

Abstract

Classification is the process that grouping the data into the class based on similarity level. Classification can be also applied in text document, to make easier act of determining whole document with certain category. There is a various way to do the classification, one of them is with the K-Nearest Neighbor method. The K-Nearest Neighbor is a popular method in classification because of easy in implementation.

But, behind in the easiness, the K-Nearest Neighbor method has a weakness if it be used in a document that has uneven distribution, because when the k value more and more bigger will appear domination by the bigger class to the smaller class. Therefore the Improved K-Nearest Neighbor method has been used for to cope with the weakness.

Precision, recall, and F1-Measure are used for evaluating the performance from the K-Nearest Neighbor method and Improved K-Nearest Neighbor method. The result shows that the Improved K-Nearest Neighbor method can eliminate the domination effect from the largest category in various kind of document training distribution.

Keywords: classification, K-Nearest Neighbor, Improved K-Nearest Neighbor