

```
In [1]: import pandas as pd
import warnings
warnings.filterwarnings("ignore")
data=pd.read_csv("/home/placement/Downloads/TelecomCustomerChurn.csv")
```

In [2]: `print(data)`

```

      customerID  gender  SeniorCitizen  Partner  Dependents  tenure  \
0      7590-VHVEG  Female              0      Yes           No         1
1      5575-GNVDE   Male              0      No            No        34
2      3668-QPYBK   Male              0      No            No         2
3      7795-CF0CW   Male              0      No            No        45
4      9237-HQITU  Female              0      No            No         2
...      ...      ...      ...      ...      ...      ...
7038  6840-RESVB   Male              0      Yes           Yes        24
7039  2234-XADUH  Female              0      Yes           Yes        72
7040  4801-JZAZL  Female              0      Yes           Yes        11
7041  8361-LTMKD   Male              1      Yes           No         4
7042  3186-AJIEK   Male              0      No            No        66

      PhoneService  MultipleLines  InternetService  OnlineSecurity  ...  \
0              No  No phone service              DSL              No  ...
1              Yes              No              DSL              Yes  ...
2              Yes              No              DSL              Yes  ...
3              No  No phone service              DSL              Yes  ...
4              Yes              No  Fiber optic              No  ...
...      ...      ...      ...      ...      ...
7038          Yes              Yes              DSL              Yes  ...
7039          Yes              Yes  Fiber optic              No  ...
7040          No  No phone service              DSL              Yes  ...
7041          Yes              Yes  Fiber optic              No  ...
7042          Yes              No  Fiber optic              Yes  ...

      DeviceProtection  TechSupport  StreamingTV  StreamingMovies  Contract  \
0              No              No              No              No  Month-to-month
1              Yes              No              No              No      One year
2              No              No              No              No  Month-to-month
3              Yes              Yes              No              No      One year
4              No              No              No              No  Month-to-month
...      ...      ...      ...      ...      ...
7038          Yes              Yes              Yes              Yes      One year
7039          Yes              No              Yes              Yes      One year
7040          No              No              No              No  Month-to-month
7041          No              No              No              No  Month-to-month
7042          Yes              Yes              Yes              Yes     Two year

```

	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	\
0	Yes	Electronic check	29.85	29.85	
1	No	Mailed check	56.95	1889.5	
2	Yes	Mailed check	53.85	108.15	
3	No	Bank transfer (automatic)	42.30	1840.75	
4	Yes	Electronic check	70.70	151.65	
...	
7038	Yes	Mailed check	84.80	1990.5	
7039	Yes	Credit card (automatic)	103.20	7362.9	
7040	Yes	Electronic check	29.60	346.45	
7041	Yes	Mailed check	74.40	306.6	
7042	Yes	Bank transfer (automatic)	105.65	6844.5	

	Churn
0	No
1	No
2	Yes
3	No
4	Yes
...	...
7038	No
7039	No
7040	No
7041	Yes
7042	No

[7043 rows x 21 columns]

```
In [3]: data['TotalCharges'] = pd.to_numeric(data['TotalCharges'], errors='coerce')
```

```
In [4]: data.describe()
```

```
Out[4]:
```

	SeniorCitizen	tenure	MonthlyCharges	TotalCharges
count	7043.000000	7043.000000	7043.000000	7032.000000
mean	0.162147	32.371149	64.761692	2283.300441
std	0.368612	24.559481	30.090047	2266.771362
min	0.000000	0.000000	18.250000	18.800000
25%	0.000000	9.000000	35.500000	401.450000
50%	0.000000	29.000000	70.350000	1397.475000
75%	0.000000	55.000000	89.850000	3794.737500
max	1.000000	72.000000	118.750000	8684.800000

In [5]: data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
#   Column                Non-Null Count  Dtype
---  -
0   customerID            7043 non-null   object
1   gender                7043 non-null   object
2   SeniorCitizen         7043 non-null   int64
3   Partner               7043 non-null   object
4   Dependents            7043 non-null   object
5   tenure                7043 non-null   int64
6   PhoneService          7043 non-null   object
7   MultipleLines         7043 non-null   object
8   InternetService       7043 non-null   object
9   OnlineSecurity        7043 non-null   object
10  OnlineBackup          7043 non-null   object
11  DeviceProtection      7043 non-null   object
12  TechSupport           7043 non-null   object
13  StreamingTV           7043 non-null   object
14  StreamingMovies       7043 non-null   object
15  Contract              7043 non-null   object
16  PaperlessBilling      7043 non-null   object
17  PaymentMethod         7043 non-null   object
18  MonthlyCharges        7043 non-null   float64
19  TotalCharges          7032 non-null   float64
20  Churn                 7043 non-null   object
dtypes: float64(2), int64(2), object(17)
memory usage: 1.1+ MB
```

```
In [6]: data.isna().sum()
```

```
Out[6]: customerID      0  
gender      0  
SeniorCitizen  0  
Partner      0  
Dependents    0  
tenure      0  
PhoneService  0  
MultipleLines  0  
InternetService  0  
OnlineSecurity  0  
OnlineBackup  0  
DeviceProtection  0  
TechSupport    0  
StreamingTV    0  
StreamingMovies  0  
Contract      0  
PaperlessBilling  0  
PaymentMethod  0  
MonthlyCharges  0  
TotalCharges   11  
Churn          0  
dtype: int64
```

```
In [7]: list(data)
```

```
Out[7]: ['customerID',  
        'gender',  
        'SeniorCitizen',  
        'Partner',  
        'Dependents',  
        'tenure',  
        'PhoneService',  
        'MultipleLines',  
        'InternetService',  
        'OnlineSecurity',  
        'OnlineBackup',  
        'DeviceProtection',  
        'TechSupport',  
        'StreamingTV',  
        'StreamingMovies',  
        'Contract',  
        'PaperlessBilling',  
        'PaymentMethod',  
        'MonthlyCharges',  
        'TotalCharges',  
        'Churn']
```

In [8]: data

Out[8]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DevicePro
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	
...	
7038	6840-RESVB	Male	0	Yes	Yes	24	Yes	Yes	DSL	Yes	...	
7039	2234-XADUH	Female	0	Yes	Yes	72	Yes	Yes	Fiber optic	No	...	
7040	4801-JJAZL	Female	0	Yes	Yes	11	No	No phone service	DSL	Yes	...	
7041	8361-LTMKD	Male	1	Yes	No	4	Yes	Yes	Fiber optic	No	...	
7042	3186-AJIEK	Male	0	No	No	66	Yes	No	Fiber optic	Yes	...	

7043 rows × 21 columns



In [9]: data1=data.fillna(data.median())


```
In [10]: data1.isna().sum()
```

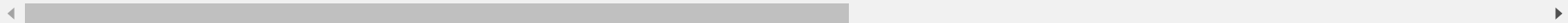
```
Out[10]: customerID      0
gender      0
SeniorCitizen  0
Partner      0
Dependents    0
tenure      0
PhoneService  0
MultipleLines  0
InternetService  0
OnlineSecurity  0
OnlineBackup  0
DeviceProtection  0
TechSupport    0
StreamingTV    0
StreamingMovies  0
Contract      0
PaperlessBilling  0
PaymentMethod  0
MonthlyCharges  0
TotalCharges  0
```

```
In [11]: x=data1.drop(['customerID','Churn'],axis=1)
y=data1['Churn']
```

```
In [12]: x.head()
```

```
Out[12]:
```

	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection
0	Female	0	Yes	No	1	No	No phone service	DSL	No	Yes	False
1	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes
2	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	False
3	Male	0	No	No	45	No	No phone service	DSL	Yes	No	Yes
4	Female	0	No	No	2	Yes	No	Fiber optic	No	No	False



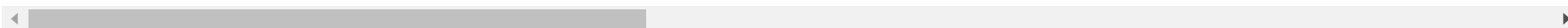
```
In [13]: x=pd.get_dummies(x)
```

In [14]: x

Out[14]:

	SeniorCitizen	tenure	MonthlyCharges	TotalCharges	gender_Female	gender_Male	Partner_No	Partner_Yes	Dependents_No	Dependents_Y
0	0	1	29.85	29.85	1	0	0	1	1	
1	0	34	56.95	1889.50	0	1	1	0	1	
2	0	2	53.85	108.15	0	1	1	0	1	
3	0	45	42.30	1840.75	0	1	1	0	1	
4	0	2	70.70	151.65	1	0	1	0	1	
...
7038	0	24	84.80	1990.50	0	1	0	1	0	
7039	0	72	103.20	7362.90	1	0	0	1	0	
7040	0	11	29.60	346.45	1	0	0	1	0	
7041	1	4	74.40	306.60	0	1	0	1	1	
7042	0	66	105.65	6844.50	0	1	1	0	1	

7043 rows × 45 columns



```
In [15]: from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.33,random_state=42)
```

```
In [16]: from sklearn.model_selection import GridSearchCV #GridSearchCV is for parameter tuning
from sklearn.ensemble import RandomForestClassifier
cls=RandomForestClassifier()
n_estimators=[25,50,75,100,125,150,175,200] #number of decision trees in the forest, default = 100
criterion=['gini','entropy'] #criteria for choosing nodes default = 'gini'
max_depth=[3,5,10] #maximum number of nodes in a tree default = None (it will go till all possible nodes)
parameters={'n_estimators': n_estimators, 'criterion': criterion, 'max_depth': max_depth} #this will undergo 8*2
RFC_cls = GridSearchCV(cls, parameters)
RFC_cls.fit(x_train,y_train)
```

```
Out[16]: GridSearchCV(estimator=RandomForestClassifier(),
                      param_grid={'criterion': ['gini', 'entropy'],
                                   'max_depth': [3, 5, 10],
                                   'n_estimators': [25, 50, 75, 100, 125, 150, 175, 200]})
```

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

```
In [17]: RFC_cls.best_params_
```

```
Out[17]: {'criterion': 'entropy', 'max_depth': 10, 'n_estimators': 200}
```

```
In [19]: cls=RandomForestClassifier(n_estimators=200,criterion='entropy',max_depth=10)
```

```
In [20]: cls.fit(x_train,y_train)
```

```
Out[20]: RandomForestClassifier(criterion='entropy', max_depth=10, n_estimators=200)
```

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

```
In [21]: rfy_pred=cls.predict(x_test)
```

```
In [22]: rfy_pred
```

```
Out[22]: array(['Yes', 'No', 'No', ..., 'Yes', 'No', 'No'], dtype=object)
```

```
In [24]: from sklearn.metrics import confusion_matrix  
confusion_matrix(y_test, rfy_pred)
```

```
Out[24]: array([[1549, 148],  
               [ 300, 328]])
```

```
In [25]: from sklearn.metrics import accuracy_score  
accuracy_score(y_test, rfy_pred)
```

```
Out[25]: 0.8073118279569892
```

```
In [ ]:
```