

Abstrak

Data mining merupakan suatu teknik yang dapat memecahkan permasalahan gudang data, sehingga tumpukan data yang tidak memiliki informasi apapun dapat digali dan dianalisa. Tumpukan data tidak dapat langsung diproses, data yang berdimensi tinggi harus *dipreprocessing* sehingga saat mining data benar-benar memberikan hasil yang maksimal.

ICA (*Independent Component Analysis*) merupakan suatu teknik reduksi pada data numerik yang dapat mengurangi dimensi, dimensi yang baru disebut *Independent Component (IC)*. ICA mereduksi dengan mencari dimensi yang saling *independent* untuk menjadi dimensi baru. Kondisi ini lebih kuat jika dibandingkan mencari ketidakterhubungan antar dimensi .

Fenomena *Curse of Dimensionality* dapat diselesaikan dengan pengurangan dimensi dengan ICA. Hasil pengujian dengan ICA dapat mereduksi data *colon tumor* dengan 2000 dimensi dapat diringkas menjadi 60 IC dan data set DLBCL dengan 4026 dimensi dapat diringkas menjadi 46 IC. Kedua dataset memiliki hasil evaluasi *clustering* terbaik pada 1 IC yaitu dengan nilai SSE sebesar 0.00 dan nilai SSE 0.97. ICA memberikan hasil evaluasi *clustering* yang lebih baik jika dibandingkan dengan PCA (*Principal Component Analysis*) yaitu dengan tingkat SSE (*Sum of Squared Error*) yang lebih kecil.

Kata kunci: *Data mining*, ICA, PCA, data numerik, dimensi tinggi, *independent*, *clustering*, SSE (*Sum of Squared Error*)