```
import pandas as pd
df=pd.read_csv("/content/German_Credit_Card_Dataset.csv")
df.info()
df.shape
df.head()
<class 'pandas.core.frame.DataFrame'>
    RangeIndex: 1000 entries, 0 to 999
    Data columns (total 14 columns):
        Column
                          Non-Null Count Dtype
         0
                                          object
                                          int64
     1
         credit_history 1000 non-null amount 1000 non-null
     2
                                          object
                                          int64
                          1000 non-null
     4
         savings_acc
                                          object
         personal_status residing at
         present_emp_since 1000 non-null
                                          object
                                          int64
                                          object
         residing_since
                          1000 non-null
                                          int64
         age
                           1000 non-null
                                          int64
     10 inst_plans
                         1000 non-null
                                          object
     11 num_credits 1000 non-null 12 job 1000 non-null
                                          int64
                                          obiect
                          1000 non-null
     13 status
                                          int64
```

memory usage: 109.5+ KB											
	checkin_acc	duration	credit_history	amount	savings_acc	<pre>present_emp_since</pre>	inst_rate	personal_status	residing_since	age	i
0	A11	6	A34	1169	A65	A75	4	A93	4	67	
1	A12	48	A32	5951	A61	A73	2	A92	2	22	
2	A14	12	A34	2096	A61	A74	2	A93	3	49	
3	A11	42	A32	7882	A61	A74	2	A93	4	45	
4	A11	24	A33	4870	A61	A73	3	A93	4	53	

Double-click (or enter) to edit

dtypes: int64(7), object(7)

df.iloc[0:5,0:7]

₹		checkin_acc	duration	credit_history	amount	savings_acc	present_emp_since	inst_rate
	0	A11	6	A34	1169	A65	A75	4
	1	A12	48	A32	5951	A61	A73	2
	2	A14	12	A34	2096	A61	A74	2
	3	A11	42	A32	7882	A61	A74	2
	4	A11	24	A33	4870	A61	A73	3

df.iloc[0:5,0:17]

_		checkin_acc	duration	credit_history	amount	savings_acc	present_emp_since	inst_rate	personal_status	residing_since	age	iı
	0	A11	6	A34	1169	A65	A75	4	A93	4	67	
	1	A12	48	A32	5951	A61	A73	2	A92	2	22	
	2	A14	12	A34	2096	A61	A74	2	A93	3	49	
	3	A11	42	A32	7882	A61	A74	2	A93	4	45	
	4	A11	24	A33	4870	A61	A73	3	A93	4	53	
	4											-

df['checkin_acc'].unique()

```
⇒ array(['A11', 'A12', 'A14', 'A13'], dtype=object)
```

```
x_features=list(df.columns)
x_features.remove('status')
encoded_df = pd.get_dummies(df[x_features],drop_first=True)
print(list(encoded_df.columns))
```

```
['duration', 'amount', 'inst_rate', 'residing_since', 'age', 'num_credits', 'checkin_acc_A12', 'checkin_acc_A13', 'checkin_acc_A14',
```

```
x=encoded_df
y=df['status']
from sklearn.model_selection import train_test_split
x\_train, x\_test, y\_train, y\_test=train\_test\_split(x, y, test\_size=0.3, random\_state=42)
from \ sklearn.tree \ import \ Decision Tree Classifier
clf=DecisionTreeClassifier(criterion='gini',max_depth=3)
clf.fit(x_train,y_train)
\rightarrow
              {\tt DecisionTreeClassifier}
      DecisionTreeClassifier(max_depth=3)
pred_y=clf.predict(x_test)
from sklearn import metrics
print("Confusion Matrix is\n",metrics.confusion_matrix(pred_y,y_test))
print("Accuracy is",metrics.accuracy_score(pred_y,y_test))
print("AUC Score is",metrics.roc_auc_score(pred_y,y_test))

→ Confusion Matrix is
       [[198 71]
       [ 11 20]]
     Accuracy is 0.726666666666667
     AUC Score is 0.6906103849382419
from sklearn.tree import export_graphviz
import pydotplus as pdot
from IPython.display import Image
export_graphviz(clf,out_file="tree.dot",feature_names=x_train.columns,filled=True)
graph=pdot.graph_from_dot_file("tree.dot")
graph.write_png("tree.png")
Image(filename="tree.png")
\rightarrow
                                                                      checkin_acc_A14 <= 0.5
gini = 0.419
samples = 700
value = [491, 209]
                                                                      True
                                                                                          False
                                                             duration <= 33.0
                                                                                      inst_plans A143 <= 0.5
                                                               gini = 0.484
                                                                                           gini = 0.222
                                                             samples = 425
                                                                                          samples = 275
                                                            value = [251, 174]
                                                                                         value = [240, 35]
                           amount <= 10975.5
                                                         savings acc A65 <= 0.5
                                                                                       residing since <= 2.5
                                                               gini = 0.464
                                                                                           gini = 0.423
samples = 46
                              gini = 0.458
                                                                                                                            gini = 0.167
                           samples = 343
value = [221, 122]
                                                                                                                           samples = 229
                                                             samples = 82
value = [30, 52]
                                                                                          value = [32, 14]
                                                                                                                          value = [208, 21]
                               gini = 0.0
                                                 gini = 0.431
                                                                    gini = 0.444
                                                                                       gini = 0.491
                                                                                                          gini = 0.287
                                                                                                                             gini = 0.48
                                                                                                                                                gini = 0.13
```

samples = 23 value = [19, 4]

samples = 15

value = [9, 6]

Start coding or generate with AI.

samples = 337

value = [221, 116]

samples = 6 value = [0, 6]

samples = 70 value = [22, 48]

samples = 12

value = [8, 4]

samples = 23

value = [13, 10]

samples = 214 value = [199, 15]