We needed to find the features that are not that important in the dataset given in the file spam-data.csv for spam detection.

For this purpose, I wrote a python code that creates a correlation matrix to see which features have the biggest correlations with the classification of the email.

The code:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Load the dataset
data = pd.read_csv("spam-data.csv")

# Compute the correlation matrix
correlation_matrix = data.corr()

print(correlation_matrix)
```

The output correlation matrix is:

	Number of Words	Number of Links	Number of Capitalized Words	Number of Spam Words	Class
Number of Words	1.000000	0.376449	0.293329	0.434065	0.223667
Number of Links	0.376449	1.000000	0.014303	0.390877	0.727843
Number of Capitalized Words	0.293329	0.014303	1.000000	0.223949	0.003442
Number of Spam Words	0.434065	0.390877	0.223949	1.000000	0.594871
Class	0.223667	0.727843	0.003442	0.594871	1.000000

We can see that the Class has the biggest correlation with the Number of Links and then after with the Number of Spam Words.

The feature with the least correlation is the Number of Words. This means that the Number of Words could be the one removed.