```python
import numpy as np
import pandas as pd
from sklearn import preprocessing
#from sklearn.neighbors import KNeighborsClassifier
from sklearn.ensemble import RandomForestClassifier
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)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 634 entries, 0 to 633
Columns: 8115 entries, <actionandroid:name="android.intent.action.FILEEXPLORE"/>
 to Binary_Type
dtypes: int64(8112), object(3)
memory usage: 39.3+ MB
```

```python
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 = RandomForestClassifier()
clf.fit(X_train, Y_train)

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

```
0.968503937007874
```