


```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings as fn
fn.filterwarnings("ignore")
```

```
df=pd.read_csv("values.csv")
```


```
df
```



	patient_id	slope_of_peak_exercise_st_segment	thal	resting_blood_pressure	chest_pain_type	num_major_vessels	fast
0	0z64un	1	normal	128	2	0	
1	ryoo3j	2	normal	110	3	0	
2	yt1s1x	1	normal	125	4	3	
3	l2xjde	1	reversible_defect	152	4	0	
4	oyt4ek	3	reversible_defect	178	1	0	
...	
175	5qfar3	2	reversible_defect	125	4	2	
176	2s2b1f	2	normal	180	4	0	
177	nsd00i	2	reversible_defect	125	3	0	
178	0xw93k	1	normal	124	3	2	
179	2nx10r	1	normal	160	3	1	


180 rows × 14 columns

```
df.head(10)
```



	patient_id	slope_of_peak_exercise_st_segment	thal	resting_blood_pressure	chest_pain_type	num_major_vessels	fastin
0	0z64un	1	normal	128	2	0	
1	ryoo3j	2	normal	110	3	0	
2	yt1s1x	1	normal	125	4	3	
3	l2xjde	1	reversible_defect	152	4	0	
4	oyt4ek	3	reversible_defect	178	1	0	
5	ldukkw	1	normal	130	3	0	
6	2gbyh9	2	reversible_defect	150	4	2	
7	daa9kp	2	fixed_defect	150	4	1	
8	3nwy2n	3	reversible_defect	170	4	0	
9	1r508r	2	normal	120	3	0	

```
df.tail(10)
```



	patient_id	slope_of_peak_exercise_st_segment	thal	resting_blood_pressure	chest_pain_type	num_major_vessels	fast
170	qhz9ye	1	reversible_defect	150	4	0	
171	u25507	1	normal	112	4	1	
172	j9tw19	2	reversible_defect	118	4	0	
173	5o32oi	1	reversible_defect	140	4	0	
174	o63ri2	1	normal	140	4	0	
175	5qfar3	2	reversible_defect	125	4	2	
176	2s2b1f	2	normal	180	4	0	
177	nsd00i	2	reversible_defect	125	3	0	
178	0xw93k	1	normal	124	3	2	
179	2nx10r	1	normal	160	3	1	

```
import sweetviz as sv
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 14 columns):
 #   Column                                Non-Null Count  Dtype
---  -
 0   patient_id                           180 non-null    object
 1   slope_of_peak_exercise_st_segment    180 non-null    int64
 2   thal                                  180 non-null    object
 3   resting_blood_pressure               180 non-null    int64
 4   chest_pain_type                     180 non-null    int64
 5   num_major_vessels                   180 non-null    int64
 6   fasting_blood_sugar_gt_120_mg_per_dl 180 non-null    int64
 7   resting_ekg_results                 180 non-null    int64
 8   serum_cholesterol_mg_per_dl         180 non-null    int64
 9   oldpeak_eq_st_depression            180 non-null    float64
10   sex                                  180 non-null    int64
11   age                                  180 non-null    int64
12   max_heart_rate_achieved              180 non-null    int64
13   exercise_induced_angina             180 non-null    int64
dtypes: float64(1), int64(11), object(2)
memory usage: 19.8+ KB
```

```
df.describe()
```

```

      slope_of_peak_exercise_st_segment  resting_blood_pressure  chest_pain_type  num_major_vessels  fasting_blood_sugar_gt_120_mg_
count                                180.000000                180.000000          180.000000          180.000000          180.
mean                                 1.550000                131.311111           3.155556           0.694444           0.
std                                 0.618838                17.010443           0.938454           0.969347           0.
min                                 1.000000                94.000000           1.000000           0.000000           0.
25%                                 1.000000                120.000000           3.000000           0.000000           0.
50%                                 1.000000                130.000000           3.000000           0.000000           0.
75%                                 2.000000                140.000000           4.000000           1.000000           0.
max                                 3.000000                180.000000           4.000000           3.000000           1.
```

```
df.num_major_vessels.value_counts()
```

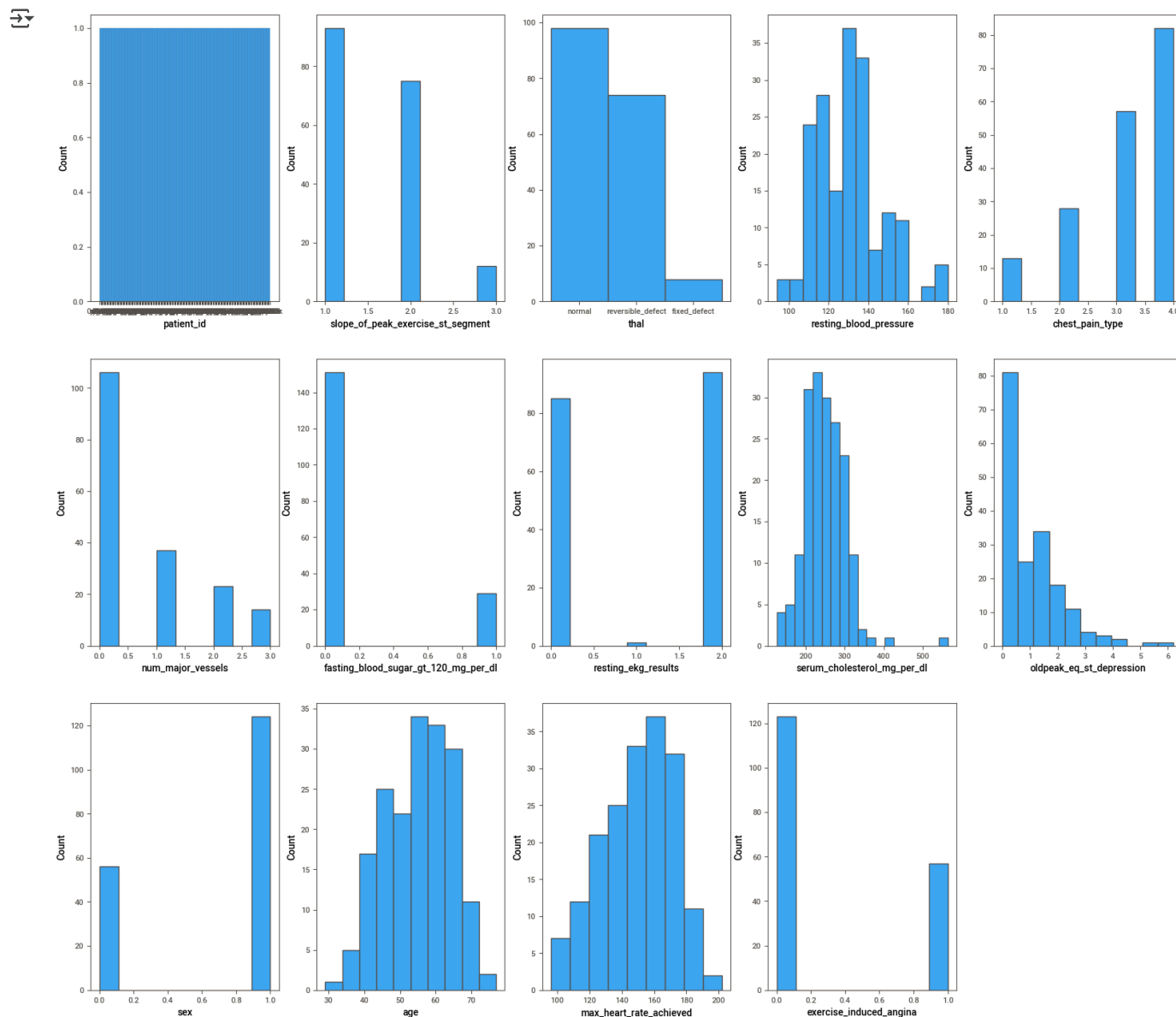
```

0    106
1     37
2     23
3     14
Name: num_major_vessels, dtype: int64
```

```
report=sv.analyze(df)
report.show_html()
```

```
Report SWEETVIZ_REPORT.html was generated! [ 0%] 00:00 -> (? left)
NOTEBOOK/COLAB USERS: the web browser MAY not pop up, regardless, the report IS saved in
```

```
plt.figure(figsize=(20,18))
plt.no=1
for i in df:
    plt.subplot(3,5,plt.no)
    sns.histplot(x=i,data=df)
    plt.no+=1
```




```
# insights
```

```
# people with st segment 3 is very low compare to the other two
# thallium stress test was normal for more people and people with thallium stress test fixed_defect was very low
# resting blood pressure was in the range of 135-140 for more people
# most of the patient have chest pain type four
# almost half of the patient has num_major_vessels colored by flourosopy of 0
# people with fasting blood sugar lesser than 120 is high
# very few peoples resting electrocardiographic results was normal
# more people has serum cholestoral in the range of 198-270, serum cholestoral has outlayers
# measure of abnormality in electrocardiograms was at 0 this proves that most of the peoples were normal, has outlayers
# many of the peoples in this dataset was male
# people in the age group of 53-68 was high
# maximum heart beat per minute was in the range of 142-179 for many
# more people does not have exercise_induced_angina
# people without heart disease is more compare to the people with heart disease
```

```
df1=pd.read_csv("labels.csv")
```

```
df1
```




	patient_id	heart_disease_present
0	0z64un	0
1	ryoo3j	0
2	yt1s1x	1
3	l2xjde	1
4	oyt4ek	0
...
175	5qfar3	1
176	2s2b1f	1
177	nsd00i	1
178	0xw93k	0
179	2nx10r	0

180 rows × 2 columns

```
df1.drop(["patient_id"],axis=1,inplace=True)
```

```
df=pd.concat([df1,df],axis=1)
```


```
df
```



	heart_disease_present	patient_id	slope_of_peak_exercise_st_segment	thal	resting_blood_pressure	chest_pain_type
0	0	0z64un	1	normal	128	2
1	0	ryoo3j	2	normal	110	3
2	1	yt1s1x	1	normal	125	4
3	1	l2xjde	1	reversible_defect	152	4
4	0	oyt4ek	3	reversible_defect	178	1
...
175	1	5qfar3	2	reversible_defect	125	4
176	1	2s2b1f	2	normal	180	4
177	1	nsd00i	2	reversible_defect	125	3
178	0	0xw93k	1	normal	124	3
179	0	2nx10r	1	normal	160	3


180 rows × 15 columns

```
report=sv.analyze(df)
report.show_html()
```



```
Report SWEETVIZ_REPORT.html was generated! | [ 0%] 00:00 -> (? left)
NOTEBOOK/COLAB USERS: the web browser MAY not pop up, regardless, the report IS saved in
```

```
df.isnull().sum()
```



```
heart_disease_present      0
patient_id                 0
slope_of_peak_exercise_st_segment  0
thal                      0
resting_blood_pressure     0
chest_pain_type            0
num_major_vessels          0
fasting_blood_sugar_gt_120_mg_per_d1  0
resting_ekg_results        0
serum_cholesterol_mg_per_dl  0
oldpeak_eq_st_depression   0
sex                        0
age                       0
max_heart_rate_achieved    0
exercise_induced_angina    0
dtype: int64
```

```
df.thal.value_counts()
```

```

normal          98
reversible_defect 74
fixed_defect     8
Name: thal, dtype: int64

```

```
df.thal=df.thal.map({"normal":0,"reversible_defect":1,"fixed_defect":2 })
```

```
df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 15 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   heart_disease_present                     180 non-null    int64
1   patient_id                               180 non-null    object
2   slope_of_peak_exercise_st_segment        180 non-null    int64
3   thal                                       180 non-null    int64
4   resting_blood_pressure                   180 non-null    int64
5   chest_pain_type                           180 non-null    int64
6   num_major_vessels                        180 non-null    int64
7   fasting_blood_sugar_gt_120_mg_per_dl    180 non-null    int64
8   resting_ekg_results                      180 non-null    int64
9   serum_cholesterol_mg_per_dl              180 non-null    int64
10  oldpeak_eq_st_depression                 180 non-null    float64
11  sex                                       180 non-null    int64
12  age                                       180 non-null    int64
13  max_heart_rate_achieved                  180 non-null    int64
14  exercise_induced_angina                  180 non-null    int64
dtypes: float64(1), int64(13), object(1)
memory usage: 21.2+ KB

```

```
df.serum_cholesterol_mg_per_dl.replace(417.0,df.serum_cholesterol_mg_per_dl.median(),inplace=True)
```

```
df.loc[df.serum_cholesterol_mg_per_dl>400]
```

```

heart_disease_present  patient_id  slope_of_peak_exercise_st_segment  thal  resting_blood_pressure  chest_pain_type  num_major_v
60                    0          rv6siv                               2    1                    115          3

```

```
df.loc[df.oldpeak_eq_st_depression>5,"oldpeak_eq_st_depression"]=df.oldpeak_eq_st_depression.median()
```

```
df
```

```

heart_disease_present  patient_id  slope_of_peak_exercise_st_segment  thal  resting_blood_pressure  chest_pain_type  num_major_
0                    0          0z64un                               1    0                    128          2
1                    0          ryoo3j                               2    0                    110          3
2                    1          yt1s1x                               1    0                    125          4
3                    1          l2xjde                               1    1                    152          4
4                    0          oyt4ek                               3    1                    178          1
...                  ...          ...                               ...  ...                    ...          ...
175                  1          5qfar3                               2    1                    125          4
176                  1          2s2b1f                               2    0                    180          4
177                  1          nsd00i                               2    1                    125          3
178                  0          0xw93k                               1    0                    124          3
179                  0          2nx10r                               1    0                    160          3

```

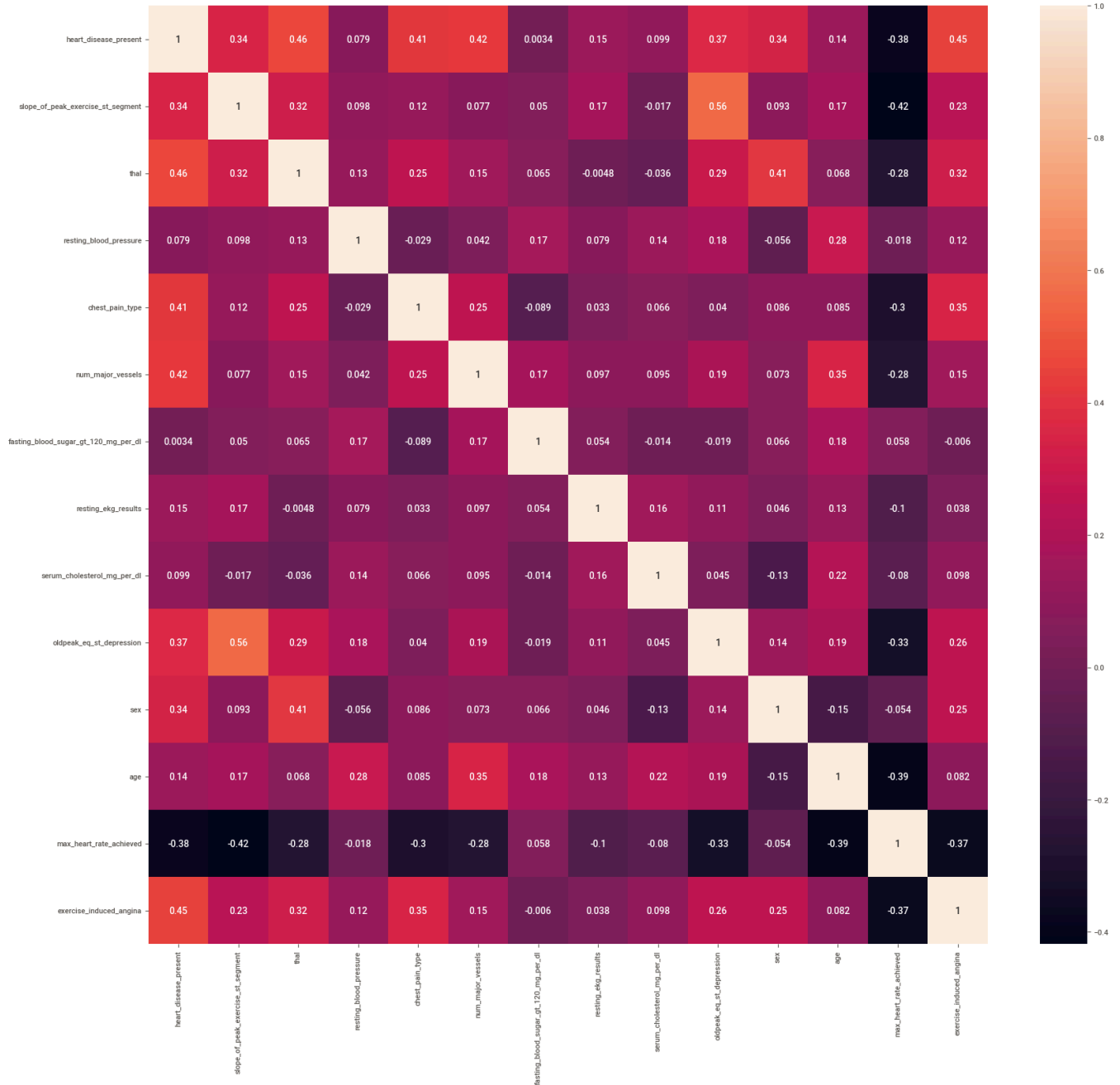
```
180 rows x 15 columns
```

```
df.describe()
```

	heart_disease_present	slope_of_peak_exercise_st_segment	thal	resting_blood_pressure	chest_pain_type	num_major_vess
count	180.000000	180.000000	180.000000	180.000000	180.000000	180.000000
mean	0.444444	1.550000	0.500000	131.311111	3.155556	0.694444
std	0.498290	0.618838	0.583765	17.010443	0.938454	0.969444
min	0.000000	1.000000	0.000000	94.000000	1.000000	0.000000
25%	0.000000	1.000000	0.000000	120.000000	3.000000	0.000000
50%	0.000000	1.000000	0.000000	130.000000	3.000000	0.000000
75%	1.000000	2.000000	1.000000	140.000000	4.000000	1.000000
max	1.000000	3.000000	2.000000	180.000000	4.000000	3.000000

```
corr=df.corr()
```

```
plt.figure(figsize=(20,18))  
sns.heatmap(corr,annot=True)
```

 <Axes: >


```

x=df.drop(["heart_disease_present","patient_id"],axis=1)

y=df["heart_disease_present"]

from sklearn.model_selection import train_test_split

xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.2,random_state=20)

#logistic Regg

from sklearn.linear_model import LogisticRegression

lr=LogisticRegression()

lr.fit(xtrain,ytrain)

```



▼ LogisticRegression
LogisticRegression()

```
pred=lr.predict(xtest)
```

```
from sklearn.metrics import classification_report
```

```
print(classification_report(pred,ytest))
```



	precision	recall	f1-score	support
0	0.71	0.83	0.77	18
1	0.80	0.67	0.73	18
accuracy			0.75	36
macro avg	0.76	0.75	0.75	36
weighted avg	0.76	0.75	0.75	36

```
from sklearn.model_selection import GridSearchCV
```

```
grid={"penalty":["l1", "l2", "elasticnet", None]
      ,"fit_intercept":[True,False],
      "solver":["lbfgs", "liblinear", "newton-cg", "newton-cholesky", "sag", "saga"],
      "max_iter":list(range(1,20)),"multi_class":["auto", "ovr", "multinomial"]}
```

```
cv=GridSearchCV(lr,param_grid=grid,refit=True,cv=3,verbose=2,scoring="f1")
```

```
cv.fit(x,y)
```


9/541

lab research google.com/drive/17siAlI95CBmrm4NSu7YXowiEMTGdGWydX#printMode=true

11/541

ab.research.google.com/drive/17sjAU95CBrMrm4NSu7XowiFMTGDGWvdX#printMode=true 13

lab research google.com/drive/17siAlI95CBrrMrm4NSu7YowwEMTGdGWvdY#printMode=true

lab.research.google.com/drive/17sjAU95CBMrm4NSu7XowiFMTGDGWvdX#printMode=true 15

lab.research.google.com/drive/17sjAU95CBmrm4NSu7XowjFMTGDGWydX#printMode=true

17/541

ah research google.com/drive/17siAlI95CBrrm4NSu7YxwiEMTGdGWvdX#printMode=true 20

lab.research.google.com/drive/17sjAU95CBrrm4NSu7XowiFMTGDGWvdX#printMode=true

lab research google.com/drive/17siAlI95CBrrm4NSu7XowiFMTGDGwYdX#printMode=true

23/541

ab.research.google.com/drive/17sjAU95CBrMrm4NSu7XowiFMTGDGWvdX#printMode=true 25

lab.research.google.com/drive/17siAlU95CBrrMrm4NSu7YowiEMTGdGWvdY#printMode=true

27/541

lab research google.com/drive/17siAlI95CBrrMrm4NSu7XowiEMTGDGWydX#printMode=true

lab.research.google.com/drive/17sjAU95CBrrm4NSu7XowiFMTGDGWvdX#printMode=true

lab research google.com/drive/17siAlI95CBrrMm4NSu7XowIEMTGdGWydX#printMode=true 30

lab.research.google.com/drive/17sjAU95CBrrm4NSu7XowiEMTGDGWvdX#printMode=true

32/541

lab.research.google.com/drive/17sjAU95CBrm4NSu7XowiFMTGDGWvdX#printMode=true

34/541

lab_research.google.com/drive/17ciAU05CBrrMm4NSu7YowwEMTGDGWydY#printMode=true

37/541

lab_research.google.com/drive/17siAU95CBrrMrm4NSu7YewiFMTGDGW/ydY#printMode=true

ab.research.google.com/drive/17sjAU95CBrrMm4NSu7XowiFMTGDGWvdX#printMode=true 39/5

lab.research.google.com/drive/17sjAU95CBmM4NSu7XowjFMTGDGWydX#printMode=true

lab.research.google.com/drive/17sjAU95CBmrm4NSu7XowiFMTGDGWvdX#printMode=true

42/541

lab.research.google.com/drive/17sjAU95CBrrm4NSu7XowiFMTGDGWvdX#printMode=true

44/541

lab.research.google.com/drive/17sjAU95CBrrm4NSu7XowiFMTGDGWvdX#printMode=true

46/541

49/541

51/541

lab.research.google.com/drive/17sjAU95CBrrm4NSu7XowiFMTGDGWvdX#printMode=true

lab.research.google.com/drive/17siAU95CBrrMm4NSu7YowiFMTGDGWydY#printMode=true 54/5

55/541

56/541

lab.research.google.com/drive/17sjAU95CBmM4NSu7XowiFMTGDGWvdX#printMode=true

ab.research.google.com/drive/17sjAU95CBmrm4NSu7XowiFMTGDGWvdX#printMode=true 58/5

lab.research.google.com/drive/17sjAU95CBrMrm4NSu7XowiFMTGDGWvdX#printMode=true

lab research google.com/drive/17siAIU95CBrrMrm4NSu7XowiEMTGdGWvdX#printMode=true

lab.research.google.com/drive/17sjAU95CBrMrm4NSu7XowiFMTGDGWvdX#printMode=true 61/5

lab research google.com/drive/17siAlI95CBrrMrm4NSu7XowiEMTGDGWydX#printMode=true

63/541

lab research google.com/drive/17siAlI95CBrrMrm4NSu7YXowiEMTGDGWydX#printMode=true

lab.research.google.com/drive/17sjAU95CBrMrm4NSu7XowiFMTGDGWvdX#printMode=true

66/541

lab.research.google.com/drive/17sjAU95CBrrm4NSu7XowiFMTGDGWvdX#printMode=true

lab research google.com/drive/17siAlU95CBmrm4NSu7XowiEMTGdGWydX#printMode=true 68/5

lab.research.google.com/drive/17sjAU95CBmM4NSu7XowiFMTGDGWvdX#printMode=true

70/541

lab.research.google.com/drive/17sjAU95CBrMrm4NSu7XowiFMTGDGWvdX#printMode=true

lab research google.com/drive/17siAIU95CBrrMrm4NSu7XowiEMTGdGWvdX#printMode=true

73/541

lab_research.google.com/drive/17ciAU05CBrrm4NSu7YowiEMTGDGWydY#printMode=true

75/541

lab research google.com/drive/17siAlI95CBrrMrm4NSu7XowiEMTGDGWvdY#printMode=true

77/541

lab_research.google.com/drive/17siAU95CBrrMrm4NSu7YewiFMTGDGWvdY#printMode=true

lab.research.google.com/drive/17sjAU95CBmM4NSu7XowiFMTGDGWvdX#printMode=true

lab research google.com/drive/17siAlI95CBrMrm4NSu7XowiFMTGDGWvdY#printMode=true 80/54

lab.research.google.com/drive/17sjAU95CBrrMrm4NSu7XowiFMTGDGWvdX#printMode=true

lab research google.com/drive/17siAlI95CBrrMrm4NSu7YxwiFMTGDGWydX#printMode=true 82/54

lab.research.google.com/drive/17sjAU95CBrrm4NSu7XowiFMTGDGWvdX#printMode=true

lab_research.google.com/drive/17siAU95CBrrMrm4NSu7YewiFMTGDGWydY#printMode=true

lab.research.google.com/drive/17sjAU95CBrrM4NSu7XowiFMTGDGWvdX#printMode=true 85/54

87/541

lab research google.com/drive/17siAlI95CBrrMrm4NSu7XowiEMTGDGWvdY#printMode=true

89/541

ab.research.google.com/drive/17sjAU95CBrrMm4NSu7XowiFMTGDGWvdX#printMode=true

lab research google.com/drive/17siAlI95CBrrMrm4NSu7XowiEMTGDGWydX#printMode=true

lab.research.google.com/drive/17sjAU95CBmM4NSu7XowiFMTGDGWvdX#printMode=true

lab.research.google.com/drive/17siAU95CBrrm4NSu7XowiFMTGDGWvdX#printMode=true 94/54

lab.research.google.com/drive/17sjAU95CBrMrm4NSu7XowiFMTGDGWvdX#printMode=true

lab.research.google.com/drive/17sjAU95CBrMrm4NSu7XowiFMTGDGWvdX#printMode=true

lab.research.google.com/drive/17sjAU95CBrMrm4NSu7XowiFMTGDGWvdX#printMode=true 97/54


```

GridSearchCV
├── estimator: LogisticRegression
│   └── LogisticRegression
└── ...

```



```
cv.best_params_
```

```
{'fit_intercept': True,
 'max_iter': 7,
 'multi_class': 'auto',
 'penalty': 'l2',
 'solver': 'newton-cg'}
```

```
lr=LogisticRegression(fit_intercept= True,
 max_iter= 7,
 multi_class='auto',
 penalty= 'l2',
 solver= 'newton-cg')
```

```
lr.fit(xtrain,ytrain)
pred=lr.predict(xtest)
```

```
print(classification_report(pred,ytest))
```

```
precision    recall  f1-score   support

0           0.76       0.84       0.80         19
1           0.80       0.71       0.75         17

 accuracy          0.78         36
 macro avg          0.78         36
 weighted avg       0.78         36
```

```
# KNN
```

```
from sklearn.neighbors import KNeighborsClassifier
```

```
kn=KNeighborsClassifier()
```

```
kn.fit(xtrain,ytrain)
pred=kn.predict(xtest)
```

```
print(classification_report(pred,ytest))
```

```
precision    recall  f1-score   support

0           0.57       0.60       0.59         20
1           0.47       0.44       0.45         16

 accuracy          0.53         36
 macro avg          0.52         36
 weighted avg       0.52         36
```

```
grid={"n_neighbors":[1,2,3,4,5,6,7,8,9],
      "weights":["uniform", "distance"],
      "algorithm":["auto", "ball_tree", "kd_tree", "brute"],
      "p":[0.1,0.2,2.1,5.3,0.9,6,9,1,7]}
```

```
cv=GridSearchCV(kn,param_grid=grid,cv=3,refit=True,verbose=2,scoring="f1")
```

```
cv.fit(x,y)
```

<https://colab.research.google.com/drive/17siAU95CBrMrm4NSu7XowjFMTGDGWydX#printMode=true>

lab_research.google.com/drive/17siAlJ05CBmM4NSu7ZYowiEMTGDGWydY#printMode=true

lab.research.google.com/drive/17sjAU95CBmrm4NSu7XowiFMTGDGWvdX#printMode=true

lab_research.google.com/drive/17siAlJ95CBmrm4NSu7YowiEMTGDGWydX#printMode=true

108/541

lab_research.google.com/drive/17siAU95CBmMrm4NSu7ZYowiEMTGDGW/ydY#printMode=true

110/541

lab research google.com/drive/17siAIJ95CBmM4NSu7XowiFMTGDGWydY#printMode=true

[illegible]

<https://colab.research.google.com/drive/17siAU95CBrMrm4NSu7XowjFMTGDGWydX#printMode=true>

114/541

lab_research google.com/drive/17siAlI95CBrMrm4NSu7YowiEMTGDGWydY#printMode=true

[illegible]

<https://colab.research.google.com/drive/17sjAU95CBrMrm4NSu7XowiFMTGDGWvdX#printMode=true>

<https://colab.research.google.com/drive/17sjAU95CBrMrm4NSu7XowiFMTGDGWvdX#printMode=true>

lab research google.com/drive/17siAU95CBrMrm4NSu7XowiEMTGDGWydX#printMode=true

120/541

lab research google.com/drive/17siAlI95CBrMrm4NSu7XowiEMTGDGWydX#printMode=true

122/541

123/541

124/541

```

[CV] END algorithm=brute, n_neighbors=9, p=0.2, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=0.2, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=0.2, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=2.1, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=2.1, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=2.1, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=2.1, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=2.1, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=2.1, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=5.3, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=5.3, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=5.3, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=5.3, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=5.3, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=0.9, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=0.9, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=0.9, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=0.9, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=0.9, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=0.9, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=6, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=6, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=6, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=6, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=6, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=6, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=9, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=9, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=9, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=9, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=9, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=9, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=1, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=1, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=1, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=1, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=1, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=1, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=7, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=7, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=7, weights=uniform; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=7, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=7, weights=distance; total time= 0.0s
[CV] END algorithm=brute, n_neighbors=9, p=7, weights=distance; total time= 0.0s

```

```

└─ GridSearchCV
  └─ estimator: KNeighborsClassifier
    └─ KNeighborsClassifier

```



```
cv.best_params_
```

```
{'algorithm': 'brute', 'n_neighbors': 3, 'p': 1, 'weights': 'uniform'}
```

```
kn=KNeighborsClassifier(algorithm= 'brute', n_neighbors= 3, p= 1, weights= 'uniform')
```

```
kn.fit(xtrain,ytrain)
pred=kn.predict(xtest)
```

```
print(classification_report(pred,ytest))
```

```

precision    recall  f1-score   support

0           0.76     0.76     0.76         21
1           0.67     0.67     0.67         15

 accuracy          0.72         36
 macro avg         0.71         36
 weighted avg      0.72         36
```

```
#svm
```

```
from sklearn.svm import SVC
svc=SVC()
svc.fit(xtrain,ytrain)
pred=svc.predict(xtest)
print(classification_report(ytest,pred))
```

```

precision    recall  f1-score   support

0           0.65     0.95     0.77         21
1           0.80     0.27     0.40         15

 accuracy          0.67         36
 macro avg         0.72         36
 weighted avg      0.71         36
```

```
from sklearn.tree import DecisionTreeClassifier
```

```
dc=DecisionTreeClassifier()
```

```
dc.fit(xtrain,ytrain)
pred=dc.predict(xtest)
```

```
print(classification_report(pred,ytest))
```

```

precision    recall  f1-score   support

0           0.67     0.67     0.67         21
1           0.53     0.53     0.53         15

 accuracy          0.61         36
 macro avg         0.60         36
 weighted avg      0.61         36
```

```
grid={"criterion":["gini", "entropy", "log_loss"],
      "splitter":["best", "random"],
      "max_depth":[1,2,3,4,5],
      "min_samples_split":[1,2,0.3,0.2,6],
      "min_samples_leaf":[10,12,13,14,15],
      "max_features":["auto", "sqrt", "log2"],
      "random_state":[11,22,33,44,55]}
```

```
cv=GridSearchCV(dc,param_grid=grid,cv=3,refit=True,verbose=2,scoring="f1")
```

```
cv.fit(x,y)
```

128/541

129/541

130/541

131/541

132/541

133/541

134/541

135/541

136/541

137/541

138/541

139/541

140/541

141/541

142/541

143/541

144/541

145/541

146/541

147/541

148/541

149/541

150/541

151/541

152/541

153/541

154/541

155/541

156/541

157/541

158/541

159/541

160/541

161/541

162/541

163/541

164/541

165/541

ab.research.google.com/drive/17sjAU95CBrMrm4NSu7XowiFMTGDGWvdX#printMode=true 166/54

167/541

168/541

169/541

170/541

171/541

172/541

173/541

174/541

175/541

176/541

177/541

lab.research.google.com/drive/17sjAU95CBmM4NSu7XowiFMTGDGWvdX#printMode=true 178/54

179/541

180/541

181/541

182/541

183/541

184/541

185/541

lab.research.google.com/drive/17sjAU95CBrMrm4NSu7XowiFMTGDGWvdX#printMode=true 186/5

187/541

188/541

189/541

190/541

191/541

192/541

193/541

194/541

195/541

196/541

197/541

198/541

199/541

200/541

201/541

202/541

203/541

204/541

205/541

206/541

207/541

208/541

209/541

210/541

211/541

212/541

213/541

214/541

215/541

216/541

217/541

218/541

219/541

220/541

221/541

222/541

223/541

224/541

225/541

[illegible]

227/541

228/541

229/541

230/541

231/541

ab.research.google.com/drive/17sjAU95CBrrMm4NSu7XowiFMTGDGWvdX#printMode=true 232/54

233/541

234/541

235/541

236/541

237/541

238/541

239/541

240/541

241/541

242/541

243/541

244/541

245/541

246/541

247/541

248/541

250/541

252/541

254/541

256/541

[illegible]

258/541

260/541

```

[CV1] END: exit after exiting main death 4, new features sent via number 1-6 10, via number add 6, random state 44, addition

```

262/541

264/541

14-00000

266/541

267/541

268/541

14-00000

270/541

[illegible]

1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819

274/541

```
[CPU] END: split_time: 0.000000, max_depth: 4, max_features: 100, min_samples_leaf: 15, min_samples_split: 2, random_state: 22, split_time: 0.000000
```

276/541

277/541

278/541

280/541

282/541

284/541

286/541

bioRxiv preprint doi: <https://doi.org/10.1101/2020.07.20.190000>; this version posted July 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

288/541

289/541

290/541

292/541

293/541

294/541

295/541

296/541

297/541

298/541

299/541

300/541

301/541

302/541

303/541

304/541

305/541

306/541

307/541

308/541

309/541

310/541

311/541

312/541

313/541

314/541

315/541

316/541

317/541

318/541

319/541

320/541

321/541

322/541

323/541

324/541

325/541

326/541

327/541

328/541

329/541

330/541

331/541

332/541

333/541

334/541

335/541

336/541

337/541

338/541

339/541

340/541

341/541

342/541

343/541

344/541

345/541

346/541

347/541

349/541

350/541

351/541

352/541

353/541

354/541

355/541

356/541

357/541

358/541

359/541

360/541

361/541

362/541

363/541

364/541

365/541

366/541

367/541

368/541

369/541

370/541

371/541

372/541

373/541

374/541

375/541

376/541

377/541

378/541

379/541

380/541

381/541

382/541

383/541

384/541

385/541

386/541

387/541

388/541

389/541

390/541

391/541

392/541

393/541

394/541

395/541

396/541

397/541

398/541

399/541

400/541

401/541

402/541

403/541

404/541

405/541

406/541

407/541

408/541

409/541

410/541

411/541

412/541

413/541

414/541

415/541

416/541

417/541

418/541

419/541

420/541

421/541

422/541

423/541

424/541

425/541

426/541

427/541

428/541

429/541

430/541

431/541

432/541

433/541

434/541

435/541

436/541

437/541

438/541

439/541

440/541

441/541

442/541

443/541

444/541

445/541

446/541

447/541

448/541

449/541

450/541

451/541

452/541

453/541

454/541

455/541

456/541

457/541

458/541

459/541

460/541

461/541

462/541

463/541

464/541

465/541

466/541

467/541

468/541

469/541

470/541

471/541

472/541

473/541

474/541

475/541

476/541

477/541

478/541

lab.research.google.com/drive/17sjAU95CBrrm4NSu7XowiFMTGDGWvdX#printMode=true 479/5

480/541

[illegible]

482/541

483/541

484/541

485/541

486/541

487/541

488/541

489/541

490/541

491/541

492/541

493/541

494/541

495/541

496/541

497/541

498/541

499/541

```
GridSearchCV
estimator: DecisionTreeClassifier
  DecisionTreeClassifier
```



```
cv.best_params_
```

```
{'criterion': 'gini',
 'max_depth': 4,
 'max_features': 'auto',
 'min_samples_leaf': 10,
 'min_samples_split': 1,
 'random_state': 11,
 'splitter': 'best'}
```

```
dc=DecisionTreeClassifier(criterion= 'gini',
 max_depth= 4,
 max_features= 'auto',
 min_samples_leaf= 10,
 min_samples_split= 1,
 random_state= 11,
 splitter= 'best')
dc.fit(xtrain,ytrain)
pred=dc.predict(xtest)
print(classification_report(pred,ytest))
```

```
precision    recall  f1-score   support

0           0.52       0.65       0.58         17
1           0.60       0.47       0.53         19

accuracy          0.56
macro avg          0.56
weighted avg       0.56
```

```
from sklearn.ensemble import RandomForestClassifier
```

```
rc=RandomForestClassifier()
```

```
rc.fit(xtrain,ytrain)
pred=rc.predict(xtest)
```

```
print(classification_report(pred,ytest))
```

```
precision    recall  f1-score   support

0           0.90       0.83       0.86         23
1           0.73       0.85       0.79         13

accuracy          0.83
macro avg          0.82
weighted avg       0.84
```

```
grid={"n_estimators":[10,20,30], "criterion":["gini","entropy"],
      "max_depth":(1,2,3,4,5), "min_samples_split":[2,3,4,5], "min_samples_leaf":[2,3,4,8], "max_features":["sqrt","log2"], "bootstrap":|
```

```
cv=GridSearchCV(rc,param_grid=grid,cv=2,verbose=3,refit=True,scoring="f1")
```

```
cv.fit(x,y)
```

513/541

514/541

ab.research.google.com/drive/17sjAU95CBrrm4NSu7XowiFMTGDGWvdX#printMode=true 515/541

516/541

517/541

518/541

519/541

520/541

521/541

522/541

[illegible]

524/541

[illegible]

526/541

527/541

528/541

529/541

530/541

531/541

532/541

533/541

534/541

535/541

536/541

537/541

538/541

539/541

540/541

541/541