ABSTRACT

Text classification is a supervised learning technique that uses labelled training data to derive a classification system (classifier) and then automatically classifies unlabelled text data using the derived classifier. This paper investigates Naïve Bayesian method (NB) and K-Nearest Neighbour algorithm (KNN) on different Arabic data sets. The bases of our comparison are the most popular text evaluation measures. The Experimental results against different Arabic text categorisation data sets reveal that NB algorithm outperforms the KNN based on Cosine Coefficient approach with regards to all measures.