Amharic Text News Classification Using Supervised Learning

Authors : Misrak Assefa

Abstract : The Amharic language is the second most widely spoken Semitic language in the world. There are several new overloaded on the web. Searching some useful documents from the web on a specific topic, which is written in the Amharic language, is a challenging task. Hence, document categorization is required for managing and filtering important information. In the classification of Amharic text news, there is still a gap in the domain of information that needs to be launch. This study attempts to design an automatic Amharic news classification using a supervised learning mechanism on four un-touch classes. To achieve this research, 4,182 news articles were used. Naive Bayes (NB) and Decision tree (j48) algorithms were used to classify the given Amharic dataset. In this paper, k-fold cross-validation is used to estimate the accuracy of the classifier. As a result, it shows those algorithms can be applicable in Amharic news categorization. The best average accuracy result is achieved by j48 decision tree and naïve Bayes is 95.2345 %, and 94.6245 % respectively using three categories. This research indicated that a typical decision tree algorithm is more applicable to Amharic news categorization.

Keywords : text categorization, supervised machine learning, naive Bayes, decision tree

Conference Title : ICAIDM 2020 : International Conference on Artificial Intelligence and Data Mining

Conference Location : Rome, Italy

Conference Dates : April 08-09, 2020

Open Science Index, Computer and Information Engineering Vol:14, No:04, 2020 publications.waset.org/abstracts/124249.pdf

International Scholarly and Scientific Research & Innovation 14(04) 2020   1   ISNI:000000091950263