CHAPTER 6
CONCLUSION

The conclusion that being obtained from the research is stated as below:

1. K-Means able to define which is Hate Speech and which is Not Hate Speech based on the dataset. This is shown when the predicted data is compared to the actual indicated data. This can be shown too by using the precision recall. The correct prediction will be counted as TP (True Positive) and TN (True Negative) in the precision recall.

2. K-Means is effective to define Hate Speech with certain amount of dataset with highest accuracy of 80%. However the accuracy value will be differ based on the amount of data and the centroid point value. The accuracy of 80% is granted by searching the most effective centroid value.

3. By using N-Gram, the results are much more stable in terms of accuracy especially on Tri-gram. This occur because of the weight distribution is different compared to the normal TF-IDF tokenize and Tri-gram.

For further research that can be conducted based on this research is that the research can use different algorithm for classification for better performance, or bigger dataset.