

**COMPARATIVE ANALYSIS OF DIMENSION REDUCTION
TECHNIQUES FOR K-MEANS CLUSTERING IN HIGH-
DIMENSIONAL DATA**

K.M.W.G.T. MEERIGAMA

Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka

A computation including high dimensional data is cumbersome due to high demand for computational power. Additionally, high dimensional data is hard to visualize. Converting the high dimensional data into lower dimensions greatly reduces the computational burden and will allow the data to be visualized easily. However, the efficiency of the processed data after reducing the dimensionality which is measured by Sum of Squared Error(SSE) method in k-means clustering algorithm should still be explored.

This research study focuses on the use of dimensionality reduction methods as a preprocessing step to improve the accuracy and the efficiency of clustering high-dimensional data. This is demonstrated by considering principal component analysis (PCA), independent component analysis (ICA) and combination of PCA and ICA (in this method, first apply PCA then ICA) as dimensionality reduction methods for k-means clustering algorithm. The observed results from PCA and ICA methods yield lower error rates compared to the use of original dimensions when clustering text data with k-means clustering algorithm.

