

REGRESI ROBUST UNTUK PEMODELAN DEFORESTASI DI INDONESIA

Ika Aprilia Rizka Azzahro

Program Studi Matematika, FMIPA, Universitas Negeri Surabaya

e-mail : ika.19057@mhs.unesa.ac.id

A'yunin Sofro

Program Studi Matematika, FMIPA, Universitas Negeri Surabaya

Penulis Korespondensi : ayuninsofro@unesa.ac.id

Abstrak

Deforestasi adalah proses penebangan atau penghilangan hutan secara besar-besaran yang menjadi masalah serius di seluruh dunia. Fenomena ini telah menjadi perhatian global karena dampak negatifnya terhadap lingkungan dan kehidupan manusia. Tinggi rendahnya tingkat deforestasi di Indonesia tentunya sangat penting, hal ini didukung dengan pentingnya memiliki data deforestasi setiap tahunnya untuk kemudian diteliti kembali deforestasi yang tidak stabil. Faktor-faktor yang diasumsikan berpengaruh terhadap deforestasi pada penelitian ini, yaitu produksi kayu bulat dan luas tanaman perkebunan. Metode yang digunakan untuk menganalisis data yang terkontaminasi oleh *outlier* dan memberikan hasil yang lebih fleksibel adalah regresi *robust*. Hasil yang diperoleh adalah model regresi *robust* dan faktor yang berpengaruh secara signifikan terhadap deforestasi di Indonesia sesuai dengan tujuan penelitian. Jenis penelitian yang dilakukan dalam penelitian ini adalah penelitian eksperimen dengan studi literatur. Variabel yang digunakan adalah data deforestasi di Indonesia, produksi kayu bulat dan luas tanaman perkebunan. Setelah itu diperoleh model regresi *robust* dengan estimasi *maximum likelihood type (M)*. Diperoleh tinggi rendahnya deforestasi tahun 2017-2021 dipengaruhi oleh luas tanaman perkebunan.

Kata Kunci: Regresi Robust, Maximum Likelihood Type (M), Outlier

Abstract

Deforestation is the process of cutting or removing forests on a large scale which is a serious problem throughout the world. This phenomenon has become a global concern because of its negative impact on the environment and human life. The level of deforestation in Indonesia is of course very important, this is supported by the importance of having deforestation data every year to then re-examine unstable deforestation. Factors that are assumed to influence deforestation in this study are log production and plantation area. The method used to analyze data contaminated by outliers and provide more flexible results is robust regression. The results obtained are robust regression models and factors that have a significant effect on deforestation in Indonesia according to the research objectives. The type of research conducted in this research is experimental research with literature studies. The variables used are deforestation data in Indonesia, log production and plantation area. After that, a robust regression model was obtained with an estimate of the maximum likelihood type (M). The level of deforestation obtained for 2017-2021 is influenced by the area of plantation crops.

Keywords: Robust Regression, Maximum Likelihood Type (M), Outlier

PENDAHULUAN

Deforestasi adalah proses penebangan atau penghilangan hutan secara besar-besaran yang menjadi masalah serius di seluruh dunia. Fenomena

ini telah menjadi perhatian global karena dampak negatifnya terhadap lingkungan dan kehidupan manusia. Deforestasi telah menyebabkan kerugian besar dalam keanekaragaman hayati, perubahan iklim, dan perubahan tata air, serta berdampak pada

kehidupan sosial dan ekonomi masyarakat di wilayah yang terkena dampaknya.

Hutan adalah salah satu ekosistem paling berharga di Bumi. Hutan berfungsi sebagai habitat bagi berbagai spesies tumbuhan dan hewan, termasuk spesies yang terancam punah. Selain itu, hutan juga berperan dalam menjaga kualitas udara, mengikat karbon dioksida, dan mempengaruhi pola iklim global. Namun, aktivitas manusia seperti penebangan liar, konversi lahan hutan menjadi lahan pertanian, dan eksploitasi sumber daya alam telah menyebabkan deforestasi yang masif. Akibatnya, luas hutan di berbagai wilayah telah menurun secara dramatis, mengancam keberlanjutan ekosistem dan mengurangi potensi penyerapan karbon oleh hutan.

Deforestasi berkontribusi secara signifikan terhadap perubahan iklim global. Hutan yang sehat berfungsi sebagai penyimpan karbon alami, tetapi ketika ditebang, karbon yang disimpan dilepaskan kembali ke atmosfer dalam bentuk gas rumah kaca. Perubahan iklim ini menyebabkan peningkatan suhu global, pola cuaca yang ekstrem, dan bencana alam yang lebih sering terjadi. Pemanasan global berdampak pada mencairnya es di kutub, naiknya permukaan air laut, dan terancamnya habitat berbagai spesies yang tergantung pada iklim tertentu.

Deforestasi tidak hanya berdampak pada lingkungan, tetapi juga pada kehidupan sosial dan ekonomi masyarakat di wilayah yang terkena dampaknya. Banyak masyarakat yang bergantung pada hutan untuk kehidupan mereka, seperti penghasilan dari hasil hutan, obat-obatan tradisional, serta sumber makanan. Hilangnya hutan berdampak pada kehidupan sosial masyarakat, perubahan mata pencaharian, dan tingkat kemiskinan yang meningkat.

Deforestasi adalah masalah global yang serius dengan dampak negatif yang luas terhadap lingkungan, iklim, dan kehidupan manusia. Upaya kolektif dari masyarakat internasional, pemerintah, organisasi lingkungan, dan masyarakat lokal diperlukan untuk mengatasi tantangan deforestasi. Perlindungan dan konservasi hutan menjadi kunci dalam memitigasi dampak deforestasi dan memastikan keberlanjutan ekosistem serta kesejahteraan manusia di masa depan. Penelitian skripsi ini bertujuan untuk memahami penyebab

deforestasi, konsekuensi dari deforestasi, serta upaya-upaya untuk mengatasi masalah tersebut guna memberikan kontribusi positif dalam pelestarian lingkungan dan keanekaragaman hayati di bumi.

Runtutan uraian di atas menyatakan bahwa deforestasi di Indonesia sangat penting, hal ini didukung dengan pentingnya memiliki data deforestasi setiap tahunnya untuk kemudian diteliti kembali deforestasi yang tidak stabil. Salah satu penelitian yang penting dan berkaitan dengan deforestasi adalah peramalan untuk tingkat deforestasi di tahun-tahun berikutnya. Semakin banyak penelitian yang dilakukan dengan menggunakan banyak metode akan semakin akurat hasil peramalan tersebut. Begitu juga dengan faktor apa saja yang mempengaruhi deforestasi. Faktor-faktor yang diasumsikan berpengaruh terhadap deforestasi pada penelitian ini, yaitu data produksi kayu bulat dan data luas tanaman perkebunan.

Penelitian tentang peramalan umumnya sangat sering berhubungan dengan *outlier*. *Outlier* dapat memberikan informasi yang tidak bisa diberikan oleh titik data lainnya, salah satunya *outlier* terjadi karena keadaan yang tidak biasa yang mungkin saja sangat penting dan perlu diselidiki lebih jauh. Sebagian penelitian menerapkan *outlier* baru ditolak jika setelah ditelusuri ternyata merupakan akibat dari kesalahan-kesalahan seperti memasukkan ukuran atau analisis yang salah, ketidaktepatan pencatatan data, dan terjadi kerusakan alat pengukuran. Bila ternyata bukan akibat dari kesalahan-kesalahan semacam itu, penyelidikan yang seksama harus dilakukan \cite*{Dra}.

Data *outlier* harus dilihat terhadap posisi dan sebaran data yang lainnya, sehingga perlu dievaluasi apakah data *outlier* tersebut perlu dihilangkan atau tidak. Menurut Soemartini terdapat beberapa metode untuk menentukan batasan *outlier* dalam sebuah analisis, yaitu *scatterplot*, *boxplot*, *leverage values*, *DfFITS* (*Difference in Fit Statndarized*), *cook's distance*, dan *DfBETA(s)* (*Diference in Beta*), dan *internal studentization* yang memiliki perhitungan hampir sama dengan *R-student*. Pada penelitian ini akan menggunakan tiga metode saja \cite*{Soe}.

Salah satu metode regresi untuk menganalisis data yang terkontaminasi oleh *outlier* dan memberikan hasil yang lebih fleksibel adalah regresi

robust. Oleh karena itu, penelitian ini akan menerapkan salah satu metode yaitu metode regresi *robust* pada kasus deforestasi di Indonesia. Hasil yang diperoleh adalah model deforestasi dengan metode regresi *robust* di Indonesia.

Regresi *robust* merupakan metode digunakan untuk menganalisa data yang dipengaruhi oleh *outlier* sehingga dihasilkan model yang *robust* atau *resistance* terhadap *outlier*. Suatu estimasi yang *resistance* adalah relatif tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data \citep*{Wid}. Prosedur regresi *robust* dirancang untuk mengurangi pengaruh dari *outlier*. Oleh karena itu, prosedur regresi *robust* cenderung untuk mengabaikan residual yang berhubungan dengan *outlier-outlier* yang besar \citep*{Rom}. Kelebihan lainnya dari metode regresi *robust* ini adalah menganalisis data yang dipengaruhi oleh *outlier* sehingga dihasilkan model yang *robust* atau resistansi terhadap *outlier*.

KAJIAN TEORI

Analisis Deskriptif

Analisis deskriptif adalah metode yang berkaitan dengan pengumpulan atau penyajian data sehingga memberi informasi yang berguna. Data yang dilihat adalah jumlah data, nilai minimum, nilai maksimum, nilai rata-rata, dan standar deviasi \citep*{Ghoz2}. Analisis deskriptif dalam penelitian ini digunakan untuk melihat gambaran data pada setiap variabel penelitian yang digunakan dalam studi kasus penelitian. Pada penelitian ini data yang akan dicari adalah nilai minimum, nilai rata-rata, nilai kuartil pertama, nilai kuartil kedua (median), nilai kuartil ketiga, nilai maksimum, dan nilai standar deviasi.

Berikut adalah rumus dan contoh perhitungan manual dengan data secara lengkap dapat dilihat pada tabel 1:

Tabel 1. Contoh Data

No	X _{1(i)}	X _{2(i)}	Y _i
1	X ₁₁	X ₂₁	Y ₁
2	X ₁₂	X ₂₂	Y ₂
3	X ₁₃	X ₂₃	Y ₃
4	X ₁₄	X ₂₄	Y ₄
5	X ₁₅	X ₂₅	Y ₅

Data yang digunakan adalah data deforestasi yang disimbolkan sebagai Y, produksi kayu bulat yang disimbolkan sebagai x1, dan luas tanaman perkebunan disimbolkan x2. Data penelitian berjumlah n data, dengan x11 adalah data produksi kayu bulat pertama, x12 adalah data kedua dan seterusnya sampai dengan data ke-n. Data x21 dan Y1 memiliki arti yang sama dengan data x11 dan seterusnya. Berikut adalah rumus yang digunakan:

Nilai rata-rata : $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ (2.1)

Nilai kuartil pertama : $Q_1 = \text{Data ke } \frac{1(n+1)}{4}$ (2.2)

Nilai kuartil kedua : $Q_2 = \text{Data ke } \frac{2(n+1)}{4}$ (2.3)

Nilai kuartil ketiga : $Q_3 = \text{Data ke } \frac{3(n+1)}{4}$ (2.4)

Standar deviasi: $s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$ (2.5)

Dengan keterangan sebagai berikut:

- x_i : Data ke-i
- n : Banyaknya data x_i
- \bar{x} : Nilai rata-rata

Data Outlier

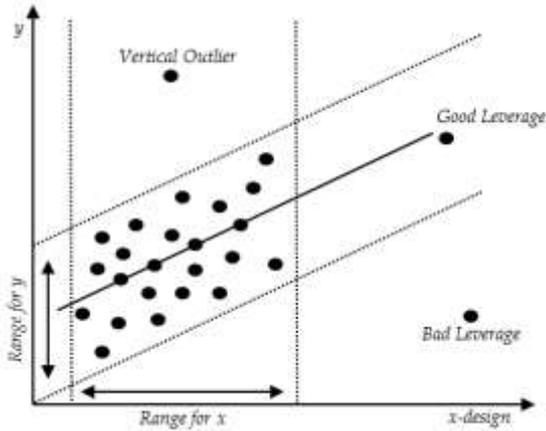
Berbagai penelitian telah diajukan untuk menolak *outlier* (memutuskan untuk menghilangkan data yang ada *outlier*-nya, setelah itu data dianalisis ulang tanpa *outlier*). Penolakan *outlier* yang begitu saja bukanlah langkah yang bijaksana. Adakalanya *outlier* dapat memberikan informasi yang tidak bisa diberikan oleh titik data lainnya, misalnya karena *outlier* timbul dari kombinasi keadaan yang tidak biasa yang mungkin saja sangat penting dan perlu diselidik lebih jauh. Sebagian penelitian menerapkan *outlier* baru ditolak jika setelah ditelusuri ternyata merupakan akibat dari kesalahan-kesalahan seperti memasukkan ukuran atau analisis yang salah, ketidaktepatan pencatatan data, dan terjadi kerusakan alat pengukuran. Bila ternyata bukan akibat dari kesalahan-kesalahan semacam itu, penyelidikan yang seksama harus dilakukan \citep*{Dra}.

Outlier menyebabkan hal-hal berikut (Soemartini, 2007):

1. Residual yang besar dari model yang terbentuk
2. Varians pada data tersebut menjadi besar

3. Taksiran interval memiliki rentang yang lebar

Pada analisis regresi, terdapat 3 tipe outlier yang berpengaruh terhadap estimasi kuadrat terkecil yaitu sebagai berikut \citep*{Soe}:



Gambar 1. Tipe Data Outlier

Sumber: (Verardi, 2008)

1. Outlier Vertikal

Outlier vertikal merupakan semua pengamatan yang terpercil pada variabel dependen, tetapi tidak terpercil pada variabel independen.

2. Good Leverage Point

Good leverage point merupakan pengamatan yang terpercil pada variabel independen tetapi terletak dekat dengan garis regresi. Hal ini berarti pengamatan menjauh tetapi dekat dengan garis regresi. Keberadaan good leverage point tidak berpengaruh terhadap estimasi kuadrat terkecil, tetapi berpengaruh terhadap inferensi statistik karena dapat meningkatkan estimasi standar error seperti gambar 1.

3. Bad leverage point

Bad leverage point merupakan pengamatan yang terpercil pada variabel independen dan terletak jauh dari garis regresi. Keberadaan bad leverage point berpengaruh signifikan terhadap estimasi kuadrat terkecil, baik terhadap intercept maupun slope dari persamaan regresi seperti gambar 1.

Data outlier harus dilihat terhadap posisi dan sebaran data yang lainnya, sehingga akan dievaluasi apakah data outlier tersebut perlu dihilangkan atau tidak. Terdapat beberapa metode

untuk menentukan batasan outlier dalam sebuah analisis, menurut (Soemartini, 2007) yaitu (1) Scatterplot, (2) Boxplot, (3) Leverage Values, DfFITS (Difference in Fit Statndarized), Cook's Distance, dan DfBETA(s) (Diference in Beta), dan (4) Internal Studentization (residu yang distudentkan) yang memiliki perhitungan hampir sama dengan R-student. Pada penelitian ini akan menggunakan ketiga metode tersebut dengan lebih terperinci, yaitu (1) Scatterplot, (2) Boxplot, dan (3) Leverage Values untuk menentukan batasan outlier.

Identifikasi Data Outlier

1. Scatterplot

Menurut (Soemartini, 2007), metode ini dilakukan dengan cara memplotkan data dengan observasi ke- i ($i = 1,2, \dots n$). Selain itu, jika sudah didapatkan model regresi maka dapat dilakukan dengan cara memplot residual (e) dengan nilai prediksi Y (\hat{Y}). Jika terdapat satu atau beberapa data yang terletak jauh dari pola kumpulan data keseluruhan, maka hal ini mengindikasikan adanya outlier.

2. Boxplot

Menurut (Soemartini, 2007), metode ini mempergunakan nilai kuartil dan jangkauan untuk mendeteksi outlier. Kuartil 1,2, dan 3 akan membagi data yang telah diurutkan sebelumnya menjadi empat bagian. Jangkauan IQR (Interquartile Range) didefinisikan sebagai selisih kuartil 1 terhadap 3.

3. Leverage Values

Metode ini mengukur pengaruh suatu observasi terhadap besarnya estimasi parameter antara lain dapat dilihat dari jarak nilai X semua observasi. Menurut (Wijaya, 2009), nilai leverage untuk linier sederhana dapat ditentukan sebagai berikut:

$$leverage(h_{ii}) = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{(n - 1)S_x^2} \tag{2.6}$$

Dengan keterangan sebagai berikut:

- h_{ii} : Leverage kasus ke- i
- n : Banyaknya data
- X_i : Nilai untuk kasus ke- i
- \bar{X} : Mean dari X

S_x^2 : Kuadrat n kasus dari simpangan X_i terhadap mean

Setelah menentukan batasan *outlier*, metode regresi yang penting untuk menganalisis data yang terkontaminasi oleh *outlier* dan memberikan hasil yang lebih fleksibel adalah regresi *robust*.

Regresi Robust

Konsep regresi pertama kali dipopulerkan oleh Sir Francis Galton pada akhir abad ke-19 dalam jurnal yang berjudul "*Regression towards mediocrity in hereditary stature*", yang dimuat dalam *Journal of the Anthropological Institute*, volume 15, hal. 246-263, tahun 1885 \citep*{Gal}. Regresi linear adalah pendekatan yang memodelkan hubungan antara respons skalar atau variabel dependen Y dan satu atau lebih variabel penjelas atau independen X . Model regresi linear dengan p variabel independen dapat dinyatakan sebagai berikut \citep*{Mon}:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.7)$$

Y_i adalah variabel dependen dari pengamatan ke- i dan $\beta_0, \beta_1, \dots, \beta_p$ adalah parameter, X_i adalah nilai variabel independen pada pengamatan ke- i dan ε_i adalah error yang saling bebas dan menyebar normal (Montgomery, Peck, & Vinning, 1982).

Model regresi linier yang menghubungkan variabel dependen dan satu variabel independen disebut dengan regresi linier sederhana. Sedangkan regresi linier berganda digunakan untuk menghubungkan variabel dependen dan dua atau lebih variabel independen. Pada penelitian ini akan menggunakan regresi linier berganda, karena variabel independen yang akan digunakan pada penelitian ini adalah lebih dari satu variabel.

Regresi linear berganda adalah model statistik yang memanfaatkan hubungan antara dua atau lebih variabel independen sehingga suatu variabel dapat diprediksi dari variabel yang lain \citep*{Net}. Model regresi linear berganda dapat dituliskan sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i(p-1)} + \varepsilon_i \quad (2.8)$$

Fungsi regresi untuk model regresi linier berganda persamaan (2.8), dengan $\varepsilon_i = 0$ adalah sebagai berikut:

$$\hat{Y} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i(p-1)} \quad (2.9)$$

Regresi *robust* merupakan alat penting untuk menganalisa data yang dipengaruhi oleh *outlier* sehingga dihasilkan model yang *robust* atau *resistance* terhadap *outlier*. Suatu estimasi yang *resistance* adalah relatif tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data (Widodo & Dewayanti, 2016).

Prosedur regresi *robust* cenderung untuk mengabaikan residual yang berhubungan dengan *outlier-outlier* yang besar. Disamping tidak sensitif jika terdapat kasus *outlier*, prosedur regresi *robust* mempunyai tingkat efisiensi yang sama dengan 90% - 95% di bawah distribusi normal. Regresi *robust* ini dikembangkan oleh Rousseeuw dan Leroy pada tahun 1987. Regresi *robust* ditujukan untuk mengakomodasi adanya keanehan data, sekaligus meniadakan identifikasi adanya data *outlier* dan juga bersifat otomatis dalam menanggulangi data *outlier*. Analisis regresi *robust* ini tidak membuat galat model menjadi normal namun model yang dihasilkan oleh metode ini memiliki tingkat keakuratan yang lebih tinggi dari model yang dihasilkan oleh Metode Kuadrat Terkeci (MKT) (Romdi, Wahyuningsih, & Yuniarti, 2015).

Pada regresi *robust* terdapat beberapa metode estimasi yang dapat digunakan, yakni (1) estimasi LMS, (2) estimasi LTS, (3) estimasi MM, (4) estimasi S, (5) estimasi M (Chen, 2002). Hanya satu estimasi yang akan digunakan pada penelitian ini, yaitu estimasi *Maximum Likelihood Type* (M), sehingga selanjutnya akan dijelaskan tentang estimasi M.

Regresi Robust dengan Estimasi *Maximum Likelihood Type* (M)

Menurut (Montgomery, Peck, & Vinning, 1982), estimasi M meminimumkan fungsi obyektif dari residualnya. Tahapan regresi *robust* dengan estimasi M adalah sebagai berikut (Susanti, Pratiwi, H., & Liana, 2014):

1. Menghitung estimasi parameter $\hat{\beta}$ sehingga didapatkan \hat{y}_i

2. Menghitung nilai residualnya dengan $e_i = Y_i - \hat{Y}_i$
3. Menghitung nilai S_m ,

$$S_m = \frac{\text{med}|e_i - \text{med}(e_i)|}{0.6745} \quad (2.10)$$

4. Menghitung nilai $u_i = \frac{e_i}{S_i}$
5. Menghitung nilai fungsi pembobot $w_i = w(u_i)$ dengan fungsi pembobot *Huber* konstanta yang digunakan adalah $c = 1.345$,

$$w(u_i) = \begin{cases} 1, & |u_i| \leq c \\ \frac{c}{|u_i|}, & |u_i| > c \end{cases} \quad (2.11)$$

6. Mengestimasi nilai $\hat{\beta}_m$ menggunakan pembobot w_i sehingga diperoleh estimasi M satu tahap. Pada setiap iterasi ke- t dihitung residual $e_i^{(t-1)}$ dan menggunakan pembobot $w_i^{(t-1)} = w(u_i^{(t-1)})$ dari iterasi sebelumnya sehingga didapatkan estimasi parameter $\hat{\beta}_m$ yang baru
7. Melakukan langkah 2 sampai langkah 6 hingga didapatkan estimasi parameter $\hat{\beta}_m$ yang konvergen

Breakdown Point

Breakdown point adalah persentase terkecil dari banyaknya data yang terkontaminasi atau banyaknya *outlier* yang menyebabkan nilai dari taksiran menjadi besar. *Breakdown point* digunakan untuk menjelaskan ukuran ke-robust-an dari teknik *robust*. Kemungkinan tertinggi *breakdown point* untuk sebuah estimator adalah 50%. Jika *breakdown point* lebih dari 50% berarti estimasi model regresi tidak dapat menggambarkan informasi dari kebanyakan data (Wulandari, 2020).

Selain pengukuran ke-robust-an, perlu dicari nilai fungsi objektif dan fungsi pembobot.

Fungsi Objektif dan Fungsi Pembobot

Fungsi objektif merupakan representasi pembobot dari residual. Fungsi ini digunakan untuk mencari fungsi pembobot pada regresi *robust*. sedangkan fungsi pembobot didapatkan dengan menggunakan fungsi objektif. Fungsi pembobot yang digunakan antara lain adalah fungsi pembobot *Huber* dan fungsi pembobot *Tukey Bisquare* (Montgomery, Peck, & Vinning, 1982).

Konstanta yang menghasilkan efisiensi tinggi dengan residual berdistribusi normal dan dapat

memberikan perlindungan terhadap *outlier* yaitu konstanta dengan nilai $c = 1.345$ untuk fungsi pembobot *Huber*. Selanjutnya akan dicari tahu ada atau tidaknya pengaruh secara bersama-sama (simultan) maupun sendiri-sendiri (parsial) variabel independen terhadap variabel dependen. Maka pada subbab selanjutnya akan dijelaskan tentang pengujian signifikansi parameter yang juga menjelaskan tentang signifikansi simultan dan parsial.

Pengujian Signifikansi Parameter

1. Pengujian Signifikansi Simultan

Menurut (Purba, Tarigan, Sinaga, & Tarigan, 2021), untuk melihat pengaruh variabel bebas secara simultan terhadap variabel terikat dilakukan uji F. Berdasarkan uji F, jika nilai F Hitung > F Tabel dan nilai signifikansi lebih kecil dari 0,05 maka disimpulkan secara simultan variabel independen berpengaruh signifikan terhadap variabel dependen. Hipotesis uji yang digunakan adalah sebagai berikut (Ratnasari, 2017):

$H_0 : \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$ (tidak ada pengaruh simultan variabel independen terhadap variabel dependen)

$H_1 : \text{tidak semua } \beta_k \text{ sama dengan nol } (k = 1, \dots, p - 1)$ (minimal ada satu variabel independen yang berpengaruh signifikan terhadap variabel dependen)

Statistik uji F didefinisikan sebagai berikut:

$$F^* = \frac{SSR/(p - 1)}{SSE/(n - p)} \quad (2.12)$$

Dimana,

$$SSR = \sum(\hat{Y}_i - \bar{Y})^2 \text{ dan } SSE = \sum(Y_i - \hat{Y}_i)^2$$

SSR adalah singkatan dari *Regression Sum of Square*, sedangkan SSE singkatan dari *Error Sum of Square*. Jika nilai $F^* > F_{1-\alpha(p-1, n-p)}$ atau $p - value < \alpha$ maka menghasilkan keputusan tolak H_0 yang berarti setidaknya terdapat satu variabel independen yang mempengaruhi variabel dependen.

2. Pengujian Signifikansi Parsial

Menurut (Purba, Tarigan, Sinaga, & Tarigan, 2021), untuk melihat pengaruh variabel independen secara sendiri-sendiri terhadap variabel dependen didapat dengan melakukan uji t. Jika hasil uji t didapat nilai t hitung > t tabel dan nilai signifikansi lebih kecil dari 0,05 maka disimpulkan terdapat pengaruh yang signifikan terhadap variabel dependen (Ratnasari, 2017).

Statistik uji t didefinisikan sebagai berikut:

$$t^* = \frac{b_k}{s\{b_k\}} \quad (2.13)$$

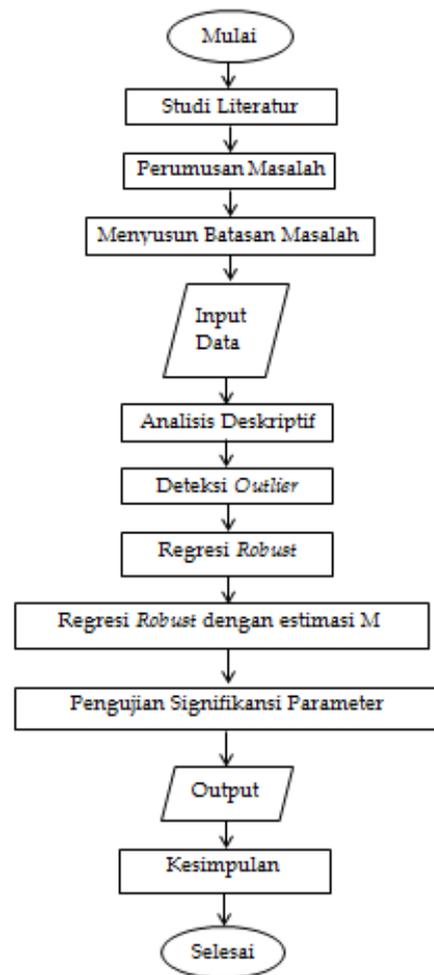
Jika $|t^*| > t_{1-\alpha/2(n-p)}$ atau $p\text{-value} < \alpha$ maka menghasilkan keputusan tolak H_0 , yang berarti variabel independen tersebut berpengaruh signifikan terhadap variabel dependen.

METODE

Data yang digunakan dalam penelitian ini adalah data sekunder. Data yang digunakan merupakan data deforestasi di Indonesia yang didapat dari website resmi publikasi Badan Pusat Statistik (<https://www.bps.go.id/>). Data yang digunakan mengambil data *update* setiap provinsi di Indonesia dalam rentang tahun 2017 sampai dengan 2021.

Diagram Alir Penelitian

Berikut adalah diagram alir berisi tahapan yang akan dilakukan pada penelitian ini:



Gambar 2. Diagram Alir Penelitian

Berdasarkan gambar 2, prosedur penelitian untuk mencapai tujuan penelitian yang ditetapkan menggunakan metode regresi *robust* adalah sebagai berikut:

1. Melakukan studi literatur

Studi literatur yang dilakukan berupa kajian pustaka terhadap sumber-sumber bacaan yang relevan sehingga mampu menunjang dalam proses analisis laju deforestasi di Indonesia menggunakan metode regresi *robust*.

2. Melakukan identifikasi dan perumusan masalah

Tahap ini dilakukan untuk membantu penelitian agar fokus terhadap pembahasan tertentu.

3. Menyusun batasan masalah

Tahap ini dilakukan untuk memberikan ruang lingkup terbatas pada permasalahan agar lebih jelas.

4. Mengumpulkan data dan melakukan analisis deskriptif

Analisis deskriptif yang dimaksud adalah metode yang berkaitan dengan pengumpulan atau penyajian data sehingga memberi informasi yang berguna. Data yang digunakan didapat dari website resmi publikasi Bank Indonesia dan Badan Pusat Statistik.

5. Menganalisis identifikasi outlier

Data outlier harus dilihat terhadap posisi dan sebaran data yang lainnya, sehingga akan dilakukan evaluasi apakah data outlier tersebut perlu dihilangkan atau tidak. Pada penelitian ini akan menggunakan ketiga metode dengan lebih terperinci, yaitu:

- a. Melakukan perhitungan statistik metode scatterplot
- b. Melakukan perhitungan statistik metode boxplot
- c. Melakukan perhitungan statistik metode leverage values

6. Menganalisis menggunakan metode regresi robust

Prosedur regresi robust cenderung untuk mengabaikan residual yang berhubungan dengan outlier-outlier yang besar. Semua tahapan yang dilakukan sebelum tahap ini merupakan syarat untuk mendapatkan model regresi robust. Pada tahap ini akan dilakukan untuk menjawab rumusan masalah dan mendapatkan tujuan penelitian yang diinginkan. Penelitian ini akan menggunakan metode regresi robust dengan estimasi M.

7. Menganalisis menggunakan fungsi objektif dan fungsi pembobot

Fungsi objektif merupakan representasi pembobot dari residual. Fungsi ini digunakan untuk mencari fungsi pembobot pada regresi robust. Sedangkan fungsi pembobot didapatkan dengan menggunakan fungsi objektif.

8. Menganalisis validitas menggunakan uji signifikan parameter

Tahap ini dilakukan untuk mengetahui ada atau tidaknya pengaruh secara simultan dan parsial

variabel independen terhadap variabel dependen. Berdasarkan nilai parameter yang signifikan, ada beberapa tahapan yang dilakukan yaitu:

- a. Melakukan uji simultan dengan menggunakan uji F
 - b. Melakukan uji parsial dengan menggunakan uji t
9. Membuat kesimpulan

Tahap akhir dari penelitian ini adalah kesimpulan yang dapat dilakukan untuk penelitian selanjutnya.

HASIL DAN PEMBAHASAN

Data yang akan digunakan penelitian ini adalah data tahun 2017 sampai dengan 2021 di Indonesia. Ada tiga data yang digunakan pada penelitian ini, diantara lain: (1) deforestasi (Y), (2) produksi kayu bulat (X1), dan (3) luas tanaman perkebunan (X2). Data yang digunakan mengambil data update setiap provinsi dari website resmi publikasi Badan Pusat Statistik.

Tabel 2. Data Penelitian Awal

Tahun	Provinsi	Y	X1	X2
2021	Aceh	2.998,8	16064,00	913,1
2021	Sumut	3.339	1689227,00	1952,9
2021	Subar	6.797	121884,59	747,5
2021	Riau	1.518	24239074,00	3640,5
...
...
2017	Papua	17.259,1	763650,42	257,2

Berdasarkan tabel data dikonversi terlebih dahulu menjadi bentuk bilangan desimal. Sehingga dihasilkan tabel 3 sebagai berikut:

Tabel 3. Data Setelah diKonversi

Tahun	Provinsi	Y	X1	X2
2021	Aceh	2998,80	16064,00	913,10
2021	Sumut	3339,00	1689227,00	1952,90
2021	Subar	6797,00	121884,59	747,50
2021	Riau	1518,00	24239074,00	3640,50
...
...
2017	Papua	17259,10	763650,42	257,20

Analisis Deskriptif

Pada penelitian ini, analisis deskriptif digunakan untuk melihat gambaran data tentang setiap variabel penelitian. Penelitian ini menggunakan variabel dependen Y yang merupakan deforestasi dan variabel independen X yaitu, produksikayu bulat (X1), dan luas tanaman perkebunan (X2). Berikut disajikan Tabel 4 untuk melihat gambaran data pada setiap variabel yang digunakan:

Tabel 4. Hasil Analisis Deskriptif

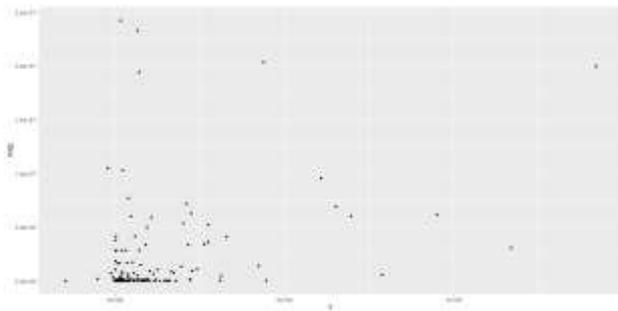
	Min.	\bar{x}	Max.	s
Y	-14802m7	9806,5	142011,1	20047,02
X1	0	1745603	24239074	4096720
X2	0	751,6	3640,5	843,6375

Berdasarkan Tabel 4 diketahui bahwa nilai minimal dari deforestasi di Indonesia setiap provinsi dalam 5 tahun terakhir adalah -14802.7 dan nilai maksimalnya adalah 142011,1. Perbedaan nilai deforestasi dapat diartikan deforestasi deforestasi dipengaruhi oleh beberapa faktor.

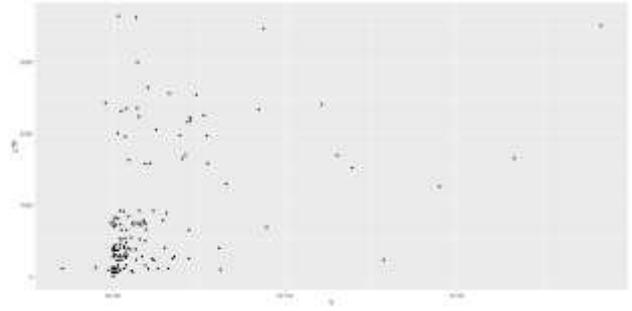
Identifikasi Data *Outlier*

1. *Scatterplot*

Scatterplot bertujuan untuk melihat apakah terdapat *outlier* pada data. Berikut adalah hasil *scatterplot* yang dilakukan dengan memplot antara data dengan observasi ke- i ($i = 1, 2, 3, \dots, n$):



Gambar 3. *Scatterplot* Y dan X1

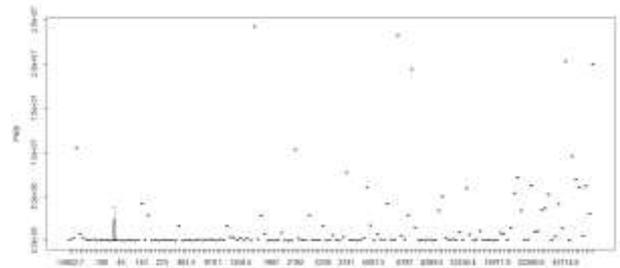


Gambar 4. *Scatterplot* Y dan X2

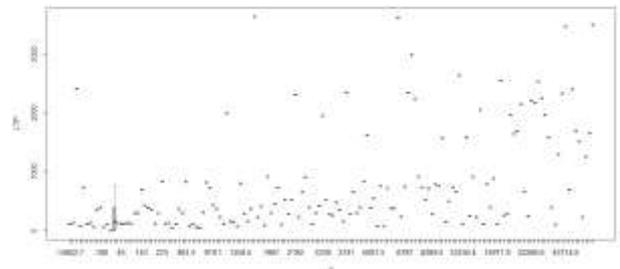
Setiap variabel independen memiliki beberapa nilai *outlier*, seperti terlihat pada Gambar 3 dan Gambar 4. *Outlier* adalah kumpulan data yang berbeda secara signifikan dari pola keseluruhan kumpulan data. Karena hanya menggunakan visualisasi grafis, metode *scatterplot* sangat bergantung pada penilaian peneliti. Akibatnya, dibutuhkan seorang ahli dalam interpretasi plot. sehingga teknik deteksi *outlier* tambahan dapat digunakan untuk mendapatkan data *outlier* yang lebih akurat.

2. *Boxplot*

Boxplot bertujuan untuk mendeteksi nilai *outlier*. Berikut adalah hasil *boxplot* yang dilakukan dengan mempergunakan nilai kuartil dan jangkauan:



Gambar 5. *Boxplot* Y dan X1



Gambar 6. *Boxplot* Y dan X2

Berdasarkan Gambar 5 dan Gambar 6 diketahui terdapat beberapa nilai *outlier* dari setiap variabel

independen X. Panjang kotak menggambarkan tingkat penyebaran data, dengan ketentuan semakin panjang berarti data semakin menyebar. Nilai outlier adalah data yang terletak lebih dari $1,5 \times$ panjang kotak (IQR). Metode ini hampir mirip dengan *scatterplot* yaitu sangat bergantung pada penilaian peneliti, karena mengandalkan visualisasi grafis. Meskipun begitu, hasil dari *boxplot* sedikit berbeda dengan *scatterplot*. Namun masih diperlukan metode pendeteksi *outlier* lainnya supaya data *outlier* yang diperoleh lebih akurat.

3. *Leverage Value*

Leverage value bertujuan untuk mengukur pengaruh suatu observasi terhadap besarnya estimasi parameter antara lain dapat dilihat dari jarak nilai X semua observasi. Apabila daerah kritis $h_{ii} > cutoff$, maka dideteksi *outlier*. Nilai $cutoff = \frac{2p}{n} = \frac{2 \cdot 2}{125} = \frac{4}{125} = 0,032$, dengan p adalah jumlah variabel independen X dan n adalah banyaknya data.

Tabel 5. Hasil *Leverage Value*

No	h_{ii}
2	0,03646
4	0,19895
6	0,03355
20	0,09086
21	0,03453
35	0,03906
37	0,18052
39	0,03554
53	0,08480
54	0,04869
68	0,04723
70	0,12698
86	0,08033
87	0,03876
101	0,05435
103	0,13250
119	0,06193

Hasil identifikasi *outlier* dengan menggunakan *leverage value* yang mempunyai nilai lebih dari *cutoff* adalah data yang disajikan pada Tabel 5, sehingga data observasi tersebut merupakan *outlier*. Diantara identifikasi data *outlier* yang digunakan, salah satu diantaranya harus menunjukkan adanya data *outlier*. Hasil yang diperoleh dari semua metode yang digunakan menunjukkan adanya data *outlier*.

Regresi Robust dengan Estimasi Maximum Likelihood Type (M)

Estimasi maximum likelihood type (M) adalah metode estimasi langsung yang dikembangkan Huber secara teoritis dan matematis. Sebagian besar *outlier* yang ditemukan dalam data diasumsikan sebagai variabel dependen dalam estimasi ini. Data yang digunakan adalah data Y atau data deforestasi dengan semua data variabel independen X1 dan X2, selanjutnya adalah hasil estimasi parameter yang konvergen setelah dilakukan iterasi, sehingga didapatkan:

Tabel 6. Hasil Regresi *Robust* dengan Estimasi M

	Nilai Estimasi	Standart Error	t-value	p-value
(Intercept)	$7,527020 \times 10^2$	620,0306	1,2140	752,7020
X1	$-5,620023 \times 10^{-4}$	0,0002	-3,3289	-0,0006
X2	7,852206	0,8198	9,5779	7,8522

Melalui Tabel 6 didapatkan model persamaan regresi *robust* estimasi M sebagai berikut:

$$Y = 7,527020 \times 10^2 - 5,620023 \times 10^{-4} X1 + 7,852206 X2 \tag{2.14}$$

Selanjutnya akan dilakukan pengujian signifikansi simultan untuk mengetahui pengaruh variabel independen secara serentak terhadap variabel dependen. Apabila $p\text{-value} < \alpha$ dengan $\alpha = 0,05$, maka ditarik kesimpulan Tolak H_0 .

Apabila pengujian signifikansi simultan menunjukkan hasil H_0 , maka tidak ada pengaruh simultan variabel independen terhadap variabel dependen. Apabila hasilnya menunjukkan H_1 , maka artinya minimal ada satu variabel independen yang berpengaruh signifikan terhadap variabel dependen.

Berdasarkan hasil pengujian didapatkan bahwa p -value X_1 sebesar $-0,0006 < \alpha$ dan X_2 sebesar $7,8522 > \alpha$ yang artinya model layak digunakan, karena terdapat variabel independen yang mempengaruhi variabel dependen.

Uji signifikansi parsial juga akan digunakan untuk melihat bagaimana pengaruh variabel independen terhadap variabel dependen secara simultan. Uji signifikansi parsial bertujuan untuk menguji pengaruh masing-masing variabel independen secara terpisah. Apabila $|t - \text{value}| > t_{\text{tabel}}$, maka ditarik kesimpulan Tolak H_0 .

Apabila pengujian signifikansi parsial menunjukkan hasil H_0 , maka variabel independen ke- k tidak memiliki pengaruh yang signifikan terhadap variabel dependen. Apabila hasilnya menunjukkan H_1 , maka artinya variabel independen ke- k memiliki pengaruh yang signifikan terhadap variabel dependen ($k=1,2,\dots,p-1$).

Berdasarkan hasil pengujian didapatkan bahwa t -value X_2 sebesar $9,5779$ mempengaruhi variabel dependen, karena secara mutlak variabel independen tersebut lebih besar dari nilai t_{tabel} yaitu $2,132$. Nilai X_1 memiliki nilai t -value sebesar $-3,3289$ yang artinya lebih kecil dari nilai t_{tabel} , maka X_1 tidak mempengaruhi variabel dependen secara parsial.

Interpretasi Hasil

Interpretasi hasil yang diperoleh adalah model deforestasi di Indonesia dengan regresi *robust* yang dilakukan dengan estimasi M. Model deforestasi di Indonesia dengan regresi *robust* yang terbentuk adalah:

$$Y = 7,527020 \times 10^2 - 5,620023 \times 10^{-4} X_1 + 7,852206 X_2 \quad (2.15)$$

Berdasarkan hasil pengujian signifikansi simultan dan parsial menunjukkan bahwa variabel independen yang berpengaruh terhadap variabel dependen adalah X_2 . X_2 adalah data luas tanaman perkebunan.

PENUTUP

SIMPULAN

Berdasarkan hasil pengujian signifikansi simultan dan parsial menunjukkan bahwa variabel

independen yang berpengaruh terhadap variabel dependen adalah X_2 . Variabel X_2 adalah data luas tanaman perkebunan. Dari hasil dan pembahasan dapat ditarik kesimpulan tinggi rendahnya deforestasi tahun 2017-2021 secara simultan dan parsial dipengaruhi oleh data luas tanaman perkebunan.

Model deforestasi di Indonesia dengan regresi *robust* yang terbentuk adalah:

$$Y = 7,527020 \times 10^2 - 5,620023 \times 10^{-4} X_1 + 7,852206 X_2 \quad (2.15)$$

SARAN

Temuan penelitian ini diharapkan dapat menjadi modal bagi peneliti lain untuk mengikuti analisis data dengan *outlier* menggunakan metode regresi *robust* dan metode estimasi *maximum likelihood type* (M). Setelah dicermati, diketahui tinggi rendahnya deforestasi tahun 2017-2021 dipengaruhi oleh luas tanaman perkebunan. Sebagai hasilnya, diharapkan para peneliti selanjutnya dapat menerapkan dan mengembangkan teknik regresi *robust* untuk studi kasus yang lebih besar dengan variabel independen yang lebih banyak dari penelitian ini.

DAFTAR PUSTAKA

- Chen, C. (2002). *Robust Regression and Outlier Detection with the Robustreg*. SUGI Proceedings: SAS institute Inc.
- Montgomery, D. C., Peck, E. A., & Vinning, G. G. (1982). *Introduction to Linear Regression Analysis*. New York: John Whilley and Sons Inc.
- Purba, D. S., Tarigan, W. J., Sinaga, M., & Tarigan, V. (2021). Pelatihan Penggunaan Software SPSS Dalam Pengolahan Regresi Linear Berganda Untuk Mahasiswa Fakultas Ekonomi Universitas Simalungun Di Masa Pandemi Covid 19. *Jurnal Karya Abdi*, 202-208.
- Ratnasari. (2017). *Pemodelan Konsumsi dan Estimasi Impor Daging Sapi untuk Keperluan Rumah Tangga di Indonesia Tahun 2016 Menggunakan Regresi Robust*. *Skripsi Sekolah Tinggi Ilmu Statistik*.
- Romdi, Wahyuningsih, S., & Yuniarti, D. (2015). Regresi Robust Linear Sederhana dengan Menggunakan Estimasi MM (Method of Moment). *Jurnal EKSPONENSIAL*, 179-186.

- Soemartini. (2007). Pencilan (Outlier). *Penelitian Universitas Padjajaran*.
- Susanti, Y., Pratiwi, H., H., S. S., & Liana, T. (2014). M Estimation, S Estimation, and MM Estimation in Robust Regression. *International Journal of Pure and Applied Mathematics*, 349-360.
- Verardi, V. (2008). *Robust Statistic In Stata*. Belgium: FNRS.
- Widodo, E., & Dewayanti, A. A. (2016). Perbandingan Metode Estimasi LTS, Estimasi M, dan stimasi MM Pada Regresi Robust. *Laporan Penelitian Universitas Islam Indonesia*.
- Wijaya, S. (2009). Taksiran Parameter pada Model Regresi Robust. *Skripsi Universitas Indonesia*.
- Wulandari, S. (2020). Perbandingan Model Regresi Robust dengan Estimasi Least Trimmed Square, Maximum Likelihood Type, dan Scale pada Data Outlier. *Skripsi Universitas Muhammadiyah Semarang*.