ABSTRACT

On the problem of high dimension and the number of data, there are methods to facilitate the search for information or categorization is Document Classification / document classification. Classification of documents itself has many methods, one of which will be discussed in this Final Project is the classification of documents with kNN (K-Nearest Neighbor) method. KNN is one of the popular methods used in document classification because the method is easy to implement. KNN has a disadvantage that is, at the time of classification, uses the k value to take as much as the nearest k, and it takes time to find the k value for optimal results. There is a case in the dataset with an unbalanced data distribution, a class of documents classified with small k values, ie classes that have lower class distributions in the training set, if using large k values, then the neighbors taken will be more neighbors Is a class with a larger class distribution, and the resulting classification may be less accurate. This dynamic k method is used to determine appropriate k-neighbor taking on the kNN classification based on the class distribution so as to make the classification better.

Keywords: Document Classification, kNN, Dynamic kNN, Improved kNN, imbalance class