

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Pendekatan regresi kerap dimanfaatkan dalam riset statistik guna menguraikan keterkaitan antara variabel bebas dan variabel terikat. Metode Kuadrat Terkecil (MKT) merupakan teknik yang paling sering digunakan. MKT bekerja dengan menaksir koefisien β sedemikian rupa sehingga menjadikan jumlah kuadrat residual terkecil, akan tetapi MKT memiliki kelemahan besar, sangat sensitif terhadap *outlier*. *Outlier* dapat merusak estimasi model, menghasilkan koefisien yang bias, prediksi yang melenceng, dan akhirnya kebijakan yang salah arah [1].

Outlier merujuk pada data pengamatan yang menyimpang secara signifikan dari pola umum data dan biasanya berada jauh dari nilai pusat distribusi. Keberadaan *outlier* bukan hanya sekadar menyimpang, tetapi juga dapat mengacaukan parameter regresi secara signifikan, sehingga mempengaruhi interpretasi dan validitas model. Sehingga, dibutuhkan metode dalam analisis regresi yang mampu mengatasi masalah *outlier*, yaitu pendekatan *robust*, baik dalam deteksi maupun estimasi. Pendekatan *robust* secara umum bertujuan untuk mengurangi pengaruh pengamatan menyimpang (*pencilan*) [2].

Dalam regresi *robust*, beberapa estimator populer antara lain *Least Trimmed Squares* (LTS), estimasi M, estimasi MM, *Least Median Of Squares* (LMS), dan estimasi S. Untuk deteksi *outlier* multivariat sebelum proses regresi, salah satu metode *robust* yang sering digunakan adalah *Minimum Covariance Determinant* (MCD), yang diperkenalkan oleh Rousseeuw pada 1984 [3].

Dari berbagai metode tersebut, LTS dan MCD menunjukkan performa yang sangat baik dalam konteks regresi multivariat yang terkontaminasi oleh *outlier*. Estimator LTS digunakan dalam regresi *robust* bekerja dengan meminimalkan jumlah residual kuadrat terkecil dari sebagian pengamatan yang dianggap tidak

mengandung *outlier*. Sementara itu, MCD digunakan dalam analisis data multivariat yang bertujuan untuk mengestimasi pusat dan kovarians dari data multivariat secara *robust*, dengan cara memilih sub-sampel data yang meminimalkan determinan matriks kovarians. Hasil penelitian [4] menunjukkan bahwa regresi *robust* dengan estimasi LTS secara efektif mampu menangani pencilan baik pada variabel independen maupun dependen. Di sisi lain, [5] menyatakan bahwa pendekatan MCD memiliki akurasi lebih tinggi dibandingkan dengan analisis faktor konvensional dalam konteks deteksi *outlier* multivariat.

Penerapan regresi *robust* tidak terbatas pada satu bidang saja, melainkan telah meluas ke berbagai ranah, termasuk bidang pendidikan. Sektor pendidikan memiliki kompleksitas data yang tinggi, terutama ketika dikaji pada level regional. Provinsi Jawa Barat yang memiliki populasi tertinggi di Indonesia, dihadapkan pada berbagai persoalan dalam pembangunan pendidikan yang merata dan berkualitas [6]. Misalnya, terdapat daerah dengan PDRB per kapita tinggi tetapi kualitas pendidikan rendah, atau wilayah dengan kepadatan penduduk tinggi namun distribusi fasilitas pendidikan yang tidak seimbang. Kasus-kasus semacam ini berpotensi menghasilkan *outlier* yang dapat merusak hasil analisis jika tidak ditangani dengan pendekatan statistik yang tepat.

Dengan latar belakang tersebut, penelitian ini mengambil kasus data pendidikan di Jawa Barat tahun 2024, dengan variabel: Y = Indeks Pendidikan, X_1 = PDRB per kapita, X_2 = tingkat kemiskinan, X_3 = jumlah sekolah, X_4 = jumlah guru, dan X_5 = kepadatan penduduk. Analisis ini dilakukan dengan dua tahap. Pertama, aplikasi *Minimum Covariance Determinant* atas variabel X_1 – X_5 untuk mendeteksi *outlier* multivariat dalam data variabel independen. Kedua, membangun model regresi antara Y dan variabel X dengan metode *Least Trimmed Squares* pada subset data yang telah dibersihkan dari pengaruh *outlier*.

Pendekatan dua tahap ini sejalan dengan literatur statistik *robust*, yang menekankan pentingnya identifikasi dan penanganan pencilan sebelum melakukan analisis regresi. Penelitian ini bertujuan untuk menghasilkan model yang lebih stabil, tidak bias, dan memiliki tingkat akurasi tinggi dalam menjelaskan faktor-faktor penentu indeks pendidikan di wilayah Jawa Barat. Dari analisis yang

dilakukan, diharapkan dapat memberi saran yang valid dan berbasis data kepada pengambil kebijakan di sektor pendidikan, khususnya dalam merumuskan strategi pemerataan dan peningkatan kualitas pendidikan secara regional.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan di atas, maka masalah yang akan dirumuskan pada penelitian ini, yaitu:

1. Pendeteksian dan penanganan *outlier* pada data yang dapat memengaruhi ketepatan hasil analisis regresi.
2. Menentukan model regresi yang paling *robust* dan akurat terhadap data yang mengandung *outlier*, melalui evaluasi dan perbandingan kinerja model berdasarkan nilai *Mean Square Error (MSE)*.

1.3 Batasan Masalah

Dalam perumusan masalah, adapun hal yang menjadi batasan dari masalah tersebut diantaranya adalah:

1. Metode pendeteksian *outlier* yang digunakan adalah pencarian titik *leverage*.
2. Metode yang digunakan untuk penanganan dan estimasi regresi *robust* terhadap *outlier* adalah *Minimum Covariance Determinant (MCD)* sebagai metode deteksi *outlier* multivariat yang dilanjutkan dengan regresi Metode Kuadrat Terkecil (MKT) setelah *outlier* dihapus, serta *Least Trimmed Squares (LTS)* sebagai metode regresi *robust* yang mengurangi pengaruh *outlier* tanpa menghapus data.
3. Evaluasi dan pemilihan model berdasarkan nilai *Mean Square Error (MSE)*.
4. Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistika (BPS) Provinsi Jawa Barat tahun 2024.

1.4 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah yang telah disampaikan maka terdapat beberapa tujuan dari penelitian ini, diantaranya :

1. Mendeteksi *outlier* pada data keseluruhan dan mengidentifikasi data termasuk *outlier* menggunakan pendekatan awal pencarian titik *leverage*. Melakukan estimasi model regresi linear berganda dengan Metode Kuadrat Terkecil (MKT) setelah penghilangan *outlier* berdasarkan *Minimum Covariance Determinant* (MCD), serta menggunakan metode *Least Trimmed Squares* (LTS) sebagai pendekatan regresi yang *robust* terhadap outlier.
2. Mendapatkan model regresi terbaik yang *robust* dan akurat dalam memodelkan data yang mengandung *outlier*, berdasarkan perbandingan nilai *Mean Square Error* (MSE) pada data pendidikan di Jawa Barat.

1.5 Metode Penelitian

Metode penelitian yang digunakan dalam penulisan ini dapat diuraikan sebagai berikut:

1. Metode yang digunakan dalam penelitian ini diperoleh dari berbagai sumber buku, artikel, jurnal, skripsi, dan thesis yang berkaitan dengan topik penelitian
2. *Software* yang digunakan untuk melakukan proses perhitungan pada setiap metode adalah *software* RStudio.

1.6 Sistematika Penulisan

Berdasarkan cara penulisan, penelitian ini menggunakan sistematika penulisan yang terdiri dari lima bab, dan setiap bab terdiri dari beberapa sub bab.

BAB I PENDAHULUAN

Bab ini berisi pendahuluan dari penelitian ini yang terdiri dari latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisi mengenai teori-teori yang berhubungan dengan topik penelitian, dan secara umum mencakup seluruh materi yang berhubungan dengan pendeteksian *outlier* meliputi model analisis regresi, Metode Kuadrat Terkecil, uji asumsi klasik, *leverage*, regresi *robust*, Mahalanobis *distance*, *robust Mahalanobis distance*, *Mean Square Error* dan RStudio.

BAB III METODE DETEKSI OUTLIER MENGGUNAKAN ESTIMASI MINIMUM COVARIANCE DETERMINANT DAN LEAST TRIMMED SQUARES

Bab ini membahas metode utama dalam penelitian ini, yaitu cara pendeteksian *outlier* menggunakan estimasi *Minimum Covariance Determinant* (MCD) dan *Least Trimmed Squares* (LTS) pada data, serta proses estimasi parameter yang terkait.

BAB IV STUDI KASUS DAN ANALISA

Bab ini berisi studi kasus dan analisa data tentang bahasan utama yang telah dijelaskan pada bab III yang diaplikasikan dalam data Pendidikan di Jawa Barat tahun 2024.

BAB V KESIMPULAN

Bab ini berisi kesimpulan dan saran yang bisa diajukan pada penelitian ini.