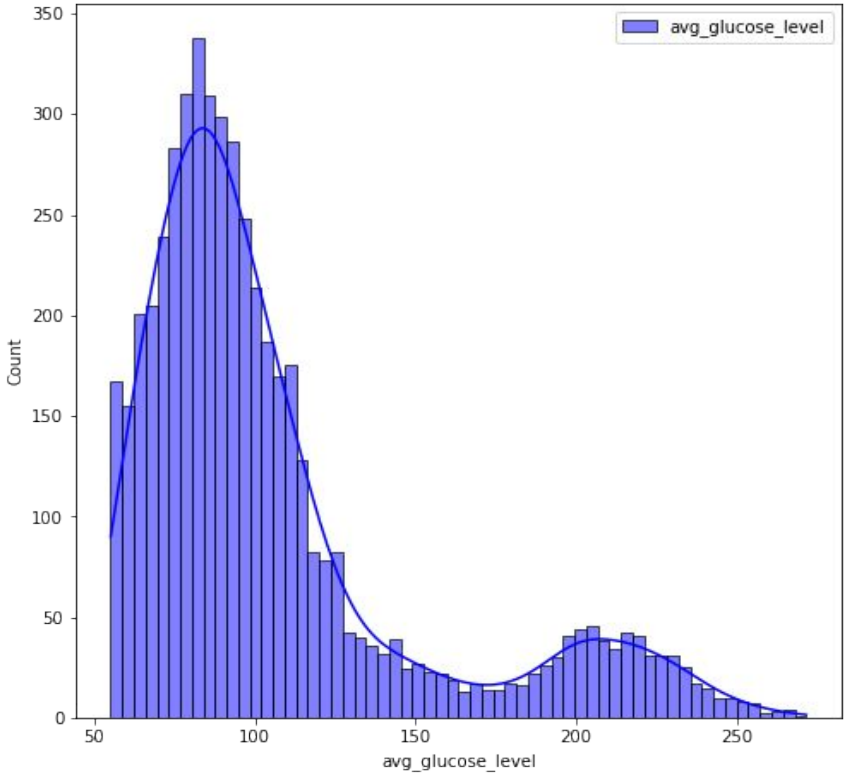
Comparison table

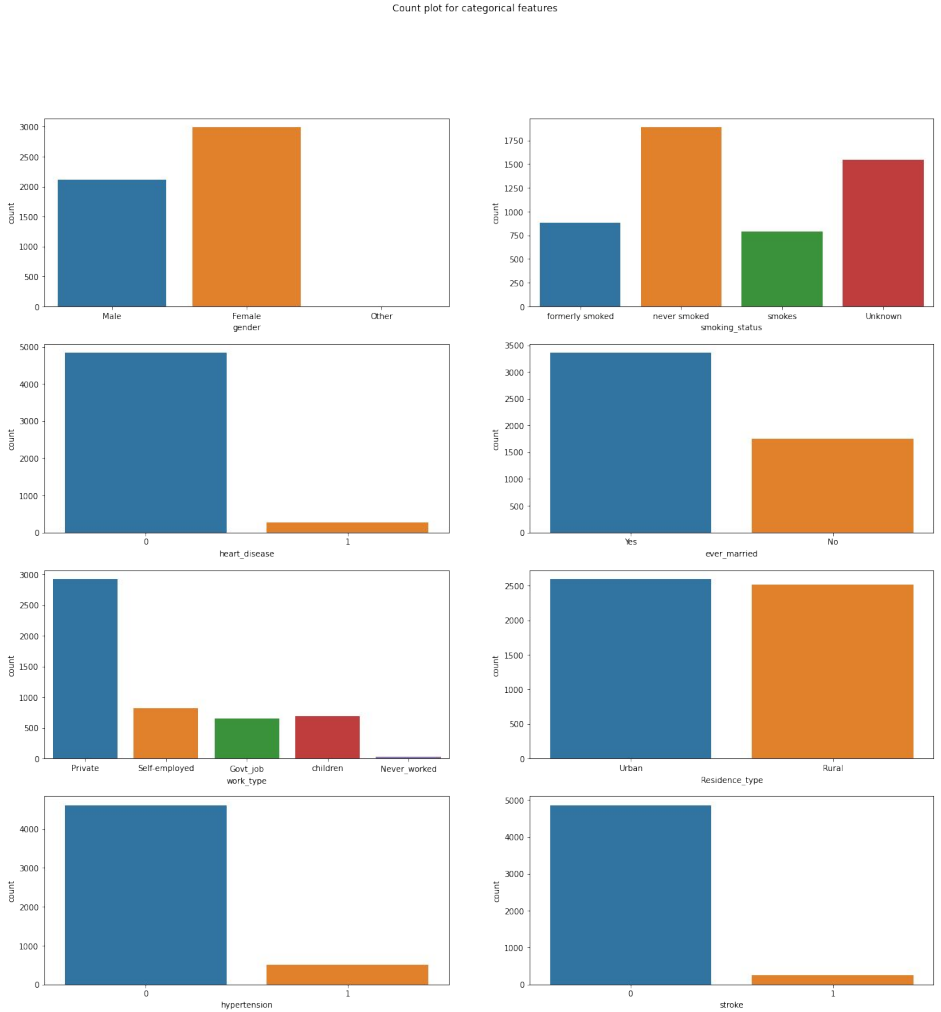
S.No	Algorithm	Executed	Research Paper
1	Random forest algorithm	92.14	96
2	Decision Tree	91.52	94
3	SVM Linear	94.13	Not implemented
4	Logistic Regression	92.88	79
5	Voting Classifier	95	91
6	KNN	95.1309	Not implemented
7	SVM Sigmoid	92.43	Not implemented
8	K- means Model	78.34	Not implemented
9	Naives Bayes	87.28	Not implemented
10	Ada Boost	98.8	Not implemented

Obtained outputs of algorithms

1) Random forest



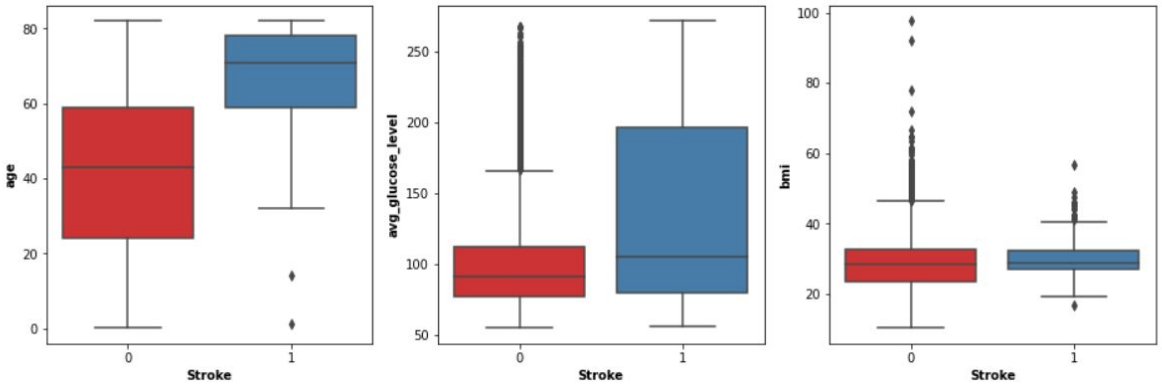

avg_glucose_level
Count
avg_glucose_level


Count plot for categorical features
count
Male
Female
Other
gender
formerly smoked
never smoked
smokes
Unknown
smoking_status
0
1
heart_disease
Yes
No
ever_married
Private
Self-employed
Govt_job
children
Never_worked
work_type
Urban
Rural
Residence_type
hypertension
stroke

```
[ ]  model.score(x_test,y_test)
```

```
0.9214953271028037
```

2) Decision tree



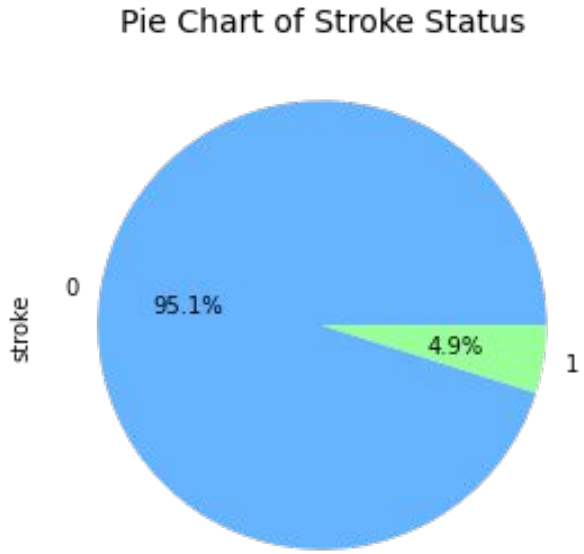

```
--------------------Decision Tree--------------------
[[1396   68]
 [  62    7]]
              precision    recall  f1-score   support

           0       0.96      0.95      0.96      1464
           1       0.09      0.10      0.10        69

    accuracy                           0.92      1533
   macro avg       0.53      0.53      0.53      1533
weighted avg       0.92      0.92      0.92      1533
```

3)SVM Linear

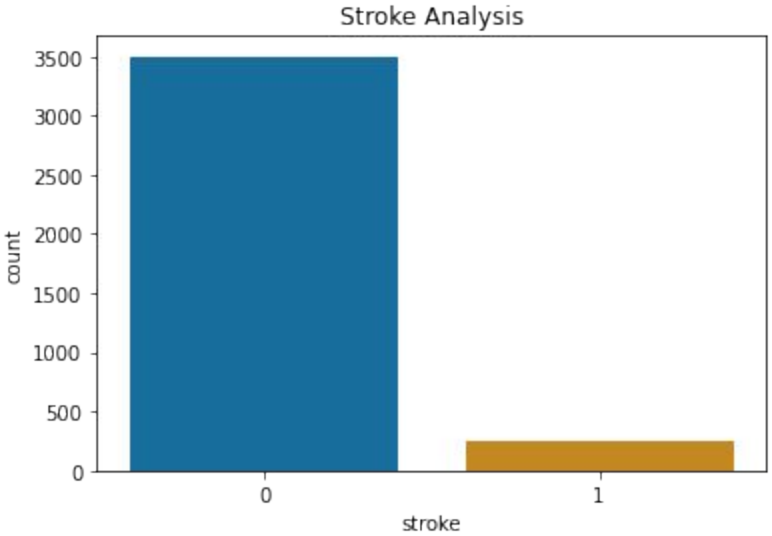
Accuracy score 0.941341

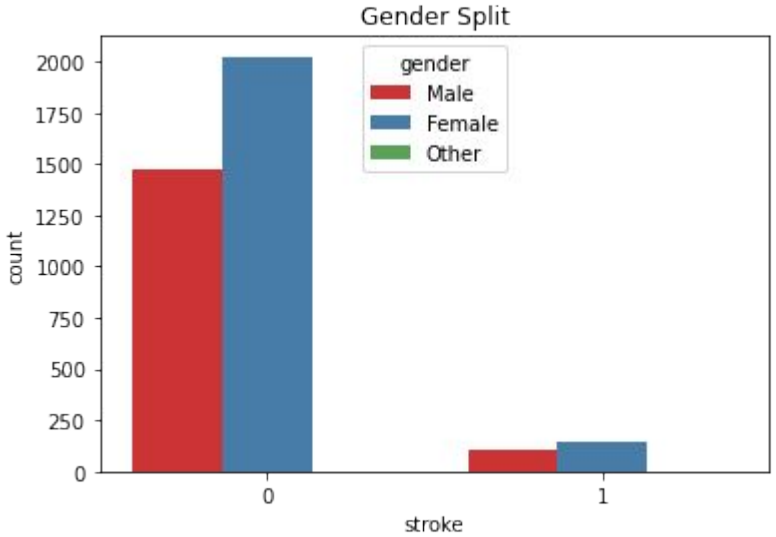
```
[ ]  acc = accuracy_score(y_test, y_pred)
     acc
```

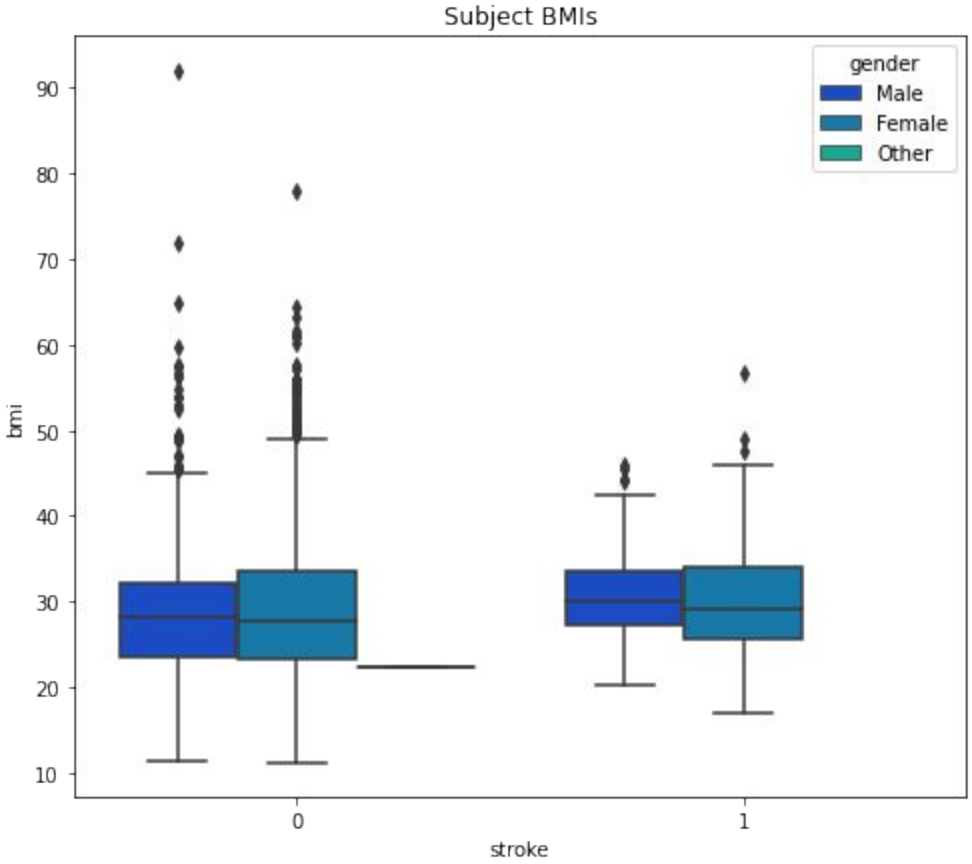
```
0.9413145539906104
```

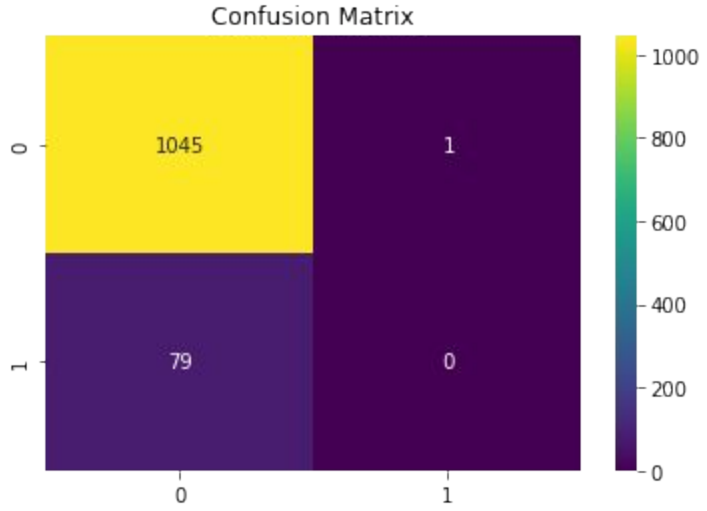

Pie Chart of Stroke Status
stroke
0
95.1%
4.9%
1

4) logistic regression

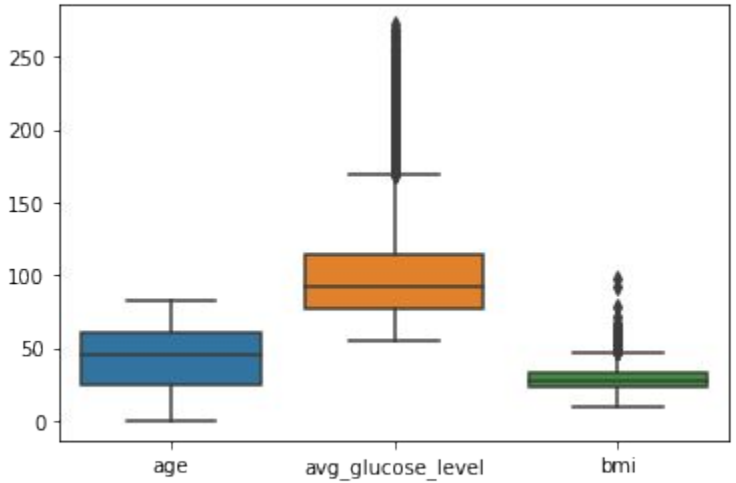


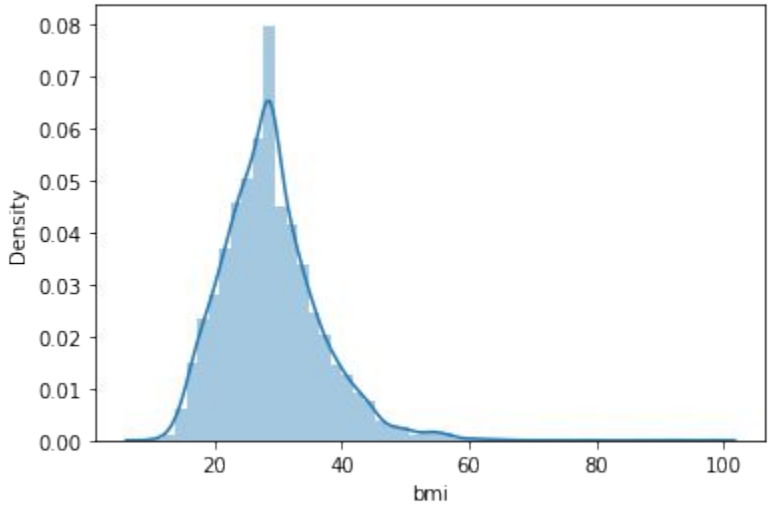


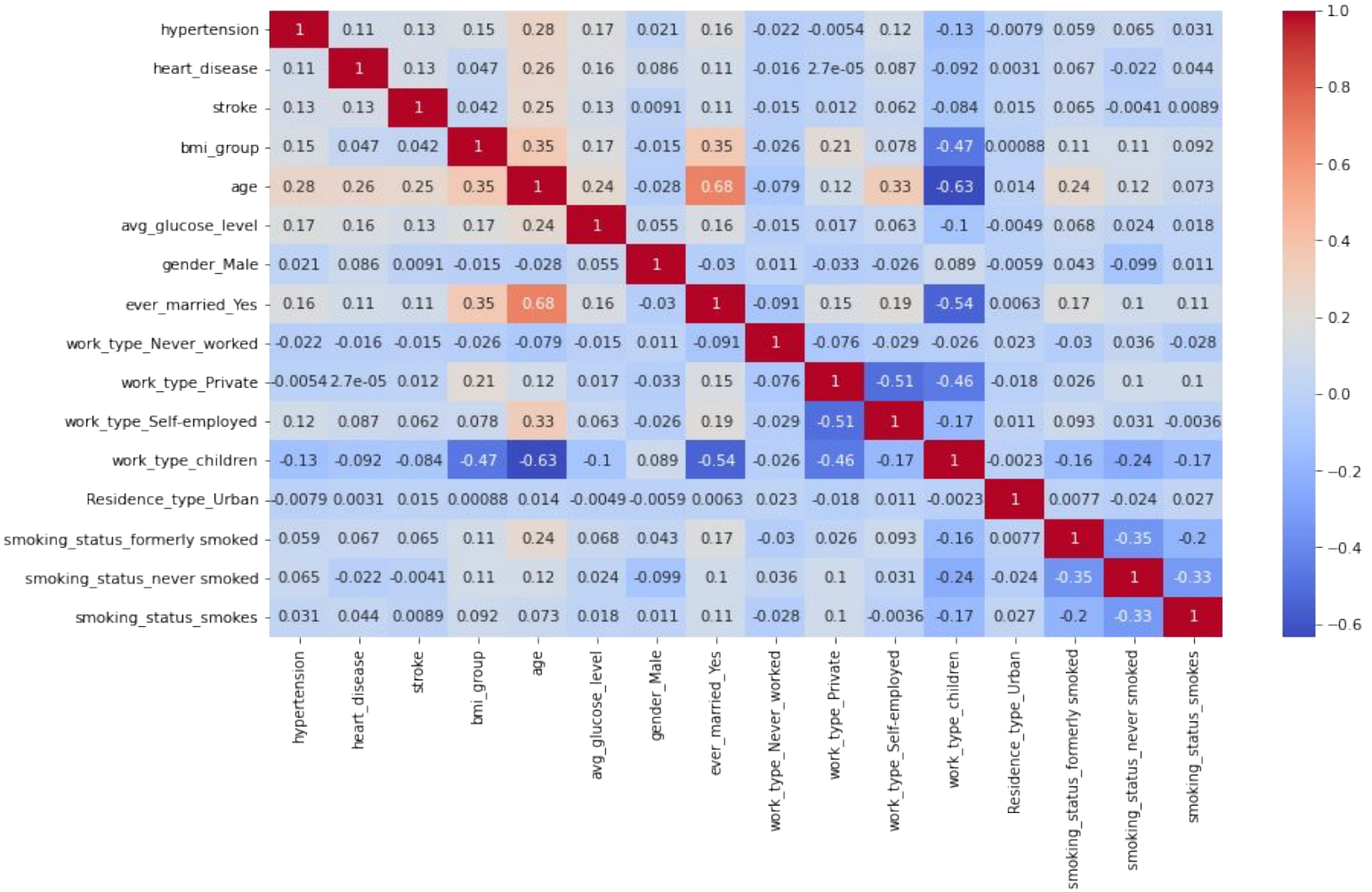

Subject BMIs
gender
Male
Female
Other
bmi
stroke
0
1
10
20
30
40
50
60
70
80
90


Confusion Matrix
1045
1
79
0
0
1
0
1
0
200
400
600
800
1000

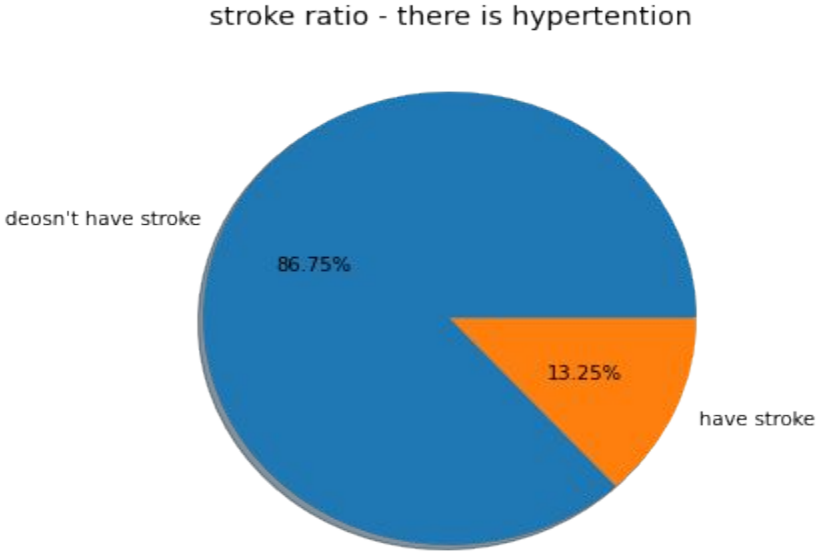
5. Voting classifier

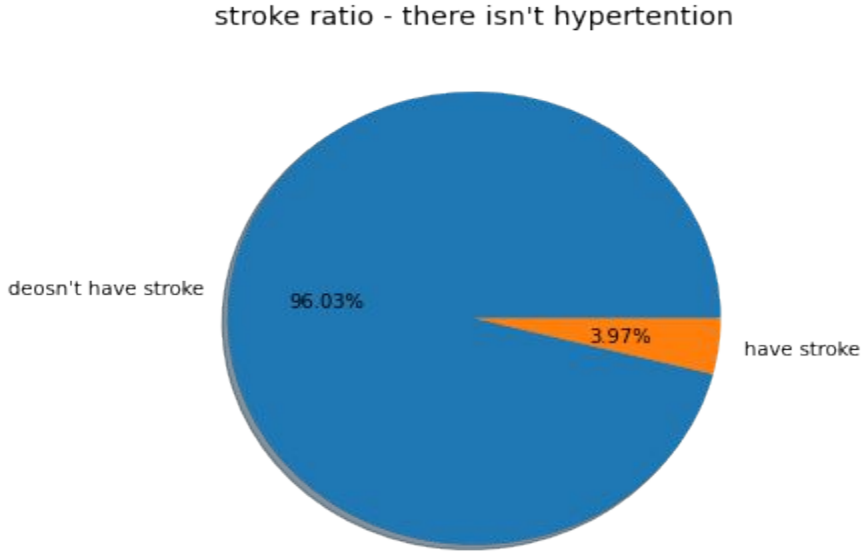


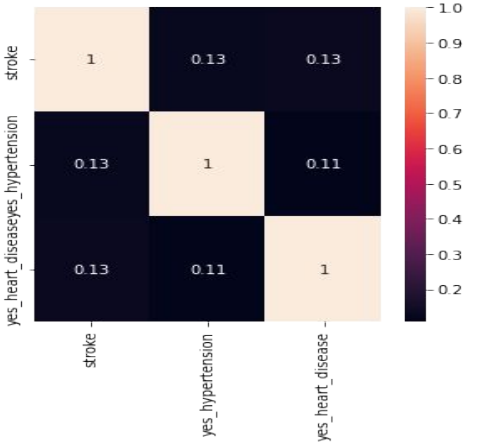


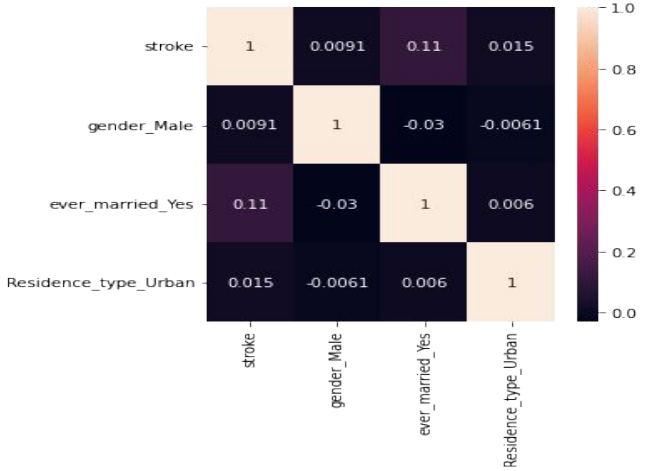

hypertension 1 0.11 0.13 0.15 0.28 0.17 0.021 0.16 -0.022 -0.0054 0.12 -0.13 -0.0079 0.059 0.065 0.031
heart_disease 0.11 1 0.13 0.047 0.26 0.16 0.086 0.11 -0.016 2.7e-05 0.087 -0.092 0.0031 0.067 -0.022 0.044
stroke 0.13 0.13 1 0.042 0.25 0.13 0.0091 0.11 -0.015 0.012 0.062 -0.084 0.015 0.065 -0.0041 0.0089
bmi_group 0.15 0.047 0.042 1 0.35 0.17 -0.015 0.35 -0.026 0.21 0.078 -0.47 0.00088 0.11 0.11 0.092
age 0.28 0.26 0.25 0.35 1 0.24 -0.028 0.68 -0.079 0.12 0.33 -0.63 0.014 0.24 0.12 0.073
avg_glucose_level 0.17 0.16 0.13 0.17 0.24 1 0.055 0.16 -0.015 0.017 0.063 -0.1 -0.0049 0.068 0.024 0.018
gender_Male 0.021 0.086 0.0091 -0.015 -0.028 0.055 1 -0.03 0.011 -0.033 -0.026 0.089 -0.0059 0.043 -0.099 0.011
ever_married_Yes 0.16 0.11 0.11 0.35 0.68 0.16 -0.03 1 -0.091 0.15 0.19 -0.54 0.0063 0.17 0.1 0.11
work_type_Never_worked -0.022 -0.016 -0.015 -0.026 -0.079 -0.015 0.011 -0.091 1 -0.076 -0.029 -0.026 0.023 -0.03 0.036 -0.028
work_type_Private -0.0054 2.7e-05 0.012 0.21 0.12 0.017 -0.033 0.15 -0.076 1 -0.51 -0.46 -0.018 0.026 0.1 0.1
work_type_Self-employed 0.12 0.087 0.062 0.078 0.33 0.063 -0.026 0.19 -0.029 -0.51 1 -0.17 0.011 0.093 0.031 -0.0036
work_type_children -0.13 -0.092 -0.084 -0.47 -0.63 -0.1 0.089 -0.54 -0.026 -0.46 -0.17 1 -0.0023 -0.16 -0.24 -0.17
Residence_type_Urban -0.0079 0.0031 0.015 0.00088 0.014 -0.0049 -0.0059 0.0063 0.023 -0.018 0.011 -0.0023 1 0.0077 -0.024 0.027
smoking_status_formerly smoked 0.059 0.067 0.065 0.11 0.24 0.068 0.043 0.17 -0.03 0.026 0.093 -0.16 0.0077 1 -0.35 -0.2
smoking_status_never smoked 0.065 -0.022 -0.0041 0.11 0.12 0.024 -0.099 0.1 0.036 0.1 0.031 -0.24 -0.024 -0.35 1 -0.33
smoking_status_smokes 0.031 0.044 0.0089 0.092 0.073 0.018 0.011 0.11 -0.028 0.1 -0.0036 -0.17 0.027 -0.2 -0.33 1
hypertension
heart_disease
stroke
bmi_group
age
avg_glucose_level
gender_Male
ever_married_Yes
work_type_Never_worked
work_type_Private
work_type_Self-employed
work_type_children
Residence_type_Urban
smoking_status_formerly smoked
smoking_status_never smoked
smoking_status_smokes
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

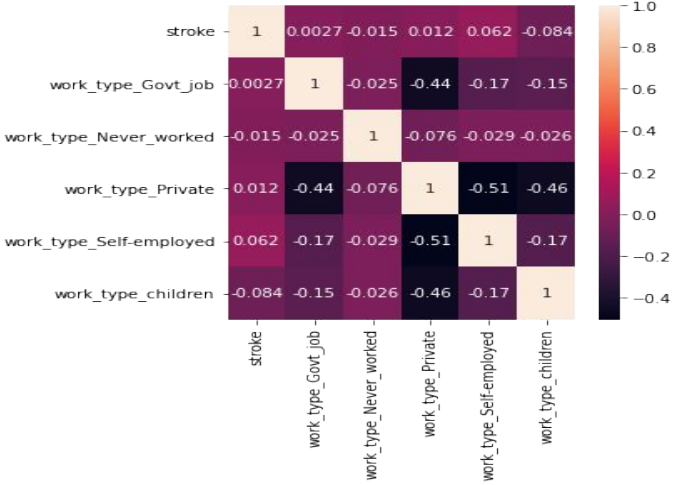
6)KNN

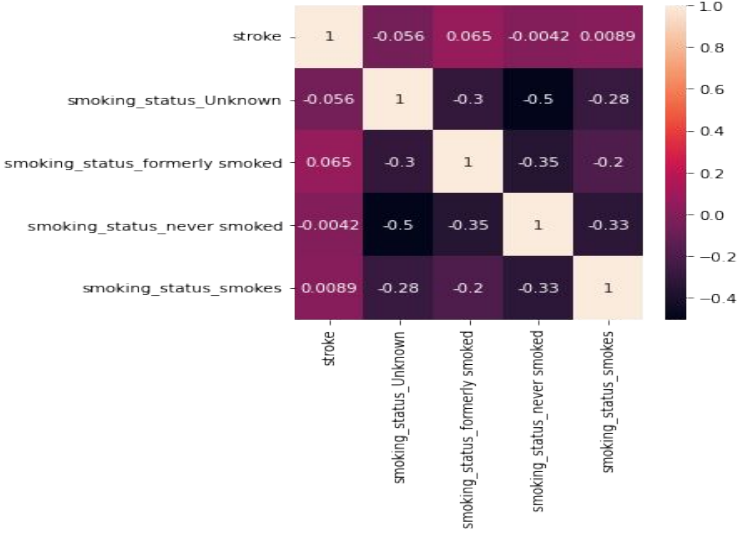

stroke ratio - there is hypertention
deosn't have stroke
86.75%
13.25%
have stroke

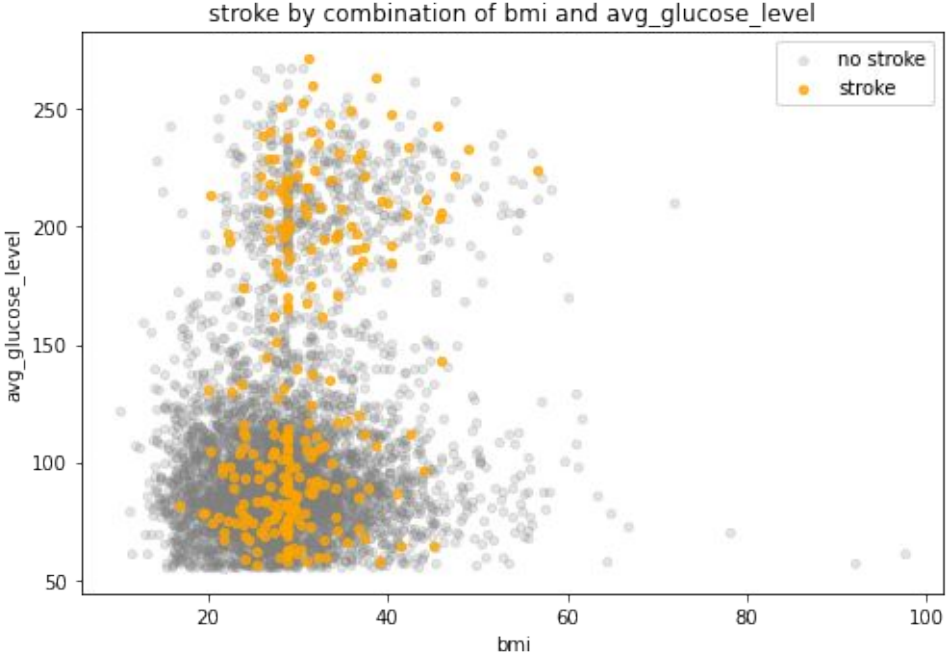

stroke ratio - there isn't hypertention
deosn't have stroke
96.03%
3.97%
have stroke

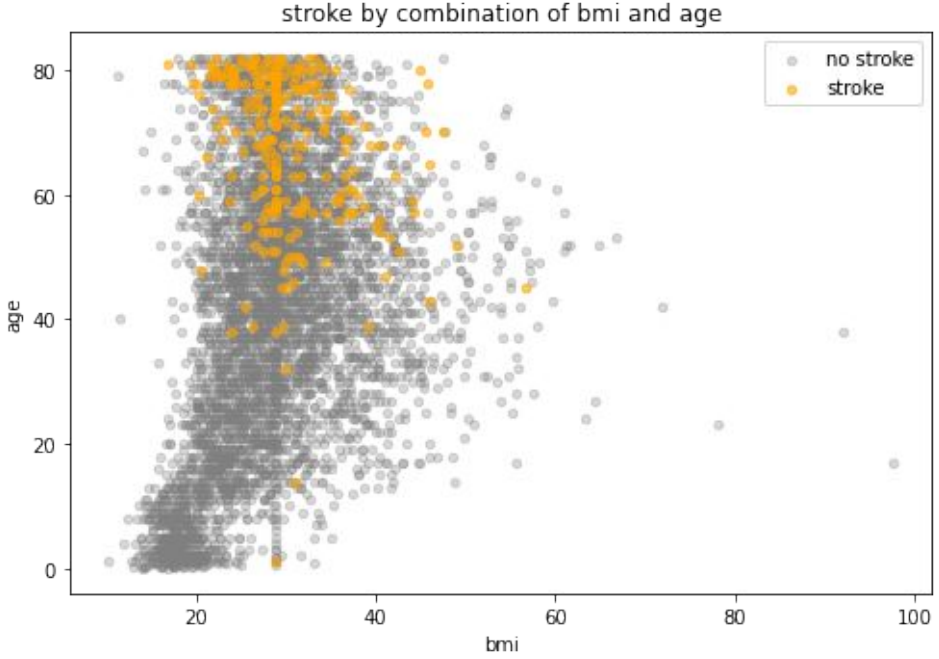

stroke
yes_hypertension
yes_heart_disease
1
0.13
0.13
0.13
1
0.11
0.13
0.11
1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2


stroke
gender_Male
ever_married_Yes
Residence_type_Urban
1
0.0091
0.11
0.015
0.0091
1
-0.03
-0.0061
0.11
-0.03
1
0.006
0.015
-0.0061
0.006
1
1.0
0.8
0.6
0.4
0.2
0.0


stroke
work_type_Govt_job
work_type_Never_worked
work_type_Private
work_type_Self-employed
work_type_children
1 0.0027 -0.015 0.012 0.062 -0.084
0.0027 1 -0.025 -0.44 -0.17 -0.15
-0.015 -0.025 1 -0.076 -0.029 -0.026
0.012 -0.44 -0.076 1 -0.51 -0.46
0.062 -0.17 -0.029 -0.51 1 -0.17
-0.084 -0.15 -0.026 -0.46 -0.17 1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4


stroke
smoking_status_Unknown
smoking_status_formerly smoked
smoking_status_never smoked
smoking_status_smokes
1 -0.056 0.065 -0.0042 0.0089
-0.056 1 -0.3 -0.5 -0.28
0.065 -0.3 1 -0.35 -0.2
-0.0042 -0.5 -0.35 1 -0.33
0.0089 -0.28 -0.2 -0.33 1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4


stroke by combination of bmi and avg_glucose_level
no stroke
stroke
avg_glucose_level
250
200
150
100
50
20
40
60
80
100
bmi


stroke by combination of bmi and age
no stroke
stroke
age
80
60
40
20
0
20
40
60
80
100
bmi

7) SVM sigmoid

```
--------------------Support Vector Machine--------------------
[[1412   52]
 [  64    5]]
              precision    recall  f1-score   support

           0       0.96      0.96      0.96      1464
           1       0.09      0.07      0.08        69

    accuracy                           0.92      1533
   macro avg       0.52      0.52      0.52      1533
weighted avg       0.92      0.92      0.92      1533
```

8) k means model

```
--------------------KMeans--------------------
[[1199  265]
 [  67    2]]
              precision    recall  f1-score   support

           0       0.95      0.82      0.88      1464
           1       0.01      0.03      0.01        69

    accuracy                           0.78      1533
   macro avg       0.48      0.42      0.45      1533
weighted avg       0.90      0.78      0.84      1533
```

9) Naive bayes

```
--------------------Naive Bayes--------------------
[[1310  154]
 [  41   28]]
              precision    recall  f1-score   support

           0       0.97      0.89      0.93      1464
           1       0.15      0.41      0.22        69

    accuracy                           0.87      1533
   macro avg       0.56      0.65      0.58      1533
weighted avg       0.93      0.87      0.90      1533
```

10) Ada boost

Accuracy 0.9885714285714285

```
# calculate and print model accuracy
print("Model Accuracy with SVC Base Estimator:",accuracy_score(y_test, y_pred))
```

```
Model Accuracy with SVC Base Estimator: 0.9885714285714285
```

Summary of Accuracy Score

Decision Tree Model: 0.9152
Logreg Model: 0.955
Random Forest Model: 0.953
Support Vector Machine Model: 0.9243
kNN Model: 0.9524
Naive Bayes Model: 0.8728
KMeans Model: 0.7834

THANK YOU