

# assingment3

August 10, 2024

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

```
[2]: dataset = pd.read_csv(r'D:\loan_detection.csv')
print(dataset)
```

	age	campaign	pdays	previous	no_previous_contact	not_working	\
0	56	1	999	0	1	0	
1	57	1	999	0	1	0	
2	37	1	999	0	1	0	
3	40	1	999	0	1	0	
4	56	1	999	0	1	0	
...	...	...	...	...	...	...	
41183	73	1	999	0	1	1	
41184	46	1	999	0	1	0	
41185	56	2	999	0	1	1	
41186	44	1	999	0	1	0	
41187	74	3	999	1	1	1	

	job_admin.	job_blue-collar	job_entrepreneur	job_housemaid	...	\
0	0	0	0	1	...	
1	0	0	0	0	...	
2	0	0	0	0	...	
3	1	0	0	0	...	
4	0	0	0	0	...	
...	...	...	...	...	...	
41183	0	0	0	0	...	
41184	0	1	0	0	...	
41185	0	0	0	0	...	
41186	0	0	0	0	...	
41187	0	0	0	0	...	

	month_sep	day_of_week_fri	day_of_week_mon	day_of_week_thu	\
0	0	0	1	0	
1	0	0	1	0	

2	0	0	1	0
3	0	0	1	0
4	0	0	1	0
...	...	...	...	...
41183	0	1	0	0
41184	0	1	0	0
41185	0	1	0	0
41186	0	1	0	0
41187	0	1	0	0

	day_of_week_tue	day_of_week_wed	poutcome_failure \
0	0	0	0
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
...	...	...	...
41183	0	0	0
41184	0	0	0
41185	0	0	0
41186	0	0	0
41187	0	0	1

	poutcome_nonexistent	poutcome_success	Loan_Status_label
0	1	0	0
1	1	0	0
2	1	0	0
3	1	0	0
4	1	0	0
...	...	...	...
41183	1	0	1
41184	1	0	0
41185	1	0	0
41186	1	0	1
41187	0	0	0

[41188 rows x 60 columns]

```
[14]: dataset.shape
```

```
[14]: (41188, 60)
```

```
[16]: dataset.head()
dataset
```

```
[16]:
```

	age	campaign	pdays	previous	no_previous_contact	not_working \
0	56	1	999	0	1	0
1	57	1	999	0	1	0

2	37	1	999	0	1	0
3	40	1	999	0	1	0
4	56	1	999	0	1	0
...	...	...	...	...	...	...
41183	73	1	999	0	1	1
41184	46	1	999	0	1	0
41185	56	2	999	0	1	1
41186	44	1	999	0	1	0
41187	74	3	999	1	1	1

	job_admin.	job_blue-collar	job_entrepreneur	job_housemaid	...	\
0	0	0	0	1	...	
1	0	0	0	0	...	
2	0	0	0	0	...	
3	1	0	0	0	...	
4	0	0	0	0	...	
...	...	...	...	...	...	
41183	0	0	0	0	...	
41184	0	1	0	0	...	
41185	0	0	0	0	...	
41186	0	0	0	0	...	
41187	0	0	0	0	...	

	month_sep	day_of_week_fri	day_of_week_mon	day_of_week_thu	\
0	0	0	1	0	
1	0	0	1	0	
2	0	0	1	0	
3	0	0	1	0	
4	0	0	1	0	
...	...	...	...	...	
41183	0	1	0	0	
41184	0	1	0	0	
41185	0	1	0	0	
41186	0	1	0	0	
41187	0	1	0	0	

	day_of_week_tue	day_of_week_wed	poutcome_failure	\
0	0	0	0	
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	0	
...	...	...	...	
41183	0	0	0	
41184	0	0	0	
41185	0	0	0	
41186	0	0	0	

41187	0	0	1
	poutcome_nonexistent	poutcome_success	Loan_Status_label
0	1	0	0
1	1	0	0
2	1	0	0
3	1	0	0
4	1	0	0
...	...	...	...
41183	1	0	1
41184	1	0	0
41185	1	0	0
41186	1	0	1
41187	0	0	0

[41188 rows x 60 columns]

```
[17]: dataset.tail()
dataset
```

```
[17]:
```

	age	campaign	pdays	previous	no_previous_contact	not_working	\
0	56	1	999	0	1	0	
1	57	1	999	0	1	0	
2	37	1	999	0	1	0	
3	40	1	999	0	1	0	
4	56	1	999	0	1	0	
...	...	...	...	...	...	...	
41183	73	1	999	0	1	1	
41184	46	1	999	0	1	0	
41185	56	2	999	0	1	1	
41186	44	1	999	0	1	0	
41187	74	3	999	1	1	1	

	job_admin.	job_blue-collar	job_entrepreneur	job_housemaid	...	\
0	0	0	0	1	...	
1	0	0	0	0	...	
2	0	0	0	0	...	
3	1	0	0	0	...	
4	0	0	0	0	...	
...	...	...	...	...	...	
41183	0	0	0	0	...	
41184	0	1	0	0	...	
41185	0	0	0	0	...	
41186	0	0	0	0	...	
41187	0	0	0	0	...	

month_sep	day_of_week_fri	day_of_week_mon	day_of_week_thu	\
-----------	-----------------	-----------------	-----------------	---

0	0	0	1	0
1	0	0	1	0
2	0	0	1	0
3	0	0	1	0
4	0	0	1	0
...	...	...	...	...
41183	0	1	0	0
41184	0	1	0	0
41185	0	1	0	0
41186	0	1	0	0
41187	0	1	0	0

	day_of_week_tue	day_of_week_wed	poutcome_failure \
0	0	0	0
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
...	...	...	...
41183	0	0	0
41184	0	0	0
41185	0	0	0
41186	0	0	0
41187	0	0	1

	poutcome_nonexistent	poutcome_success	Loan_Status_label
0	1	0	0
1	1	0	0
2	1	0	0
3	1	0	0
4	1	0	0
...	...	...	...
41183	1	0	1
41184	1	0	0
41185	1	0	0
41186	1	0	1
41187	0	0	0

[41188 rows x 60 columns]

```
[18]: dataset.info()
dataset
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 41188 entries, 0 to 41187
```

```
Data columns (total 60 columns):
```

#	Column	Non-Null Count	Dtype
---	-----	-----	-----

0	age	41188	non-null	int64
1	campaign	41188	non-null	int64
2	pdays	41188	non-null	int64
3	previous	41188	non-null	int64
4	no_previous_contact	41188	non-null	int64
5	not_working	41188	non-null	int64
6	job_admin.	41188	non-null	int64
7	job_blue-collar	41188	non-null	int64
8	job_entrepreneur	41188	non-null	int64
9	job_housemaid	41188	non-null	int64
10	job_management	41188	non-null	int64
11	job_retired	41188	non-null	int64
12	job_self-employed	41188	non-null	int64
13	job_services	41188	non-null	int64
14	job_student	41188	non-null	int64
15	job_technician	41188	non-null	int64
16	job_unemployed	41188	non-null	int64
17	job_unknown	41188	non-null	int64
18	marital_divorced	41188	non-null	int64
19	marital_married	41188	non-null	int64
20	marital_single	41188	non-null	int64
21	marital_unknown	41188	non-null	int64
22	education_basic.4y	41188	non-null	int64
23	education_basic.6y	41188	non-null	int64
24	education_basic.9y	41188	non-null	int64
25	education_high.school	41188	non-null	int64
26	education_illiterate	41188	non-null	int64
27	education_professional.course	41188	non-null	int64
28	education_university.degree	41188	non-null	int64
29	education_unknown	41188	non-null	int64
30	default_no	41188	non-null	int64
31	default_unknown	41188	non-null	int64
32	default_yes	41188	non-null	int64
33	housing_no	41188	non-null	int64
34	housing_unknown	41188	non-null	int64
35	housing_yes	41188	non-null	int64
36	loan_no	41188	non-null	int64
37	loan_unknown	41188	non-null	int64
38	loan_yes	41188	non-null	int64
39	contact_cellular	41188	non-null	int64
40	contact_telephone	41188	non-null	int64
41	month_apr	41188	non-null	int64
42	month_aug	41188	non-null	int64
43	month_dec	41188	non-null	int64
44	month_jul	41188	non-null	int64
45	month_jun	41188	non-null	int64
46	month_mar	41188	non-null	int64
47	month_may	41188	non-null	int64

48	month_nov	41188	non-null	int64
49	month_oct	41188	non-null	int64
50	month_sep	41188	non-null	int64
51	day_of_week_fri	41188	non-null	int64
52	day_of_week_mon	41188	non-null	int64
53	day_of_week_thu	41188	non-null	int64
54	day_of_week_tue	41188	non-null	int64
55	day_of_week_wed	41188	non-null	int64
56	poutcome_failure	41188	non-null	int64
57	poutcome_nonexistent	41188	non-null	int64
58	poutcome_success	41188	non-null	int64
59	Loan_Status_label	41188	non-null	int64

dtypes: int64(60)

memory usage: 18.9 MB

```
[18]:
```

	age	campaign	pdays	previous	no_previous_contact	not_working	\
0	56	1	999	0	1	0	
1	57	1	999	0	1	0	
2	37	1	999	0	1	0	
3	40	1	999	0	1	0	
4	56	1	999	0	1	0	
...	...	...	...	...	...	...	
41183	73	1	999	0	1	1	
41184	46	1	999	0	1	0	
41185	56	2	999	0	1	1	
41186	44	1	999	0	1	0	
41187	74	3	999	1	1	1	

	job_admin.	job_blue-collar	job_entrepreneur	job_housemaid	...	\
0	0	0	0	1	...	
1	0	0	0	0	...	
2	0	0	0	0	...	
3	1	0	0	0	...	
4	0	0	0	0	...	
...	...	...	...	...	...	
41183	0	0	0	0	...	
41184	0	1	0	0	...	
41185	0	0	0	0	...	
41186	0	0	0	0	...	
41187	0	0	0	0	...	

	month_sep	day_of_week_fri	day_of_week_mon	day_of_week_thu	\
0	0	0	1	0	
1	0	0	1	0	
2	0	0	1	0	
3	0	0	1	0	
4	0	0	1	0	

...	...	...	...	...
41183	0	1	0	0
41184	0	1	0	0
41185	0	1	0	0
41186	0	1	0	0
41187	0	1	0	0

	day_of_week_tue	day_of_week_wed	poutcome_failure	\
0	0	0	0	
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	0	
...	...	...	...	
41183	0	0	0	
41184	0	0	0	
41185	0	0	0	
41186	0	0	0	
41187	0	0	1	

	poutcome_nonexistent	poutcome_success	Loan_Status_label
0	1	0	0
1	1	0	0
2	1	0	0
3	1	0	0
4	1	0	0
...	...	...	...
41183	1	0	1
41184	1	0	0
41185	1	0	0
41186	1	0	1
41187	0	0	0

[41188 rows x 60 columns]

```
[19]: dataset.describe()
dataset
```

```
[19]:
```

	age	campaign	pdays	previous	no_previous_contact	not_working	\
0	56	1	999	0	1	0	
1	57	1	999	0	1	0	
2	37	1	999	0	1	0	
3	40	1	999	0	1	0	
4	56	1	999	0	1	0	
...	...	...	...	...	...	...	
41183	73	1	999	0	1	1	
41184	46	1	999	0	1	0	



41185	56	2	999	0	1	1
41186	44	1	999	0	1	0
41187	74	3	999	1	1	1

	job_admin.	job_blue-collar	job_entrepreneur	job_housemaid	...	\
0	0	0	0	1	...	
1	0	0	0	0	...	
2	0	0	0	0	...	
3	1	0	0	0	...	
4	0	0	0	0	...	
...	...	...	...	...	...	
41183	0	0	0	0	...	
41184	0	1	0	0	...	
41185	0	0	0	0	...	
41186	0	0	0	0	...	
41187	0	0	0	0	...	

	month_sep	day_of_week_fri	day_of_week_mon	day_of_week_thu	\
0	0	0	1	0	
1	0	0	1	0	
2	0	0	1	0	
3	0	0	1	0	
4	0	0	1	0	
...	...	...	...	...	
41183	0	1	0	0	
41184	0	1	0	0	
41185	0	1	0	0	
41186	0	1	0	0	
41187	0	1	0	0	

	day_of_week_tue	day_of_week_wed	poutcome_failure	\
0	0	0	0	
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	0	
...	...	...	...	
41183	0	0	0	
41184	0	0	0	
41185	0	0	0	
41186	0	0	0	
41187	0	0	1	

	poutcome_nonexistent	poutcome_success	Loan_Status_label
0	1	0	0
1	1	0	0
2	1	0	0

3	1	0	0
4	1	0	0
...	...	...	...
41183	1	0	1
41184	1	0	0
41185	1	0	0
41186	1	0	1
41187	0	0	0

[41188 rows x 60 columns]

```
[3]: import numpy as np
import pandas as pd

import xgboost
from xgboost import XGBClassifier

from sklearn.model_selection import train_test_split
from sklearn.metrics import \
    confusion_matrix, accuracy_score, classification_report
```

```
[4]: dataset = pd.read_csv(r'D:\loan_detection.csv')
print(dataset)
```

	age	campaign	pdays	previous	no_previous_contact	not_working	\
0	56	1	999	0	1	0	
1	57	1	999	0	1	0	
2	37	1	999	0	1	0	
3	40	1	999	0	1	0	
4	56	1	999	0	1	0	
...	...	...	...	...	...	...	
41183	73	1	999	0	1	1	
41184	46	1	999	0	1	0	
41185	56	2	999	0	1	1	
41186	44	1	999	0	1	0	
41187	74	3	999	1	1	1	

	job_admin.	job_blue-collar	job_entrepreneur	job_housemaid	...	\
0	0	0	0	1	...	
1	0	0	0	0	...	
2	0	0	0	0	...	
3	1	0	0	0	...	
4	0	0	0	0	...	
...	...	...	...	...	...	
41183	0	0	0	0	...	
41184	0	1	0	0	...	
41185	0	0	0	0	...	
41186	0	0	0	0	...	

41187	0	0	0	0	...
-------	---	---	---	---	-----

	month_sep	day_of_week_fri	day_of_week_mon	day_of_week_thu	\
0	0	0	1	0	
1	0	0	1	0	
2	0	0	1	0	
3	0	0	1	0	
4	0	0	1	0	
...	...	...	...	...	
41183	0	1	0	0	
41184	0	1	0	0	
41185	0	1	0	0	
41186	0	1	0	0	
41187	0	1	0	0	

	day_of_week_tue	day_of_week_wed	poutcome_failure	\
0	0	0	0	
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	0	
...	...	...	...	
41183	0	0	0	
41184	0	0	0	
41185	0	0	0	
41186	0	0	0	
41187	0	0	1	

	poutcome_nonexistent	poutcome_success	Loan_Status_label
0	1	0	0
1	1	0	0
2	1	0	0
3	1	0	0
4	1	0	0
...	...	...	...
41183	1	0	1
41184	1	0	0
41185	1	0	0
41186	1	0	1
41187	0	0	0

[41188 rows x 60 columns]

```
[5]: dataset.columns
```

```
[5]: Index(['age', 'campaign', 'pdays', 'previous', 'no_previous_contact',
          'not_working', 'job_admin.', 'job_blue-collar', 'job_entrepreneur',
```

```

'job_housemaid', 'job_management', 'job_retired', 'job_self-employed',
'job_services', 'job_student', 'job_technician', 'job_unemployed',
'job_unknown', 'marital_divorced', 'marital_married', 'marital_single',
'marital_unknown', 'education_basic.4y', 'education_basic.6y',
'education_basic.9y', 'education_high.school', 'education_illiterate',
'education_professional.course', 'education_university.degree',
'education_unknown', 'default_no', 'default_unknown', 'default_yes',
'housing_no', 'housing_unknown', 'housing_yes', 'loan_no',
'loan_unknown', 'loan_yes', 'contact_cellular', 'contact_telephone',
'month_apr', 'month_aug', 'month_dec', 'month_jul', 'month_jun',
'month_mar', 'month_may', 'month_nov', 'month_oct', 'month_sep',
'day_of_week_fri', 'day_of_week_mon', 'day_of_week_thu',
'day_of_week_tue', 'day_of_week_wed', 'poutcome_failure',
'poutcome_nonexistent', 'poutcome_success', 'Loan_Status_label'],
dtype='object')

```

```

[6]: X=dataset.iloc[:,1:]
X

```

```

[6]:
    campaign  pdays  previous  no_previous_contact  not_working  \
0           1    999         0                     1           0
1           1    999         0                     1           0
2           1    999         0                     1           0
3           1    999         0                     1           0
4           1    999         0                     1           0
...         ...    ...         ...                 ...           ...
41183        1    999         0                     1           1
41184        1    999         0                     1           0
41185        2    999         0                     1           1
41186        1    999         0                     1           0
41187        3    999         1                     1           1

    job_admin.  job_blue-collar  job_entrepreneur  job_housemaid  \
0             0                0                 0              1
1             0                0                 0              0
2             0                0                 0              0
3             1                0                 0              0
4             0                0                 0              0
...         ...                ...                 ...              ...
41183         0                0                 0              0
41184         0                1                 0              0
41185         0                0                 0              0
41186         0                0                 0              0
41187         0                0                 0              0

    job_management  ...  month_sep  day_of_week_fri  day_of_week_mon  \
0                 0  ...         0                 0                 1

```

1	0	...	0	0	1
2	0	...	0	0	1
3	0	...	0	0	1
4	0	...	0	0	1
...	...	...	...	...	...
41183	0	...	0	1	0
41184	0	...	0	1	0
41185	0	...	0	1	0
41186	0	...	0	1	0
41187	0	...	0	1	0

	day_of_week_thu	day_of_week_tue	day_of_week_wed	poutcome_failure	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	
...	...	...	...	...	
41183	0	0	0	0	0
41184	0	0	0	0	0
41185	0	0	0	0	0
41186	0	0	0	0	0
41187	0	0	0	0	1

	poutcome_nonexistent	poutcome_success	Loan_Status_label
0	1	0	0
1	1	0	0
2	1	0	0
3	1	0	0
4	1	0	0
...	...	...	...
41183	1	0	1
41184	1	0	0
41185	1	0	0
41186	1	0	1
41187	0	0	0

[41188 rows x 59 columns]

```
[7]: y=dataset['job_housemaid']
```

```
[8]: y
```

```
[8]: 0    1
      1    0
      2    0
      3    0
```

```

4          0
..
41183      0
41184      0
41185      0
41186      0
41187      0
Name: job_housemaid, Length: 41188, dtype: int64

```

```
[9]: X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.
      ↪2,random_state=45)
```

```
[10]: X_train
```

```
[10]:
      campaign  pdays  previous  no_previous_contact  not_working  \
6125          3    999          0                    1            0
7608          2    999          0                    1            0
14291         3    999          0                    1            0
29172         1    999          0                    1            0
21909         1    999          0                    1            0
...         ...    ...         ...                  ...          ...
32469         4    999          1                    1            0
16317         5    999          0                    1            0
12895         2    999          0                    1            0
6012          1    999          0                    1            0
6558          1    999          0                    1            0

      job_admin.  job_blue-collar  job_entrepreneur  job_housemaid  \
6125            0                0                  0              0
7608            0                1                  0              0
14291           0                0                  0              0
29172           0                1                  0              0
21909           0                0                  0              0
...         ...                ...                  ...          ...
32469           0                0                  0              0
16317           1                0                  0              0
12895           1                0                  0              0
6012            0                1                  0              0
6558            0                1                  0              0

      job_management  ...  month_sep  day_of_week_fri  day_of_week_mon  \
6125                0  ...          0                0                0
7608                0  ...          0                1                0
14291               0  ...          0                0                1
29172               0  ...          0                1                0
21909               0  ...          0                0                0
...                ...  ...         ...              ...              ...

```

32469	0	...	0	1	0
16317	0	...	0	0	0
12895	0	...	0	0	0
6012	0	...	0	0	0
6558	0	...	0	0	0

	day_of_week_thu	day_of_week_tue	day_of_week_wed	poutcome_failure	\
6125	0	1	0	0	
7608	0	0	0	0	
14291	0	0	0	0	
29172	0	0	0	0	
21909	0	0	1	0	
...	...	...	...	...	
32469	0	0	0	1	
16317	0	1	0	0	
12895	0	1	0	0	
6012	0	1	0	0	
6558	0	0	1	0	

	poutcome_nonexistent	poutcome_success	Loan_Status_label
6125	1	0	0
7608	1	0	0
14291	1	0	0
29172	1	0	1
21909	1	0	0
...	...	...	...
32469	0	0	0
16317	1	0	0
12895	1	0	0
6012	1	0	0
6558	1	0	0

[32950 rows x 59 columns]

```
[11]: X_test
```

```
[11]:
```

	campaign	pdays	previous	no_previous_contact	not_working	\
40598	2	9	1	0	1	
39325	1	3	4	0	1	
8430	2	999	0	1	0	
14127	2	999	0	1	0	
24296	1	999	0	1	0	
...	...	...	...	...	...	
8668	5	999	0	1	0	
17752	3	999	0	1	1	
6326	2	999	0	1	0	
32909	2	999	0	1	0	

33600	2	999	1	1	0
-------	---	-----	---	---	---

	job_admin.	job_blue-collar	job_entrepreneur	job_housemaid	\
40598	0	0	0	0	
39325	0	0	0	0	
8430	0	0	0	0	
14127	0	0	0	0	
24296	0	1	0	0	
...	...	...	...	...	
8668	0	1	0	0	
17752	0	0	0	0	
6326	0	0	0	0	
32909	1	0	0	0	
33600	0	1	0	0	

	job_management	...	month_sep	day_of_week_fri	day_of_week_mon	\
40598	0	...	1	1	0	
39325	0	...	0	0	1	
8430	0	...	0	0	0	
14127	0	...	0	0	1	
24296	0	...	0	0	1	
...	...	...	...	...	...	
8668	0	...	0	0	0	
17752	0	...	0	0	0	
6326	1	...	0	0	0	
32909	0	...	0	0	1	
33600	0	...	0	0	0	

	day_of_week_thu	day_of_week_tue	day_of_week_wed	poutcome_failure	\
40598	0	0	0	0	
39325	0	0	0	0	
8430	0	1	0	0	
14127	0	0	0	0	
24296	0	0	0	0	
...	...	...	...	...	
8668	0	0	1	0	
17752	0	1	0	0	
6326	0	1	0	0	
32909	0	0	0	0	
33600	0	1	0	1	

	poutcome_nonexistent	poutcome_success	Loan_Status_label
40598	0	1	1
39325	0	1	1
8430	1	0	0
14127	1	0	0
24296	1	0	0



...	...	...	...
8668	1	0	0
17752	1	0	0
6326	1	0	0
32909	1	0	0
33600	0	0	0

[8238 rows x 59 columns]

```
[12]: xgb=XGBClassifier()
      xgb.fit(X_train,y_train)
```

```
[12]: XGBClassifier(base_score=None, booster=None, callbacks=None,
                    colsample_bylevel=None, colsample_bynode=None,
                    colsample_bytree=None, device=None, early_stopping_rounds=None,
                    enable_categorical=False, eval_metric=None, feature_types=None,
                    gamma=None, grow_policy=None, importance_type=None,
                    interaction_constraints=None, learning_rate=None, max_bin=None,
                    max_cat_threshold=None, max_cat_to_onehot=None,
                    max_delta_step=None, max_depth=None, max_leaves=None,
                    min_child_weight=None, missing=nan, monotone_constraints=None,
                    multi_strategy=None, n_estimators=None, n_jobs=None,
                    num_parallel_tree=None, random_state=None, ...)
```

```
[13]: xgb.score(X_train,y_train)
```

```
[13]: 1.0
```

```
[14]: xgb.fit(X_test,y_test)
```

```
[14]: XGBClassifier(base_score=None, booster=None, callbacks=None,
                    colsample_bylevel=None, colsample_bynode=None,
                    colsample_bytree=None, device=None, early_stopping_rounds=None,
                    enable_categorical=False, eval_metric=None, feature_types=None,
                    gamma=None, grow_policy=None, importance_type=None,
                    interaction_constraints=None, learning_rate=None, max_bin=None,
                    max_cat_threshold=None, max_cat_to_onehot=None,
                    max_delta_step=None, max_depth=None, max_leaves=None,
                    min_child_weight=None, missing=nan, monotone_constraints=None,
                    multi_strategy=None, n_estimators=None, n_jobs=None,
                    num_parallel_tree=None, random_state=None, ...)
```

```
[15]: y_pred_train=xgb.predict(X_train)
      y_pred_test=xgb.predict(X_test)
```

```
[16]: y_train
```

```
[16]: 6125      0
      7608      0
      14291     0
      29172     0
      21909     0
      ..
      32469     0
      16317     0
      12895     0
      6012      0
      6558      0
      Name: job_housemaid, Length: 32950, dtype: int64
```

```
[17]: y_pred_train
```

```
[17]: array([0, 0, 0, ..., 0, 0, 0])
```

```
[18]: #Evaluation
```

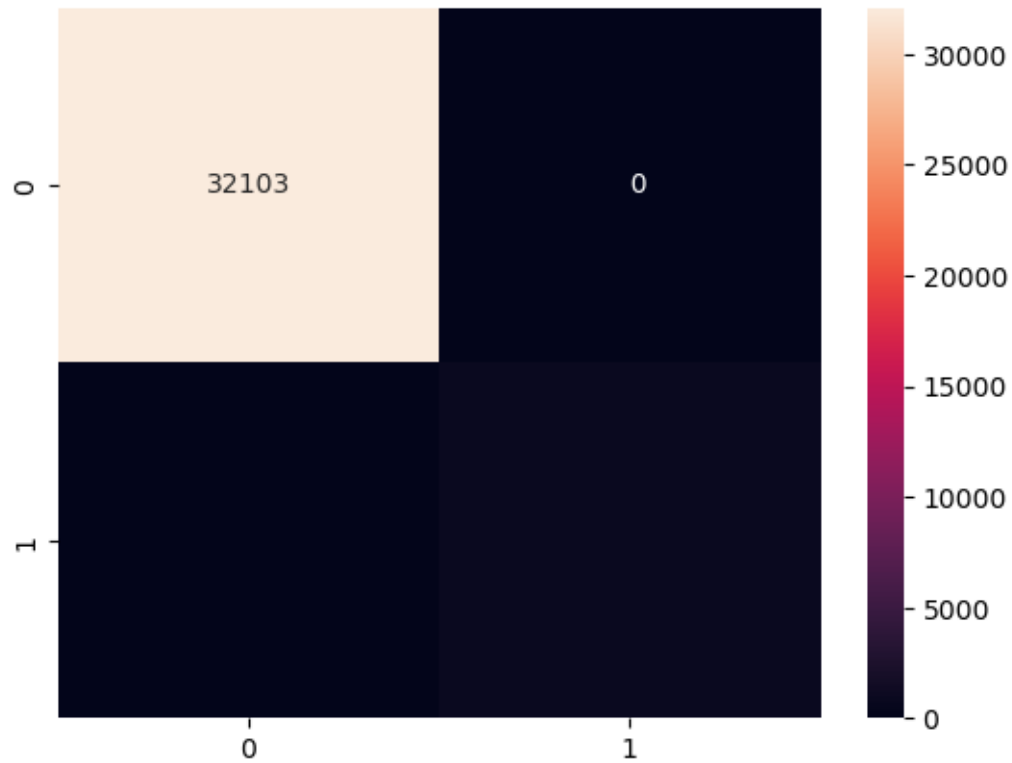
```
[19]: import seaborn as sns
```

```
[20]: confusion_matrix(y_train,y_pred_train)
```

```
[20]: array([[32103,      0],
        [      0,   847]], dtype=int64)
```

```
[21]: sns.heatmap(confusion_matrix(y_train,y_pred_train),annot=True,fmt='3.0f')
```

```
[21]: <Axes: >
```



```
[22]: accuracy_score(y_train,y_pred_train)
```

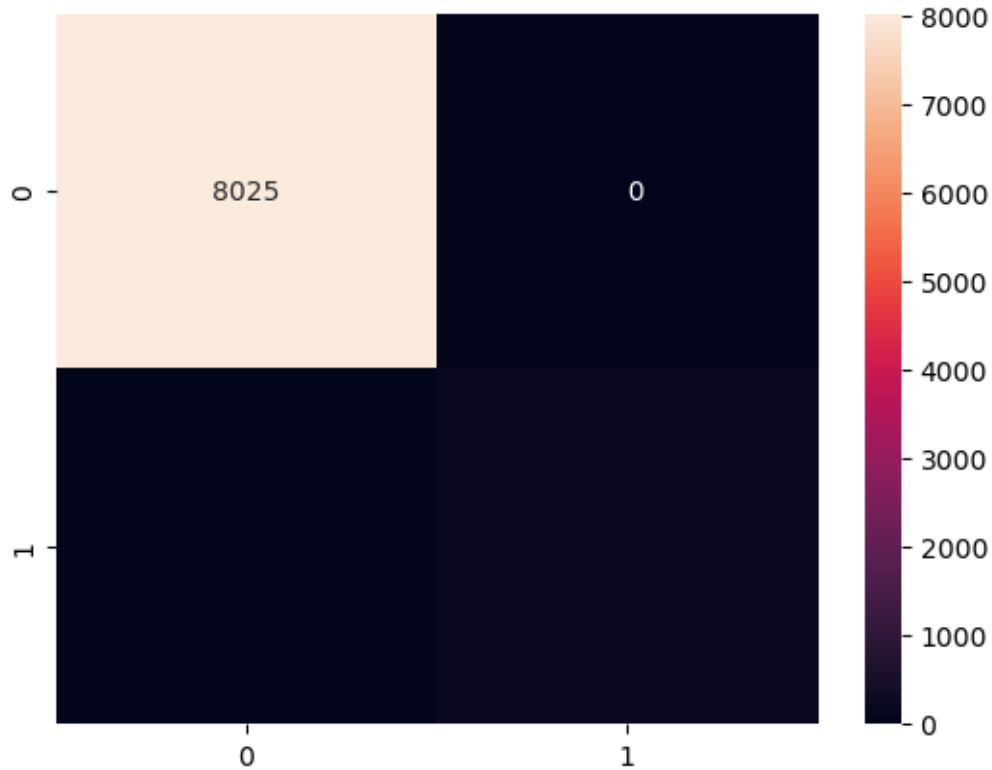
```
[22]: 1.0
```

```
[23]: confusion_matrix(y_test,y_pred_test)
```

```
[23]: array([[8025,    0],
          [    0,  213]], dtype=int64)
```

```
[24]: sns.heatmap(confusion_matrix(y_test,y_pred_test),annot=True,fmt='3.0f')
```

```
[24]: <Axes: >
```



```
[25]: accuracy_score(y_test,y_pred_test)
```

```
[25]: 1.0
```

```
[26]: xg=XGBClassifier(gamma=0.5,reg_alpha=0.6,reg_lambda=0.3)
      xg.fit(X_train,y_train)
```

```
[26]: XGBClassifier(base_score=None, booster=None, callbacks=None,
                    colsample_bylevel=None, colsample_bynode=None,
                    colsample_bytree=None, device=None, early_stopping_rounds=None,
                    enable_categorical=False, eval_metric=None, feature_types=None,
                    gamma=0.5, grow_policy=None, importance_type=None,
                    interaction_constraints=None, learning_rate=None, max_bin=None,
                    max_cat_threshold=None, max_cat_to_onehot=None,
                    max_delta_step=None, max_depth=None, max_leaves=None,
                    min_child_weight=None, missing=nan, monotone_constraints=None,
                    multi_strategy=None, n_estimators=None, n_jobs=None,
                    num_parallel_tree=None, random_state=None, ...)
```

```
[27]: xg.score(X_train,y_train)
```

```
[27]: 1.0
```

```
[28]: xg.score(X_test,y_test)
```

```
[28]: 1.0
```

```
[30]: import numpy as np
import pandas as pd

from sklearn.model_selection import train_test_split
from sklearn.ensemble import AdaBoostClassifier
from sklearn.metrics import \
    confusion_matrix, accuracy_score, classification_report
```

```
[31]: dataset = pd.read_csv(r'D:\loan_detection.csv')
print(dataset)
```

	age	campaign	pdays	previous	no_previous_contact	not_working	\
0	56	1	999	0	1	0	
1	57	1	999	0	1	0	
2	37	1	999	0	1	0	
3	40	1	999	0	1	0	
4	56	1	999	0	1	0	
...	...	...	...	...	...	...	
41183	73	1	999	0	1	1	
41184	46	1	999	0	1	0	
41185	56	2	999	0	1	1	
41186	44	1	999	0	1	0	
41187	74	3	999	1	1	1	

	job_admin.	job_blue-collar	job_entrepreneur	job_housemaid	...	\
0	0	0	0	1	...	
1	0	0	0	0	...	
2	0	0	0	0	...	
3	1	0	0	0	...	
4	0	0	0	0	...	
...	...	...	...	...	...	
41183	0	0	0	0	...	
41184	0	1	0	0	...	
41185	0	0	0	0	...	
41186	0	0	0	0	...	
41187	0	0	0	0	...	

	month_sep	day_of_week_fri	day_of_week_mon	day_of_week_thu	\
0	0	0	1	0	
1	0	0	1	0	
2	0	0	1	0	
3	0	0	1	0	
4	0	0	1	0	
...	...	...	...	...	

41183	0	1	0	0
41184	0	1	0	0
41185	0	1	0	0
41186	0	1	0	0
41187	0	1	0	0

	day_of_week_tue	day_of_week_wed	poutcome_failure	\
0	0	0	0	
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	0	
...	...	...	...	
41183	0	0	0	
41184	0	0	0	
41185	0	0	0	
41186	0	0	0	
41187	0	0	1	

	poutcome_nonexistent	poutcome_success	Loan_Status_label
0	1	0	0
1	1	0	0
2	1	0	0
3	1	0	0
4	1	0	0
...	...	...	...
41183	1	0	1
41184	1	0	0
41185	1	0	0
41186	1	0	1
41187	0	0	0

[41188 rows x 60 columns]

```
[32]: X=dataset.iloc[:,1:]
X
```

```
[32]:
```

	campaign	pdays	previous	no_previous_contact	not_working	\
0	1	999	0	1	0	
1	1	999	0	1	0	
2	1	999	0	1	0	
3	1	999	0	1	0	
4	1	999	0	1	0	
...	...	...	...	...	...	
41183	1	999	0	1	1	
41184	1	999	0	1	0	
41185	2	999	0	1	1	

41186	1	999	0	1	0
41187	3	999	1	1	1

	job_admin.	job_blue-collar	job_entrepreneur	job_housemaid	\
0	0	0	0	1	
1	0	0	0	0	
2	0	0	0	0	
3	1	0	0	0	
4	0	0	0	0	
...	...	...	...	...	
41183	0	0	0	0	
41184	0	1	0	0	
41185	0	0	0	0	
41186	0	0	0	0	
41187	0	0	0	0	

	job_management	...	month_sep	day_of_week_fri	day_of_week_mon	\
0	0	...	0	0	1	
1	0	...	0	0	1	
2	0	...	0	0	1	
3	0	...	0	0	1	
4	0	...	0	0	1	
...	...	...	...	...	...	
41183	0	...	0	1	0	
41184	0	...	0	1	0	
41185	0	...	0	1	0	
41186	0	...	0	1	0	
41187	0	...	0	1	0	

	day_of_week_thu	day_of_week_tue	day_of_week_wed	poutcome_failure	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	
...	...	...	...	...	
41183	0	0	0	0	
41184	0	0	0	0	
41185	0	0	0	0	
41186	0	0	0	0	
41187	0	0	0	1	

	poutcome_nonexistent	poutcome_success	Loan_Status_label
0	1	0	0
1	1	0	0
2	1	0	0
3	1	0	0

4		1		0		0
...		...		...		...
41183		1		0		1
41184		1		0		0
41185		1		0		0
41186		1		0		1
41187		0		0		0

[41188 rows x 59 columns]

```
[33]: y=dataset['Loan_Status_label']
      y
```

```
[33]: 0      0
      1      0
      2      0
      3      0
      4      0
      ..
      41183    1
      41184    0
      41185    0
      41186    1
      41187    0
      Name: Loan_Status_label, Length: 41188, dtype: int64
```

```
[34]: clf=AdaBoostClassifier()
      clf.fit(X_train,y_train)
```

```
[34]: AdaBoostClassifier()
```

```
[35]: clf.score(X_train,y_train)
```

```
[35]: 1.0
```

```
[37]: clf.score(X_test,y_test)
```

```
[37]: 1.0
```

```
[39]: y_pred_train=clf.predict(X_train)
      y_pred_test=clf.predict(X_test)
```

```
[40]: y_train
```

```
[40]: 6125    0
      7608    0
      14291   0
      29172   0
```



```

21909    0
      ..
32469    0
16317    0
12895    0
6012     0
6558     0
Name: job_housemaid, Length: 32950, dtype: int64

```

```
[41]: y_pred_train
```

```
[41]: array([0, 0, 0, ..., 0, 0, 0], dtype=int64)
```

```
[42]: clf=AdaBoostClassifier(n_estimators=100,random_state=44)
      clf.fit(X_train,y_train)
```

```
[42]: AdaBoostClassifier(n_estimators=100, random_state=44)
```

```
[43]: clf.score(X_train,y_train)
```

```
[43]: 1.0
```

```
[44]: clf.score(X_test,y_test)
```

```
[44]: 1.0
```

```
[45]: from sklearn.ensemble import RandomForestClassifier
```

```
[52]: clf = AdaBoostClassifier(
      base_estimator=RandomForestClassifier(criterion='entropy'),
      n_estimators=100,
      random_state=44
    )
      clf.fit(X_train,y_train)
```

```

C:\ProgramData\anaconda3\Lib\site-packages\sklearn\ensemble\_base.py:166:
FutureWarning: `base_estimator` was renamed to `estimator` in version 1.2 and
will be removed in 1.4.
  warnings.warn(

```

```
[52]: AdaBoostClassifier(base_estimator=RandomForestClassifier(criterion='entropy'),
      n_estimators=100, random_state=44)
```

```
[53]: #using svc
```

```
[55]: from sklearn.svm import SVC
```

```
[57]: svc=SVC()
      svc.fit(X_train,y_train)
      svc_pred_train=svc.predict(X_train)
      svc_pred_test=svc.predict(X_test)
```

```
[58]: print(confusion_matrix(y_train,svc_pred_train))
      print()
      print(accuracy_score(y_train,svc_pred_train))
```

```
[[32103    0]
 [  847    0]]
```

0.9742943854324735

```
[60]: print(confusion_matrix(y_test,svc_pred_test))
      print()
      print(accuracy_score(y_test,svc_pred_test))
```

```
[[8025    0]
 [  213    0]]
```

0.9741442097596504

```
[61]: #logistic regression
```

```
[63]: from sklearn.linear_model import LogisticRegression
```

```
[64]: lr=LogisticRegression()
      lr.fit(X_train,y_train)
      lr_pred_train=lr.predict(X_train)
      lr_pred_test=lr.predict(X_test)
```

C:\ProgramData\anaconda3\Lib\site-packages\sklearn\linear\_model\\_logistic.py:458: ConvergenceWarning: lbfgs failed to converge (status=1):  
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max\_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

[https://scikit-learn.org/stable/modules/linear\\_model.html#logistic-regression](https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
```

```
[65]: logreg = LogisticRegression(max_iter=200) # Increase from the default (100)
```

```
[66]: from sklearn.preprocessing import StandardScaler

      # Create a scaler
```

```

scaler = StandardScaler()

# Fit the scaler on the training data and transform both training and testing
↳ data
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# Now fit the model on the scaled data
logreg = LogisticRegression(max_iter=200)
logreg.fit(X_train_scaled, y_train)

```

```
[66]: LogisticRegression(max_iter=200)
```

```
[67]: logreg = LogisticRegression(solver='liblinear', max_iter=200)
```

```
[68]: lr=LogisticRegression()
lr.fit(X_train,y_train)
lr_pred_train=lr.predict(X_train)
lr_pred_test=lr.predict(X_test)
```

C:\ProgramData\anaconda3\Lib\site-packages\sklearn\linear\_model\\_logistic.py:458: ConvergenceWarning: lbfgs failed to converge (status=1):  
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max\_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

[https://scikit-learn.org/stable/modules/linear\\_model.html#logistic-regression](https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
```

```
[ ]:
```