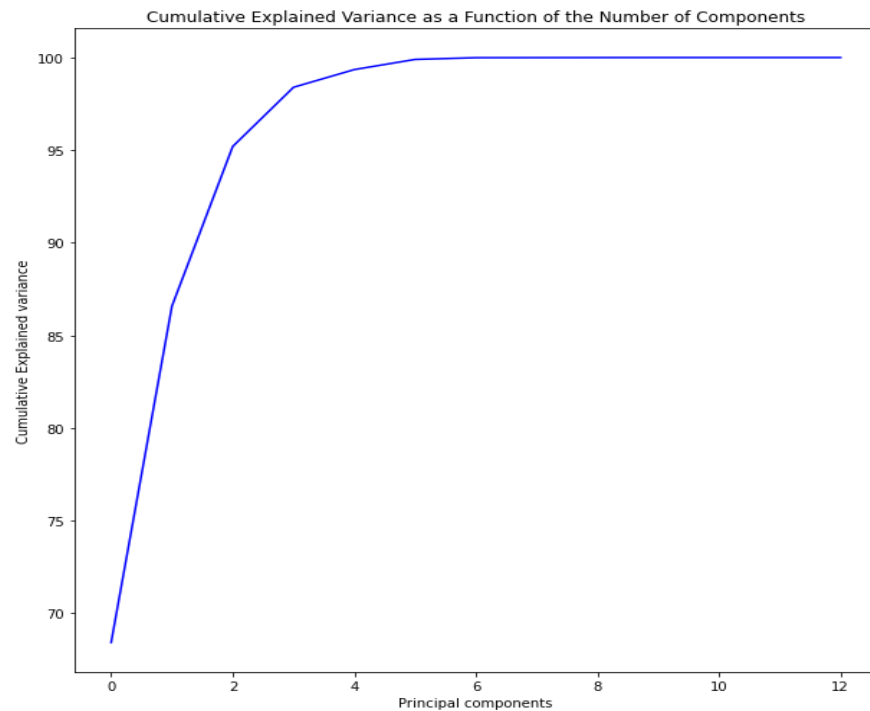


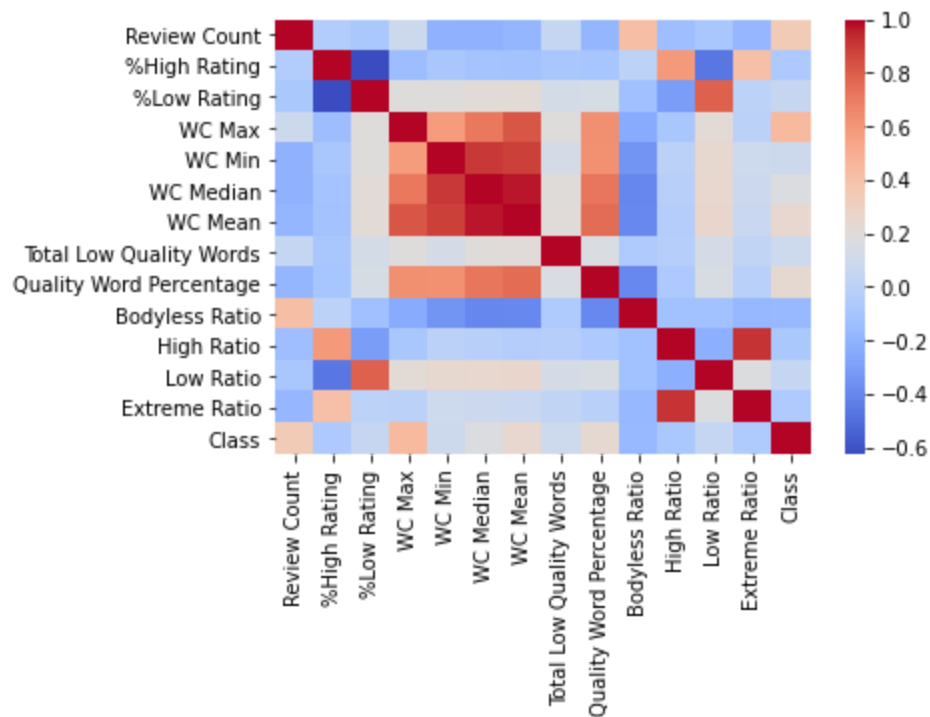
# Author Reliability: Binary Classifier

A binary classifier was implemented using sklearn's implementation of SVM classifier with various values of regularization parameter (C), and kernel types ('RBF' and 'Linear').

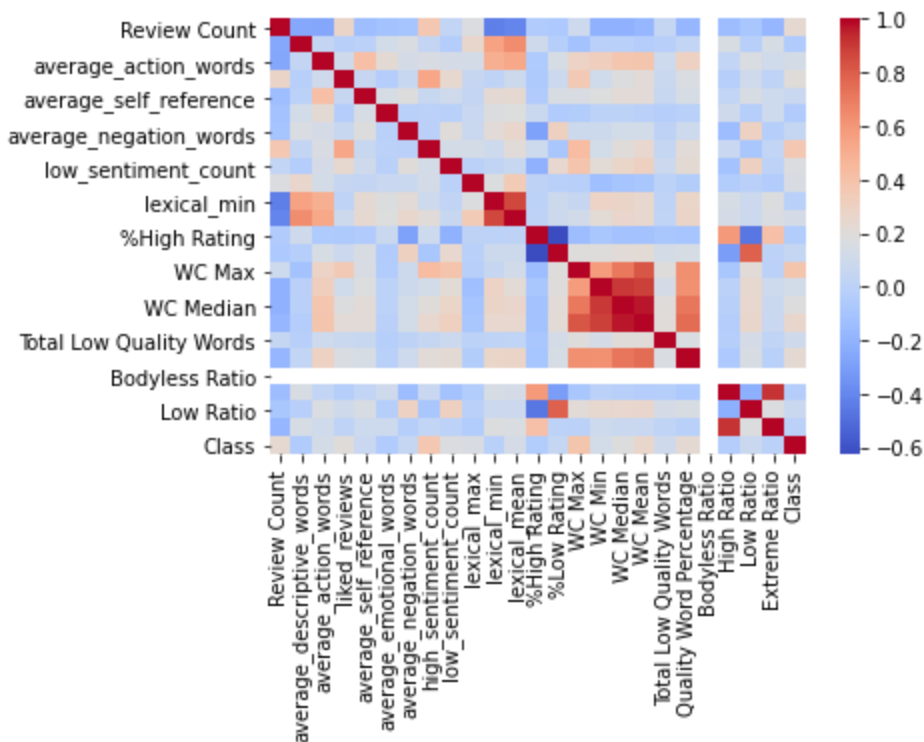


## Scree Plot

From the scree plot, we observe that the first four principal components describe more than 99% variance in the data. It means that our features are highly correlated.



**Correlation HeatMap of Original Features.**



**Correlation HeatMap of updated features.**

As we observe from the correlation heatmaps, none of the features have a strong correlation with our output ('class'). This may explain why we are not able to achieve high accuracy in the present situation.

Presently, we are at close to 50% accuracy in multi-class classification and 80% accuracy in binary classification.

### **What it means and where we can go from here?**

One option is to go ahead with binary classification and 80% accuracy (*with RBF kernel, and  $C = 1000$* ). While this is not much, in case we want something at this very moment, this is our best bet. The crux of the problem that appears right now is the data and the features that we have at this moment. As we can see from the Scree plot (below), 4 principal components are able to explain more than 99% of variance of the data which has 13 features. It means that our features in general are not explaining any variance in the data, or in other words can be thought of as highly correlated. Therefore, changing the model and/or hyper-parameters will not magically increase this number. We **have** to come up with other features and hope that they give us better results.