

Correlation between attributes

May 20, 2021

This technique is to find whether there are correlations between the features and to ensure that each feature is sharing a unique information

```
[1]: import pandas as pd

dataset = pd.read_csv('dataset2.csv')
X= dataset.drop(columns='Result')
Y= dataset['Result']
X.head()
```

```
[1]:   Links_in_tags  SFH  Submitting_to_email  Abnormal_URL  Redirect  \
0              1   -1                -1             -1         0
1             -1   -1                 1              1         0
2             -1   -1                -1             -1         0
3              0   -1                 1              1         0
4              0   -1                 1              1         0

      onmouseover  RightClick  popUpWidnow  Iframe
0              1           1           1        1
1              1           1           1        1
2              1           1           1        1
3              1           1           1        1
4             -1           1          -1        1
```

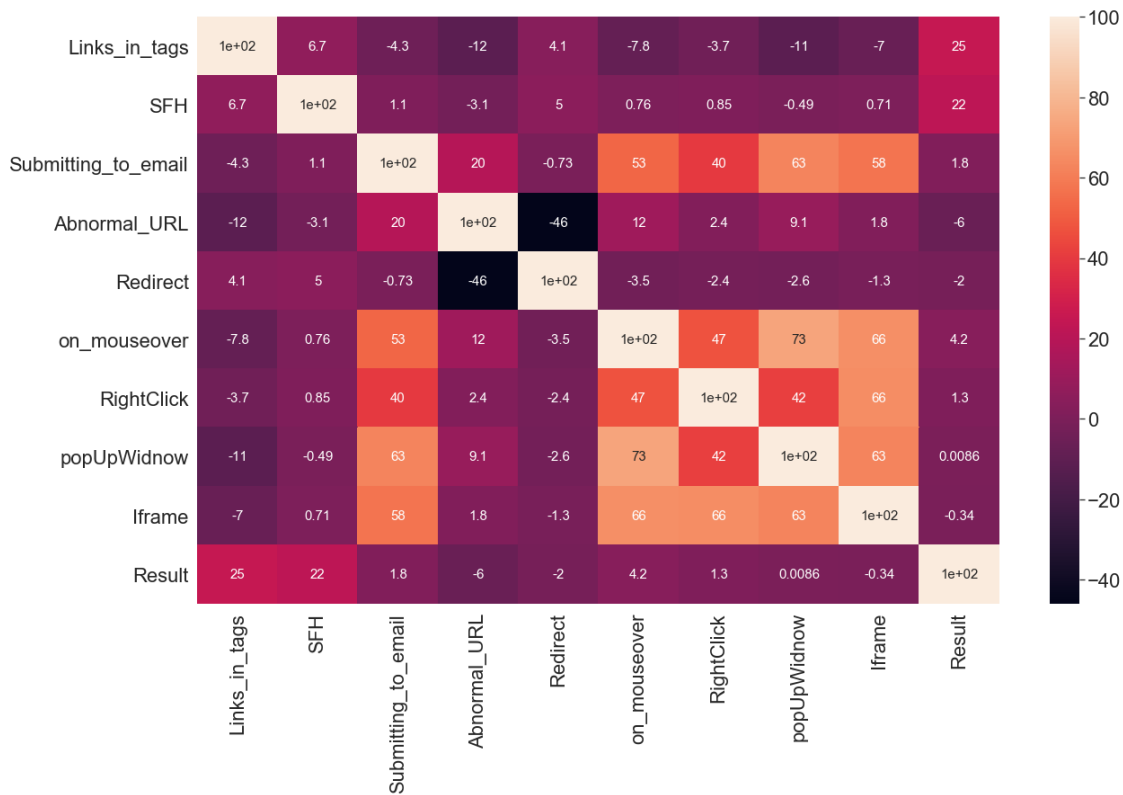
```
[2]: Y.head()
```

```
[2]: 0    -1
1    -1
2    -1
3    -1
4     1
Name: Result, dtype: int64
```

```
[14]: import matplotlib.pyplot as plt
import seaborn as sns

# correlation matrix
cor = dataset.corr()*100
```

```
# plotting heatmap
plt.figure(figsize=(20,12))
sns.heatmap(cor, annot = True, annot_kws={"size": 15})
sns.set(font_scale=2)
```



According to the above heat map, the maximum correlation is between “on mouse hover” feature and the “pop up window” feature, the value is 73. However all the correlations are not closer to 100. Therefore, it shows each attribute is independent and not sharing the same information.

[]: