

Abstrak

Data umumnya memiliki jumlah fitur yang sangat besar, fitur pada data dapat juga disebut sebagai atribut atau dimensi. Penggunaan fitur data yang besar dapat mengakibatkan pemrosesan *learning* yang kurang efektif, bahkan dapat menghasilkan kualitas klaster yang kurang baik. Oleh sebab itu, dilakukan proses seleksi fitur untuk mengurangi jumlah fitur dari data. Pada Tugas Akhir ini dilakukan seleksi fitur pada data yang akan dilakukan proses klasterisasi sebagai *learningnya*. Proses seleksi fitur dapat mempertahankan atau bahkan meningkatkan kualitas klaster yang dihasilkan. Metode seleksi fitur dibagi menjadi dua, yaitu: metode *wrapper* dan metode *filter*. Metode *wrapper* melakukan seleksi fitur dengan menggunakan kinerja dari algoritma *learning*, sedangkan metode *filter* melakukan proses seleksi fitur terlebih dahulu kemudian melakukan algoritma *learningnya*. Proses seleksi fitur ini dilakukan dengan menggunakan metode *wrapper*. Metode *wrapper* dibagi menjadi tiga proses, yaitu proses pencarian, proses klasterisasi dan proses evaluasi fitur. Pada Tugas akhir ini menggunakan strategi pencarian *sequential forward search* yang bersifat *greedy*. Proses klasterisasi menggunakan algoritma EM dan klaster yang dihasilkan dievaluasi dengan menggunakan nilai evaluasi klaster *log likelihood*.

Kata kunci: *feature selection*, klasterisasi, *log likelihood*, *sequential forward search*, *wrapper*.