ABSTRACT

Roni Novettio Chairullah. Classification of Documents in Bahasa Indonesia using DCS-LA with Inverse Distance Weighting. Supervised by AHMAD RIDHA.

Dynamic Classifier Selection with Local Accuracy (DCS-LA) is a document classification method that combines several classification methods and k-NN. In this study, we implemented the DCS-LA with Inverse Distance Weighting for documents written in Bahasa Indonesia as well as comparing between the DCS-LA with Inverse Distance Weighting and DCS-LA without Inverse Distance Weighting. We used four classifiers: Rocchio, Naïve Bayes, Bernoulli, and Poisson Naïve Bayes as classifiers in the DCS-LA. For the data, we used agriculture documents consisting of 174 training documents and 75 test documents, and news documents consisting of 500 training documents and 250 test documents. This method can yield an accuracy of 66% and 96% for agriculture documents and news documents, respectively. Without Inverse Distance Weighting, DCS-LA only yields an accuracy of 56% and 86% for agriculture documents and news documents, respectively. Therefore, Inverse Distance Weighting can improve the accuracy of the DCS-LA in classifying text documents in Bahasa Indonesia.

Keywords: Document classification, DCS-LA, Rocchio, Naïve Bayes, Bernoulli, Poisson Naïve Bayes, Inverse Distance Weighting.