PEMODELAN ANALISIS KOMPONEN UTAMA *ROBUST* MENGGUNAKAN METODE *MINIMUM COVARIAN DETERMINANT-LEAST TRIMMED SQUARE* PADA DATA PRODUK DOMESTIK REGIONAL BRUTO

SKRIPSI

WAODE SITTI AMNI

H051191080

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

NOVEMBER 2023

PEMODELAN ANALISIS KOMPONEN UTAMA *ROBUST* MENGGUNAKAN METODE *MINIMUM COVARIAN DETERMINANT-LEAST TRIMMED SQUARE* PADA DATA PRODUK DOMESTIK REGIONAL BRUTO

SKRIPSI

Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Sains
Pada Program Studi Statistika Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Hasanuddin

WAODE SITTI AMNI
H051191080

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
NOVEMBER 2023

Universitas Hasanuddin

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

Pemodelan Analisis Komponen Utama *Robust* Menggunakan Metode *Minimum Covarian Determinant-Least Trimmed Square* Pada Data Produk Domestik Regional Bruto

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, 8 November 2023

METERAI TEMPEL 10000 7I5B3AJX781759131

WAODE SITTI AMNI
H051191080

PEMODELAN ANALISIS KOMPONEN UTAMA *ROBUST* MENGGUNAKAN METODE *MINIMUM COVARIAN DETERMINANT-LEAST TRIMMED SQUARE* PADA DATA PRODUK DOMESTIK REGIONAL BRUTO

Disetujui Oleh

Pembimbing Utama	Pembimbing Pertama
Andi Kresna Jaya, S.Si., M.Si.	**Dr. Nirwan, M.Si.**
NIP. 197312282000031001	**NIP. 196303061987021002**

Ketua Program Studi

Dr. Anna Islamiyati, S.Si., M.Si.
NIP. 197708082005012002

Pada 8 November 2023

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama	: Waode Sitti Amni
NIM	: H051191080
Program Studi	: Statistika
Judul Skripsi	: Pemodelan Analisis Komponen Utama *Robust* Menggunakan Metode *Minimum Covarian Determinant-Least Trimmed Square* Pada Data Produk Domestik Regional Bruto

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematikan dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

1. Ketua : Andi Kresna Jaya, S.Si., M.Si. (....................)
2. Sekretaris : Dr. Nirwan, M.Si. (....................)
3. Anggota : Dr. Anna Islamiyati, S.Si., M.Si. (....................)
4. Anggota : Sitti Sahriman, S.Si., M.Si. (....................)

Ditetapkan di : Makassar

Tanggal : 8 November 2023

KATA PENGANTAR

Dengan penuh rasa syukur, penulis mengucapkan terima kasih kepada Allah Subhanahu Wa Ta'ala atas segala anugerah, rahmat, dan petunjuk-Nya yang telah diberikan hingga saat ini, sehingga penulis dapat menyelesaikan tugas akhir ini. Shalawat serta salam senantiasa tercurahkan kepada Nabi Besar, Rasulullah Shallallahu ‘Alaihi Wa Sallam serta kepada keluarga dan para sahabatnya. Dengan nikmat kemudahan serta kekuatan yang diberikan oleh Allah Subhanahu Wa Ta'ala tugas akhir ini dapat terselesaikan dengan baik. Tugas akhir dengan judul "**Pemodelan Analisis Komponen Utama Robust Menggunakan Metode *Minimum Covarian Determinant-Least Trimmed Square* pada Data Produk Domestik Regional Bruto**" ini disusun sebagai salah satu syarat akademik untuk memperoleh gelar sarjana di Program Studi Statistika, Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin.

Dengan sepenuh hati, penulis menyadari bahwa penyelesaian tugas akhir ini tidak akan terwujud tanpa bantuan serta dorongan yang konsisten dari berbagai pihak yang telah memberikan dukungan moral dan materiil. Walaupun penulis memiliki keterbatasan dalam kemampuan dan pengetahuan, namun berkat bantuan serta dukungan yang diberikan, penulis berhasil menyelesaikan skripsi ini dengan baik. Oleh karena itu, penulis ingin mengungkapkan rasa terima kasih yang tak terhingga kepada almarhum ayahanda tercinta **Laode Resi** dan almarhumah ibunda tercinta **Haini** atas semua pengorbanan dan cinta yang telah kalian berikan. Terima kasih atas segala hal yang kalian berikan. Kenangan indah bersama kalian akan selalu ada dalam hati penulis. Terimakasih ayah dan ibu, penulis selalu merindukan kalian setiap hari. Ucapan terima kasih juga penulis haturkan kepada Kakak tersayang penulis **Waode Nur Karima** beliau penulis anggap sebagai pengganti kedua orang, terimakasi kepada kakak penulis tersayang yang senantiasa memberikan bantuan dan selalu memotivasi penulis dalam segala hal. Penulis ucapkan terima kasih yang sangat spesial kepada adik-adik tercinta penulis **Laode Muhammad Salamun**, **Laode Achmad Khatami**, dan **Waode Nur Aiba** yang telah menjadi sumber inspirasi utama dan motivasi yang luar biasa bagi penulis dalam menyelesaikan tugas akhir ini.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada:

1. Terima kasih kepada **Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.,** selaku Rektor Universitas Hasanuddin beserta seluruh staf jajarannya.
2. Terima kasih kepada Bapak **Dr. Eng. Amiruddin,** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh staf jajarannya.
3. Terima kasih kepada Ibu **Dr. Anna Islamiyati, S.Si., M.Si.,** sebagai Ketua Departemen Statistika dan sekaligus dosen penguji penulis yang telah bersedia meluangkan waktunya untuk memberikan penilaian dan masukan terhadap skripsi ini.
4. Terima kasih kepada Bapak **Andi Kresna Jaya, S.Si., M.Si.,** selaku Pembimbing Utama dan penasihat akademik penulis dengan penuh kesabaran telah meluangkan waktunya dan pemikiranya serta memberikan arahan, dorongan semangat dan motivasi selama dalam menyelesaikan skripsi.
5. Terima kasih kepada Bapak **Dr. Nirwan, M.Si.,** sebagai pembimbing pertama penulis yang dengan penuh kesabaran telah meluangkan waktu dan pikirannya untuk membimbing penulis dalam menyelesaikan skripsi dan memberikan dorongan semangat selama menjadi mahasiswa di Departemen Statistika.
6. Terima Kasih kepada Ibu **Sitti Sahriman S.Si., M.Si.,** sebagai dosen penguji yang telah bersedia meluangkan waktunya untuk memberikan penilaian dan masukan terhadap skripsi ini.
7. Terima kasih kepada segenap jajaran **Dosen Pengajar** dan **Staf Departemen Statistika** yang telah banyak membantu, memberikan ilmu-ilmunya, serta berbagai kemudahan lainnya yang diberikan selama menempuh pendidikan sarjana di Departemen Statistika.

8. Terima kasih kepada **Arif Rahman Nur** yang selalu memberikan dukungan dan mendampingi penulis dalam setiap langkah perjalanan ini dari suka hingga duka, selama proses perkuliahan di Universitas Hasanuddin.
9. Terimakasih kepada **Waode Ambrida** dan **Laode Zakahar** yang telah menjadi orang tua penulis selama di Makassar yang selalu memberi semangat dalam menyelesaikan Pendidikan di Universitas Hasanuddin.
10. Terima kasih kepada **Refa Joyce Semida** selaku sahabat terbaik yang telah memberikan masukan dan bantuan dalam menyelesaikan perkuliahan.
11. Terimakasi kepada **Dian Ayu Permata Rusdy**, **Muhammad Faturrahman** dan **Sapriyadi Rasyid** yang telah memberikan bantuan dan dukungannya selama mengerjakan skripsi.
12. Terima kasih kepada sahabat penulis semasa SMA yang telah memberikan semangat dan motivasi dalam menyelesaikan skripsi ini. Terima kasih: **Pitri, Windry, Intan T, Intan N, Devi, Dinar, Devira dan Risma.**
13. Terima kasih kepada sepupu tercinta penulis, **Laode Muhammad Samiruddin**, **Laode Muhammad Samtun**, **Olivia Sukma Saputri**, dan **Waode Salsa Dewanti** terima kasih atas semangat dan bantuan yang diberikan kepada penulis.
14. Terima kasih kepada semua teman-teman **Angkatan Statistika 2019** yang telah menerima penulis yang memiliki banyak kekurangan ini dengan tulus.
15. Terima kasih yang setinggi-tingginya kepada **seluruh pihak** yang mungkin tidak sempat penulis sebutkan satu persatu. Terima kasih atas segala dukungan, partisipasi, dan apresiasinya yang diberikan kepada penulis.

Penulis juga menyadari bahwa skripsi ini masih jauh dari kata sempurna, namun ini hasil terbaik yang dapat diberikan oleh penulis pada penelitian ini. Oleh karena itu dengan segala kerendahan hati penulis mengucapkan permohonan maaf yang sebesar-besarnya. Akhir kata, semoga tulisan ini dapat memberikan manfaat untuk berbagai pihak.

Makassar, 8 November 2023

WAODE SITTI AMNI
NIM. H05119108

Universitas Hasanuddin

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIK

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Waode Sitti Amni

NIM : H051191080

Program Studi : Statistika

Departemen : Statistika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty- Free Right)** atas tugas akhir saya yang berjudul:

"Pemodelan Analisis Komponen Utama *Robust* Menggunakan Metode *Minimum Covarian Determinant-Least Trimmed Square* Pada Data Produk Domestik Regional Bruto"

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (database), merawat dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar, 8 November 2023

Yang Menyatakan,

WAODE SITTI AMNI

NIM. H051191080

ABSTRAK

Analisis regresi merupakan salah satu metode analisis data yang digunakan untuk mempelajari hubungan antara dua atau lebih variabel dalam menentukan pengaruh antara variabel tersebut. Dalam analisis regresi linear sering kali muncul masalah multikolinearitas antara variabel prediktor. Hal ini dapat menyebabkan masalah interpretasi yang sulit dan mengurangi keandalan hasil estimasi koefisien regresi. Salah satu metode yang efektif untuk menangani multikolineritas adalah analsisi komponen utama (AKU). AKU klasik memiliki kelemahan ketika data mengandung *outlier* karena vektor rata-rata dan matriks kovarian atau korelasi sampel sangat sensitif terhadap *outlier*. Begitu pula dengan metode *ordinary least square* (OLS) pada regresi komponen utama (RKU) dapat menghasilkan estimasi dari parameter regresi yang bias ketika terdapat *outlier* dalam data. Oleh karena itu akan dikembangkan menjadi regresi *robust* yang menerapkan metode *robust* pada kedua tahap tersebut. Regresi *robust* merupakan metode regresi yang digunakan pada saat distribusi dari galat tidak normal dalam model. Metode *robust* yang digunakan dalam penelitian ini adalah metode *minimum covariance determinant* (MCD) pada saat melakukan AKU dan metode estimasi *least trimmed square* (LTS) pada saat melakukan RKU. Pada penelitian ini, produk domestik regional bruto di Indonesia tahun 2020 akan dimodelkan dengan metode AKU *robust* menggunakan MCD-LTS berdasarkan variable jumlah produk domestik regional bruto, panjang jalan, distribusi listrik, infrastruktur kesehatan, infrastruktur pendidikan, infrastruktur pariwisata, infrastruktur perumahan, dan fasilitas industri. Salah satu ukuran untuk melihat kebaikan model regresi adalah dengan menggunakan nilai *adjusted* R^2 dan nilai MSE. Model regresi AKU *robust menggunakan MCD-LTS* yang telah diuji memiliki nilai *adjusted* R^2 sebesar 99.03% dan nilai MSE sebesar 0,0226%.

Kata kunci: *adjusted* R^2, AKU, LTS, MCD, MSE, Multikolinearitas, *Outlier*, PDRB

ABSTRACT

Regression analysis is a data analysis method used to study the relationship between two or more variables in determining the influence between these variables. In linear regression analysis, multicollinearity problems often arise between predictor variables. This can cause difficult interpretation problems and reduce the reliability of the regression coefficient estimation results. One effective method for dealing with multicollinearity is principal component analysis (AKU). Classical AKU has a weakness when the data contains outliers because the mean vector and sample covariance or correlation matrix are very sensitive to outliers. Likewise, the ordinary least squares (OLS) method in principal component regression (RKU) can produce biased estimators of regression parameters when there are outliers in the data. Therefore, it will be developed into a robust regression that applies robust methods at both stages. Robust regression is a regression method used when the distribution of errors is not normal in the model. The robust method used in this research is the minimum covariance determinant (MCD) method when carrying out AKU and the least trimmed square (LTS) estimation method when carrying out RKU. In this research, gross regional domestic product in Indonesia in 2020 will be modeled using the AKU robust method using MCD-LTS based on variables total gross regional domestic product, road length, electricity distribution, health infrastructure, education infrastructure, tourism infrastructure, housing infrastructure and facilities. industry. One measure to see the goodness of the regression model is the adjusted R^2*value and the MSE value. The AKU robust regression model using MCD-LTS which has been tested has an adjusted* R^2 *value of 99.03% and an MSE value of 0.0226%.*

Keywords*: adjusted* R^2*, AKU, LTS, MCD, MSE, Multicollinearity, Outliers, PDRB.*

DAFTAR ISI

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB 1
PENDAHULUAN

1.1 Latar Belakang

Produk domestrik regional bruto (PDRB) adalah salah satu indikator penting untuk mengetahui kondisi ekonomi disuatu daerah dalam satu periode tertentu, baik atas dasar harga berlaku maupun atas dasar harga konstan. Pada dasarnya PDRB merupakan jumlah nilai tambah yang dihasilkan oleh seluruh unit usaha dalam suatu daerah tertentu, atau merupakan jumlah nilai barang dan jasa akhir (netto) yang dihasilkan oleh seluruh unit ekonomi (Putri, 2020). Menurut Kusuma (2019) batasan-batasan infrastruktur terbagi atas tiga bagian. Pertama, infrastruktur ekonomi berupa tenaga listrik, telekomunikasi, air, sanitasi, gas, jalan, bendungan, jembatan, kanal, irigasi, drainase, pusat bisnis dan perdagangan, sektor transportasi, dan fasilitas indutri. Kedua infrastruktur sosial seperti pendidikan, kesehatan, perumahan, dan rekreasi. Dan ketiga, infrastruktur administrasi berupa penegakan hukum, kontrol administrasi dan koordinasi. Oleh karena itu, untuk menentukan faktor-faktor yang berpengaruh terhadap PDRB akan dilakukan analisis regresi.

Analisis regresi merupakan salah satu metode analisis data yang digunakan untuk mempelajari hubungan antara dua atau lebih variabel dan menentukan pengaruh antara variabel tersebut (Subandriyo, 2020). Apabila hubungan yang diselidiki terdiri dari satu variabel respon dan satu variabel prediktor maka disebut dengan analisis regresi linear sederhana. Sedangkan analisis regresi linear berganda terdiri dari satu variabel respon dan lebih dari satu variabel prediktor (Padilah dan Adam, 2019). Pada metode regresi linear berganda terdapat beberapa asumsi yang harus terpenuhi untuk mendapakan model yang baik, diantaranya adalah asumsi kenormalan, homoskedastisitas, tidak adanya autokorelasi dan tidak adanya multikolinearitas (Shodiqin dkk., 2018).

Dalam analisis regresi linear berganda, sering kali muncul masalah multikolinearitas antara variabel prediktor. Multikolinearitas terjadi ketika terdapat korelasi yang kuat antara variabel prediktor (Larasati dkk., 2020). Metode statistika yang dapat digunakan untuk mengatasi pengaruh multikolinearitas, seperti mengumpulkan data tambahan, perbandingan dan evaluasi estimator, spesifikasi model dengan menghapus suatu variabel yang berkorelasi, metode regresi *ridge* dan metode regresi analisis komponen utama (AKU) (Montgomery dan Peck, 2012).

AKU adalah suatu teknik analisis statistik untuk menyelesaikan pelanggaran asumsi multikolinearitas yang terjadi pada regresi linear berganda. Metode AKU mengatasi pelanggaran asumsi multikolinearitas dengan cara mereduksi data berdimensi besar, yakni mentransformasikan variabel yang saling berkorelasi dari satu variabel dengan variabel lain yang akan menghasilkan variabel baru yang bebas dari korelasi. Variabel baru yang terbentuk dinamakan dengan *principal componen* (Pendi, 2021). Terbentuknya metode AKU memiliki dua langkah yaitu, pertama *principal componen* tebentuk berdasarkan matriks varian kovarian atau matriks korelasi variabel prediktor dengan memanfaatkan vektor eigen dan yang kedua *principal component* yang terpililih akan diregresikan dengan variabel respon menggunakan regresi komponen utama (RKU) dengan metode *ordinary least square* (OLS) (Fabiana, 2019).

AKU klasik memiliki kelemahan ketika data mengandung *outlier* karena vektor rata-rata dan matriks kovarian atau korelasi sampel sangat sensitif terhadap *outlier*. Begitu pula dengan metode OLS pada analisis RKU yang dapat menghasilkan estimator dari parameter regresi menjadi tidak akurat dan tidak efisien ketika terdapat *outlier* dalam data (Filzmoser, 1999). Oleh karena itu akan dikembangkan menjadi regresi *robust* yang menerapkan metode *robust* pada kedua tahap tersebut. Regresi *robust* merupakan metode regresi yang digunakan pada saat distribusi dari *error* yang tidak normal dalam model. Metode *robust* yang digunakan dalam penelitian ini adalah metode *minimum covariance determinant* (MCD) pada saat melakukan AKU dan metode estimasi *least trimmed square* (LTS) pada saat melakukan analisis RKU.

Metode MCD digunakan karena efisien dalam perhitungan statistik, serta keandalannya dalam menentukan lokasi dan sebaran multivariat. Metode MCD mempertimbangkan sebaran data dan memberikan bobot yang lebih besar pada data yang memiliki sebaran yang lebih kecil, sehingga lebih tahan terhadap pengaruh data *outlier* (Rousseeuw dan Driessen, 1999). Serta Estimasi LTS dipilih karena menggunakan algoritma yang relatif lebih sederhana dibandingkan metode lain, sehingga lebih mudah diimplementasikan. Prinsip dasar dari estimasi LTS adalah meminimalkan jumlah kuadrat *error* yang dihasilkan (Dea dan Susanti, 2021).

Susilowati dkk (2020) menjelaskan tentang metode AKU *robust* dan metode *clustering large area* (Clara) dalam meminimalisisr *outlier* pada sebuah data tentang indeks kebahagiaan dunia tahun 2018. Dari hasil yang diperoleh bahwasanya metode AKU *robust* lebih efisien dalam meminimalisisr kasus *outlier* dari pada metode AKU klasik. Selain itu pada penelitian Siburian (2019) AKU *robust* juga diterapkan pada data jumlah pengangguran di Jawa Tengah menggunakan estimasi-S yang menunjukkan hasil bahwa model yang lebih baik untuk menjelaskan jumlah pengangguran Jawa Tengah adalah model RKU *robust*. Penelitian selanjutnya mengenai perbandingan AKU klasik dengan AKU *robust* pada data pengeluaran perkapita, dapat disimpulkan bahwa metode AKU *robust* memberikan performa yang lebih baik dari pada AKU klasik, hal ini dapat dilihat model AKU *robust* mampu menghasilkan jumlah komponen utama yang lebih sedikit daripada variabel asalnya (Indra dan Sihombing, 2021). Berdasarkan penelitian terdahulu dapat disumpulkan bahwa metode AKU robust dapat meminimalisir outlier yang terjadi pada data, sehingga pada penelitian ini akan menggunakan AKU *robust* dalam menangani *outlier*.

Adanya beberapa faktor yang saling berkaitan dan berpengaruh terhadap distribusi listrik, infrastruktur kesehatan, infrastruktur pendidikan, infrastruktur pariwisata dan vasilitas industri, dapat memungkinkan terjadi korelasi yang tinggi diantara variabel tersebut sehingga dapat menyebabakan terjadinya multikolinearitas. Kemudian jumlah infrastruktur di Indonesia yang terdapat disetiap provinsi berdasarkan atas harga berlaku maupun atas dasar harga konstan memiliki jumlah nilai yang berbeda disetiap provinsi sehingga dapat menyebabkan terjadinya *outlier* pada data yang diperoleh. Berdasarkan hal tersebut, maka penelitian ini menuliskan cara menghilangkan multikolinearitas dan *outlier* menggunakan AKU *robust* metode yang digunakan pada regresi *robust* ini yaitu metode MCD yang akan diterapkan pada analisis AKU dan metode LTS yang akan diterapkan pada analisis RKU. Hasil penelitian ini dapat menjadi acuan bagi pemerintah maupun masyarakat dalam upaya meningkatkan pertumbuhan ekonomi di Indonesia.

1.2 Rumusan Masalah

Rumusan masalah yang dibahas pada penelitian ini adalah sebagai berikut:

1. Bagaimana kinerja analisis komponen utama *robust* menggunakan metode MCD dengan estimasi LTS dalam mengidentifikasi dan meminimalisir data *outliers*?
2. Bagaimana model analisis komponen utama *robust* menggunakan metode MCD dengan estimasi LTS pada kasus PDRB di Indonesia?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah data kasus PDRB di Indonesia tahun 2020.
2. Penentuan *outlier* pada variabel prediktor menggunakan nilai jarak mahalanobis.
3. Penentuan *outlier* pada variabel repson menggunakan nilai DFFITS.
4. Penentuan model regresi terbaik berdasarkan nilai MSE nilai $adjusted\ R^2$.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Dapat mengetahui kinerja analisis komponen utama *robust* menggunakan metode MCD dengan estimasi LTS dalam mengidentifikasi dan meminimalisir data *outliers*.
2. Mendapatkan model komponen utama *robust* menggunakan metode MCD dengan estimator LTS pada kasus PDRB di Indonesia.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari hasil penelitian ini adalah sebagai berikut:

1. Sebagai sumber informasi terkait faktor yang mempengaruhi PDRB di Indonesia sehingga dapat menjadi acuan bagi pemerintah maupun masyarakat untuk meningkatkan pertumbuhan ekonomi di indonesia.
2. Memberikan informasi mengenai cara MCD dalam mengatasi *oulier* pada data sehingga memberikan perkiraan komponen utama yang lebih stabil dan lebih baik yang mewakili variabilitas data.
3. Memberikan informasi mengenai cara LTS dalam mengatasi *outlier* pada metode RKU sehingga memberikan model regresi yang lebih andal dan akurat.

BAB II
TINJAUAN PUSTAKA

2.1 Analisis Regresi Berganda

Regresi linear berganda adalah metode statistik yang digunakan untuk memodelkan hubungan antara satu variabel respon dan dua atau lebih variabel prediktor. Analisis ini bertujuan untuk menentukan arah hubungan antara variabel-variabel tersebut, masing-masing variabel prediktor bisa berhubungan positif atau negatif, serta untuk memprediksi nilai variabel respon yang mengalami kenaikan atau penurunan berdasarkan perubahan pada variabel prediktor (Adiguno dkk., 2022). Dengan bertambahnya variabel prediktor maka bentuk umum dari persamaan regresi linear berganda yang mencakup dua atau lebih variabel prediktor dapat ditulis sebagai berikut: (Ningsih dan Dukalang, 2019).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon \quad (2.1)$$

Berdasarkan Persamaan (2.1) maka bentuk matriks dari model regresi linear berganda dapat dituliskan sebagai berikut:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Atau dapat ditulis dengan $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, dengan:

$\boldsymbol{Y}$ adalah vektor pengamatan berukuran $n \times 1$

$\boldsymbol{X}$ adalah matriks variabel prediktor ukuran $n \times (p+1)$

$\boldsymbol{\beta}$ adalah vektor parameter yang akan ditaksir berukuran $(p+1) \times 1$

$\boldsymbol{\varepsilon}$ adalah vektor random *error* berukuran $n \times 1$

2.2 Estimasi Parameter Model

Menurut Montgomery dan Peck (2012) estimasi parameter dapat diperoleh dengan menggunakan metode *ordinary least squares* (OLS) yaitu dengan meminimumkan jumlah kuadrat *error* sesuai persamaan (2.2) berikut:

$$S(\beta) = \Sigma_{i=1}^{n} \varepsilon_i^2 = \varepsilon^T \varepsilon = Y^T Y - 2\beta^T X^T Y + \beta^T X^T X \beta \quad (2.2)$$

Untuk mendapatkan estimator OLS $(\hat{\beta})$ yang meminimumkan $S(\beta)$ disyaratkan bahwa:

$$\left.\frac{\partial S(\beta)}{\partial \beta}\right|_{\beta = \hat{\beta}} = 0$$

Maka

$$\left.\frac{\partial S(\beta)}{\partial \beta}\right|_{\beta=\hat{\beta}} = 0$$

$$-2X^TY + 2X^TX\hat{\beta} = 0$$

$$X^TX\hat{\beta} = X^Ty$$

$$\hat{\beta} = (X^TX)^{-1}(X^TY) \quad (2.3)$$

Perbedaan unit satuan pada model regresi yang tidak distandarkan dapat menyebabkan koefisien regresi tidak bisa dibandingkan. Oleh karena itu, perlu dilakukan standarisasi menggunakan rumus sebagai berikut:

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_y} \text{ dan } X_i^* = \frac{X_{ji} - \bar{X}_j}{S_j} \quad (2.4)$$

Dengan $S_y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}}$ dan $S_j = \sqrt{\frac{\sum_{i=1}^{n}(Y_{ji} - \bar{X}_j)^2}{n-1}}$

Keterangan:

y_i^* : Y standarisasi

x_i^* : X standarisasi

S_y : Standar deviasi Y

S_j : Standar deviasi X

Sehingga diperoleh model regresi standar sebagai berikut:

$$Y_1^* = \beta_1^* X_{1i}^* + \beta_2^* X_{2i}^* + \cdots + \beta_k^* X_{ki}^* + \varepsilon_i^*$$

2.3 Pengujian Model Regresi

Menurut Montgomery dan Peck (2012), dalam analisis regresi berganda ada beberapa uji signifikansi yang berguna untuk mengukur ketepatan model, antara lain sebagai berikut:

a. Uji Signifikansi Model Regresi

Uji ini dimaksudkan untuk menentukan apakah ada hubungan linier antara variabel respon Y dengan variabel prediktor $X_1, X_2, \dots, X_p$ atau tidak. Adanya hubungan linier antara variabel respon dengan variabel prediktor menandakan bahwa model regresi yang terbentuk sesuai. Berikut langkah-langkahnya:

Hipotesis:

H_0: $\beta_1 = \beta_2 = \cdots = \beta_p = 0$ (model regresi tidak sesuai)

H_1: ada minimal satu $\beta_j \neq 0$, dengan $j = 1, 2, \dots, p$ (model regresi sesuai)

Statistik uji:

$$F_0 = \frac{JKR/p}{JKE/(n-p-1)}, JKR = \Sigma_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2;\ JKE = \Sigma_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \qquad (2.5)$$

Kriteria uji:

H_0 ditolak jika $F_0 > F_{tabel} = F_{(a;p;n-p-1)}$ atau $p - value < a$

b. Uji Signifikansi Koefisien Regresi Secara Individu

Uji ini digunakan untuk menguji ada tidaknya pengaruh masing-masing variabel prediktor terhadap variabel respon. Adapun langkah-langkahnya sebagai berikut:

Hipotesis:

$H_0: \beta_1 = 0$ (koefisien regresi X_j tidak signifikan)

$H_1: \beta_j \neq 0$, dengan $j = 1, 2, \dots, p$ (koefisien regresi X_j signifikan)

Statistika uji:

$$t_0 = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \text{ dengan } se(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 C_{jj}} \qquad (2.6)$$

Dengan C_{jj} adalah elemen diagonal dari $(X^T X)^{-1}$ dan $\hat{\sigma}^2 = \frac{\Sigma_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-p-1}$

Kriteria uji:

H_0 ditolak jika $|t_0| > t_{tabel} = t_{(\frac{a}{2};n-p-1)}$ atau $p - value < a$

2.4 Uji Ukuran Kebaikan Model Regresi

Metode yang digunakan untuk menentukan ukuran kebaikan model dilakukan dengan menghitung nilai *Mean Square Error* (MSE). Perhitungan MSE dilakukan menggunakan persamaan berikut: (Montgomery dan Peck., 2012)

$$MSE = \frac{JKE}{n-p-1} \qquad (2.7)$$

dengan JKE adalah jumlah kuadrat residual, n adalah jumlah sampel, dan p adalah banyaknya variabel prediktor. Jika nilai MSE semakin kecil hingga mendekati nol maka dapat dikatakan bahwa model regresi semakin baik.

Selain metode tersebut, untuk menentukan ukuran kebaikan model juga dapat menggunakan koefisien determinasi *adjusted* R^2. Kecocokan model lebih baik jika nilai *adjusted* R^2 semakin mendekati satu (Gujarati, 2003). Adapun rumus untuk mendapatkan nilai *adjusted* R^2 sebagai berikut:

$$R_{adj}^2 = 1 - \frac{\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2/(n-p-1)}{\Sigma_{i=1}^{n}(y_i - \bar{y})^2/(n-1)} \qquad (2.8)$$

dengan $0 \leq adjusted\, R^2 \leq 1$, apabila $adjusted\ \ R^2 = 0$ artinya tidak ada hubungan di antara X dan Y atau secara model regresi yang terbentuk tidak tepat untuk medefinisikan Y dan apabila nilai $adjusted\, R^2 = 1$ artinya garis regresi yang terbentuk dapat mendefinisikan Y secara sempurna.

2.5 Multikolinearitas

Multikolinearitas terjadi ketika terdapat keterkaitan linear yang signifikan antara variabel prediktor dalam model regresi berganda. Umumnya, keberadaan multikolinearitas tidak diinginkan dalam model regresi yang baik. Multikolinearitas dapat mengakibatkan beberapa masalah serius dalam analisis regresi. Hubungan linear yang kuat antara variabel prediktor dapat menyebabkan estimasi parameter menjadi tidak stabil dan sulit untuk diinterpretasikan dengan baik (Maubanu, 2018). Multikolinearitas dalam model regresi linear berganda menghasilkan kondisi yang kurang baik (*ill condition*) atau mendekati singularitas pada matriks $\boldsymbol{X^T X}$. Akibatnya, estimasi varian parameter regresi menjadi lebih tinggi, dan terdapat sedikit variabel prediktor yang signifikan atau bahkan tidak signifikan.

Menurut Montgomery dan Peck (2012) cara mendeteksi adanya multikolinearitas pada model regresi dapat dilakukan dengan menggunakan nilai *variance inflation factory* (VIF). VIF yaitu faktor perubahan variansi dalam variabel prediktor ke-i. Pada VIF besarnya nilai bergantung pada nilai koefisian determinasi R^2 yang dihasilkan. Berikut persamaan yang digunakan dalam menghitung VIF:

$$VIF = \frac{1}{(1-R_j^2)} = \frac{1}{Tolerance} \tag{2.9}$$

$$R_{adj}^2 = 1 - \frac{JKE/(n-k-1)}{JKR/(n-1)}$$

Keterangan:

JKE = Jumlah kuadrat *error*

JKR = Jumlah kuadrat regresi

Dengan R_j^2 adalah koefisien determinasi ke-j, di mana $j = 1,2,3, \dots, p$. *Tolerance* dalam perhitungan VIF memiliki ketentuan semakin rendah nilai yang diperolehnya maka semakin besar kemungkinan terjadinya multikolinearitas antar variabel. Batas nilai yang digunakan untuk VIF adalah 10 dan batas nilai dari *tolerance* adalah 0.1. Apabila nilai yang diperoleh $VIF \geq 10$ dan $tolerance <$

0.1 maka terjadi multikolinearitas yang kuat diatara variabel prediktor dan sebaliknya.

2.6 Outlier

Outlier adalah data yang tidak mengikuti pola yang umumnya diikuti oleh model atau data yang berada di luar model yang dibentuk, serta tidak berada dalam daerah selang kepercayaan yang telah ditentukan. (Neter dkk., 1988). Keberadaan *outlier* dalam data dapat mengganggu asumsi-asumsi dasar regresi, seperti asumsi normalitas *error*, homoskedastisitas dan juga dapat memiliki dampak signifikan pada hasil estimasi model regresi.

Adanya outlier pada variabel prediktor dapat dideteksi dengan menghitung jarak mahalanobis. Untuk mengukur jarak mahalanobis digunakan vektor rata-rata dan matriks kovarian. Sebuah pengamatan X_i dideteksi sebagai *outlier* jika jarak mahalanobisnya:

$$d_{MD}^2 = (X_i - \bar{X})^T S^{-1} > X_{(p;a)}^2 \quad (2.10)$$

dengan $\bar{X}$ dan S adalah vector rata-rata dan matriks kovarian dari data.

Selanjutnya, untuk mendeteksi adanya outlier pada variabel respon dapat dilakukan dengan melihat *error* dari model regresi. Salah satu metode yang dapat digunakan adalah metode DFFITS (*Difference in Fit Standardized*). Perhitungan DFFITS dapat ditulis sebagai berikut (Montgomery dan Peck, 2012):

$$DFFITS_i = t_i \left(\frac{h_i}{1-h_i}\right)^{1/2} \quad (2.11)$$

dengan

$$t_i = \varepsilon_i \sqrt{\frac{n-p-1}{JKE(1-h_i) - \varepsilon_i^2}}$$

Keterangan:

t_i = *Studentized deleted* error untuk kasus ke-i,

ε_i = *Error* ke-i

h_i = Elemen baris ke-i kolom ke-i dari matriks $\hat{\beta} = X(X^T X)^{-1} X^T$

JKE = Jumlah kuadrat *error*

Sebuah data bisa dikatakan *outlier* apabila $|DEFITS| > 2\sqrt{\frac{k}{n}}$ dengan $k = p + 1$ adalah banyaknya variabel data dan n adalah banyaknya jumlah observasi.

2.7 Matriks

2.7.1 Matriks Varian-Kovarian dan Matriks Korelasi

Vektor mean sampel dari matriks varians-kovarians sampel dapat diperoleh dengan cara sebagai berikut. Nilai varian-kovarian dapat diperoleh dengan mengikuti persamaan sebagai berikut (Pratiwi dkk., 2020):

$$[\boldsymbol{X}]_{\boldsymbol{n}\times\boldsymbol{p}}$$

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2p} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{np} & x_{np} & x_{np} & \cdots & x_{np} \end{bmatrix}$$

Matriks rata-rata peubah prediktor:

$$[\overline{\boldsymbol{X}}]_{\boldsymbol{n}\times\boldsymbol{p}}$$

$$\overline{\boldsymbol{X}} = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \bar{x}_3 & \cdots & \bar{x}_p \\ \bar{x}_1 & \bar{x}_2 & \bar{x}_3 & \cdots & \bar{x}_p \\ \bar{x}_1 & \bar{x}_2 & \bar{x}_3 & \cdots & \bar{x}_p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{x}_1 & \bar{x}_2 & \bar{x}_3 & \cdots & \bar{x}_p \end{bmatrix}$$

Selisih antara matriks peubah prediktor dan matriks rata-rata peubah prediktor:

$$[\boldsymbol{X} - \overline{\boldsymbol{X}}]_{\boldsymbol{n}\times\boldsymbol{p}}$$

$$\boldsymbol{X} - \overline{\boldsymbol{X}} = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & x_{13} - \bar{x}_3 & \dots & x_{1p} - \bar{x}_p \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & x_{23} - \bar{x}_3 & \dots & x_{2p} - \bar{x}_p \\ x_{31} - \bar{x}_1 & x_{32} - \bar{x}_2 & x_{33} - \bar{x}_3 & \cdots & x_{3p} - \bar{x}_p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & x_{n3} - \bar{x}_3 & \cdots & x_{np} - \bar{x}_p \end{bmatrix}$$

Transpose matriks peubah prediktor dan matriks rata-rata peubah prediktor:

$$[\boldsymbol{X} - \overline{\boldsymbol{X}}^{\boldsymbol{T}}]_{\boldsymbol{p}\times\boldsymbol{n}}$$

$$(\boldsymbol{X} - \overline{\boldsymbol{X}})^{\boldsymbol{T}} = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{21} - \bar{x}_1 & x_{31} - \bar{x}_1 & \dots & x_{n1} - \bar{x}_1 \\ x_{12} - \bar{x}_2 & x_{22} - \bar{x}_2 & x_{32} - \bar{x}_2 & \dots & x_{n2} - \bar{x}_2 \\ x_{13} - \bar{x}_3 & x_{23} - \bar{x}_3 & x_{33} - \bar{x}_3 & \cdots & x_{n3} - \bar{x}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1p} - \bar{x}_p & x_{2p} - \bar{x}_p & x_{3p} - \bar{x}_p & \cdots & x_{np} - \bar{x}_p \end{bmatrix}$$

Persamaan matriks varian-kovarian adalah perkalian antara matriks $(\boldsymbol{X} - \overline{\boldsymbol{X}})^{\boldsymbol{T}}$ dengan matriks $(\boldsymbol{X} - \overline{\boldsymbol{X}})$:

$$\boldsymbol{S} = \frac{1}{n-1}[(X - \bar{X})^T(X - \bar{X})] \tag{2.12}$$

$$p \times p \qquad p \times n \qquad n \times p$$

$$\boldsymbol{S} = \frac{1}{n-1} = \begin{bmatrix} x_{11}-\bar{x}_1 & x_{21}-\bar{x}_1 & x_{31}-\bar{x}_1 & \dots & x_{n1}-\bar{x}_1 \\ x_{12}-\bar{x}_2 & x_{22}-\bar{x}_2 & x_{32}-\bar{x}_2 & \dots & x_{n2}-\bar{x}_2 \\ x_{13}-\bar{x}_3 & x_{23}-\bar{x}_3 & x_{33}-\bar{x}_3 & \cdots & x_{n3}-\bar{x}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1p}-\bar{x}_p & x_{2p}-\bar{x}_p & x_{3p}-\bar{x}_p & \cdots & x_{np}-\bar{x}_p \end{bmatrix}$$

$$\times \begin{bmatrix} x_{11}-\bar{x}_1 & x_{12}-\bar{x}_2 & x_{13}-\bar{x}_3 & \dots & x_{1p}-\bar{x}_p \\ x_{21}-\bar{x}_1 & x_{22}-\bar{x}_2 & x_{23}-\bar{x}_3 & \dots & x_{2p}-\bar{x}_p \\ x_{31}-\bar{x}_1 & x_{32}-\bar{x}_2 & x_{33}-\bar{x}_3 & \cdots & x_{3p}-\bar{x}_p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1}-\bar{x}_1 & x_{n2}-\bar{x}_2 & x_{n3}-\bar{x}_3 & \cdots & x_{np}-\bar{x}_p \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{n-1}\Sigma_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{i1}-\bar{x}_1) & \frac{1}{n-1}\Sigma_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2) & & \frac{1}{n-1}\Sigma_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{ip}-\bar{x}_p) \\ \frac{1}{n-1}\Sigma_{i=1}^{n}(x_{i2}-\bar{x}_2)(x_{i1}-\bar{x}_1) & \frac{1}{n-1}\Sigma_{i=1}^{n}(x_{i2}-\bar{x}_2)(x_{i2}-\bar{x}_2) & \cdots & \frac{1}{n-1}\Sigma_{i=1}^{n}(x_{i2}-\bar{x}_2)(x_{ip}-\bar{x}_p) \\ \frac{1}{n-1}\Sigma_{i=1}^{n}(x_{i3}-\bar{x}_3)(x_{i1}-\bar{x}_1) & \frac{1}{n-1}\Sigma_{i=1}^{n}(x_{i3}-\bar{x}_3)(x_{i2}-\bar{x}_2) & \ddots & \frac{1}{n-1}\Sigma_{i=1}^{n}(x_{i3}-\bar{x}_3)(x_{ip}-\bar{x}_p) \\ \vdots & \vdots & \cdots & \vdots \\ \frac{1}{n-1}\Sigma_{i=1}^{n}(x_{ip}-\bar{x}_p)(x_{i1}-\bar{x}_1) & \frac{1}{n-1}\Sigma_{i=1}^{n}(x_{ip}-\bar{x}_p)(x_{i2}-\bar{x}_2) & & \frac{1}{n-1}\Sigma_{i=1}^{n}(x_{ip}-\bar{x}_p)(x_{ip}-\bar{x}_p) \end{bmatrix}$$

$$\boldsymbol{S} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

Matriks koefisien korelasi adalah matriks simetri berukuran $p \times p$ yang dapat dinyatakan dalam persamaan berikut ini:

$$\boldsymbol{S_{pp}} = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{21}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{11}}} & \cdots & \frac{\sigma_{p1}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{11}}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{p2}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{22}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \end{bmatrix}$$

$$\boldsymbol{S_{pp}} = \begin{bmatrix} 1 & \rho_{21} & \cdots & \rho_{p1} \\ \rho_{12} & 1 & \cdots & \rho_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix} \tag{2.13}$$

2.7.2 Nilai Eigen dan Vektor Eigen

Misalkan matriks **S** adalah matriks berukuran $p \times p$, terdapat suatu skalar λ vektor tak nol v sehingga memenuhi persamaan berikut (Kosasih, 2021)

$$Sa = \lambda a \tag{2.14}$$

Skalar λ disebut nilai eigen dari $\boldsymbol{S}$ dan a disebut vektor eigen dari $\boldsymbol{S}$ yang bersesuaian dengan λ. Untuk memperoleh nilai eigen pada persamaan (2.14) dapat dituliskan menjadi:

$$Sa = \lambda a \text{ ; dengan } v \neq 0$$
$$Sa - \lambda a = 0,$$
$$Sa - \lambda I = 0,$$
$$(S - \lambda I)a = 0.$$

Agar λ menjadi nilai eigen, maka harus terdapat solusi tak nol dari persamaan (2.15). Persamaan tersebut akan memiliki penyelesaian tak nol jika dan hanya jika:

$$|S - \lambda I| = 0 \tag{2.15}$$

2.8 Analisis Komponen Utama

Analisis komponen utama (AKU) adalah suatu teknik analisis statistik yang digunakan untuk mentransformasi sekumpulan variabel yang saling berkorelasi menjadi satu set variabel baru yang bebas dari korelasi (Sriningsih dkk., 2018). Secara umum, tujuan dari AKU adalah melakukan reduksi dimensi data. AKU memberikan solusi untuk proses pengumpulan data yang terdiri dari banyak variabel, dengan menghasilkan variabel baru yang jumlahnya lebih sedikit namun tetap mampu menjelaskan keragaman data yang ada. (Johnson dan Wichern, 2007).

Menurut Larasati dkk (2020) proses penyelesaian masalah multikolinearitas dalam model regresi menggunakan metode AKU dengan menggunakan vektor eigen dari matriks varian-kovarian. Komponen utama tersebut kemudian diregresikan terhadap variabel respon menggunakan metode *least square*. Komponen utama merupakan suatu kombinasi linear berdasarkan pada skala pengukuran variabel acak $X_1, X_2, \dots, X_p$ yang sama dan memiliki struktur matriks varian kovarian $\boldsymbol{S}$ yang kemudian akan dihasilkan nilai $eigen$ λ sebanyak p ($\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_p$). Maka komponen utama yang merupakan kombinasi linear dari X^T variabel asal dan e sebagai vektor eigen didefinisikan sebagai berikut:

$$\begin{aligned} w_1 &= a_1^T X = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\ w_2 &= a_2^T X = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\ &\vdots \\ w_p &= a_p^T X = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p \end{aligned} \tag{2.16}$$

Dengan matriks sebagai berikut:

$$\boldsymbol{a_j^T X} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

dengan w_1 adalah komponen pertama yang memenuhi maksimum nilai $a_1^T S a_1 = \lambda_1$, w_2 adalah komponen kedua yang memenuhi sisa keragaman selain komponen kedua dengan memaksimumkan nilai $a_2^T S a_2 = \lambda_2$, dan w_p adalah komponen ke p yang memenuhi sisa keragaman selain dari $w_1, w_2, \dots, w_p$ dengan memaksimumkan nilai $a_j^T S a_j = \lambda_j$. Syarat untuk membentuk komponen utama yang merupakan kombinasi linear dari variabel X agar mempunyai variansi maksimum adalah dengan memilih vektor eigen yaitu $a = (a_1, a_2, \dots, a_p)$ sedemikian hingga $var(w_p) = a_j^T S a_j$ maksimum dan $a_j^T a_j = 1$ (Hair, 2014).

Kriteria pemilihan komponen utama menggunakan persentase variansi kumulatif terhadap total variansi. Persentase variansi kumulatif yang dijelaskan komponen utama ke-j dapat dihitung menggunakan persamaan berikut ini (Kurnia, 2021)

$$\frac{\lambda_j}{\sum_{j=1}^{p} \lambda_j} \times 100\% = \frac{\lambda_j}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \times 100\%, \text{ untuk } j = 1, 2, \cdots, p \quad (2.17)$$

Komponen utama dapat ditentukan dengan melihat proporsi kumulatif varians dan mampu menerangkan total variansi data sekitar 70% sampai 80%.

2.9 Regresi komponen utama

Metode regresi komponen utama (RKU) adalah suatu pendekatan analisis yang menggabungkan analisis regresi dengan AKU. Dalam metode ini, analisis regresi digunakan untuk menentukan keberadaan hubungan antara variabel respon dan variabel prediktor, sedangkan AKU digunakan untuk menyederhanakan variabel pengamatan dengan mereduksi dimensinya (Astuti, 2021). Metode RKU difungsikan untuk meminimumkan masalah multikolinearitas yang terjadi di dalam model regresi linear berganda dengan cara tidak menghilangkan variabel prediktor yang bersinggungan dengan kolinieritas.

Menurut (Siburian dkk., 2019) Metode RKU dilakukan setelah melakukan analisis menggunakan metode AKU untuk mendapatkan nilai-nilai komponen utama. Kemudian komponen utama yang dipilih akan diregresikan dalam model regresi bersama variabel respon menggunakan metode OLS.

2.10 Regresi *Robust*

Regresi *robust* adalah metode regresi yang digunakan ketika distribusi dari *error* tidak normal atau adanya beberapa *outlier* yang berpengaruh pada model (Nugrahani dkk., 2021). Tujuan dari metode regresi *robust* adalah mendeteksi *outlier* dan memberikan hasil estimasi yang resisten atau stabil dengan adanya *outlier*. Metode regresi *robust* menangani kasus *outlier* dengan cara mengabaikan identifikasi adanya *outlier* dan tidak menghilangkan data yang mengandung *outlier* didalam proses penanganannya (Chen, 2002). Metode yang telah dikembangkan dalam regresi *robust* untuk mengatasi masalah *outlier* meliputi beberapa pendekatan seperti estimasi-M, *least median square* (LMS), *least trimmed square* (LTS), estimasi-S, dan estimasi-MM (Rahayu dkk., 2023)

2.11 *Minimum Covarian Determinant*

Metode *minimum covarian determinant* (MCD) merupakan penaksir *robust* untuk rata-rata dan matriks kovarian dengan mencari sebagian data yang mempunyai kovarian minimum yang digunakan untuk mengidentifikasi *outlier*, menentukan jarak dan residu *robust* yang akan digunakan dalam pembobotan data dan penentuan parameter regresi (Astuti, 2021). Metode MCD bersifat resisten terhadap keberadaan *outlier* di dalam data pengamatan, sehingga sangat berguna untuk mendeteksi *outlier*. Tujuan MCD untuk mendapatkan suatu subsampel berukuran h dari keseluruhan pengamatan n, yang matriks varian kovariannya memiliki determinan terkecil diantara semua kombinasi kemungkinan data, dengan menggunakan persamaan sebagai berikut:

$$h = \frac{n+p+1}{2} \tag{2.18}$$

Metode estimasi MCD mudah dihitung dan ditemukan apabila jumlah n pengamatan kecil. Apabila jumlah n pengamatan besar maka akan banyak kombinasi subsampel dari h yang akan dicari dan ditemukan. Oleh karena itu, metode MCD menggunakan persamaan sebagai berikut:

$$\bar{X}_{MCD} = \frac{1}{h}\sum_{i \in h}^{h} X_i \tag{2.19}$$

dan

$$S_{MCD} = \frac{1}{h}\sum_{i\in h}^{h}[X_i - \bar{X}_{MCD}]^T[X_i - \bar{X}_{MCD}] \quad (2.20)$$

Keterangan:

h : Elemen pengamatan

$\bar{X}_{MCD}$: Vektor rata-rata MCD

S_{MCD} : Matriks varian kovarian MCD (Nisa dkk., 2019)

Proses perhitungan berikutnya adalah memilih himpunan yang memiliki determinan S_{MCD} terkecil dari himpunan h terpilih.

2.12 *Least Trimmed Square*

Least trimmed squares (LTS) adalah metode estimasi regresi yang digunakan untuk mengatasi dampak data *outlier* pada model regresi. LTS mengambil pendekatan yang berbeda dengan metode OLS dengan mempertimbangkan hanya sebagian kecil data yang paling cocok dengan model dan mengabaikan data yang dianggap sebagai *outlier* (Andriany dan Susanti, 2021). LTS adalah salah satu metode pendugaan parameter pada regresi *robust* yang sangat tahan terhadap keberadaan *outlier* dan memiliki *high breakdown point*. Metode ini memilih subhimpunan data dengan ukuran s, penentuan ukuran s menggunakan rumus ($\frac{(3h+w+1)}{4} \leq s \leq h$), di mana s merupakan jumlah pengamatan yang dianggap bukan sebagai *outlier*. Kemudian, metode ini meminimalkan jumlah kuadrat *error* hanya dari subhimpunan data tersebut. Dengan demikian, metode LTS memberikan bobot yang lebih rendah pada data yang dianggap sebagai *outlier* (Setyowati dkk., 2021).

Pendekatan ini memungkinkan metode LTS untuk menjadi lebih tahan terhadap keberadaan *outlier*. Dengan hanya mempertimbangkan sebagian kecil data yang dianggap sebagai data yang tidak *outlier*, metode LTS dapat memberikan estimasi parameter yang lebih konsisten dan *robust* terhadap pengaruh *outlier*. Metode LTS menduga koefisien regresi dengan menggunakan metode OLS terhadap subhimpunan data berukuran s dengan menggunakan persamaan sebagai berikut:

$$\hat{\beta} = arg\ min_{\beta}\left(\Sigma_{i=1}^{h}\varepsilon_i^2\right) = \arg min_{\beta}\left(\Sigma_{i=1}^{h}\left(Y_i - \hat{Y}_i\right)^2\right), \frac{(3h+w+1)}{4} \leq s \leq h \quad (2.21)$$

Keterangan:

ε_i^2 : kuadrat *error*, ε_i^2 diurutkan dari terkecil ke terbesar ($\varepsilon_1^2 < \varepsilon_2^2 <, \ldots,$ $< \varepsilon_n^2$

h : banyaknya observasi

w : banyaknya parameter

Solusi $\hat{\beta}$ pada persamaan dapat diperoleh dengan menggunakan turunan/differensial seperti pada penyelesaian pendugaan metode OLS, hanya pada LTS persamaan tersebut dihitung pada subhimpunan s terbaik. Subhimpunan s yang diperoleh merupakan sebaran data yang sudah terpangkas (Chen, 2002).

2.13 Produk Domestik Regional Bruto Indonesia

Produk Domestik Regional Bruto (PDRB) merupakan nilai tambah bruto seluruh barang dan jasa yang tercipta atau dihasilkan di wilayah domestik suatu negara yang timbul akibat berbagai aktivitas ekonomi dalam suatu periode tertentu tanpa memperhatikan faktor produksi yang dimiliki residen atau nonresident (Suryani, 2019). Penyusunan PDRB dapat dilakukan melalui 3 (tiga) pendekatan yaitu pendekatan produksi, pengeluaran, dan pendapatan yang disajikan atas dasar harga berlaku dan harga konstan (riil). PDRB dihitung dalam dua cara, yaitu atas dasar harga berlaku dan atas dasar harga konstan. Dalam menghitung PDRB atas dasar harga berlaku menggunakan harga barang dan jasa tahun berjalan, sedangkan pada PDRB atas dasar harga konstan menggunakan harga pada suatu tahun tertentu (tahun dasar) (Bunyanun, 2018).

Kertersediaan infrastruktur dapat memberikan pengaruh pada peningkatan akses masyarakat terhadap sumberdaya sehingga meningkatkan akses produktivitas sumberdaya yang pada akhirnya mendorong pertumbuhan ekonomi. Infrastruktur atau prasarana dan sarana fisik, di samping memiliki keterkaitan yang sangat kuat dengan kesejahteraan sosial dan kualitas lingkungan, juga terhadap proses pertumbuhan ekonomi suatu wilayah. Hal tersebut dapat ditunjukkan dengan indikasi bahwa wilayah yang memiliki kelengkapan sistem infrastruktur lebih baik biasanya mempunyai tingkat kesejahteraan sosial dan kualitas lingkungan serta pertumbuhan ekonomi yang lebih baik pula. Dengan demikian, pertumbuhan ekonomi yang tinggi tidak mungkin dicapai apabila tidak ada ketersediaan infrastruktur yang memadai atau dengan kata lain infrastruktur merupakan *basic determinant* atau kunci bagi perkembangan ekonomi (Ristiyanto, 2020).

Pada dasarnya, infrastruktur merupakan tanggung jawab dan kewajiban pemerintah pusat dan atau daerah, namun karena keterbatasan kemampuan dan dana yang semakin besar dalam Anggaran Pendapatan dan Belanja Negara (APBN) dan Anggaran Pendapatan dan Belanja Daerah (APBD), dibanding kebutuhan masyarakat dalam jumlah dan kualitas infrastruktur, maka berbagai persoalan pun timbul sehingga perlu dicarikan solusinya. Untuk kepentingan itu diperlukan pendekatan terpadu, mulai dari proses perencanaan, pembangunan sampai pada pengelolaan dan pelayanannya untuk masyarakat, utamanya dengan cara mensinergiskan fungsi dan tanggung jawab antar *stake holder*, antar sektor, antar daerah atau wilayah, dengan pelibatan pihak masyarakat, dunia usaha dan pihak-pihak lainnya (Kusuma, 2019). Infrastruktur tersebut menurut klasifikasi *World Bank* dapat dikelompokkan menjadi tiga yaitu, (1) Infrastruktur ekonomi adalah infrastruktur fisik yang diperlukan untuk menunjang aktivitas ekonomi, meliputi *public utilities* (listrik, telekomunikasi, sanitasi, gas), *public work* (jalan, bendungan, kanal, irigasi dan *drainase*) dan sektor transportasi (jalan rel, pelabuhan, lapangan terbang dan sebagainya). (2) infrastruktur sosial meliputi pendidikan, kesehatan, perumahan dan rekreasi. (3) infrastruktur administrasi meliputi penegakan hukum, kontrol administrasi dan koordinasi. Berdasarkan klasifikasi *World Bank* tersebut maka pada penelitian ini akan dilihat perbandingan pengaruh antara infrastruktur ekonomi (listrik, telekomunikasi, irigasi dan jalan) dan infrastruktur sosial (pendidikan dan kesehatan) terhadap pertumbuhan ekonomi.