

Abstrak

Data mining adalah suatu proses analisis untuk mendapatkan informasi yang penting dan terpendam dari sekumpulan data yang belum diketahui sebelumnya. *Clustering* merupakan salah satu fungsionalitas data mining yang digunakan untuk melakukan pengelompokan data ke dalam suatu kelas atau *cluster*. Prinsip dasar *clustering* adalah mengelompokkan objek pada suatu kelas yang memiliki kemiripan sangat besar dengan objek lain pada kelas yang sama (*similarity*), tetapi sangat tidak mirip dengan objek pada kelas lain (*dissimilarity*). Kebanyakan algoritma *clustering* hanya bisa menangani data yang memiliki dimensi rendah sehingga hasil *clustering* yang didapat kurang berkualitas. *Subspace clustering* adalah metode yang dapat digunakan untuk menentukan *cluster* yang terbentuk pada *subspace* yang berbeda sehingga lebih baik dalam menangani data multidimensi dibandingkan dengan cara *clustering* pada umumnya. Algoritma yang menerapkan *subspace clustering* antara lain: COSA, FINDIT, ENCLUS, CLIQUE, MAFIA, dan lain-lain.

Pada Tugas Akhir ini diimplementasikan algoritma MAFIA (*Merging of Adaptive Finite Intervals*). MAFIA adalah algoritma *clustering* yang menerapkan metode *subspace clustering* dengan mengintegrasikan metode *density-based* dan *adaptive-grid based* sehingga algoritma ini cukup baik untuk menangani data multidimensional. MAFIA termasuk *density-based methods* karena melakukan *clustering* berdasarkan densitas dalam suatu data yang ada dan menggunakan histogram untuk menganalisis jarak antar data dalam setiap dimensi. Algoritma ini menggunakan pendekatan *bottom up search* untuk menemukan *dense unit* berdasarkan dimensi untuk mengurangi ruang pencarian. MAFIA tidak membutuhkan inputan berupa jumlah *cluster* yang akan dicari dalam proses *clustering*. Sehingga algoritma ini lebih fleksibel dalam pencarian *cluster* dari dataset.

Algoritma MAFIA diuji menggunakan data buatan yang di-*generate* secara online di www.datasetgenerator.com. Hasil uji menunjukkan bahwa algoritma MAFIA baik digunakan untuk *clustering* data multidimensi. Kualitas ini tercermin dari maksimum akurasi yang ditemukan yaitu sebesar 0.89.

Kata kunci : *clustering, MAFIA algorithm, data mining, subspace clustering.*