

Conferences > 2020 5th International Confer... ⓘ

Text Document Clustering Using K-means Algorithm with Dimension Reduction Techniques

Publisher: IEEE

Cite This

PDF

Rutuja Kumbhar ; Snehal Mhamane ; Harshada Patil ; Sukruta Patil ; Shubhangi Kale All Authors

1Paper Citation

138Full Text Views



Alerts

- Manage Content
- Alerts
- Add to Citation
- Alerts

More Like This

A Regularized Singular Value Decomposition-Based Approach for Failure Pattern Classification on Fail Bit Map in a DRAM Wafer
IEEE Transactions on Semiconductor Manufacturing
Published: 2015

Generalizing discriminant analysis using the generalized singular value decomposition
IEEE Transactions on Pattern Analysis and Machine Intelligence
Published: 2004

Show More

Abstract



Document Sections

- I. Introduction
- II. Literature Survey
- III. Proposed System
- IV. Proposed Methodology
- V. Experimental Results

Show Full Outline

- Authors
- Figures
- References
- Citations

Metadata

Abstract:

With growing technologies commercial sites, social media, organizations generate lots of data. However, this huge amount of information needs to be organized properly. For this, the text mining process used as extracting relevant features and knowledge of the given unstructured text documents. The document clustering method in text mining is used for retrieving interesting features. In which similar documents organize into different groups called clusters. The large dimensions of data become a barrier in the extraction of useful information, so using dimensionality reduction (DR) technique for reducing the dimensions of the data matrices. Further data divide into groups using the k-means clustering algorithm. This study implements TF-IDF, singular value decomposition (SVD), non-negative matrix factorization (NMF), and k-means clustering. Finally, the results of the comparison of scores of kmeans, SVD with kmeans, and NMF with k-means are shown by graphical representation. The system uses 20 newsgroup datasets for simulating results.