نویسندگان:

Ali Allahverdipour - *Department of Computer Engineering, Urmia Branch, Islamic Azad University, Urmia, Iran*

Farhad Soleimanian Gharehchopogh - *Department of Computer Engineering, Urmia Branch, Islamic Azad University, Urmia, Iran*

خلاصه مقاله:

The Internet provides easy access to a kind of library resources. However, classification of documents from a large amount of data is still an issue and demands time and energy to find certain documents. Classification of similar documents in specific classes of data can reduce the time for searching the required data, particularly text documents. This is further facilitated by using Artificial Intelligence (AI) and optimization algorithms which are highly potential in Feature Selection (FS) and words extraction. In this paper Crow Search Algorithm (CSA) is used for FS and K-Nearest Neighbor (KNN) for classification. Additionally, TF technique is proposed for counting words and calculating the words' frequency. Analysis is performed on Reuters-21578, Webkb and Cade 12 datasets. The results indicate that the proposed model is more accurate in classification than KNN model and, show greater F-Measure compared to KNN and C4.5. Moreover, by using FS, the proposed model promotes classification accuracy by %27, compared to KNN.

کلمات کلیدی:

Text Documents Classification, Crow Search Algorithm, K-Nearest Neighbor

لینک ثابت مقاله در پایگاه سیویلیکا:

https://civilica.com/doc/966275