```
import numpy as np
import pandas as pd
from sklearn import preprocessing
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
import requests
import io
#from google.colab import drive
#drive.mount('/content')
#from google.colab import files
#uploaded = files.upload()

#url = "https://www.kaggle.com/xwolf12/datasetandroidpermissions/download/b3nIEWm6zcroShVi
#s = requests.get(url).content
#df = pd.read_csv(io.StringIO(s.decode('utf-8')), error_bad_lines=False)

df = pd.read_csv('/Testing_sample10mb')
df.describe()
#df.info()
```

```
/usr/local/lib/python3.6/dist-packages/IPython/core/interactiveshell.py:2718: DtypeWa
  interactivity=interactivity, compiler=compiler, result=result)
```

	<actionandroid:name="android.intent.action.FILEEXPLORE"/>\n	<actionandroid:n
count	634.0	
mean	0.0	
std	0.0	
min	0.0	
25%	0.0	
50%	0.0	
75%	0.0	
max	0.0	

8 rows × 8112 columns

```
X = np.array(df.drop(columns=[ '<family>','<category>','<MD5>','Binary_Type']))
Y = np.array(df['Binary_Type'])

X_train, X_test, Y_train, Y_test = train_test_split(X,Y,test_size=0.2)
clf = KNeighborsClassifier()
clf.fit(X_train, Y_train)

accuracy = clf.score(X_test, Y_test)
print(accuracy)
```

```
0.8267716535433071
```