

METODE PENDETEKSIAN *OUTLIER* DI DALAM DATA MULTIVARIAT

Oleh
Benni Setyawan
NIM.023114009

ABSTRAK

Outlier atau pencilan adalah titik sampel yang memiliki karakteristik unik dan dapat diidentifikasi secara jelas karena berbeda dengan mayoritas titik sampel lainnya. Pada umumnya, *outlier* dapat terjadi karena kesalahan manusia, kesalahan instrumen, perilaku curang, perubahan perilaku sistem atau kesalahan sistem, dan penyimpangan alami di dalam populasi. Kehadiran *outlier* sering kali berdampak buruk terhadap analisis data, karena *outlier* mampu menyimpangkan tes-tes statistik yang didasarkan pada dua penaksir klasik yaitu rerata sampel dan kovariansi sampel. Dalam satu atau dua dimensi, *outlier* dapat diidentifikasi secara mudah melalui grafik, tetapi dalam analisis multivariat dengan tiga dimensi atau lebih *outlier* menjadi sulit diidentifikasi.

Penulisan ini bertujuan untuk memperoleh kejelasan tentang cara-cara pendeteksian *outlier* di dalam data multivariat, menjelaskan cara-cara analisis data yang memuat *outlier* dan dapat menerapkan metode pendeteksian *outlier* di dalam data multivariat. Konsep dasar dalam pendeteksian *outlier* multivariat adalah dengan mengukur jarak setiap titik ke pusat datanya. Titik sampel yang memencil akan memiliki nilai jarak yang besar relatif terhadap mayoritas titik yang lain. Titik yang memencil ini patut dicurigai sebagai *outlier*.

Pendeteksian *outlier* di dalam data multivariat dapat dilakukan dengan metode jarak Mahalanobis, metode *Minimum Volume Ellipsoid* (MVE), dan metode *Fast Minimum Covariance Determinant* (*Fast MCD*). Analisis data yang memuat *outlier* dapat dilakukan dengan cara melakukan koreksi terhadap *outlier*, membuang *outlier*, atau menyertakan *outlier* di dalam analisis. Bila *outlier* disertakan dalam analisis sebaiknya peneliti memakai metode statistik tegar. Dalam penulisan ini, diambil dua contoh kasus pendeteksian *outlier* di dalam data *Hematology* dan *Track Record*. Dari hasil analisis diketahui bahwa di dalam data *Hematology* dan *Track Record* ditemui adanya *outlier*. Efek *masking* yang menunjukkan kelemahan metode jarak Mahalanobis dalam analisis data *Hematology* dan *Track Record* dapat diatasi dengan baik dengan metode MVE dan *Fast MCD*.