

Abstrak

Clustering adalah salah satu teknik dalam *data mining* yang melakukan distribusi terhadap data ke dalam kelompok yang anggota-anggotanya memiliki kemiripan. Salah satu algoritma *clustering* yang populer adalah k-means.

Pada tugas akhir ini, *clustering* yang dilakukan mengalami *preprocessing* terlebih dahulu menggunakan metode *Kernel Principal Component Analysis* (KPCA). KPCA menghasilkan sejumlah *feature* atau disebut juga sebagai *principal component* dalam jumlah yang lebih banyak daripada atribut data input. Setelah terbentuk sekumpulan *principal component* yang baru, maka dilakukan pemilihan *principal component* berdasarkan besarnya variansi. Semakin besar variansi yang dimiliki oleh suatu *principal component*, maka *principal component* makin mencirikan data. Setelah tahap pemilihan *feature* selesai, maka langkah yang selanjutnya adalah *clustering* menggunakan *modified* k-means. Pada *modified* k-means ini, akan dilakukan proses penentuan *centroid*, berbeda dengan algoritma k-means asli yang penentuan *centroid*nya dilakukan secara acak. Dari hasil pengujian dan analisis, didapatkan bahwa penentuan banyaknya *feature* yang digunakan dalam proses *clustering* mempengaruhi kualitas *cluster*. Selain itu pula, pemilihan terhadap jenis kernel dan nilai parameter kernel yang digunakan juga berpengaruh terhadap hasil *cluster*.

Kata Kunci : *data mining, clustering, KPCA, modified k-means*