Formosa Journal of Applied Sciences (FJAS)
Vol. 2, No. 1, 2023: 185 - 202

Bootstrap Estimation of Confidence Intervals of Multiple Regression Model Parameters in the Presence of Multicollinearity Using Principal Component Analysis

Chindy Eskana Nababan[1]*, Elmanani Simamora[2]
Universitas Negeri Medan
Corresponding Author: Chindy Eskana Nababan nababan.chin11@gmail.com

ARTICLEINFO

Keywords: Bootstrap Confidence Interval, Classical Confidence Interval, Multicollinearity, Principal Component Regression

Received : 20 November
Revised : 20 December
Accepted: 20 January

©2023 Nababan, Simamora: This is an open-access article distributed under the terms of the Creative Commons Atribusi 4.0 Internasional.

ABSTRACT

Multicollinearity in multiple regression can result in biased parameter estimators and increase the risk of accepting the null hypothesis of the regression model as an insignificant variable. This study aims to determine the use of the Principal Component Analysis method in overcoming multicollinearity problems that occur in Facebook metric data and to implement the use of the PCA (Principal Component Analysis) method with Bootstrap in overcoming multicollinearity problems that occur in Facebook metric data. One method that can be used to overcome multicollinearity is principal component regression analysis. Principal component regression will produce a point estimate. To measure the accuracy of the point estimation, the bootstrap method can be used, which generates confidence intervals by resampling the data with returns. The results of this study indicate that the problem of multicollinearity in Facebook metric data can be resolved using Principal component analysis and point estimation and classical confidence intervals in principal component regression are not significantly different from the results of estimated means and bootstrap confidence intervals.

DOI: https://doi.org/10.55927/fjas.v2i1.2851
ISSN-E: 2962 - 6447
https://journal.formosapublisher.org/index.php/fjas

Estimasi *Bootstrap* untuk Interval Kepercayaan Parameter Model Regresi Berganda dengan Adanya Multikolinearitas Menggunakan *Principal Component Analysis*

Chindy Eskana Nababan[1]*, Elmanani Simamora[2]
Universitas Negeri Medan
Corresponding Author: Chindy Eskana Nababan nababan.chin11@gmail.com

ARTICLEINFO

Kata Kunci: Interval Kepercayaan *Bootstrap,* Interval Kepercayaan Klasik, Multikolinearitas, Regresi Komponen Utama

Received : 20 November
Revised : 20 Desember
Accepted: 20 Januari

©2023 Nababan, Simamora: This is an open-access article distributed under the terms of the Creative Commons Atribusi 4.0 Internasional.

ABSTRAK

Multikolinearitas pada regresi berganda dapat mengakibatkan penduga parameter yang bias dan meningkatkan resiko diterimanya hipotesis nol dari model regresi sebagai variabel tidak signifikan. Penelitian ini bertujuan untuk mengetahui penggunaan metode *Principal Component Analysis* dalam mengatasi masalah multikolinearitas yang terjadi pada data metrik *facebook* dan Mengimplementasikan penggunaan metode PCA (*Principal Component Analysis*) dengan *Bootstrap* dalam mengatasi masalah multikolinearitas yang terjadi pada data metrik *facebook*. Salah satu metode yang dapat digunakan untuk mengatasi multikolinearitas adalah analisis regresi komponen utama. Regresi komponen utama akan menghasilkan estimasi titik. Untuk mengukur keakurasian estimasi titik tersebut, maka dapat digunakan metode *bootstrap,* yang menghasilkan interval kepercayaan dengan cara *resampling* data dengan pengembalian. Hasil penelitian ini menunjukkan bahwa masalah multikolinearitas pada data metrik *facebook* dapat teratasi menggunakan *Principal component analysis* dan estimasi titik dan interval kepercayaan klasik pada regresi komponen utama tidak berbeda signifikan dengan hasil estimasi rataan dan interval kepercayaan *bootstrap.*

PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi pada abad ke 21 menjadi tanda adanya peningkatan penggunaan teknologi yang semakin maju. Kebutuhan informasi pada masa ini tidak hanya pada masyarakat kelas-kelas tertentu, akan tetapi semua elemen masyarakat dapat memanfaatkan informasi yang tanpa batas. Masyarakat dapat mengakses setiap informasi dalam bentuk jejaring internet secara online. Dengan keterbukaan dan keterluasan informasi yang terdapat di internet maka muncul budaya baru seperti penggunaan media sosial/virtual sebagai bentuk pemanfaatan informasi dan interaksi dalam berkomunikasi. Dalam artian untuk mendapatkan suatu informasi tentulah para pengguna akan melihat terhadap akun-akun yang ada di media sosial seperti facebook, twitter, instagram, tiktok dan sebagainya, kemudian dari beberapa akun media sosial tersebut seorang pengguna akan mengikuti tautan yang terdapat dalam akun pilihannya. Dari alur ini maka muncul suatu metrics untuk mengukur kreadibilitas akun media sosial tersebut terhadap informasi yang dimanfaatkan oleh para pengguna media social. (Bronman,2014).

Regresi linear (*linear regression*) merupakan metode yang digunakan untuk memperoleh model hubungan antara satu variabel terikat dengan satu variabel bebas. Harlan (2018:5) menyebutkan bahwa Variabel terikat pada regresi linear disebut juga sebagai respons, sedangkan variabel bebas dikenal sebagai prediktor atau regresor. Jika hanya terdapat satu variabel terikat dalam model maka teknik ini disebut sebagai regresi linear sederhana (*simple linear regression*), sedangkan jika terdapat beberapa variabel bebas, teknik ini disebut regresi linear berganda (*multiple linear regression*). Analisis regresi memiliki 3 kegunaan, yaitu untuk tujuan deskripsi fenomena data atau kasus yang sedang diteliti, untuk tujuan kontrol dan untuk tujuan prediksi.

Bawono dan Shina (2018:20) mengatakan model regresi linier disebut valid apabila model tersebut memenuhi asumsi klasik. uji asumsi klasik merupakan uji prasyarat yang dilakukan sebelum melakukan analisis lebih lanjut terhadap data yang telah dikumpulkan. Pengujian asumsi klasik ini ditujukan agar dapat menghasilkan model regresi yang memenuhi kriteria BLUE (*Best Linier Unbiased Estimator*) yang digunakan sebagai estimator terpercaya, dimana estimator tersebut dinyatakan tidak bias, konsisten, berdistribusi normal dan juga efisien. Untuk mengetahui apakah model regresi yang akan digunakan telah memenuhi kriteria *BLUE* maka perlu dilakukan serangkaian pengujian yaitu Uji normalitas, uji multikolinieritas, uji heteroskedastisitas, dan uji autokorelasi. Ghozali (2016) menyatakan bahwa normalitas terjadi ketika variabel pengganggu atau residual dalam persamaan regresi memiliki distribusi normal, multikolinearitas terjadi ketika ada korelasi diantara variabel bebas, heteroskedastisitas terjadi jika varian dari *error* suatu pengamatan ke pengamatan lain terjadi ketidaksamaan (tidak konstan), sedangkan autokorelasi terjadi ketika nilai korelasi dalam persamaan regresi bernilai nol. Asumsi yang akan dibahas dalam penelitian ini adalah asumsi multikolinearitas.

Kusrini Setiawan (2010) mendefenisikan bahwa multikolinearitas merupakan adanya hubungan korelasi linear yang tinggi diantara beberapa atau semua variabel bebas dari model regresi berganda. Multikolinearitas terjadi apabila terdapat hubungan linear diantara beberapa variabel atau seluruh variabel bebas. Semakin tinggi multikolinearitas antara variabel, semakin tinggi galat koefisien regresinya. Galat yang tinggi akan menghasilkan penduga parameter yang bias. Untuk mengatasi multikolinearitas, salah satu metode yang dapat digunakan adalah PCA(*Principal Component Analysis*) atau sering disebut dengan metode analis komponen utama.

Penelitian sebelumnya,Ryan,Iskandar dkk (2013) mengkaji perbandingan metode *bootstrap* dan metode *jackknife* dalam menaksir parameter regresi untuk mengatasi multikolinearitas, data yang digunakan adalah seratus ekor ikan yang sama spesiesnya dari studi perikanan di Universitas Mustafa Kemal (Turkey) dimana panjang sirip ikan dan panjang ekor ikan tidak saling berkorelasi sebagai variabel bebasnya yang menjelaskan variasi umur ikan. Hasil penelitiannya menunjukkan bahwa bias parameter, standar error, dan interval konfidensi Jackknife lebih besar dibandingkan Bootstrap. Sehingga metode Bootstrap lebih efisien dalam menduga parameter regresi dibandingkan metode Jackknife ketika terjadi multikolinearitas.

Berdasarkan uraian latar belakang di atas, penulis tertarik ingin melakukan penelitian mengenai "Estimasi *Bootstrap* Untuk Interval Kepercayaan Parameter Model Regresi Berganda Dengan Adanya Multikolinearitas Menggunakan *Principal Component Analysis* "

TINJAUAN PUSTAKA

Analisis Regresi Berganda

Analisis regresi pertama kali diperkenalkan sebagai metode analisis data statistik pada tahun 1877 oleh Sir Francis Galton (1822-1911) yang meneliti hubungan antara tinggi badan ayah dengan anaknya. hasilnya terdapat kecenderungan orang tua yang tinggi akan memiliki anak yang tinggi dan juga sebaliknya, tetapi penyebaran rata-rata tinggi badan dari generasi ke generasi adalah tetap. Hasil analisis tersebut disempurnakan oleh Karl Pearson dengan mengambil sampel lebih dari 1000 pengamatan. Hasilnya untuk kelompok anak dan ayah yang tinggi, ternyata tinggi badan anak lebih pendek dari ayahnya, sedangkan dari kelompok anak dan ayah yang pendek, ternyata tinggi badan anak lebih tinggi dari ayahnya. Kecenderungan ini dianggap sebagai kebalikan (regress) dari nilai tengah masing-masing kelompok tinggi. Galton mengembangkan deskripsi matematis dari kecenderungan regresi tersebut yang merupakan awal dari model regresi saat ini.

Multikolinearitas

Istilah multikolinearitas atau kolinearitas ganda pertama kali dikenalkan oleh Ragner Frish yang berarti, adanya hubungan linear (korelasi) yang sangat tinggi antar variabel-variabel bebas dalam model regresi. Multikolinearitas menimbulkan masalah dalam model regresi. Korelasi antar variabel bebas yang sangat tinggi menghasilkan penduga model regresi yang berbias, tidak stabil, dan mungkin jauh dari nilai prediksinya.

Firdaus (2019:166) menjelaskan bahwa apabila terjadi multikolinearitas sempurna maka koefisien regresi dari variabel X tidak dapat ditentukan (*indeterminate*) dan standar erornya tak terhingga (*infinite).* Jika multikolinearitas kurang sempurna, walau koefisien regresi dari variabel X dapat ditentukan (*determinate*), tetapi standar erornya tinggi, yang berarti koefisien regresi tidak dapat diperkirakan dengan tingkat ketelitian yang tinggi. Jadi, semakin kecil korelasi di antara variabel bebasnya maka semakin baik model regresi yang akan diperoleh.

Pada analisis regresi, multikoliniearitas dikatakan ada apabila beberapa kondisi berikut dipenuhi :

1. Jika variabel bebas tersebut berkorelasi sempurna yaitu koefisien korelasinya mendekati -1 atau 1.
2. Menggunakan *Variation Inflation Factor.Variation Inflation Factor* (VIF) dirumuskan sebagai berikut :

$$VIF = \frac{1}{1-R_j^2}$$

Dengan:
VIF = *Variance Inflation Factor* (VIF).
R_j^2 = Koefisien determinasi antara X_j dengan variabel bebas lainnya pada persamaan/model dugaan.
j = 1, 2,..,k.

Jika nilai VIF lebih besar dari 10 maka dapat diidentifikasikan dalam variabel bebas terdapat multikolinearitas pada data.

Metode PCA (*Principal Component Analysis)*

G.L.Marcus (2012) mendefenisikan bahwa analisis komponen utama merupakan Teknik statistik yang dapat digunakan untuk menjelaskan struktur variansi-kovariansi dari sekumpulan variabel melalui beberapa variabel baru dimana variabel baru ini saling bebas dan merupakan kombinasi linier dari variabel asalnya

Analisis komponen utama bertujuan untuk menyederhanakan variabel yang diamati dengan cara mereduksi dimensi data sehingga lebih muda untuk menginterpretasikan data-data tersebut.Hal ini dilakukan dengan menghilangkan korelasi variabel melalui transformasi variabel asal ke variabel baru yang tidak berkorelasi.

Jika didefinisikan A sebagai matriks konstan berukuran kxk , maka komponen utama didefinisikan sebagai variabel baru (F) yang merupakan hasil transformasi dari variabel asal yang modelnya dapat ditulis dalam bentuk matriks :

$$\boldsymbol{F = AX}$$

Dengan :
A = Matriks yang melakukan transformasi terhadap variabel asal x sehingga diperoleh vektor komponen F.
F = Komponen utama

Algoritma PCA

PCA sensitif terhadap skala pengukuran sehingga variabel harus dinormalisasi sebelum menerapkan PCA.

- Tahap pertama proses PCA adalah input data. Data disiapkan dalam bentuk matriks ukuran mxn, dimana jumlah variabel n akan berkurang menjadi k jumlah *principal component* yang dipertahankan dimana setiap k jumlah *principal component* saling ortogonal dan saling bebas satu sama lain.
- Tahap selanjutnya adalah pre-PCA yaitu dengan melakukan standarisasi setiap atribut dan hitung varians, simpangan baku dengan menggunakan rumus; $Var(x) = \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(\tilde{z}_{ij} - \mu_j)^2$
- Selanjutnya dihitung kovarians untuk mencari hubungan antara dua kelas, dimana nilai nol menunjukkan bahwa tidak ada hubungan antara dua dimensi;

$$Cov(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \mu_{xj})\,(y_{ij} - \mu_{yj})$$

Dengan μ_{xj}dan μ_{yj} merupakan rata – rata (mean) sampel dari variabel x dan y, dimana x_{ij} dan y_{ij}merupakan nilai observasi ke- i, j dari variabel x dan x. maka dari data nilai yang digunakan, diperoleh matrik kovarian berukuran $n\ x\ n$.

- Mencari *eigenvalues* dan *eigenvector* dari matrik kovarian yang telah diperoleh;

  Dalam pendekatan aljabar nilai eigen dan vektor eigen dapat ditentukan dengan menggunakan definisi berikut ini:

$$(S - \lambda I)a = 0,$$

dengan S merupakan matriks kovarians pada data, λ merupakan nilai eigen, I merupakan matriks identitas, dan a merupakan vektor eigen. Sehingga diperoleh kombinasi linear yaitu:

a. $\lambda_1, \lambda_2, \lambda_3 \ldots . \lambda_n$ adalah eigenvalue matrik S

b. $x_1, x_2, x_3 \ldots . x_n$ adalah eigenvector sesuai eigenvaluenya (λ_n)

Menentukan variabel baru (komponen utama) dengan mengalikan variabel asli dengan matrik vektor eigen (Jolliffe,2002)

Eigenvalues (λ) adalah bilangan skalar dan A adalah matrik dengan ukuran nxn untuk memperoleh nilai n *eigenvalues* (λ1, λ2……..λn) maka memenuhi persamaan berikut: $Det(A - \lambda I) = 0$

A = matrik nxn

λ = nilai eigen value

I = matriks identitas merupakan matriks persegi dengan elemen diagonal utama bernilai 1, sedangkan elemen lain bernilai nol

- Menentukan variabel baru *(principal component*) dengan mengalikan variabel asli dengan matrik eigenvector melalui persamaan:

Z_{ki} = $U_{1k}\ X_{1i} + U_{1k}\ X_{1i} + \ldots + U_{Pk}\ X_{Pi}$

Z_{ki} = matriks nxn dari principal component dengan koordinat objek ke-i pada posisi ke-k pada *principal component*

U = matriks pxk (matrik *eigenvector*)

X = matriks nxn (variabel asli)

Kriteria pemilihan komponen utama

Salah satu tujuan dari analisis komponen utama adalah mereduksi dimensi data asal yang semula. Jika terdapat k variabel bebas menjadi p komponen utama (dimana $p < k$). Maka kriteria pemilihan p, yaitu :

1. Didasarkan pada akar ciri yang lebih besar dari satu.
2. Proporsi kumulatif variansi data asal yang dijelaskan oleh p komponen utama minimal 80 %, dan proporsi total variansi populasi bernilai cukup besar.

Metode *Bootstrap*

Metode *bootstrap* merupakan suatu metode berbasis komputer yang sangat potensial untuk dipergunakan pada masalah ketakstabilan dan keakurasian, khususnya dalam menentukan interval. Istilah *bootstrap* berasal dari "*pull oneself up by one's bootstrap*", yang berarti berpijak di atas kaki sendiri, berusaha dengan sumber daya minimal. Dalam sudut pandang statistika, sumber daya yang minimal adalah data yang sedikit, data yang menyimpang dari asumsi tertentu, atau data yang tidak mempunyai asumsi apapun tentang distribusi populasinya.Metode *bootstrap* diperkenalkan pertama kali oleh Efron pada tahun 1979. *Bootstrap* adalah metode yang didasarkan pada simulasi data untuk keperluan inferensi statistik (Efron, 1993). Metode *bootstrap* digunakan untuk mencari distribusi sampling dari suatu estimator dengan prosedur *resampling* dengan pengembalian dari data asli. Metode bootstrap dapat digunakan tanpa membutuhkan asumsi distribusi karena penggunaan sampel asli sebagai populasi. Hasil penyampelan ulang dari populasi tersebut akan digunakan sebagai sampel b*ootstrap*. Sampel *bootstrap* tersebut yang akan digunakan sebagai solusi dalam menyelesaikan statistika inferensi (Sungkono, 2013).

Interval Kepercayaan *Bootstrap*

Mengestimasi interval kepercayaan dengan *bootstrap* dapat digunakan tiga metode yaitu metode bootstrap standar, metode bootstrap-t dan metode *bootstrap* persentil. Dalam Penelitian ini metode yang digunakan adalam metode metode *bootstrap* persentil.

- Metode *Bootstrap* persentil

  Persentil bootstrap merupakan salah satu metode pendugaan selang kepercayaan dengan menetapkan batas bawah dan atas selang berdasarkan persentase dari replikasi bootstrap. Penelitian ini bertujuan untuk menduga selang kepercayaan persentil bootstrap untuk parameter model regresi linier satu dan dua peubah bebas dengan melakukan beberapa variasi jumlah sampel *bootstrap* dan jumlah pengulangan pendugaan parameter. Data yang disimulasikan adalah data riil agar dapat dipastikan ada hubungan fungsionalnya antara peubah-peubah bebas dan peubah tak bebas.

Interval kepercayaan bootstrap yang ditingkatkan didasarkan pada kuantil distribusi replika bootstrap. Teknik ini memiliki keuntungan lebih stabil daripada *bootstrap -t,* dan juga memilik sifat cakupan teoretis yang lebih baik (Efron dan Tibshirani, 1993). interval kepercayaan persentil *bootstrap* adalah:

$$(\hat{\theta}_B{}^{*(\alpha/2)}, \hat{\theta}_B{}^{*(1-\alpha/2)})$$

dimana
$\hat{\theta}_B{}^{*(\alpha/2)}$ adalah kuantil
$\alpha/2$ dalam distribusi bootstrap dari $\hat{\theta}^*$.

METODOLOGI

Penelitian dilakukan di Perpustakaan Digital Library untuk mengumpulkan studi literatur terkait tema penelitian, Perpustakaan FMIPA Universitas Negeri Medan untuk mengumpulkan penelitian terdahulu terkait tema penelitian dan tempat lainnya yang lebih memadai selama kurang lebih satu bulan. Jenis penelitian yang digunakan dalam penulisan ini adalah penelitian kepustakaan atau riset kepustakaan (*Library riset) dan* Bahan penelitian atau data yang dipakai dalam penelitian ini adalah data sekunder yaitu data metrik kinerja *facebook* yang diperoleh melalui website UCI *Machine Learning* dengan pengolahan data mengunakan software R-Programing.

Agar penelitian dapat berjalan dengan baik, maka diperlukan beberapa prosedur penelitian sebagai berikut:

1. Melakukan study kepustakaan dengan mencari dan membaca bahan-bahan Pustaka yang relevan dengan topik penelitian baik berupa buku, jurnal, paper dan artikel.
2. Pengambilan data metrik *facabook* melalui website UCI *Machine Learning*
3. Melakukan analisis data eksploratif dengan memahami isi dan komponen penyusun data melalui grafik, statistik deskriptif dan pola sebaran data dan untuk memeriksa asumsi, melihat hubungan antara variabel, mendeteksi ketidaknormalan, mendeteksi kesalahan dan outlier pada data. dengan demikian, akan diperoleh pemahaman tentang karakteristik data
4. Data processing dimana proses persiapan data sebelum dilakukannya proses *machine learning* untuk meningkatkan kualitas perhitungan model.
5. Menentukan Analisis Komponen Utama (PCA)
6. Membandingkan estimasi interval kepercayaan *Bootstrap* dan estimasi interval kepercayaan klasik
7. Menarik Kesimpulan

HASIL DAN PEMBAHASAN

Analisis Data Eksploratif

Pada penelitian ini data yang digunakan adalah data metrik kinerja *facebook* dengan 19 (Sembilan belas) variabel yaitu 18 (delapan belas) variabel bebas (X) dan 1 (satu) variabel terikat (Y).

Tabel.1 Data Metrik Kinerja *Facebook* dengan 19 (Sembilan Belas) Variabel

Var	***Nama Variabel***	***Tipe Data***
X_1	*Page total like*	Integer
X_2	*Type*	Obyek
X_3	*Category*	Integer
X_4	*Post Month*	Integer
X_5	*Post Weekday*	Integer
X_6	*Post Hour*	Integer
X_7	*Paid*	Floating
X_8	*Lifetime post total reach*	Integer
X_9	*Lifetime post total impression*	Integer
X_{10}	*Lifetime enganged users*	Integer
X_{11}	*Lifetime post consumers*	Integer
X_{12}	*Lifetime post Consumptions*	Integer
X_{13}	*Lifetime Post Impressions by people who have liked your Page*	Integer
X_{14}	*Lifetime Post reach by people who like your Page*	Integer
X_{15}	*Lifetime People who have liked your Page and engaged with your post*	Integer
X_{16}	*Comment*	Integer
X_{17}	*Like*	Floating
X_{18}	*Share*	Floating
Y	*Total Interaction*	Integer

Selanjutnya ekplorasi data dalam penelitian ini adalah dengan melihat hasil deskriptif data dari variabel prediktor. Yang akan ditunjukkan pada Tabel 2 dan Tabel 3.

Tabel 2. Statistika Deskriptif Variabel Prediktor

Nama Variabel	Mean	Median	Min	Max
Page total like	123194	129600	81370	139441
Type	1.88	2.00	1.00	3.00
Category	7.038	7.000	1.000	12.000
Post Month	4.15	4.00	1.00	7.00
Post Weekday	7.84	9.00	1.00	23.00
Post Hour	0.2786	0.0000	0.0000	1.0000
Paid	13903	5281	238	180480
Lifetime post total reach	29586	9051	570	1110282
Lifetime post total impression	920.3	625.5	9.0	11452.0
Lifetime enganged users	798.8	551.5	9.0	11328.0

Lifetime post consumers	1415.1	851.0	9.0	19779.0
Lifetime post Consumptions	16766	6256	567	1107833
Lifetime Post Impressions by people who have liked your Page	6585	3417	236	51456
Lifetime Post reach by people who like your Page	610.0	412.0	9.0	4376.0
Lifetime People who have liked your Page and engaged with your post	7.482	3.000	0.000	372.000
Comment	177.9	101.0	0.0	5172.0
Like	27.27	19.00	0.00	790.00
Share	212.1	123.5	0.0	6334.0
Total Interaction				

Tabel 3. Perbandingan Kategori Variabel Prediktor

Nama Variabel	Kategori	Jumlah
Paid	0	361
	1	139
Category	1	215
	2	130
	3	155
Type	Link	22
	Photo	426
	Status	45
	Video	7

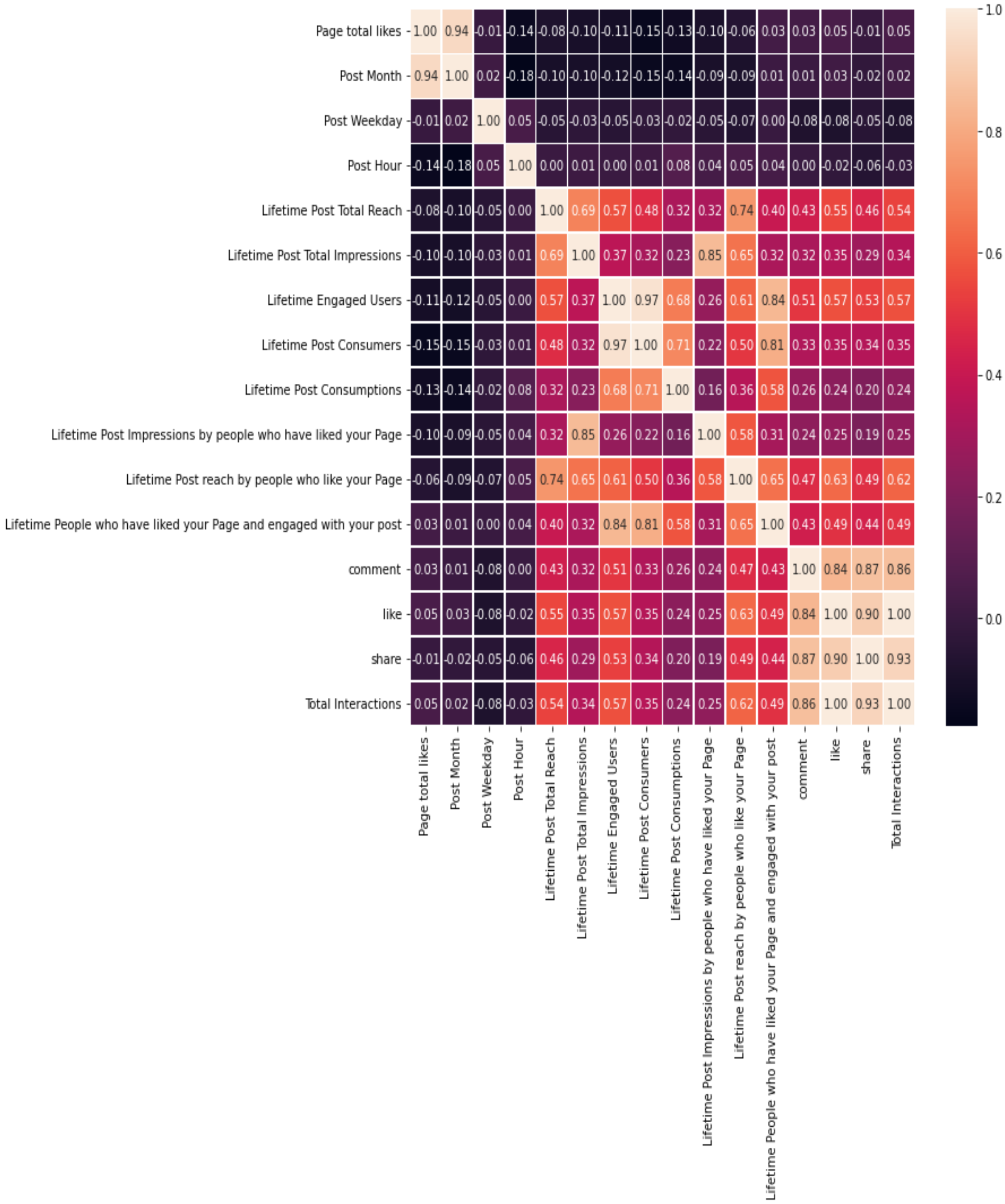
Uji Outlier

Outlier merupakan pengamatan yang jauh berbeda (ekstrim) dari data pengamatan lainnya, atau dapat diartikan data yang tidak mengikuti pola umum model. Adakalanya outlier memberikan informasi yang tidak dapat diberikan oleh data yang lainnya. Karena itulah outlier tidak boleh begitu saja dihilangkan. Outlier juga merupakan pengamatan berpengaruh. Banyak sekali metode yang dapat digunakan untuk mendeteksi adanya outlier.

Setelah Outlier ada juga masalah yang terdapat data metrik *faceebok* yaitu masalah multikolinearitas, dimana masalah ini akan di selesaikan dengan menggunakan *Principal Component Analysis (PCA)*

Mendeteksi Multikolinearitas dengan VIF

Untuk mendeteksi apakah suatu model memiliki gejala multikolinearitas akan digunakan ukuran nilai VIF dan juga dapat dilihat dari nilai koefisien korelasi setiap variabel. Jika koefisien korelasi diantara masing-masing variabel prediktor lebih dari 0,7 maka terjadi multikolinearitas dan sebaliknya, jika koefisien korelasi antara masing-masing variabel prediktor kurang dari 0,7 maka tidak terjadi multikolinearitas (Cohen 1998). Apabila nilai dari VIF lebih dari 10 , maka model tersebut mengandung multikolinearitas begitu juga sebaliknya , apabila nilai VIF kurang dari 10 , maka model tersebut tidak mengandung multikolinearitas. Koefisien korelasi variabel prediktor dapat dilihat pada gambar dibawah ini :



Gambar 1. Koefisien Korelasi Variabel Prediktor

Analisis Komponen Utama (PCA)

Matriks Kovarian

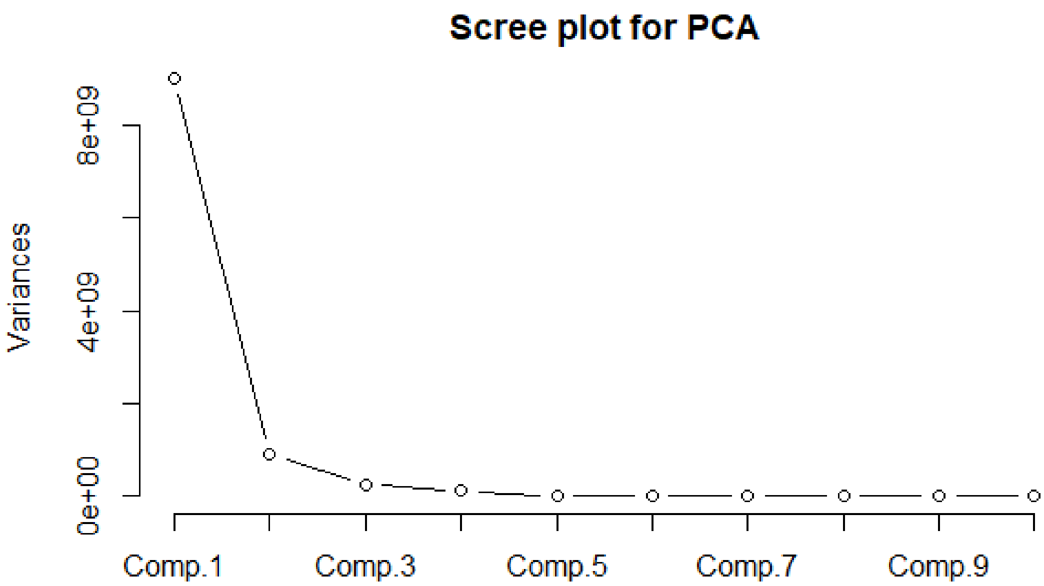
Matriks kovarian dipilih karena dilakukan standarisasi pada data penelitian dan berfungsi sebagai nilai masukan untuk mendapatkan nilai eigen dan vektor eigen. Berdasarkan hasil tahapan preprocessing yaitu 16 variabel prediktor maka matriks kovarian yang dibentuk dari data menghasilkan square matrix berukuran 16 x 16. Berikut merupakan hasil matriks kovarian dari data menggunakan R Programing.

$$A=\begin{bmatrix}
1 & 0.94 & -0.01 & -0.14 & 0.01 & -0.08 & -0.10 & -0.11 & -0.15 & -0.13 & -0.10 & -0.06 & 0.03 & 0.03 & 0.05 & -0.01 \\
0.94 & 1 & 0.02 & -0.18 & -0.02 & -0.10 & -0.10 & -0.12 & -0.15 & -0.14 & -0.09 & -0.09 & 0.01 & 0.01 & 0.03 & -0.02 \\
-0.01 & 0.02 & 1 & 0.05 & 0.00 & -0.05 & -0.03 & -0.05 & -0.03 & -0.02 & -0.05 & -0.07 & 0.00 & -0.08 & -0.08 & -0.05 \\
-0.14 & -0.18 & 0.05 & 1 & -0.07 & 0.00 & 0.01 & 0.00 & 0.01 & 0.08 & 0.04 & 0.05 & 0.04 & 0.00 & -0.02 & -0.06 \\
0.01 & -0.02 & 0.00 & -0.07 & 1 & 0.15 & 0.06 & 0.12 & 0.10 & 0.10 & 0.00 & 0.11 & 0.05 & 0.08 & 0.11 & 0.08 \\
-0.08 & -0.10 & -0.05 & 0.00 & 0.15 & 1 & 0.69 & -0.57 & 0.48 & 0.32 & 0.32 & 0.74 & 0.40 & 0.43 & 0.55 & 0.46 \\
-0.10 & -0.10 & -0.03 & 0.01 & 0.06 & 0.69 & 1 & 0.37 & 0.32 & 0.23 & 0.85 & 0.65 & 0.32 & 0.32 & 0.35 & 0.29 \\
-0.11 & -0.12 & -0.05 & 0.00 & 0.12 & 0.57 & 0.37 & 1 & 0.97 & 0.68 & 0.26 & 0.61 & 0.84 & 0.51 & 0.57 & 0.53 \\
-0.15 & -0.15 & -0.03 & 0.01 & 0.10 & 0.48 & 0.32 & 0.97 & 1 & 0.71 & 0.22 & 0.50 & 0.81 & 0.33 & 0.35 & 0.34 \\
-0.13 & -0.14 & -0.02 & 0.08 & 0.10 & 0.32 & 0.23 & 0.68 & 0.71 & 1 & 0.16 & 0.36 & 0.58 & 0.26 & 0.24 & 0.20 \\
-0.10 & -0.09 & -0.05 & 0.04 & 0.00 & 0.32 & 0.85 & 0.26 & 0.22 & 0.16 & 1 & 0.58 & 0.31 & 0.24 & 0.25 & 0.19 \\
-0.06 & -0.09 & -0.07 & 0.05 & 0.11 & 0.74 & 0.65 & 0.61 & 0.50 & 0.36 & 0.58 & 1 & 0.65 & 0.47 & 0.63 & 0.49 \\
0.03 & 0.01 & 0.00 & 0.04 & 0.05 & 0.40 & 0.32 & 0.84 & 0.81 & 0.58 & 0.31 & 0.65 & 1 & 0.43 & 0.49 & 0.44 \\
0.03 & 0.01 & -0.08 & 0.00 & 0.08 & 0.43 & 0.32 & 0.51 & 0.33 & 0.26 & 0.24 & 0.47 & 0.43 & 1 & 0.84 & 0.87 \\
0.05 & 0.03 & -0.08 & -0.02 & 0.11 & 0.55 & 0.35 & 0.57 & 0.35 & 0.24 & 0.25 & 0.63 & 0.49 & 0.84 & 1 & 0.90 \\
-0.01 & -0.02 & -0.05 & -0.06 & 0.08 & 0.46 & 0.29 & 0.53 & 0.34 & 0.20 & 0.19 & 0.49 & 0.44 & 0.87 & 0.90 & 1
\end{bmatrix}$$

Tabel.4 Dekomposisi Nilai Eigen

Komponen Utama	Nilai eigen (λ_n)	Vekor eigen (x_n)
1	9004342700	-0.018452,0.000000,-0.000001, 0.000001, 0.000001,0.000000, 0.144608, 0.795746, 0.004623, 0.585335
2	903531897	-0.007152,0.000000,-0.000003, 0.000001,-0.000006,0.000002, 0.542463, 0.428189, 0.009948,-0.720972
3	261525570	0.999055,0.000006,-0.000005, 0.000001,-0.000039,0.000000, 0.023170, 0,033181, -0.013608, -0.007808
4	117218633	0.033189, 0.000005, -0.000006, -0.000016, 0.000020, 0.000004, 0.717889, -0,404569, 0.047937, 0.331403
5	8209073	-0.014659, 0.000030, 0.000011, 0.000027, 0.000063, 0.000001, -0.410679, 0.136577, 0.076562, -0.166340
6	3435178	-0.012964, -0.000052, 0.000055, -0.000006, -0.000117, -0.000012, 0.007512, -0.001031, -0.986648, -0.001850
7	143041.8	-0.003663, 0.000430, 0.000471, 0.000332, -0.000291, 0.000064, 0.014433,-0.003224, -0.134393, 0.004586
8	18.5	0.000036, 0.004926, -0.023206, 0.030864, 0.999213, -0.007558, 0.000019, 0.000001, -

		0.000162, 0.000000
9	4.05	0.000002, -0.000367, -0.035136, 0.998876, -0031643, 0.003267, 0.000017, -0.000009, 0.000038, 0.000009
10	0.65	0.000009, 0.028213, 0.998680, 0.035809, 0.022014, 0.008694, 0.000004, -0.000001, 0.000114, 0.000000
11	0.19	0.00001, -0.027531, 0.009520, 0.003363, -0.007307, -0.999543, 0.000002, 0.000000, 0.000023, 0.000000
12	0.13	0.000005, -0.999210, 0.027834, 0.000704, 0.005760, 0.027747, -0.000002, 0.000001, -0.000003, -0.000001



Scree Plot Komponen Utama

Pembentukan Komponen Utama

Total komponen utama yang terbentuk sama dengan banyak variabel yang dianalisis, yaitu jumlah proporsi varians kovarians maksimum yang mampu menjelaskan varians kovarians dari atribut asli. Persentase varians total dianggap cukup mewakili varians jika data 75% atau lebih (Arista 2015). Secara subjektif, hasil perhitungan proporsi kumulatif varians dari seluruh komponen utama dapat dilihat pada Tabel 5

Tabel 5. Proporsi Kumulatif Varians

Komponen Utama	Nilai eigen (λ_n)	(%) Variansi	(%) Kumulatif
1	9004342700	87 %	87 %
2	903531897	9 %	96 %
3	261525570	3 %	99 %
4	117218633	1 %	100 %
5	8209073	0 %	100 %
6	3435178	0 %	100 %
7	143041.8	0 %	100 %
8	18.5	0 %	100 %
9	4.05	0 %	100 %
10	0.65	0 %	100 %
11	0.19	0 %	100 %
12	0.13	0 %	100 %

Berdasarkan perhitungan proporsi kumulatif varians sampai komponen utama kedua (PC2) adalah sekitar 96 % sehingga PC1, PC2 merupakan komponen utama yang dianggap cukup untuk menjelaskan struktur data, sehingga komponen-komponen utama tersebut digunakan untuk mewakili variabel-variabel prediktor awal. Data komponen utama dapat dilihat pada Tabel 4.4

Tabel 6. Data Komponen Utama

No	PCA_1	PCA_2	Y
1	29689.2833	7042.2881	100
2	12158.9476	2866.9168	164
3	30472.7136	7348.6322	80
⋮	⋮	⋮	⋮
500	26167.098	5436.2928	119

Berdasarkan Tabel 6. menunjukkan bahwa pada data Metrik kinerja *facebook* cukup digunakan dua buah komponen utama, yaitu komponen utama pertama dan komponen utama kedua, yang telah mampu menerangkan struktur data yaitu sebesar 96 %. Dengan demikian, untuk analisis selanjutnya hanya digunakan dua buah komponen (F_1 dan F_2), sehingga dapat dinyatakan dalam persamaan berikut :

$$\hat{Y} = 212.120 + (-0.0013)F_1 + (-0.0031)F_2$$

Jadi dapat disimpulkan bahwa komponen utama pertama dan komponen utama kedua secara simultan memiliki pengaruh signifikan terhadap total interaksi(y).

Selanjutnya untuk melihat kontribusi variabel prediktor di dalam dua atribut komponen utama maka dilakukan perhitungan bobot atribut menggunakan *vector eigen*. Hasil pembobotan atribut dapat di lihat pada tabel 7 .

Tabel 7. Kontribusi Variabel PCA

Variabel	PC1	PC2
Page total like	0.03	0.01
Type	0.00	0.00
Category	0.00	0.00
Post Weekday	0.00	0.00
Post Hour	0.00	0.00
Paid	0.00	0.00
Lifetime Post Total Reach	2.09	29.43
Lifetime Post Total Impressions	63.32	18.33
Lifetime Post Consumptions	0.00	0.01
Lifetime Post Impressions by people who have liked your Page	34.26	51.98
Lifetime Post reach by people who like your Page	0.29	0.24
Lifetime People who have liked your Page and engaged with your post	0.00	0.00

Dari Tabel 5. dapat diperoleh variabel untuk PC1 berdasarkan nilai yang terbesar yaitu *Page total like,Type, Category, Post weekday, Post Hour, Paid, Lifetime Post Consumptions* dan juga *Lifetime People who have liked your Page and engaged with your post*. Sedangkan Variabel untuk PC2 yaitu *Lifetime Post Total Reach, Lifetime Post Total Impressions, Lifetime Post Impressions by people who have liked your Page.*

Metode *Bootsrap*

Metode *bootstrap* merupakan metode yang digunakan untuk mengestimasi suatu distribusi populasi yang tidak diketahui dengan distribusi empiris yang diperoleh dari proses penyampelan ulang (Efron dan Tibshirani, 1998).

Selanjutnya menentukan penduga parameter dengan menggunakan resampling data untuk setiap replikasi (50, 100, 150 dan 200) digunakan menggunakan software R programing untuk memudahkan proses bootstrap. Hasil program R dapat dilihat pada tabel 8. Setelah penduga parameter boostrap rataan diperoleh, kemudian menghitung variansi penduga bootstrap yang selanjutnya akan digunakan untuk menghitung nilai standar deviasi. Adapun hasil yang di peroleh dapat dilihat pada tabel 8.

Tabel 8. Penduga Rataan dan Standar Deviasi *Bootstrap*

		$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$
Regresi Komponen Utama	Estimasi Titik	212.120	-0.0013	-0.0031
	Std. Deviasi	216.039	0.00048	0.001
Replikasi 50	Rataan	182,8	-0.0011	-0.002
	Std. Deviasi	1.009	0.000	0.000
Replikasi 100	Rataan	176,8	-0.001	-0.0021
	Std. Deviasi	1.475	0.000	0.000
Replikasi 150	Rataan	177,3	-0.0009	-0.0019
	Std.Deviasi	1.492	0.000	0.000
Replikasi 200	Rataan	178,5	-0.001	-0.0022
	Std. Deviasi	1.985	0.000	0.001

Berdasarkan Tabel 8. tampak bahwa nilai estimasi titik dan standar deviasi parameter regresi komponen utama tidak berbeda jauh dengan hasil estimasi rataan dan standar deviasi *bootstrap* yang diperoleh untuk setiap replikasi.

Hasil perhitungan *bootstrap* yang telah dilakukan, selanjutnya akan digunakan untuk menghitung interval kepercayaan *bootstrap* dengan menggunakan persamaan (3.6), berikut hasil yang diperoleh pada Tabel 9.

Tabel 9. Regresi Komponen Utama (PCA) dan *Bootstrap*

Parameter	Replikasi	Rataan *Bootstrap*	Regresi PCA	Interval Kepercayaan *Bootstrap*	Interval Kepercayaan PCA
β_0	50	182,8	212.120	(173,4 ;191,2)	(208,645 ; 216,040)
	100	176,8		(172,4 ;181,3)	
	150	177,3		(173,4 ;181,3)	
	200	178,5		(174,4 ; 182,6)	
β_1	50	- 0,0011	-0.0013	(-0,0021 ; - 0,0001)	(-0,00069; 0,00048)
	100	-0,001		(-0,0019 ; - 0,0001)	
	150	-0,0009		(-0,0018 ; - 0,0001)	
	200	-0,001		(-0,0018 ; - 0,0002)	
β_2	50	-0,002	-0.0031	(-0,0036 ; - 0,0004)	(-0,00150 ; 0.00133)
	100	-0,0021		(-0,0039 ; -0,0003)	
	150	-0,0019		(-0,0035 ; -0,0003)	
	200	-0,0022		(-0,0042 ; -0,0002)	

Berdasarkan Tabel 9, Tampak bahwa hasil estimasi titik regresi komponen utama tidak berbeda jauh dengan hasil rataan bootstrap untuk setiap replikasi (50, 100, 150,200). Dan hasil estimasi interval kepercayaan regresi komponen utama tidak berbeda jauh dengan interval kepercayaan *bootstrap*. Hal ini menunjukkan bahwa penerapan regresi komponen utama dan interval kepercayaan *bootstrap* untuk analisis data metrik kinerja *facebook* menunjukkan bahwa semua faktor, dalam hal ini komponen utama kedua, yaitu variabel *Page total like* (x_1), *Type* (x_2), *Category* (x_3) , *Post weekday* (x_4) , *Post Hour, Paid* (x_6) , *Lifetime Post Consumptions*(x_9) dan *Lifetime People who have liked your Page and engaged with your post* (x_{12}) dan komponen utama kedua yaitu *Lifetime Post Total Reach*(x_7), *Lifetime Post Total Impressions*(x_8), *Lifetime Post Impressions by people who have liked your Page* (x_{10}) Signifikan untuk mengukur data yang mengandung multikolinearitas pada data metrik *facebook.*

KESIMPULAN

Kesimpulan Berdasarkan hasil penelitian didapatkan bahwa:

1. Masalah multikolinearitas yang terdapat pada data Metrik kinerja *facebook* dapat diselesaikan menggunakan metode *Principal Component Analysis (PCA*) dengan proporsi varians sebesar 96 % dan cukup digunakan dua buah komponen utama, yaitu komponen utama pertama dan komponen utama kedua (PC1 dan PC2) yang telah mampu menerangkan struktur data. Dengan demikian, untuk analisis selanjutnya hanya digunakan dua buah komponen, sehingga dapat dinyatakan dalam persamaan berikut :

$$\hat{Y} = 212.120 + (-0.0013)F_1 + (-0.003)F_2$$

2. Penggunaan Metode *Principal Component Analysis (PCA*) dengan *Bootstrap* menghasilkan nilai estiimasi interval kepercayaan *bootstrap* yang dibangun berdasarkan *resampling* skor komponen utama sebagai variabel bebas dan variabel terikat secara berpasangan. Hasil estimasi titik dan interval kepercayaan klasik yang diterapkan pada regresi komponen utama tidak berbeda signifikan dengan hasil estimasi rataan dan interval kepercayaan *bootstrap*, dimana rataan *bootstrap* berada dalam interval kepercayaan regresi komponen utama klasik dan estimasi titik regresi komponen utama berada dalam interval kepercayaan *bootstrap.*

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada semua pihak yang telah membantu dan memberikan masukan serta arahan kepada penulis dalam penyelesaian artikel ini.

DAFTAR PUSTAKA

A.Conlin, E. Martin, A. Morris (2000)Confidence limits for contribution plots, Journal of Chemometrics,725–736.

Bawono, A. & Shina, A. (2018). Ekonometrika Terapan. Salatiga: LP2M IAIN Salatiga.

Bronman L. Do Almetrics Point to the Broader Impact of Research? An Overview of Benefits and Disadvantages of Almetrics. Journal of Infometrics, 8 (4), 2014.

DiCiccio dan Efron. (1996). Bootstrap confidence intervals. Stat.Sci. 11(3):189–212.

Efron, B., dan Tibshirani, R. J., (1993): An Introduction to the Bootstrap, Chapman and Hall, New York.

Fisher,N.I.(1995).Statistical Analysis of Circular Data. Cambridge: Cambridge University Press.

Hall, P. (1988). On Symetric Bootstrap Confidence Intervals. Journal of the Royal Statistical Society. Series B (methodological), 50 : 35– 45.

Hillier, F. S. & Lieberman, G. J. (2001). Introduction to Operations Research, Seventh Edition, The McGraw-Hill Companies, Inc., New York.

Iskandar, R., M.N.Mara dan N.Satyahadewi.(2013).Perbandingan Metode Bootstrap dan Jackknife dalam Menaksir Parameter Regresi untuk Mengatasi Multikolinearitas. Buletin Ilmiah Matematika Statistika dan Terapannya (Bimaster), 2 (2) : 137-146.

Jollife, I.T. (2002). Principal Component Analysis, Second Edition. New York : Springer.

Marzuki,dkk. (2010).Pendugaan Selang Kepercayaan Persentil Bootstrap Nonparametrik untuk Parameter Regresi ,Jurnal Statistika.10(1) :13-23.

Mega Kusuma.,F.Arief.W. (2017). Principal Component Analysis (PCA) untuk mengatasi multikolinearitas terhadap faktor angka kejadian Pneumonia balita di jawa timur. Jurnal Biometrika dan Kependudukan.6(2):89-97.

Montgomery, D.C. & E.A. Peck. 1991. Introduction to Linear Regression Analysis, Second Edition. New York: John Wiley and Sons, Inc.

Sahinler, S., dan Topuz, D., (2007): Bootstrap and Jackknife Resampling Algorithms for Estimation of Regression Parameters, JAQM, (2), 188-199.

Soemartini. (2008). Principal Component Analysis sebagai salah Satu metode mengatasi masalah multikolinearitas. Jatinangor: FMIPA Universitas Padjadjaran.

Solih,A. (2015). Estimasi Confidence Interval Bootstrap Untuk Analisis Data Sampel Terbatas.9(1).

Sudjarat,A. (2016).PCA dalam mengatasi multikolinearitas pada faktor yang mempengarui AHH Penduduk Jawa Timur.Surabaya:FKM,Universitas Airlangga.hal.34-35.

Suhaeni, C.,Made,S. & Anik,D. (2012). Bootstrap Confidence Interval Estimation of Mean Direction for Circular Data,Journal of statistics.17(2) :1-8.

Sungkono, J., (2013): Resampling Bootstrap pada R, Magistra, (84 Th XXV).

Witten. I. H, Fran, E., & Hall, M. A. 2011. Data Mining Practical Machine Learning Tools and Techniques (3rd ed). USA: Elsevier. ISBN 978-0-12-374856-0.