

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

```
In [2]: df=pd.read_csv(r"C:\Users\user\Downloads\C8_loan-train.csv")
df
```

Out[2]:

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Am
0	LP001002	Male	No	0	Graduate	No	5849	0.0	NaN	
1	LP001003	Male	Yes	1	Graduate	No	4583	1508.0	128.0	
2	LP001005	Male	Yes	0	Graduate	Yes	3000	0.0	66.0	
3	LP001006	Male	Yes	0	Not Graduate	No	2583	2358.0	120.0	
4	LP001008	Male	No	0	Graduate	No	6000	0.0	141.0	
...	...	...	...	...	...	...	...	...	...	
609	LP002978	Female	No	0	Graduate	No	2900	0.0	71.0	
610	LP002979	Male	Yes	3+	Graduate	No	4106	0.0	40.0	
611	LP002983	Male	Yes	1	Graduate	No	8072	240.0	253.0	
612	LP002984	Male	Yes	2	Graduate	No	7583	0.0	187.0	
613	LP002990	Female	No	0	Graduate	Yes	4583	0.0	133.0	

614 rows × 13 columns

```
In [3]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 614 entries, 0 to 613
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Loan_ID                614 non-null    object
1   Gender                 601 non-null    object
2   Married                611 non-null    object
3   Dependents             599 non-null    object
4   Education              614 non-null    object
5   Self_Employed          582 non-null    object
6   ApplicantIncome        614 non-null    int64
7   CoapplicantIncome      614 non-null    float64
8   LoanAmount             592 non-null    float64
9   Loan_Amount_Term       600 non-null    float64
10  Credit_History         564 non-null    float64
11  Property_Area          614 non-null    object
12  Loan_Status            614 non-null    object
dtypes: float64(4), int64(1), object(8)
memory usage: 62.5+ KB
```

```
In [4]: df=df.dropna()
```

```
In [5]: df.isnull().sum()
```

```
Out[5]: Loan_ID      0
Gender      0
Married     0
Dependents  0
Education   0
Self_Employed  0
ApplicantIncome  0
CoapplicantIncome  0
LoanAmount   0
Loan_Amount_Term  0
Credit_History  0
Property_Area  0
Loan_Status  0
dtype: int64
```

```
In [6]: df.describe()
```

```
Out[6]:
```

	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History
count	480.000000	480.000000	480.000000	480.000000	480.000000
mean	5364.231250	1581.093583	144.735417	342.050000	0.854167
std	5668.251251	2617.692267	80.508164	65.212401	0.353307
min	150.000000	0.000000	9.000000	36.000000	0.000000
25%	2898.750000	0.000000	100.000000	360.000000	1.000000
50%	3859.000000	1084.500000	128.000000	360.000000	1.000000
75%	5852.500000	2253.250000	170.000000	360.000000	1.000000
max	81000.000000	33837.000000	600.000000	480.000000	1.000000

```
In [7]: df.columns
```

```
Out[7]: Index(['Loan_ID', 'Gender', 'Married', 'Dependents', 'Education',
              'Self_Employed', 'ApplicantIncome', 'CoapplicantIncome', 'LoanAmount',
              'Loan_Amount_Term', 'Credit_History', 'Property_Area', 'Loan_Status'],
              dtype='object')
```

```
In [8]: df['Gender'].value_counts()
```

```
Out[8]: Male      394
Female    86
Name: Gender, dtype: int64
```

```
In [9]: g1={"Gender":{"Female":1,'Male':2}}
df=df.replace(g1)
print(df)
```

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	\
1	LP001003	2	Yes	1	Graduate	No	
2	LP001005	2	Yes	0	Graduate	Yes	
3	LP001006	2	Yes	0	Not Graduate	No	
4	LP001008	2	No	0	Graduate	No	
5	LP001011	2	Yes	2	Graduate	Yes	
..	...	...	...	...	...	...	
609	LP002978	1	No	0	Graduate	No	
610	LP002979	2	Yes	3+	Graduate	No	
611	LP002983	2	Yes	1	Graduate	No	
612	LP002984	2	Yes	2	Graduate	No	
613	LP002990	1	No	0	Graduate	Yes	

	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	\
1	4583	1508.0	128.0	360.0	
2	3000	0.0	66.0	360.0	
3	2583	2358.0	120.0	360.0	
4	6000	0.0	141.0	360.0	
5	5417	4196.0	267.0	360.0	
..	...	...	...	...	
609	2900	0.0	71.0	360.0	
610	4106	0.0	40.0	180.0	
611	8072	240.0	253.0	360.0	
612	7583	0.0	187.0	360.0	
613	4583	0.0	133.0	360.0	

	Credit_History	Property_Area	Loan_Status
1	1.0	Rural	N
2	1.0	Urban	Y
3	1.0	Urban	Y
4	1.0	Urban	Y
5	1.0	Urban	Y
..	...	...	...
609	1.0	Rural	Y
610	1.0	Rural	Y
611	1.0	Urban	Y
612	1.0	Urban	Y
613	0.0	Semiurban	N

[480 rows x 13 columns]

```
In [10]: x=df.drop(["Gender","Loan_ID","Married","Dependents","Education","Self_Employed","Property_Area","Loan_Status"])
y=df["Gender"]
```

```
In [11]: from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,train_size=0.70)
```

```
In [12]: from sklearn.ensemble import RandomForestClassifier
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

```
Out[12]: RandomForestClassifier()
```

```
In [13]: parameters={'max_depth':[1,2,3,4,5],
                    'min_samples_leaf':[5,10,15,20,25],
                    'n_estimators':[10,20,30,40,50]}
```

```
In [14]: from sklearn.model_selection import GridSearchCV
grid_search=GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="accuracy")
grid_search.fit(x_train,y_train)
```

```
Out[14]: GridSearchCV(cv=2, estimator=RandomForestClassifier(),
                    param_grid={'max_depth': [1, 2, 3, 4, 5],
                                'min_samples_leaf': [5, 10, 15, 20, 25],
                                'n_estimators': [10, 20, 30, 40, 50]},
                    scoring='accuracy')
```

```
In [15]: grid_search.best_score_
```

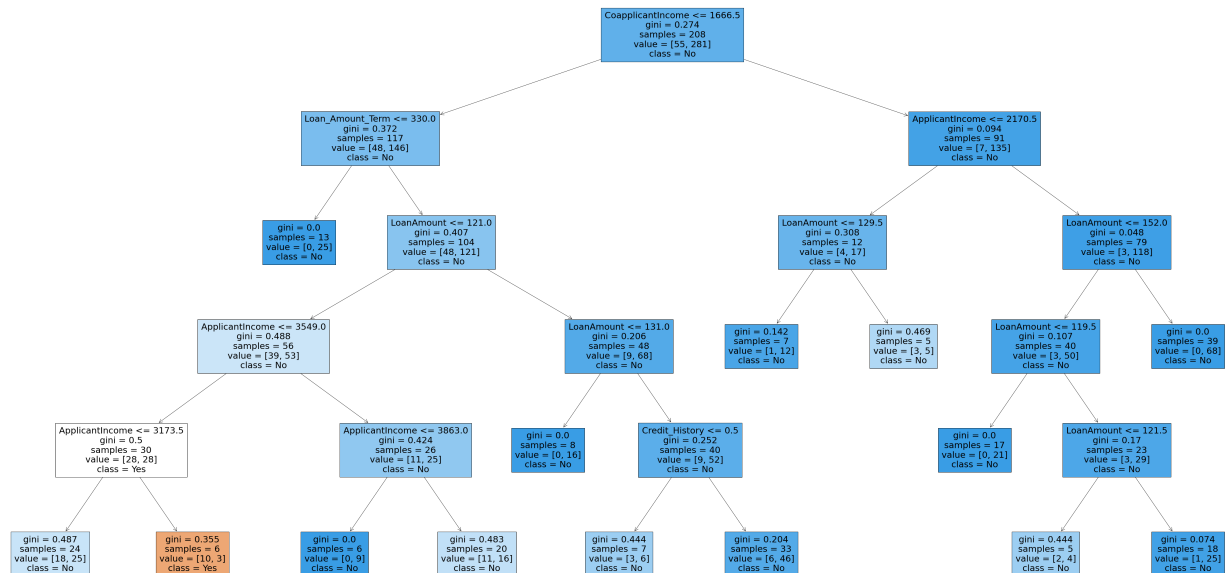
```
Out[15]: 0.8184523809523809
```

```
In [16]: parameters={'max_depth':[1,2,3,4,5],  
                    'min_samples_leaf':[5,10,15,20,25],  
                    'n_estimators':[10,20,30,40,50]}
```

```
In [17]: rfc_best=grid_search.best_estimator_
```

```
In [18]: from sklearn.tree import plot_tree
plt.figure(figsize=(80,40))
plot_tree(rfc_best.estimators_[5],feature_names=x.columns,class_names=['Yes','No'],filled=True)
```

```
Out[18]: [Text(2418.0, 1993.2, 'CoapplicantIncome <= 1666.5\ngini = 0.274\nsamples = 208\nvalue = [55, 281]\nnclass = No'),
Text(1364.0, 1630.8000000000002, 'Loan_Amount_Term <= 330.0\ngini = 0.372\nsamples = 117\nvalue = [48, 146]\nnclass = No'),
Text(1116.0, 1268.4, 'gini = 0.0\nsamples = 13\nvalue = [0, 25]\nnclass = No'),
Text(1612.0, 1268.4, 'LoanAmount <= 121.0\ngini = 0.407\nsamples = 104\nvalue = [48, 121]\nnclass = No'),
Text(992.0, 906.0, 'ApplicantIncome <= 3549.0\ngini = 0.488\nsamples = 56\nvalue = [39, 53]\nnclass = No'),
Text(496.0, 543.5999999999999, 'ApplicantIncome <= 3173.5\ngini = 0.5\nsamples = 30\nvalue = [28, 28]\nnclass = Yes'),
Text(248.0, 181.19999999999998, 'gini = 0.487\nsamples = 24\nvalue = [18, 25]\nnclass = No'),
Text(744.0, 181.19999999999998, 'gini = 0.355\nsamples = 6\nvalue = [10, 3]\nnclass = Yes'),
Text(1488.0, 543.5999999999999, 'ApplicantIncome <= 3863.0\ngini = 0.424\nsamples = 26\nvalue = [11, 25]\nnclass = No'),
Text(1240.0, 181.19999999999998, 'gini = 0.0\nsamples = 6\nvalue = [0, 9]\nnclass = No'),
Text(1736.0, 181.19999999999998, 'gini = 0.483\nsamples = 20\nvalue = [11, 16]\nnclass = No'),
Text(2232.0, 906.0, 'LoanAmount <= 131.0\ngini = 0.206\nsamples = 48\nvalue = [9, 68]\nnclass = No'),
Text(1984.0, 543.5999999999999, 'gini = 0.0\nsamples = 8\nvalue = [0, 16]\nnclass = No'),
Text(2480.0, 543.5999999999999, 'Credit_History <= 0.5\ngini = 0.252\nsamples = 40\nvalue = [9, 52]\nnclass = No'),
Text(2232.0, 181.19999999999998, 'gini = 0.444\nsamples = 7\nvalue = [3, 6]\nnclass = No'),
Text(2728.0, 181.19999999999998, 'gini = 0.204\nsamples = 33\nvalue = [6, 46]\nnclass = No'),
Text(3472.0, 1630.8000000000002, 'ApplicantIncome <= 2170.5\ngini = 0.094\nsamples = 91\nvalue = [7, 135]\nnclass = No'),
Text(2976.0, 1268.4, 'LoanAmount <= 129.5\ngini = 0.308\nsamples = 12\nvalue = [4, 17]\nnclass = No'),
Text(2728.0, 906.0, 'gini = 0.142\nsamples = 7\nvalue = [1, 12]\nnclass = No'),
Text(3224.0, 906.0, 'gini = 0.469\nsamples = 5\nvalue = [3, 5]\nnclass = No'),
Text(3968.0, 1268.4, 'LoanAmount <= 152.0\ngini = 0.048\nsamples = 79\nvalue = [3, 118]\nnclass = No'),
Text(3720.0, 906.0, 'LoanAmount <= 119.5\ngini = 0.107\nsamples = 40\nvalue = [3, 50]\nnclass = No'),
Text(3472.0, 543.5999999999999, 'gini = 0.0\nsamples = 17\nvalue = [0, 21]\nnclass = No'),
Text(3968.0, 543.5999999999999, 'LoanAmount <= 121.5\ngini = 0.17\nsamples = 23\nvalue = [3, 29]\nnclass = No'),
Text(3720.0, 181.19999999999998, 'gini = 0.444\nsamples = 5\nvalue = [2, 4]\nnclass = No'),
Text(4216.0, 181.19999999999998, 'gini = 0.074\nsamples = 18\nvalue = [1, 25]\nnclass = No'),
Text(4216.0, 906.0, 'gini = 0.0\nsamples = 39\nvalue = [0, 68]\nnclass = No')]
```



In [ ]: