

BAB I

PENDAHULUAN

1.1 Latar Belakang

Permasalahan yang sering muncul saat mengelola data salah satunya adalah *outlier*. *Outlier* dapat disebabkan kesalahan pengukuran, kesalahan entri data, atau data alami itu sendiri. Keberadaan *outlier* mengakibatkan hasil analisis menjadi menyimpang sehingga ketidak akuratan hasil analisis dalam penelitian. Adanya *outlier* juga mengakibatkan varians pada data tersebut menjadi besar.[6] Oleh karenanya mengidentifikasi adanya *outlier* pada data menjadi penting dan solusi jika adanya *outlier* pada data.[1]

Outlier atau pencilan, adalah data menyimpang dari perilaku global mayoritas data.[2] *Outlier* dapat mempengaruhi hasil analisis dan interpretasi data jika tidak dikelola dengan baik. Keberadaan *outlier* sering kali berada pada data berdimensi tinggi daripada dalam data berdimensi rendah. Ini karena semakin banyak dimensi yang ada, semakin sulit untuk menemukan atau mengidentifikasi *outlier*.

Dalam data berdimensi tinggi penyebaran data menjadi lebih besar, sehingga lebih mungkin ada titik-titik data yang berada jauh dari sebagian besar data lainnya. Akibatnya, kemungkinan adanya *outlier* meningkat karena ada lebih banyak celah di antara pengamatan. Sehingga data dalam penelitian ini menggunakan data berdimensi tinggi. Yang dimaksud dengan dimensi data yaitu atribut dari data tersebut.

Pada penelitian ini atribut yang dipakai meliputi wilayah, penduduk, pendidikan, keuangan, kesejahteraan rakyat di Kabupaten dan Kota di Provinsi Jawa Barat. Sebagai contoh dalam hal pendidikan, di mana ada Kabupaten atau Kota yang menonjol dengan tingkat pendidikan yang jauh lebih tinggi atau rendah dari Kabupaten atau Kota lain di Provinsi Jawa Barat. Kabupaten atau Kota yang memiliki pendidikan jauh lebih tinggi atau rendah itulah yang disebut sebagai *outlier*.

Jika dikaitkan dengan Al-Qur'an hal ini erat kaitannya dengan QS. Al-Maidah : 8 sebagai berikut.

يَا أَيُّهَا الَّذِينَ آمَنُوا كُونُوا قَوَّامِينَ لِلَّهِ شُهَدَاءَ بِالْقِسْطِ وَلَا يَجْرِمَنَّكُمْ شَنَا نُ قَوْمٍ عَلَىٰ آلَا تَعْدِلُونَ ۗ اٰعِدِلُوْا هُوَ اٰقْرَبُ لِلتَّقْوٰى وَاتَّقُوا اللّٰهَ اِنَّ اللّٰهَ خَبِيْرٌۢ بِمَا تَعْمَلُوْنَ

Artinya : Wahai orang beriman, jadilah kamu penegak (kebenaran) karena Allah (dan) saksi-saksi (yang bertindak) adil. Janganlah kebencianmu terhadap suatu kaum mendorong kamu untuk berlaku tidak adil. Berlakulah adil karena (adil) itu lebih dekat pada takwa. Bertakwalah kepada Allah. Sesungguhnya Allah Mahateliti terhadap apa yang kamu kerjakan.

Kedepannya, Kabupaten dan Kota yang termasuk sebagai *outlier*, baik dalam bidang pendidikan, kesejahteraan, maupun bidang lainnya, perlu mendapat perhatian dan perbaikan yang sesuai. Hal ini bertujuan untuk memastikan bahwa mereka dapat sejajar dan berkembang sebagaimana Kabupaten dan Kota lainnya di Provinsi Jawa Barat, sehingga tidak ada lagi data yang menyimpang dari pola mayoritas data. Dengan demikian, tercapailah keadilan dalam berbagai aspek, termasuk kesejahteraan, pendidikan, dan bidang lainnya di Kabupaten dan Kota Provinsi Jawa Barat.

Beberapa metode digunakan untuk mengidentifikasi pencilan dalam data. Metode yang umum dipakai salah satunya adalah jarak mahalanobis. Metode jarak mahalanobis merupakan metode yang menggunakan vektor rata-rata dan matriks kovarians.[3] Salah satu variasi dari metode jarak mahalanobis yaitu robust mahalanobis distance. Pendekatan yang umum digunakan pada robust mahalanobis distance adalah dengan menggunakan kovarian yang lebih robust, seperti *Minimum Covariance Determinant* (MCD). Metode *Minimum Covariance Determinant* (MCD) didasarkan pada nilai determinan matriks kovarians yang nantinya digunakan untuk mendeteksi pencilan. Pada prinsipnya metode *Minimum Covariance Determinant* yaitu mengambil subhimpunan dari semua pengamatan yang determinan matriks kovariansnya terkecil diantara semua pengamatan yang ada.

Tetapi pada tahun 2017, Maronna dan Yohai [14][15] menunjukkan bahwa bias dari estimator MCD cepat meningkat seiring dengan bertambahnya dimensi data, dan kemudian mengusulkan estimator Rocke. Estimator Rocke adalah metode statistik yang dikembangkan oleh David Rocke yang digunakan dalam analisis

multivariat untuk memperkirakan matriks kovarians dari sekumpulan variabel. Metode ini adalah modifikasi dari estimator *Scale* yang tahan terhadap *outlier* dan pengamatan yang berpengaruh.

Beberapa penelitian lain pernah dilakukan untuk mendeteksi *outlier* menggunakan estimator *Minimum Covariance Determinan* (MCD) diantaranya membandingkan Metode Kuadrat Terkecil (MKT) dengan Regresi *Robust Minimum Covariance Determinan* (MCD) oleh Nur Kamilah Sa'diyah (2019). Sehingga pengembangan penelitian kali ini menggunakan estimator Roche sebagai perbandingannya.

1.2 Rumusan Masalah

Rumusan masalah pada skripsi ini adalah:

1. Manakah estimator terbaik antara estimator *MINIMUM COVARIANCE DETERMINANT* (MCD) dan estimator Roche dalam mendeteksi data yang mengandung *outlier*?
2. Bagaimana penerapan mendeteksi *outlier* pada Kabupaten dan Kota yang berada di Provinsi Jawa Barat?

1.3 Batasan Masalah

Batasan masalah dalam penelitian skripsi ini diantaranya:

1. Pendekatan yang dipakai dalam *robust mahalanobis distance* dengan menggunakan estimator MCD dan estimator Roche.
2. Data yang dipakai dalam penelitian adalah data sekunder yang diperoleh dari Badan Pusat Statistika (BPS) yang berjudul Provinsi Jawa Barat dalam angka.

1.4 Tujuan Penelitian

Tujuan dalam penelitian skripsi ini diantaranya:

1. Dapat mengetahui estimator mana yang lebih baik dalam mendeteksi data yang mengandung *outlier*.
2. Mengetahui penerapan dari mendeteksi *outlier* pada Kabupaten dan Kota yang berada di Provinsi Jawa Barat.

1.5 Metode Penelitian

Metode yang dilakukan dalam menyelesaikan skripsi ini bersifat studi literatur. Dengan cara mengumpulkan berbagai data berupa teori yang berhubungan dengan mendeteksi *outlier* menggunakan estimator *MINIMUM COVARIANCE DETERMINANT* (MCD) dan estimator Roche yang diperoleh dari berbagai sumber berupa buku, jurnal, dan skripsi. Selanjutnya adalah pengambilan data sekunder yang diperoleh dari Kabupaten dan Kota yang berada di Provinsi Jawa Barat. Kemudian melakukan analisis, dengan membandingkan hasil deteksi *outlier* menggunakan estimator *MINIMUM COVARIANCE DETERMINANT* (MCD) dan estimator Roche pada data tersebut.

1.6 Sistematika Penulisan

Berdasarkan sistematika penulisannya, tugas akhir ini terdiri atas lima bab serta daftar pustaka, dimana dalam setiap bab terdapat beberapa subbab.

BAB I : PENDAHULUAN

Bab ini membahas latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, metodologi penelitian, dan sistematika penelitian.

BAB II : LANDASAN TEORI

Bab ini berisi materi tentang hal-hal yang menjadi landasan masalah dan teori-teori yang digunakan sebagai pedoman untuk menyelesaikan penelitian. Meliputi matriks, langkah pengolahan data, *outlier*, karakteristik estimator, jarak mahalanobis, *robust mahalanobis distance*, dan estimator scale.

BAB III : METODE PENELITIAN

Pada bab ini akan dijelaskan mengenai inti penelitian yang dilakukan, berupa pembahasan rinci tentang metode yang digunakan, baik secara teoritis maupun analisisnya, dan langkah-langkah dalam penerapannya.

BAB IV : STUDI KASUS DAN ANALISA

Pada bab ini akan dilakukan studi kasus dan analisis hasil menggunakan estimator MCD dan estimator rocke.

BAB V : KESIMPULAN DAN SARAN

Pada bab ini memuat kesimpulan dari penelitian yang dilakukan serta saran untuk penelitian berikutnya.

