

ABSTRAKSI

Data Mining adalah proses pencarian pola dan kecenderungan yang menarik dari basisdata berukuran besar. Klasterisasi merupakan salah satu fungsionalitas data mining yang dimanfaatkan untuk mengelompokkan objek-objek ke dalam klaster-klaster dimana objek-objek di dalam satu klaster yang sama mempunyai kesamaan yang tinggi dan mempunyai perbedaan yang tinggi pada objek-objek antar klaster yang berbeda.

Masalah klasterisasi telah diketahui dalam literatur basisdata untuk kegunaannya dalam banyak aplikasi seperti segmentasi pasar, klasifikasi, dan analisis kecenderungan. Pada data berdimensi tinggi tidak semua dimensi relevan dengan klasternya. Salah satu cara untuk menanganinya adalah dengan memilih dimensi-dimensi yang secara dekat berkorelasi dan menemukan klaster-klaster di dalam *subspace* atau subruang yang bersesuaian. Algoritma seleksi *feature* tradisional mencoba untuk mencapainya.

Namun, pendekatan dengan seleksi *feature* atau reduksi *feature* dapat menghilangkan variabel-variabel dan tidak efektif mengidentifikasi klaster-klaster pada subruang-subruang yang berbeda. Sekumpulan titik data dapat dikelompokkan lebih baik pada subset dimensi yang berbeda. Jumlah dimensi di setiap subruang juga bisa bervariasi. Maka diperlukan suatu konsep yang disebut *subspace* atau *projected clustering*, dimana subset-subset dimensi yang dipilih tersebut spesifik terhadap klasternya sendiri. Tugas akhir ini membahas dan menganalisa bagaimana algoritma PROCLUS melakukan klasterisasi pada kasus *subspace clustering* untuk data berdimensi tinggi.

Kata Kunci : data mining, algoritma PROCLUS, *subspace clustering*, *projected clustering*, data berdimensi tinggi.