In [1]:
```
import pandas as pd
import fsspec
import numpy as np
from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import chi2
```

In [6]:
```
data = pd.read_csv("C://Users//Vaibhav//Downloads//test.csv")
```

In [7]:
```
data.head()
```

Out[7]:

	id	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt	...
0	1	1043	1	1.8	1	14	0	5	0.1	193	...
1	2	841	1	0.5	1	4	1	61	0.8	191	...
2	3	1807	1	2.8	0	1	0	27	0.9	186	...
3	4	1546	0	0.5	1	18	1	25	0.5	96	...
4	5	1434	0	1.4	0	11	1	49	0.5	108	...

5 rows × 21 columns

In [8]:
```
data.tail()
```

Out[8]:

	id	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt
995	996	1700	1	1.9	0	0	1	54	0.5	170
996	997	609	0	1.8	1	0	0	13	0.9	186
997	998	1185	0	1.4	0	1	1	8	0.5	80
998	999	1533	1	0.5	1	0	0	50	0.4	171
999	1000	1270	1	0.5	0	4	1	35	0.1	140

5 rows × 21 columns

In [9]:
```
x = data.iloc[:, 1:20]
y = data.iloc[:, -1]
```

In [11]:
```
print(x)
```

```
   battery_power  blue  clock_speed  dual_sim  fc  four_g  int_memory  \
0           1043     1          1.8         1  14       0           5
1            841     1          0.5         1   4       1          61
2           1807     1          2.8         0   1       0          27
3           1546     0          0.5         1  18       1          25
```

```
4            1434     0          1.4          0  11       1          49
..            ...   ...          ...        ...  ..     ...         ...
995          1700     1          1.9          0   0       1          54
996           609     0          1.8          1   0       0          13
997          1185     0          1.4          0   1       1           8
998          1533     1          0.5          1   0       0          50
999          1270     1          0.5          0   4       1          35

     m_dep  mobile_wt  n_cores  pc  px_height  px_width   ram  sc_h  sc_w  \
0      0.1        193        3  16        226      1412  3476    12     7
1      0.8        191        5  12        746       857  3895     6     0
2      0.9        186        3   4       1270      1366  2396    17    10
3      0.5         96        8  20        295      1752  3893    10     0
4      0.5        108        6  18        749       810  1773    15     8
..     ...        ...      ...  ..        ...       ...   ...   ...   ...
995    0.5        170        7  17        644       913  2121    14     8
996    0.9        186        4   2       1152      1632  1933     8     1
997    0.5         80        1  12        477       825  1223     5     0
998    0.4        171        2  12         38       832  2509    15    11
999    0.1        140        6  19        457       608  2828     9     2

     talk_time  three_g  touch_screen
0            2        0             1
1            7        1             0
2           10        0             1
3            7        1             1
4            7        1             0
..         ...      ...           ...
995         15        1             1
996         19        0             1
997         14        1             0
998          6        0             1
999          3        1             0

[1000 rows x 19 columns]
```

In [12]:
```
print ( and )
```

```
0      0
1      0
2      1
3      0
4      1
      ..
995    0
996    1
997    0
998    0
999    1
Name: wifi, Length: 1000, dtype: int64
```

In [13]:
```
bestfeatures = SelectKBest(score_func = chi2, k=13)
```

In [14]:
```
fit = bestfeatures.fit(x,y)
```

In [15]:
```
print(fit)
```

```
SelectKBest(k=13, score_func=<function chi2 at 0x0000029244AFE1F0>)
```

In [23]:

```
dfscores = pd.DataFrame(fit.scores_)
```

In [18]:
```
dfcolumns = pd.DataFrame(x.columns)
```

In [19]:
```
print(dfscores)
```

```
             0
0     0.025709
1     0.316392
2     1.052762
3     0.480637
4    15.793117
5     0.652040
6     1.372252
7     0.240068
8    42.328627
9     0.063620
10   11.148155
11   46.347162
12  852.914979
13  562.837207
14    0.013941
15    0.809077
16    0.760553
17    0.148205
18    0.338066
```

In [20]:
```
print(dfcolumns)
```

```
                0
0   battery_power
1            blue
2     clock_speed
3        dual_sim
4              fc
5          four_g
6      int_memory
7           m_dep
8       mobile_wt
9         n_cores
10             pc
11      px_height
12       px_width
13            ram
14           sc_h
15           sc_w
16      talk_time
17        three_g
18   touch_screen
```

In [26]:
```
#concating two dataframes for better visualization
featureScores = pd.concat([dfcolumns, dfscores], axis = 1)
print(featureScores)
```

```
                0           0
0   battery_power    0.025709
1            blue    0.316392
2     clock_speed    1.052762
3        dual_sim    0.480637
4              fc   15.793117
```

```
5          four_g    0.652040
6      int_memory    1.372252
7           m_dep    0.240068
8       mobile_wt   42.328627
9         n_cores    0.063620
10             pc   11.148155
11      px_height   46.347162
12       px_width  852.914979
13            ram  562.837207
14           sc_h    0.013941
15           sc_w    0.809077
16      talk_time    0.760553
17        three_g    0.148205
18   touch_screen    0.338066
```

In [28]:

```
featureScores.columns = ["specs", "score"] #namming the dataframe colums
print(featureScores.nlargest(10, "score")) #printing 10 best features
```

```
          specs       score
12     px_width  852.914979
13          ram  562.837207
11    px_height   46.347162
8     mobile_wt   42.328627
4            fc   15.793117
10           pc   11.148155
6    int_memory    1.372252
2   clock_speed    1.052762
15         sc_w    0.809077
16    talk_time    0.760553
```

In [ ]: