

# REPORT

## What is the procedure for selecting the initial set of centroids in your implementation?

In my program for the purpose of experiment, I am giving hardcoded values for selecting the documents.

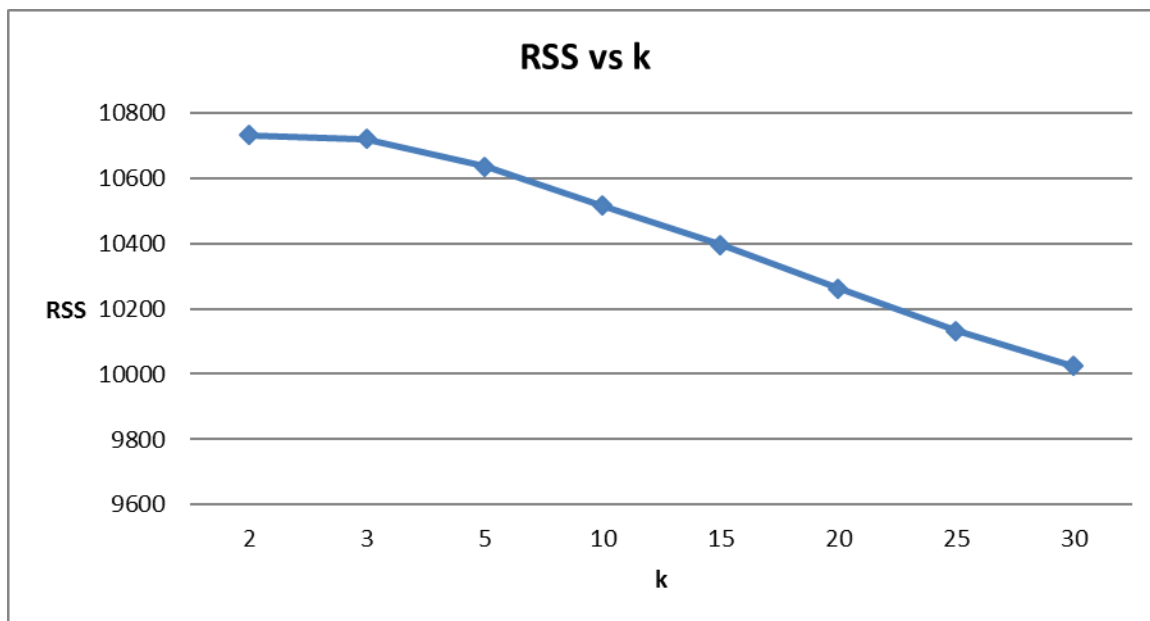
I did this because the change in the number of documents in the cluster and the RSS values were not changing after a set of iterations, the comparison wouldn't be possible.

## What is the stopping condition in your implementation?

I have set a parameter "iter", where in its value can be changed.

My program will run for those many iterations.

## From the plot, what is the value of 'k' that provides a good tradeoff with change in RSS?



From the plot, we can see that the RSS value decreases with the increase in 'k' value.

There is a steady decrease in the Residual Sum of Squares as value of k changes from k=2 to k= 5.

We can say from the experiment that, the value of k should minimum be equal to or greater than 5.

So, it can be said that k=5 provides a good tradeoff as value of RSS changes significantly.