

**REGRESI *ELASTIC NET* DENGAN PERINGKASAN LUAS
UNTUK MENGUKUR KEAKURATAN ALAT *NON-INVASIVE*
PRODUK TAHUN 2017 DAN 2019**

FARIZ MUFTI RUSDANA



**DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2021**

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Regresi *Elastic Net* dengan Peringkasan Luas untuk Mengukur Keakuratan Alat *Non-Invasive* produk tahun 2017 dan 2019” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2021

Fariz Mufti Rusdana

NIM G14160071

ABSTRAK

FARIZ MUFTI RUSDANA. Regresi *Elastic Net* dengan Peringkasan Luas untuk Mengukur Keakuratan Alat *Non-Invasive* Produk Tahun 2017 dan 2019. Dibimbing oleh ITASIA DINA SULVIANI dan ERFIANI.

Diabetes melitus merupakan salah satu penyakit yang berbahaya karena sulit untuk disembuhkan. Oleh karena itu, pemeriksaan kadar glukosa darah harus rutin dilakukan agar tidak menaikkan tingkat keparahan penyakit diabetes melitus. Tim *Non-Invasive Biomarking* IPB sedang mengembangkan purwarupa alat ukur kadar glukosa darah secara *non-invasive*. Saat ini, tim *Non-Invasive Biomarking* IPB sudah menciptakan 2 alat, yaitu produk tahun 2017 dan tahun 2019 dengan output berbentuk spektrum residu intensitas terhadap *time-domain*. Oleh karena itu, dibutuhkan pemodelan kalibrasi untuk memprediksi kadar glukosa darah. Metode pemodelan kalibrasi terbaik pada produk tahun 2017 ditemukan oleh Herianti (2020) dengan menggunakan regresi *elastic net* dengan algoritma *detect deviating cell* (DDC) untuk mengatasi pencilan. Pada tahun 2019 dilakukan pengukuran kadar gula darah *non-invasive* dengan alat yang berbeda. Penelitian ini bertujuan untuk menentukan alat yang lebih stabil dalam mengukur kadar gula darah secara *non-invasive* dari 2 alat yang tersedia, serta untuk menentukan metode peringkasan dari spektrum residu intensitas yang lebih akurat. Alat yang lebih stabil untuk mengukur kadar gula darah secara *non-invasive* adalah produk tahun 2017. Peringkasan pada penelitian ini menggunakan pendekatan peringkasan luas trapesium dan peringkasan 3 angka. Hasil penelitian menunjukkan bahwa, kedua metode peringkasan tidak memiliki perbedaan akurasi yang signifikan.

Kata kunci : algoritma *detect deviating cell*, *elastic net*, *non-invasive*, peringkasan 3 angka, peringkasan luas

ABSTRACT

FARIZ MUFTI RUSDANA. Elastic Net Regression with Area Summarization for measuring the accuracy of Non-Invasive device 2017's and 2019's products. Supervised by ITASIA DINA SULVIANI and ERFIANI.

Diabetes melitus is one of dangerous disease because it's hard to be cured. This is shows it's important for everyone to always control and checking their blood glucose levels to prevent make the diabetes melitus is getting worse. Non-invasive biomarking team from IPB currently developing blood glucose device measurement with non-invasive method. Now, the non-invasive biomarking team from IPB already created 2 products, design product for 2017's and 2019's with the output in the form of a residual intensity spectrum with respect to the time-domain. Therefore, calibration modeling is needed to predict blood glucose level. The best calibration modeling method for 2017's device discovered by Herianti (2020) with elastic net regression and DDC algorithm for resolve the outlier. In 2019, measuring the blood glucose level were using different tools. This research aims to determine a more stable tool for measuring the blood glucose level with non-invasive method from 2 available tools, and to determine a more accurate summarization method of the intensity residual spectrum. More stable tool for measuring the blood glucose level is a 2017's device. The summarization method in this research uses a trapezoidal area and 3 digit summarization approach. The result showed that the 2 summarization method didn't have a significant different in accuracy.

Keywords : 3 digit summarization, area summarization, , detect deviating cell, elastic net, non-invasive

© Hak Cipta milik IPB, tahun 2021¹
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

**REGRESI ELASTIC NET DENGAN PERINGKASAN LUAS
UNTUK MENGUKUR KEAKURATAN ALAT *NON-INVASIVE*
PRODUK TAHUN 2017 DAN 2019**

FARIZ MUFTI RUSDANA

Skripsi
Sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Statistika dan Sains Data

**DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2021**

Tim Penguji pada Ujian Skripsi:

- 1 Nama lengkap dan gelar
- 2 Nama lengkap dan gelar

Judul Skripsi: Regresi *Elastic Net* dengan Peringkasan Luas untuk Mengukur
Keakuratan Alat *Non-Invasive* Produk Tahun 2017 dan 2019

Nama : Fariz Mufti Rusdana

NIM : G14160071

Disetujui oleh

Pembimbing 1:

Dra. Itasia Dina Sulvianti, M.Si

Pembimbing 2:

Dr. Ir. Erfiani, M.Si

Diketahui oleh

Ketua Departemen:

Dr. Anang Kurnia, M.Si.

NIP. 197308241997021001

Tanggal Ujian:

Tanggal Lulus:

PRAKATA

Puji syukur penulis panjatkan kepada Allah SWT yang telah memberikan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan karya ilmiah dengan judul “Regresi *Elastic Net* dengan Peringkasan Luas untuk Mengukur Keakuratan Alat *Non-Invasive* Produk Tahun 2017 dan 2019”.

Terima kasih penulis ucapkan kepada semua pihak yang telah membantu dalam proses menyelesaikan karya ilmiah ini, diantaranya:

1. Ibu Dra. Itasia Dina Sulvianti, M.Si dan Ibu Dr. Ir. Erfiani, M.Si selaku komisi pembimbing yang telah memberi saran, bimbingan, dan waktunya dalam penyusunan karya ilmiah.
2. Seluruh Staf Pengajar dan Administrasi Departemen Statistika FMIPA IPB yang telah memberikan sarana penunjang selama masa perkuliahan.
3. Orang tua tercinta yang senantiasa memberikan dukungan, semangat, kasih sayang, dan doa untuk penulis.
4. Astra, Arif, Kania dan teman-teman lainnya selaku rekan sebangkuan yang bersama-sama mencurahkan waktu dan pikirannya untuk saling membantu dalam penyusunan karya ilmiah mengenai topik terkait .
5. Mahasiswa/i Departemen Statistika FMIPA IPB Angkatan 53.
6. Seluruh pihak terkait yang telah membantu penulis menyusun yang tidak dapat dituliskan satu persatu.

Penulis menyadari bahwa karya ilmiah ini tidak luput dari kekurangan maupun kesalahan. Sehubungan dengan itu, penulis sangat terbuka menerima kritik dan saran yang dapat menyempurnakan karya ilmiah ini. Semoga karya ilmiah ini dapat bermanfaat dan digunakan sebagaimana mestinya.

Bogor, Juli 2021

Fariz Mufti Rusdana

DAFTAR ISI

DAFTAR ISI	vii
DAFTAR TABEL	viii
DAFTAR GAMBAR	viii
I. PENDAHULUAN	1
1. 1 Latar Belakang	1
1. 2 Rumusan Masalah	2
1. 3 Tujuan Penelitian	2
1. 4 Manfaat	2
II. TINJAUAN PUSTAKA	3
2. 1 Alat ukur Kadar Glukosa Darah <i>Non-Invasive</i>	3
2. 2 Regresi <i>Elastic Net</i>	3
2. 3 <i>Cross-Validation</i>	4
2. 4 Kenormalan Ganda	4
2. 5 <i>Detect Deviating Cell</i>	5
III. METODE	7
3. 1 Data	7
3. 2 Prosedur Analisis Data	8
IV. HASIL DAN PEMBAHASAN	9
4. 1 Eksplorasi Data Kadar Glukosa Darah <i>Invasive</i>	9
4. 2 Eksplorasi Data Kadar Glukosa Dara <i>Non-Invasive</i>	10
4. 3 Praproses Data Residu Intensitas	11
4. 4 Deteksi Multikolinearitas	12
4. 5 Pengujian Kenormalan Ganda	13
4. 6 <i>Detect Deviating Cells</i>	14
4. 7 Pemodelan Regresi <i>Elastic Net</i>	16
V. SIMPULAN DAN SARAN	18
5. 1 Simpulan	18
5. 2 Saran	18
DAFTAR PUSTAKA	19
LAMPIRAN	21

DAFTAR TABEL

1	Matriks korelasi antar peubah X data hasil pengukuran produk tahun 2017 dengan metode peringkasan 3 angka	10
2	Matriks korelasi antar peubah X data hasil pengukuran produk tahun 2017 dengan metode peringkasan luas	11
3	Matriks korelasi antar peubah X data hasil pengukuran produk tahun 2019 dengan metode peringkasan luas	11
4	Ilustrasi koefisien hasil permodelan regresi <i>elastic net</i>	14

DAFTAR GAMBAR

1	Ilustrasi prosedur <i>k-folds cross-validation</i>	2
2	Diagram kotak garis hasil pengukuran kadar glukosa darah secara <i>invasive</i> dengan produk tahun 2017 dan 2019	7
3	Plot <i>time-domain</i> terhadap residu intensitas yang memiliki kadar glukosa sama hasil pengukuran dengan produk tahun 2017 dan tahun 2019	8
4	Ilustrasi metode peringkasan luas berdasarkan periode	9
5	Ilustrasi metode peringkasan 3 angka berdasarkan puncak	10
6	Plot kuantil khi-kuadrat data produk tahun 2017 dengan peringkasan 3 angka, tahun 2017 peringkasan luas, dan tahun 2019 peringkasan luas	12
7	Plot hasil deteksi amatan pencilan menggunakan algoritma DDC pada data dengan produk tahun 2017 peringkasan 3 angka, tahun 2017 peringkasan luas, dan tahun 2019 peringkasan luas	13
8	Diagram kotak garis RMSEP hasil permodelan	15

DAFTAR LAMPIRAN

1	<i>Corrplot</i> matriks korelasi antar peubah data pengukuran produk tahun 2017 dengan peringkasan 3 angka	19
2	<i>Corrplot</i> matriks korelasi antar peubah data pengukuran produk tahun 2017 dengan peringkasan luas	19
3	<i>Corrplot</i> matriks korelasi antar peubah data pengukuran produk tahun 2019 dengan peringkasan luas	20
4	Nilai α , λ , dan RMSEP produk tahun 2017 dengan peringkasan 3 angka	20
5	Nilai α , λ , dan RMSEP produk tahun 2017 dengan peringkasan luas	22
6	Nilai α , λ , dan RMSEP produk tahun 2019 dengan peringkasan luas	25

I PENDAHULUAN

1.1 Latar Belakang

Diabetes melitus merupakan salah satu penyakit yang sangat berbahaya karena penyakit ini sulit untuk disembuhkan. IDF Diabetes Atlas Edisi ke-9 (2019) menyatakan bahwa sebanyak 463 juta jiwa mengidap penyakit diabetes melitus dan 4,2 juta jiwa di antaranya meninggal dunia. WHO (2006) juga mengatakan bahwa banyaknya orang dewasa yang menderita penyakit diabetes melitus diperkirakan akan meningkat sebanyak 300 juta jiwa pada tahun 2025. Hal ini menunjukkan pentingnya bagi setiap orang untuk selalu mengontrol dan memeriksa kadar glukosa darah untuk mencegah peningkatan keparahan penyakit diabetes melitus.

Saat ini terdapat 2 cara untuk mengukur kadar glukosa darah dalam tubuh yaitu secara *invasive* dan *non-invasive*. Metode pemeriksaan kadar glukosa darah yang paling populer untuk digunakan adalah metode *invasive* yaitu pengambilan sampel darah dengan melukai tubuh pasien. Namun, metode ini dapat menimbulkan rasa ketidaknyamanan karena bersifat melukai. Biaya yang dibutuhkan untuk mengukur kadar glukosa darah juga termasuk biaya yang tidak murah. Oleh karena itu, tim *Non-Invasive Biomarking* Institut Pertanian Bogor (IPB) saat ini sedang merancang purwarupa alat ukur kadar glukosa darah secara *non-invasive*. Pemeriksaan kadar glukosa darah *non-invasive* tidak perlu melukai tubuh pasien karena metode ini menggunakan konsep spektroskopi inframerah dengan biaya yang murah.

Purwarupa alat ini sudah dikembangkan dari tahun ke tahun sampai sekarang dengan harapan mendapatkan hasil yang lebih akurat dan lebih stabil sehingga dapat menggantikan metode pemeriksaan kadar glukosa darah secara *invasive*. Saat ini, tim *Non-Invasive Biomarking* IPB sudah menciptakan 2 alat, yaitu produk tahun 2017 dan tahun 2019 dengan output berbentuk spektrum residu intensitas terhadap *time-domain*. Oleh karena itu, dibutuhkan pemodelan kalibrasi untuk memprediksi kadar glukosa darah. Model kalibrasi merupakan suatu fungsi yang menghubungkan antara suatu pengukuran yang memiliki proses yang relatif lebih mudah dan murah dengan hasil pengukuran yang memiliki proses yang lama dan lebih mahal (Naes *et al.* 2002). Berdasarkan hal ini, penulis melakukan penelitian untuk membandingkan keakuratan purwarupa alat ukur kadar glukosa darah yang telah dikembangkan oleh tim *Non-Invasive Biomarking* IPB pada tahun 2017 dan tahun 2019. Pada penelitian sebelumnya, untuk pemodelan produk tahun 2017 Herianti (2020) menemukan bahwa pemodelan menggunakan pendekatan regresi *elastic net* dan algoritma *Detect Deviating Cell* (DDC) dengan peringkasan data 3 angka memiliki keakuratan yang lebih tinggi dibandingkan dengan model lain. Akan tetapi, metode ini masing kurang efektif dikarenakan pada peringkasan data 3 angka dapat membuang sebagian informasi dari data. Oleh karena itu, penelitian ini juga akan membuat model kalibrasi menggunakan pendekatan metode regresi *elastic net* dan algoritma DDC dengan peringkasan data menghitung luasan di bawah kurva spektrum.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah menentukan alat yang paling stabil dari 2 alat yang telah diproduksi oleh tim *Non-Invasive Biomarking* IPB serta membandingkan metode peringkasan 3 angka dengan metode peringkasan luas untuk menentukan metode peringkasan yang akurat.

1.3 Tujuan Penelitian

Berdasarkan latar belakang yang telah dijelaskan, maka penelitian ini bertujuan untuk:

1. Membandingkan kestabilan purwarupa alat ukur kadar glukosa darah yang telah dikembangkan oleh tim *Non-Invasive Biomarking* IPB pada tahun 2017 dan 2019.
2. Menentukan metode peringkasan data spektrum residu intensitas yang memiliki akurasi lebih tinggi.

1.4 Manfaat

Berdasarkan tujuan penelitian, manfaat dari penelitian ini adalah memberikan rekomendasi alat yang stabil dan akurat untuk mengukur kadar gula darah secara *non-invasive*. Hasil penelitian ini juga diharapkan dapat digunakan sebagai referensi bagi penelitian-penelitian selanjutnya.

II TINJAUAN PUSTAKA

2.1 Alat Ukur Kadar Glukosa Darah *Non-Invasive*

Menurut Umar dan Amin (2019), pengukuran kadar gula darah dapat dilakukan dengan 3 metode, yaitu secara *invasive*, minimal *invasive*, dan *non-invasive*. Metode pengukuran yang paling banyak digunakan adalah metode pengukuran secara *invasive* dengan dua jenis alat ukur. Pertama, alat ukur glukometer (*silver standard*), yaitu alat ukur dengan cara pengambilan sampel darah di ujung jari telunjuk lalu darah tersebut akan diteteskan pada strip tes untuk dimasukkan ke dalam alat ukur glukometer. Alat tersebut memiliki tingkat akurasi sebesar 74%-82% (Kermani *et al.* 2017). Kedua, pengambilan sampel darah pada pembuluh vena lalu dilakukan pemeriksaan secara kimiawi di laboratorium (*gold standard*). Akan tetapi metode ini memiliki banyak kekurangan yaitu membutuhkan biaya yang besar.

Pengukuran kadar glukosa darah secara *non-invasive* dilakukan dengan menggunakan prinsip spektroskopi inframerah. Alat ini bekerja dengan melewatkan sinar inframerah melewati sampel dan diserap oleh sensor (Stuart 2004). Sensor yang ditangkap akan melewati tahap konversi dan transformasi terlebih dahulu sebelum akhirnya mendapatkan spektrum intensitas residu terhadap *time-domain*. Setelah itu, hasil spektrum akan dimodelkan untuk memprediksi kadar glukosa darah.

2.2 Regresi *Elastic Net*

Metode regularisasi merupakan metode pendugaan dengan menyusutkan koefisien regresi peubah yang saling berkorelasi untuk mengatasi masalah multikolinearitas (Sundberg 2000). Regresi *elastic net* merupakan teknik regularisasi baru yang dapat menyusutkan koefisien regresi tepat nol (Zhou and Hastie 2005). Penyusutan koefisien regresi dilakukan pada koefisien yang memiliki korelasi sehingga dapat digunakan untuk menyeleksi peubah penjelas agar model yang didapatkan lebih ringkas. Penalti bagi regresi *elastic net* merupakan gabungan dari penalti regresi gulud dan regresi LASSO. Menurut Zhou dan Hastie (2005) pendugaan koefisien regresi pada regresi gulud akan menambahkan penalti:

$$\sum_{j=1}^p \beta_j^2 \leq t, \quad t \geq 0$$

dan LASSO menambahkan penalti:

$$\sum_{j=1}^p |\beta_j| \leq t, \quad t \geq 0$$

sehingga penalti regresi *elastic net* adalah sebagai berikut:

$$\lambda \sum_{j=1}^p (\alpha \beta_j^2 + (1 - \alpha) |\beta_j|)$$

dengan β_j merupakan koefisien regresi peubah ke- j , t adalah suatu besaran yang mengendalikan besaran penyusutan, λ adalah parameter penyusutan, α adalah koefisien regresi gulud, dan $(1 - \alpha)$ adalah koefisien regresi LASSO. Jika $\alpha = 0$, maka penalti *elastic net* akan sama dengan penalti dari regresi LASSO, sedangkan jika $\alpha = 1$ maka penalti *elastic net* akan menjadi penalti dari regresi gulud, dan jika $0 < \alpha < 1$ maka akan menjadi pinalti *elastic net*. Hastie *et al.* (2008) memperkenalkan penduga koefisien pada regresi *elastic net* sebagai berikut:

$$\hat{\beta} = \underset{\beta \in R}{\operatorname{argmin}} \left[\sum_{i=1}^t (y_i - x'_i \beta)^2 + \lambda \sum_{j=1}^p (\alpha \beta_j^2 + (1 - \alpha) |\beta_j|) \right]$$

2.3 Cross-Validation

Validasi silang (*Cross-validation/CV*) merupakan metode yang paling sederhana dan memiliki banyak kegunaan dalam penyeleksian model serta memperkirakan kesalahan prediksi (Hastie 2008). Salah satu metode validasi silang adalah *k-folds cross-validation*. Pada pengujian *k-folds cross-validation* data akan dibagi dalam k lipatan secara acak tanpa pengembalian, dimana $k-1$ lipatan akan digunakan sebagai model *training*, dan 1 lipatannya akan digunakan sebagai evaluasi hasil pemodelan. Prosedur ini akan diulang sebanyak k kali sehingga akan didapatkan model dan evaluasi hasil pemodelan sebanyak k (Raschka 2017). Berikut merupakan ilustrasi dari prosedur *k-folds cross-validation*:



Gambar 1. Ilustrasi prosedur *k-folds cross-validation*.

Ukuran kebaikan model dapat dilakukan dengan cara melihat nilai *Root Mean Square Error Prediction* (RMSEP). Semakin kecil nilai RMSEP artinya model yang diperoleh semakin sesuai (Naes *et al.* 2002). Menurut Shao dan Yadong (2004), nilai RMSEP dapat diperoleh dengan rumus:

$$\text{RMSEP}_j = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$$

Keterangan

$i = 1, 2, \dots, n$

- $j = 1, 2, \dots, k$
 y_i = nilai aktual hasil pengukuran respon ke-i
 \hat{y}_i = nilai dugaan respon berdasarkan model pada pengamatan ke-i
 n = banyaknya pengamatan pada data

2.4 Kenormalan Ganda

Asumsi kenormalan ganda adalah satu-satunya asumsi yang harus dipenuhi ketika menggunakan algoritma DDC. Oleh karena itu, dibutuhkan uji untuk menunjukkan terpenuhinya asumsi kenormalan ganda. Menurut Johson dan Wichern (2007), cara paling sederhana yang dapat digunakan untuk menguji kenormalan ganda adalah sebagai berikut:

1. Menghitung nilai d_j^2 dengan rumus

$$d_j^2 = (x_j - \bar{x})S^{-1}(x_j - \bar{x}), j = 1, 2, \dots, n$$

Keterangan :

d_j^2 : jarak mahalanobis ke-j

x_j : vektor x ke-j

\bar{x} : vektor rata-rata x

S^{-1} : invers dari matriks ragam peragam

n : banyaknya amatan

Keputusan untuk mengatakan bahwa data menyebar normal ganda adalah jika terdapat lebih dari 50% nilai dari d_j^2 yang kurang dari atau sama dengan $\chi_{0.5, p}^2$ dengan p adalah banyaknya peubah penjelas

2. Selain itu, dapat dilihat melalui plot kuantil χ^2 antara jarak mahalanobis dengan plot kuantilnya. Data dapat dikatakan mengikuti sebaran normal ganda apabila plot kuantil χ^2 dengan jarak mahalanobis mengikuti garis lurus.

2.5 Detect Deviating Cell

Pencilan merupakan suatu data yang memiliki karakteristik yang jauh berbeda dibandingkan dengan keseluruhan data (Soemartini 2005). Menurut Wilems and Aelst (2005), metode kuadrat terkecil untuk model regresi linear merupakan metode yang sensitif terhadap pencilan dalam data. Akan tetapi, pencilan tidak dapat dihilangkan begitu saja dikarenakan pencilan dapat memberikan pengaruh positif maupun negatif dalam pengambilan keputusan. Algoritma *Detect Deviating Cell* (DDC) merupakan salah satu solusi atas permasalahan ini. Algoritma DDC digunakan untuk mendeteksi dan mengatasi pencilan yang dikembangkan oleh Rousseeuw dan Van den Bossche (2018). Algoritma ini menggunakan pendekatan korelasi antar peubah untuk mendeteksi dan memprediksi data yang menyimpang dalam setiap sel-sel pada matriks data peubah ganda. Rousseeuw dan Van den Bossche (2018) menjelaskan bahwa terdapat beberapa langkah dalam algoritma DDC yaitu:

1. Standarisasi matriks pada setiap kolom j dari matriks Y

Pertama, akan dilakukan pendugaan *robLoc* atau penduga kekar lokasi serta *robScale* atau penduga kekar skala untuk setiap kolom y_j . Pendugaan *robust estimator* dapat dihitung dengan rumus:

$$m_j = \text{robLoc}_j(Y) = \frac{\sum_{i=1}^n w_i y_{ij}}{\sum_{i=1}^n w_i}$$

dengan bobot yang diberikan:

$$w_i = W\left(\frac{y_i - m_{1j}}{s_1}\right)$$

dimana m_1 dan s_1 didapatkan dengan rumus:

$$m_{1j} = \text{med}_{i=1}^n(y_{ij})$$

$$s_{1j} = \text{med}_{i=1}^n |y_{ij} - m_{1j}|$$

W merupakan fungsi pembobot Tukey's Biweight:

$$W(t) = \left(1 - \left(\frac{t}{c}\right)^2\right)^2$$

dengan kondisi ($|t| \leq c$) dan nilai konstanta $c = 3$. Apabila $|t| = 0$, maka $W(t) = 0$.

Pendugaan skala kekar akan menggunakan fungsi $\rho(t) = \min(t^2, b^2)$ dengan $b = 2,5$. Nilai dari s_2 akan diduga dengan $s_2 = \text{med}_j(|y_{ij}|)$, selanjutnya penduga skala kekar akan dihitung menggunakan rumus:

$$s_j = \text{robScale}_j(y_{ij}) = s_2 \sqrt{\frac{1}{\delta} \text{ave}_{i=1}^n \rho\left(\frac{y_{ij}}{s_2}\right)}$$

dengan nilai konstanta $\delta = 0,845$. Matriks Z merupakan hasil standarisasi dari matriks Y .

$$z_{ij} = \frac{x_{ij} - m_j}{s_j}$$

2. Mendeteksi pencilan pada setiap peubah

Setelah melakukan standarisasi untuk setiap kolom, akan ditetapkan matriks U dengan entri-entri diperoleh dari:

$$u_{ij} = \begin{cases} z_{ij} & \text{if } |z_{ij}| \leq c \\ NA & \text{if } |z_{ij}| > c \end{cases}$$

Nilai c merupakan nilai *cut-off*:

$$c = \sqrt{\chi_{1,p}^2}$$

$\chi^2_{1,p}$ merupakan persentil ke-99 dari sebaran khi-kuadrat berderajat bebas 1.

3. Menghitung korelasi antar peubah

Perhitungan korelasi dilakukan untuk seluruh variable dengan rumus:

$$cor_{jh} = robCorr_i(u_{ij}, u_{ih})$$

peubah $j \neq h$.

RobCorr merupakan metode korelasi kekar yang diperoleh dari:

$$robCorr_i(u_{ij}, u_{ih}) = \hat{\rho}_{jh} = \frac{(robScale_i(z_{ij} + z_{ih}))^2 - (robScale_i(z_{ij} - z_{ih}))^2}{4}$$

Jika nilai $|cor_{jh}| \geq 0.5$, maka pasangan peubah (j,h) disebut variabel *connected*. Sebaliknya, jika nilai $|cor_{jh}| < 0.5$ maka pasang peubah (j,h) disebut dengan variabel *standalone*. Selanjutnya, pasangan peubah (j,h) yang termasuk variabel *connected* akan dihitung dugaan kemiringan tanpa intersep dengan menggunakan rumus penduga kemiringan kekar:

$$b_{jh} = robSlope_i(u_{ij}|u_{ih}) = med_{i=1}^n \left(\frac{u_{ij}}{u_{ih}} \right)$$

4. Memprediksi setiap entri pada matriks U

Prediksi setiap entri pada matriks U akan menggunakan besarnya korelasi dan kemiringan yang sudah diperoleh dari langkah ke-3. Nilai hasil prediksi akan menghasilkan matriks \hat{Z} dengan rumus sebagai berikut:

$$\hat{z}_{ij} = \frac{\sum_{h \in H_j} |\hat{\rho}_{jh}| b_{jh} u_{jh}}{\sum_{h \in H_j} |\hat{\rho}_{jh}|}$$

dimana H_j merupakan himpunan pasangan peubah variable *connected*.

5. Pada tahap ini akan mengganti seluruh nilai prediksi \hat{z}_{ij} dengan nilai dari $a_j \hat{z}_{ij}$. a_j merupakan hasil dari regresi antara observasi z_j dengan hasil prediksi \hat{z}_j .

$$a_j = robSlope_{i'}(z_{i'j}|\hat{z}_{i'j})$$

6. Menandai pencilan *cellwise*

$$r_{ij} = \frac{z_{ij} - \hat{z}_{ij}}{robScale(z_{i'j} - \hat{z}_{i'j})}$$

Sebuah entri/sel akan ditandai sebagai pencilan apabila $|r_{ij}| > c$ dimana nilai c diperoleh dari langkah ke-2. Selanjutnya membuat matriks Z_{imp} yang berisi matriks Z tetapi seluruh pencilan *cellwise* dan semua NA akan digantikan dengan nilai prediksi \hat{z}_{ij} .

7. Membuat matriks X_{imp}

Matriks X_{imp} diperoleh dengan melakukan destandarisasi pada matriks Z_{imp} dengan menggunakan persamaan pada langkah ke-1. Hasil dari matriks X_{imp} adalah output dari algoritma DDC.

III METODE

3.1 Data

Data yang digunakan dalam penelitian ini adalah data primer hasil pengukuran kadar glukosa darah secara *invasive* maupun *non-invasive* yang diambil pada tahun 2017 dan 2019. Data ini merupakan hasil dari penelitian “Pengembangan dan Uji Klinis Purwarupa Alat Pemantauan Kadar Glukosa Darah *Non-Invasive Biomarking* Institut Pertanian Bogor”. Banyaknya responden yang digunakan pada penelitian ini adalah 118 responden pada tahun 2017 dan 74 responden pada tahun 2019. Pengukuran kadar glukosa darah dilakukan secara *invasive* dan *non-invasive*. Pengukuran secara *non-invasive* menggunakan purwarupa alat yang saat ini dikembangkan. Alat ini mengukur kadar glukosa darah dengan cara menyinari jari tangan menggunakan lampu inframerah 1600 nm. Sinar yang ditembakkan sebagian intensitas akan diserap dan sebagian dilewatkan. Intensitas yang terlewatkan kemudian akan ditangkap oleh sensor. Selanjutnya, intensitas yang tertangkap akan dikonversi menggunakan *Analog Digital Converter* (ADC) menjadi tegangan digital. Setelah itu, tegangan digital akan di transformasi menggunakan algoritma *Fast Fourier Transform* dan menghasilkan spektrum residu intensitas terhadap *time-domain*. Peubah yang akan dipakai pada penelitian ini adalah data pengukuran kadar glukosa darah hasil pemeriksaan di laboratorium (*gold standard*) dan spektrum residu hasil pengukuran purwarupa alat pemantauan kadar glukosa darah *non-invasive*.

3.2 Prosedur Analisis Data

Tahapan analisis yang dilakukan pada penelitian ini dengan menggunakan data kadar glukosa darah *invasive gold standard* dan *non-invasive* desain alat tahun 2017 dan 2019 adalah:

1. Melakukan eksplorasi data hasil pemantauan kadar glukosa darah *invasive* dengan melihat diagram kotak garis dari seluruh responden.
2. Melakukan prapemrosesan data hasil pemantauan kadar glukosa darah *non-invasive* dengan tahapan sebagai berikut:
 - a. Melakukan peringkasan data dengan menghitung luas trapesium dibawah kurva spektrum residu intensitas masing-masing *time-domain* pada data hasil pengukuran produk tahun 2017 dan 2019 dengan rumus sebagai berikut:

$$X = \frac{1}{2} \sum_{i=1}^{n-1} (t_{i+1} - t_i)(y_t + y_{i+1}) \quad (17)$$

X = Nilai luas kurva spektrum 1 ulangan modulasi

t_i = *Time domain* ke- i

y_i = Nilai residu intensitas ke- i

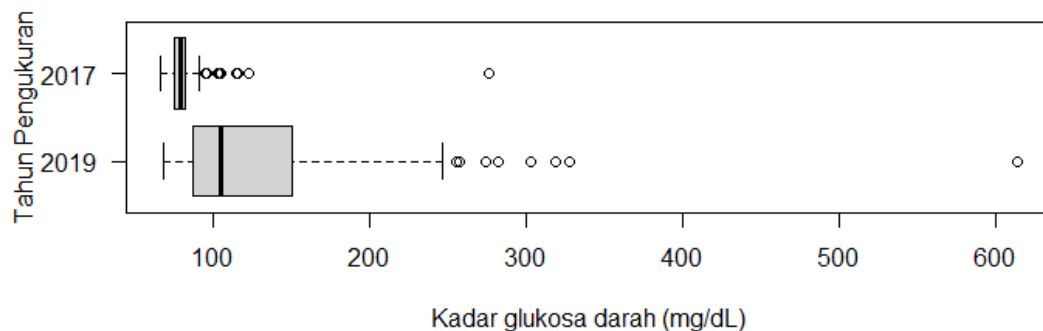
n = Banyak rentang *time domain*

- b. Melakukan peringkasan data 3 angka setiap puncak data spektrum residu intensitas masing-masing *time-domain* pada data hasil pengukuran produk tahun 2017.

3. Pemeriksaan multikolinearitas pada peubah penjelas X yang sudah didapat pada langkah (2b dan 2c) menggunakan matriks korelasi.
4. Melakukan pengujian kenormalan ganda pada data menggunakan plot kuantil χ^2 dan membandingkan nilai d_j^2 dengan nilai $\chi_{0.5, p}^2$.
5. Melakukan penerapan algoritma DDC pada peubah penjelas X . Algoritma DDC akan menghasilkan plot deteksi pencilan serta data hasil prediksi.
6. Membagi data secara acak menjadi dua bagian, yaitu 80% digunakan sebagai data *training* dan 20% sisanya digunakan sebagai data *testing*.
7. Melakukan pemodelan kalibrasi untuk memprediksi kadar glukosa darah dengan menggunakan pendekatan metode analisis regresi *elastic net* dan memilih model yang memiliki nilai λ dan α optimum dengan validasi silang lipat-10.
8. Melakukan validasi dengan menggunakan data *testing* dan diulang sebanyak 100 kali.
9. Memilih metode peringkasan terbaik dari kedua metode peringkasan pada data desain alat tahun 2017 dengan membandingkan nilai median RMSEP dari model yang telah didapat.
10. Melakukan evaluasi desain alat tahun 2017 dan 2019 dengan membandingkan median RMSEP dari setiap model.

IV HASIL DAN PEMBAHASAN

4.1 Eksplorasi Data Kadar Glukosa Darah *Invasive*

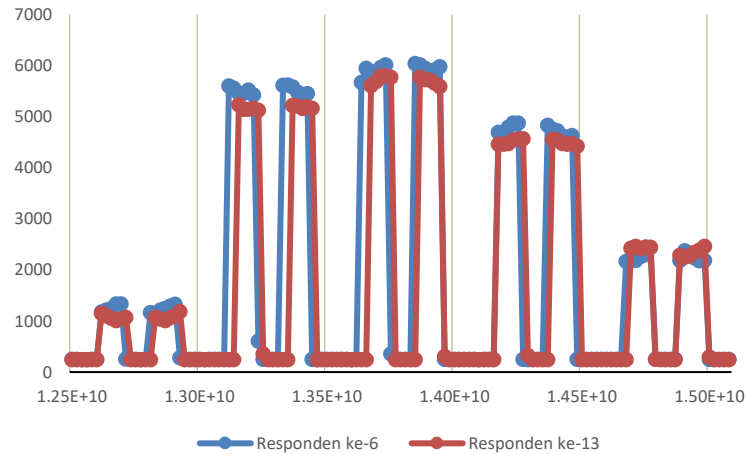


Gambar 2. Diagram kotak garis hasil pengukuran kadar glukosa darah secara *invasive* dengan produk tahun 2017 dan 2019

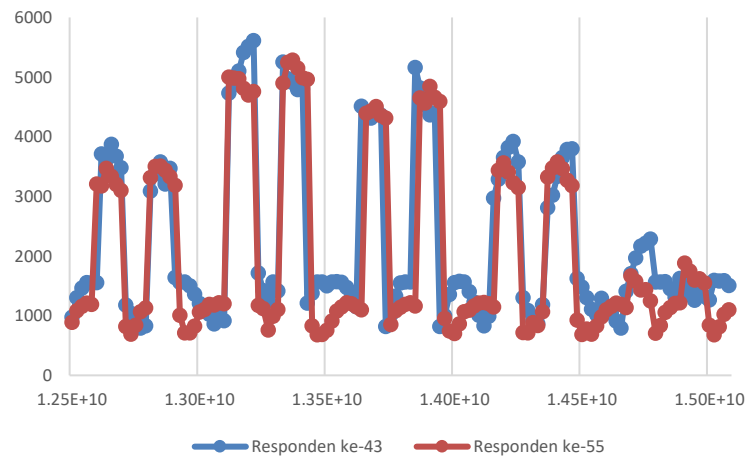
Pertama akan dilakukan eksplorasi data kadar glukosa darah tahun 2017 dan 2019 yang diperoleh melalui pengukuran *gold standard*. Eksplorasi dilakukan dengan melihat diagram kotak garis dari kedua data. Pada Gambar 1, terlihat bahwa data tahun 2017 memiliki rentang data yang lebih sempit dibandingkan data tahun 2019. Terlihat juga bahwa kedua data masing-masing memiliki pencilan dengan 1 pencilan ekstrim. Pada data tahun 2017 terdapat pencilan sebanyak 12 amatan dengan pencilan ekstrim memiliki kadar sebesar 276 mg/dL. Sedangkan pada data tahun 2019 terdapat pencilan sebanyak 8 amatan dengan pencilan ekstrim memiliki

kadar sebesar 614 mg/dL. Oleh karena itu, pada penelitian ini akan menggunakan algoritma DDC untuk mengatasi masalah pencilan pada data.

4.2 Eksplorasi Data Kadar Glukosa Darah *Non-Invasive*



(a)



(b)

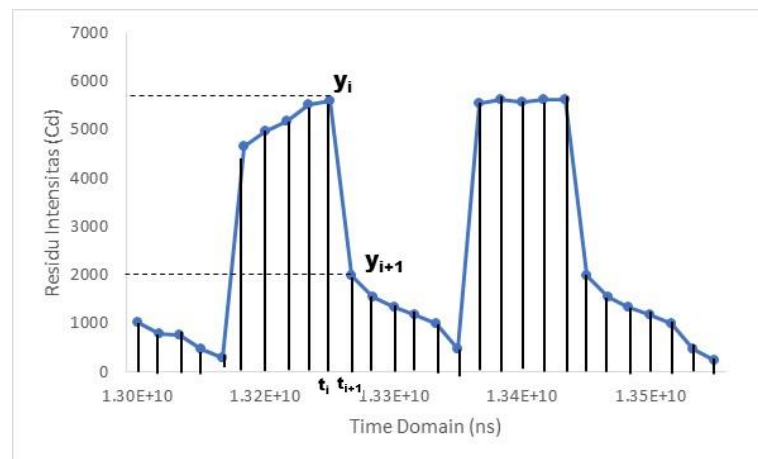
Gambar 3. Plot *time-domain* terhadap residu intensitas yang memiliki kadar glukosa sama hasil pengukuran alat (a) tahun 2017, dan (b) tahun 2019

Eksplorasi dilakukan dengan melihat grafik hasil pengukuran alat *non-invasive* tahun 2017 dan tahun 2019. Terlihat pada grafik desain alat tahun 2017 lebih stabil dibandingkan dengan desain alat tahun 2019 yang lebih fluktuatif. Pemisahan puncak data desain alat tahun 2019 akan lebih sulit dilakukan dan bersifat subjektif karena sulit untuk menentukan titik awal dari puncak modulasi. Pemisahan puncak secara subjektif ditakutkan akan mempengaruhi dari hasil analisis yang akan dilakukan. Aurelia (2020) mengatakan bahwa, pada data desain alat tahun 2019 peringkasan berdasarkan periode lebih tepat untuk digunakan. Oleh

karena itu, pendekatan peringkasan luas yang akan digunakan pada penelitian ini adalah peringkasan luas berdasarkan periode.

4.3 Praproses Data Residu Intensitas

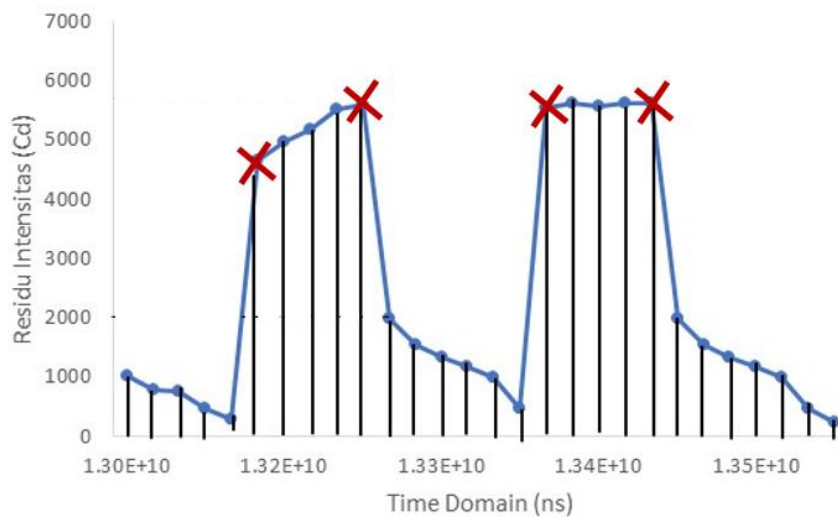
Setiap data puncak residu intensitas memiliki banyaknya data residu intensitas antara 5 sampai 7 data. Banyaknya data residu intensitas setiap puncaknya berbeda-beda, hal ini dapat menyebabkan banyaknya amatan setiap peubah X berbeda-beda. Oleh karena itu akan dilakukan peringkasan pada kedua data hasil pengukuran alat *non-invasive* tahun 2017 dan 2019. Pada penelitian ini akan dilakukan 2 metode peringkasan, yaitu peringkasan 3 angka dan peringkasan luas. Peringkasan data dengan menggunakan pendekatan luas trapesium diterapkan pada kedua dataset. Nilai grafik residu intensitas yang membentuk puncak residu intensitas yang terlihat pada Gambar 2 tidak seluruhnya terbentuk pada *time-domain* yang sama. Berdasarkan penelitian Aurelia (2020) menunjukkan bahwa peringkasan berdasarkan periode memiliki hasil yang lebih baik dibandingkan peringkasan berdasarkan puncak.



Gambar 4. Ilustrasi metode peringkasan luas berdasarkan periode

Konsep yang digunakan adalah dengan menjumlahkan beberapa luas trapesium dibawah grafik yang berada dalam satu periode yang diilustrasikan dalam gambar 4. Keunggulan dari peringkasan luas yaitu dapat memanfaatkan seluruh informasi dari data tanpa membuang satupun informasi.

Peringkasan 3 angka akan diterapkan pada data desain alat tahun 2017 untuk mencari metode peringkasan data yang terbaik. titik awal dan titik akhir pada setiap puncak akan dibuang karena dikhawatirkan mengandung *noise* pada saat sinar inframerah alat ukur *non-invasive* dinyalakan dan dimatikan. Selanjutnya, akan diambil 3 titik pada setiap puncak residu intensitas yaitu data awal, data tengah dan data akhir. Data spektrum puncak yang terdiri atas 3 titik akan diambil titik pertama, titik kedua, dan titik ketiga, apabila terdiri atas 4 titik akan diambil titik pertama, titik kedua dan titik ketiga dijumlahkan lalu dibagi dua, dan titik keempat, sedangkan jika data puncak terdiri atas 5 titik yang akan diambil adalah titik pertama, titik ketiga dan titik kelima.



Gambar 5. Ilustrasi metode peringkasan 3 angka berdasarkan puncak

Hasil dari kedua metode peringkasan ini akan membentuk 3 dataset yaitu data desain alat tahun 2017 dengan peringkasan 3 angka, data desain alat tahun 2017 dengan peringkasan luas dan data desain alat tahun 2019 dengan peringkasan luas.

4. 4 Deteksi Multikolinearitas

Pemodelan kalibrasi yang akan digunakan menggunakan model regresi sering kali menjadi masalah apabila terdapat multikolinearitas. Multikolinearitas dapat menyebabkan pendugaan *over estimate* pada regresi penduga MKT (Setiawan 2007). Akan tetapi, korelasi antar peubah yang tinggi akan membuat algoritma DDC bekerja lebih baik. Algoritma DDC melakukan pendeteksian pencilan dengan menggunakan korelasi antar peubah. Oleh karena itu, algoritma DDC akan bekerja lebih baik jika terdapat korelasi tinggi antar peubah. Pendeteksian multikolinearitas dapat dilakukan dengan melihat nilai matriks korelasi setiap peubah penjelas X . Korelasi antar peubah yang tinggi menunjukkan bahwa adanya multikolinearitas. Matriks korelasi peubah penjelas yang memiliki nilai koefisien korelasi mendekati -1 dan 1 menunjukkan adanya korelasi yang tinggi.

Tabel 1 Matriks korelasi antar peubah X data hasil pengukuran produk tahun 2017 peringkasan 3 angka

	X1	X2	X3	X4	X5	...	X150
X1	1	0.91	0.83	0.83	0.88	...	0.78
X2	0.91	1	0.94	0.84	0.88	...	0.84
X3	0.83	0.94	1	0.79	0.82	...	0.80
X4	0.83	0.84	0.79	1	0.89	...	0.80
X5	0.88	0.88	0.82	0.89	1	...	0.78
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
X148	0.76	0.81	0.78	0.80	0.78	...	0.91
X149	0.79	0.85	0.83	0.78	0.79	...	0.94

X150 0.78 0.84 0.80 0.80 0.78 ... 1

Tabel 2 Matriks korelasi antar peubah X data hasil pengukuran produk tahun 2017 dengan peringkasan luas

	X1	X2	X3	X4	X5	...	X25
X1	1	0.72	0.71	0.72	0.79	...	0.78
X2	0.72	1	0.82	0.79	0.76	...	0.78
X3	0.71	0.82	1	0.84	0.80	...	0.82
X4	0.72	0.79	0.84	1	0.89	...	0.87
X5	0.79	0.76	0.80	0.89	1	...	0.94
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
X23	0.65	0.75	0.80	0.84	0.83	...	0.83
X24	0.72	0.81	0.83	0.88	0.86	...	0.88
X25	0.78	0.78	0.82	0.87	0.94	...	1

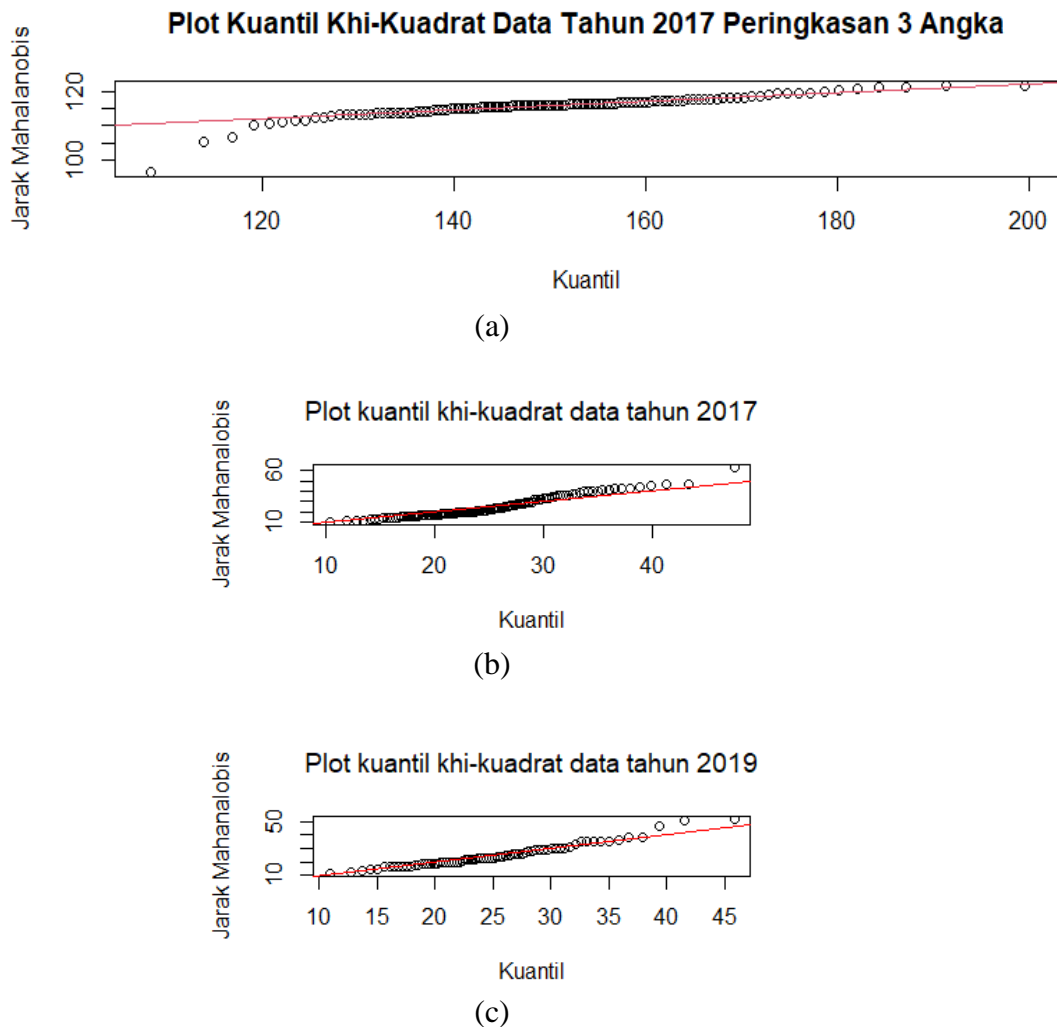
Tabel 3 Matriks korelasi antar peubah X data hasil pengukuran produk tahun 2019 dengan peringkasan luas

	X1	X2	X3	X4	X5	...	X25
X1	1	0.93	0.91	0.87	0.64	...	0.14
X2	0.93	1	0.93	0.89	0.57	...	0.03
X3	0.91	0.93	1	0.87	0.53	...	-0.05
X4	0.87	0.89	0.87	1	0.69	...	0.13
X5	0.64	0.57	0.53	0.69	1	...	0.59
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
X23	0.82	0.80	0.75	0.78	0.66	...	0.47
X24	0.60	0.56	0.45	0.57	0.75	...	0.76
X25	0.14	0.03	-0.05	0.13	0.59	...	1

Tabel 1, 2 dan 3 menunjukkan bahwa korelasi antar peubah X memiliki nilai yang cukup tinggi karena mendekati nilai 1. Oleh karena itu, pemodelan menggunakan pendekatan regresi *elastic net* akan digunakan untuk mengatasi permasalahan multikolinearitas. Selain itu, algoritma DDC dapat bekerja lebih baik karena terdapat korelasi tinggi antar peubahnya.

4.5 Pengujian Kenormalan Ganda

Penggunaan algoritma DDC untuk mengatasi pencilan diperlukannya terpenuhi asumsi kenormalan ganda. Oleh karena itu dilakukan uji kenormalan ganda pada data peubah penjelas X pada masing-masing dataset. Berikut merupakan hasil uji asumsi kenormalan ganda:



Gambar 6. Plot kuantil khi-kuadrat data desain alat (a) tahun 2017 peringkasan 3 angka, (b) data tahun 2017 peringkasan luas, (c) data tahun 2019 peringkasan luas

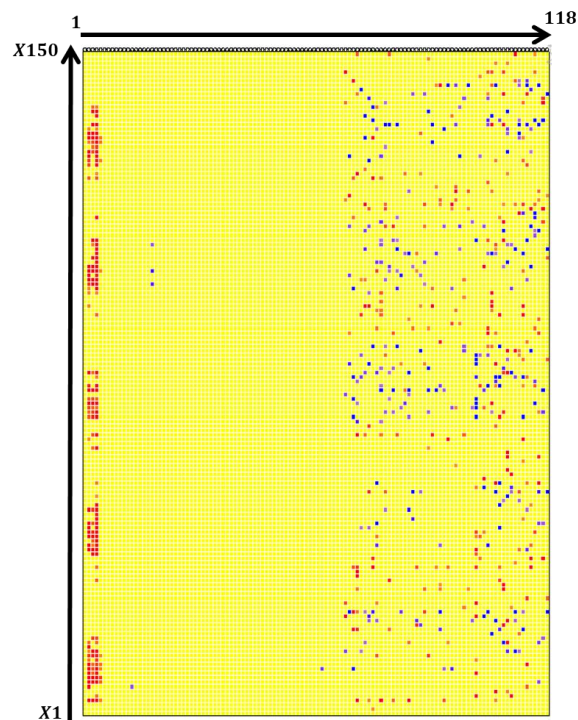
Gambar 6 memperlihatkan bahwa ketiga plot yang dihasilkan mengikuti garis lurus. Oleh karena itu, secara eksploratif dapat dikatakan bahwa kedua data penelitian mengikuti sebaran normal ganda. Berdasarkan hasil uji kenormalan ganda dengan membandingkan nilai d_j^2 dengan nilai $\chi_{0.5,p}^2$ diperoleh bahwa banyaknya nilai d_j^2 yang memenuhi kriteria keputusan adalah sebesar 100% untuk data tahun 2017 dengan peringkasan 3 angka, 60% untuk data tahun 2017 dengan peringkasan luas, sedangkan data tahun 2019 dengan peringkasan luas sebesar 59%. Nilai tersebut menunjukkan bahwa asumsi kenormalan ganda untuk ketiga data telah terpenuhi karena terdapat lebih dari 50% nilai yang memenuhi kriteria.

4.6 Detect Deviating Cells

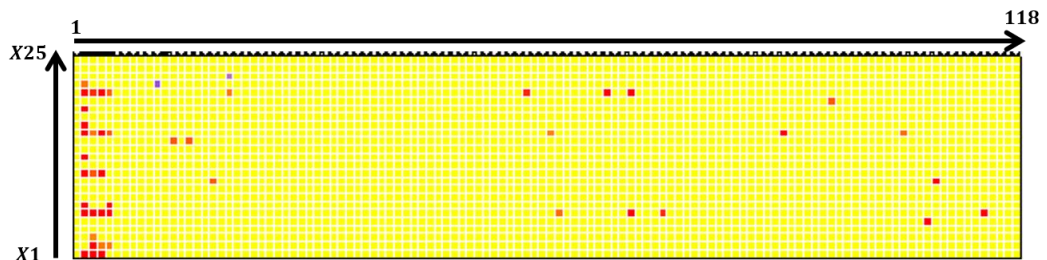
Pencilan merupakan suatu data yang menyimpang dari sekumpulan data yang lain. Umumnya pencilan dapat ditemukan dalam setiap data. Pencilan dapat terjadi karena kesalahan alat ukur, kesalahan pengumpulan data, dan kesalahan pencatatan pada saat proses pengambilan data. Dampak dari adanya pencilan dapat mempengaruhi hasil analisis, oleh karena itu deteksi pencilan sangat penting untuk

dilakukan. Namun, pendeteksian pencilan dengan peubah berdimensi tinggi sulit untuk dilakukan. Rousseeuw dan Bossche (2018) telah mengembangkan algoritma DDC untuk mendeteksi pencilan dengan peubah berdimensi tinggi serta mengatasinya.

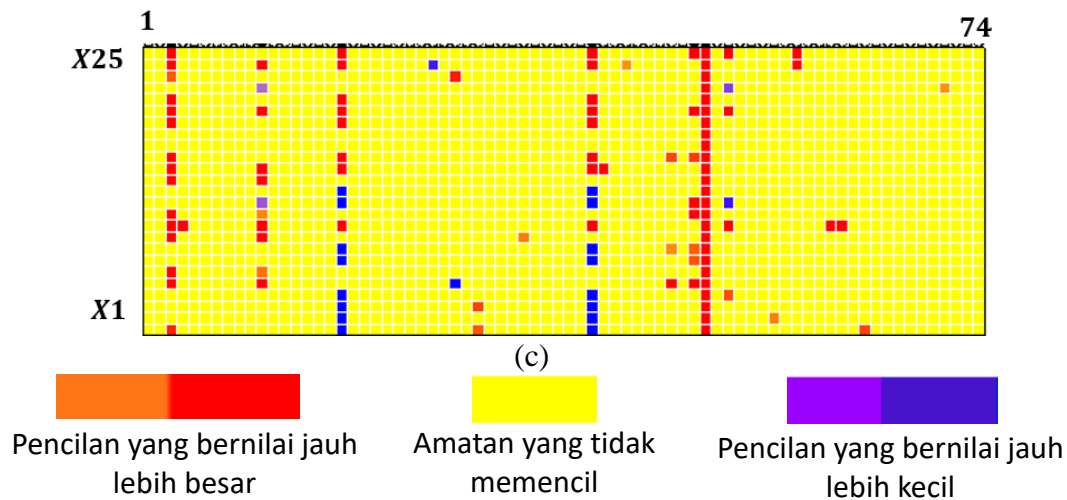
Banyaknya peubah penjelas pada penelitian ini sebanyak 25 peubah untuk alat produk tahun 2017 dan 2019 dengan peringkasan luas, sedangkan pada desain alat tahun 2017 dengan peringkasan 3 angka memiliki 150 peubah penjelas serta 1 peubah respon pada masing-masing dataset hasil pengukuran *invasive*. Banyaknya amatan pada data desain alat tahun 2017 adalah sebanyak 118 amatan dan 74 amatan pada desain alat tahun 2019. Algoritma DDC akan diterapkan pada seluruh peubah penjelas X , hal ini dikarenakan apabila melakukan pendeteksian serta memprediksi pada peubah respon Y akan mempengaruhi hasil pendugaan nanti. Pendeteksian pencilan akan dilakukan dengan menggunakan *software* R dan *cellWise package*.



(a)



(b)



Gambar 7. Plot hasil deteksi amatan pencilan menggunakan algoritma DDC pada data produk (a) tahun 2017 peringkasan 3 angka, (b) tahun 2017 peringkasan luas, dan (c) tahun 2019 peringkasan luas

Gambar 7 Menunjukkan plot dengan beberapa jenis warna yang berbeda yaitu merah, oranye, kuning, ungu muda, ungu tua, biru tua, dan kuning. Sel berwarna merah dan oranye menunjukkan nilai residu intensitas yang jauh lebih tinggi daripada pengamatan lainnya, sedangkan sel berwarna ungu muda, ungu tua, dan biru tua merupakan nilai residu intensitas yang jauh lebih rendah dari pengamatan lainnya. Sel berwarna kuning menunjukkan sel tidak mengandung pencilan. Terlihat bahwa terdapat pencilan pada ketiga dataset.

4.7 Pemodelan Regresi *Elastic Net*

Model kalibrasi akan dibentuk menggunakan ketiga dataset tanpa pencilan hasil prediksi menggunakan algoritma DDC. Konsep yang digunakan pada pemodelan regresi *elastic net* adalah menggabungkan penalti gulud β_j^2 dan penalti LASSO $|\beta_j|$ dengan memberikan bobot untuk setiap penaltinya. Bobot yang digunakan pada penalti gulud adalah $(1 - \alpha)$ sedangkan pada penalti LASSO adalah α , dengan nilai α berada pada kisaran $0 < \alpha < 1$. Setiap dataset akan dibagi secara acak menjadi 80% sebagai data untuk membangun model dan 20% untuk validasi model. Selanjutnya akan dilakukan *Cross Validation* (CV) dengan validasi silang *k-fold* (k lipatan) untuk mendapatkan parameter λ dan α yang paling optimum. Banyaknya lipatan yang digunakan adalah 10 lipatan dikarenakan banyaknya amatan yang dimiliki pada penelitian sedikit. Hasil pemodelan yang digunakan setelah metode CV adalah model yang memiliki nilai RMSEP yang paling kecil. Langkah pengacakan dan validasi model akan di ulang sebanyak 100x.

Pemodelan regresi *elastic net* akan menggunakan *software* R dengan *package caret* dan fungsi *glmnet*. Ilustrasi koefisien regresi *elastic net* ditunjukkan pada tabel 4. Koefisien regresi *elastic net* dengan algoritma DDC mengalami penyusutan dan terdapat beberapa peubah yang mengalami penyusutan hingga sampai 0.

Tabel 4 Ilustrasi koefisien hasil permodelan regresi *elastic net*

Peubah yang terseleksi	Koefisien Regresi <i>Elastic Net</i>		
	Data Desain Alat Tahun 2017	Data Desain Alat Tahun 2017	Data Desain Alat Tahun 2019
	Peringkasan 3 Angka	Peringkasan Luas	Peringkasan Luas
Intersept	80.11	74.20	145.44
X_3	0	0	0
X_4	9.1×10^{-3}	0	0
X_{10}	0	0	0
X_{13}	0	1.5×10^{-11}	0
X_{15}	0	0	0
X_{17}	-2.4×10^{-3}	-1.3×10^{-12}	-0.88
X_{21}	0	0	0
X_{24}	0	0	0

Persamaan model terbaik yang didapatkan untuk setiap dataset adalah sebagai berikut:

Persamaan regresi *elastic net* dengan algoritma DDC pada data desain alat tahun 2017 peringkasan 3 angka:

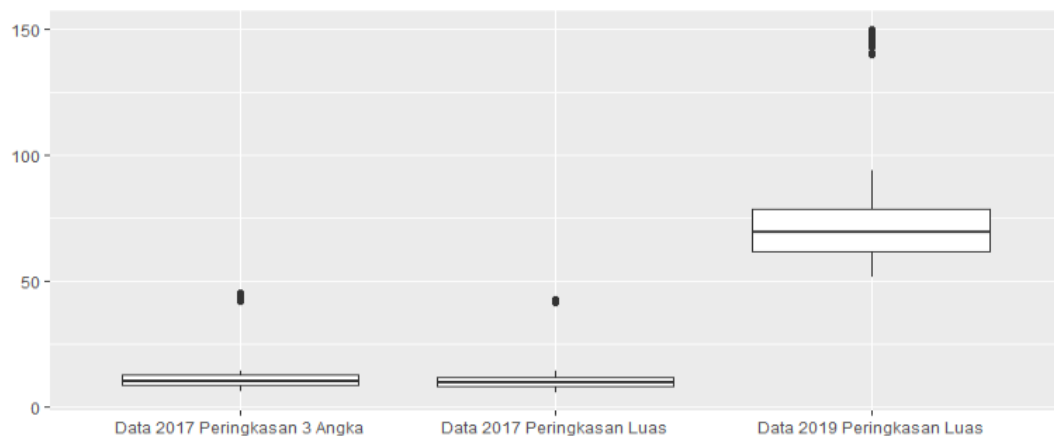
$$Y = 79.91 + 0.005X_{33} - 0.00063 X_{136}$$

Persamaan regresi *elastic net* dengan algoritma DDC pada data desain alat tahun 2017 peringkasan luas:

$$Y = 72.45 - 2.52 \times 10^{-19} X_2 + 6.42 \times 10^{-11} X_6 - 6.18 \times 10^{-12} X_{14}$$

Persamaan regresi *elastic net* dengan algoritma DDC pada data desain alat tahun 2019 peringkasan luas:

$$Y = 147.1 - 2.57X_{11} - 1.62 X_{17}$$



Gambar 8. Diagram kotak garis RMSEP hasil pemodelan

Tingkat akurasi yang akan diukur pada pemodelan ini menggunakan nilai *Root Mean Square Error Prediction* (RMSEP) setiap set data. Diagram kotak garis menunjukkan bahwa tidak terdapat perbedaan nilai RMSEP untuk data produk tahun 2017 dengan peringkasan luas maupun peringkasan 3 angka. RMSEP kedua metode peringkasan tersebut memiliki letak pencilan dan lebar kotak yang mirip. Nilai median dari RMSEP kedua metode peringkasan memiliki nilai RMSEP yang

relatif sama, yaitu 10,27 untuk data produk tahun 2017 dengan peringkasan 3 angka dan 9.58 untuk data produk tahun 2017 dengan peringkasan luas.

Terkait peringkasan luas dilihat dari diagram kotak garis pada data produk tahun 2017 dan tahun 2019 dengan peringkasan luas. Terlihat bahwa produk tahun 2019 memiliki nilai RMSEP yang berbeda dengan tahun 2017. RMSEP yang dimiliki produk tahun 2019 memiliki nilai yang jauh lebih besar. Hal ini ditunjukkan dengan letak diagram kotak garis dari produk tahun 2019 yang lebih tinggi dari produk tahun 2017 dengan nilai median dari RMSEP adalah 69.31. Produk tahun 2017 memiliki hasil yang lebih stabil dibandingkan dengan tahun 2019, dapat dilihat dari lebar kotaknya yang lebih kecil dibandingkan lebar kotak dari tahun 2019.

V SIMPULAN DAN SARAN

5.1 Simpulan

Metode peringkasan 3 angka dan metode peringkasan luas memberikan hasil yang relatif sama. Namun metode peringkasan luas memiliki keunggulan yaitu tidak adanya informasi yang terbuang karena seluruh hasil pengukuran digunakan untuk pemodelan. Sedangkan, pada peringkasan 3 angka memungkinkan adanya informasi yang terbuang karena terdapat beberapa amatan yang dibuang ketika proses peringkasan.

Produk tahun 2017 memiliki akurasi lebih akurat dibandingkan dengan produk tahun 2019 karena memiliki RMSEP yang lebih kecil. Selain itu, data 2017 juga memiliki grafik spektrum serta hasil pendugaan yang lebih stabil. Hal ini menunjukkan bahwa produk tahun 2017 memberikan hasil yang lebih baik dibandingkan dengan produk tahun 2019.

5.2 Saran

Hasil spektrum residu intensitas yang terbaca hasil pengukuran produk tahun 2017 terbaca lebih baik dibandingkan dengan yang terbaca pada produk tahun 2019. Alat yang baik adalah alat yang memiliki akurasi yang lebih tinggi dan menghasilkan pendugaan yang lebih stabil dengan keragaman yang kecil. Oleh karena itu, alat yang dapat digunakan dan dipertahankan untuk pengembangan alat selanjutnya adalah produk tahun 2017. Saran penelitian selanjutnya dapat mengkaji peringkasan data yang terbaik dengan menggunakan metode pemodelan yang berbeda.

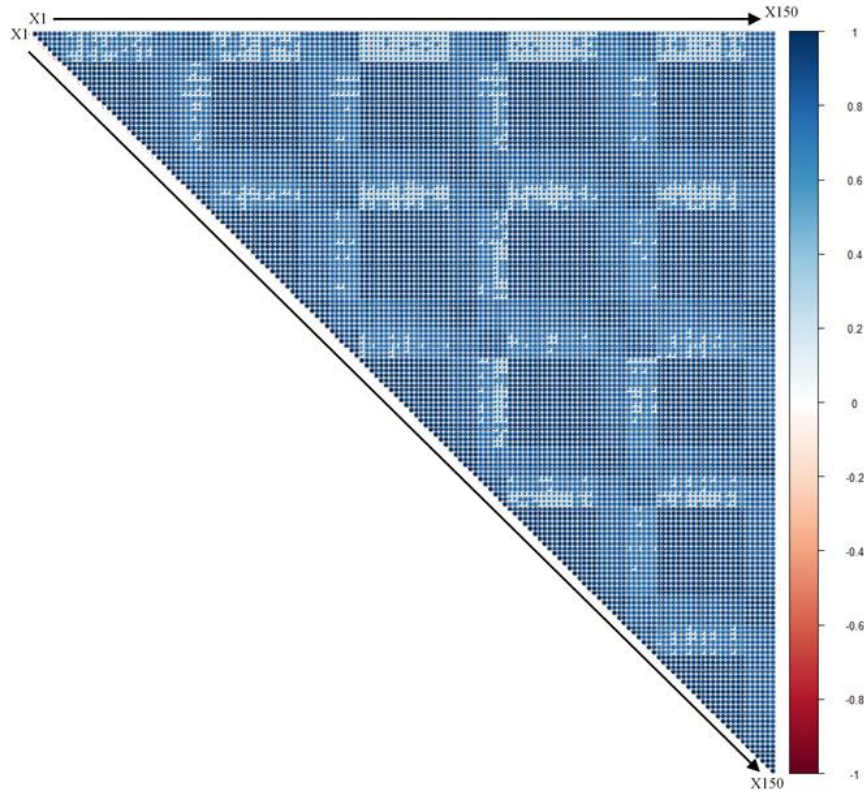
DAFTAR PUSTAKA

- Diyah NW, Ambarwati A, Warsito GM, Niken G, Eriza T. Heriwiyan, Windysari R, Prismawan D, Hartasari RF, Purwanto. 2016. Evaluasi Kandungan Glukosa dan Indeks Glikemik Beberapa Sumber Karbohidrat dalam Upaya Penggalan Pangan Ber-Indeks Glikemik Rendah. *Jurnal Farmasi dan Ilmu Kefarmasian Indonesia*. 3(2).
- Hastie T, Tibshirani R, Friedman J. 2008. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second Edition. New York (US): Springer.
- Herianti. 2020. Pemodelan Kalibrasi yang Kekar Terhadap Pencilan [thesis]. Bogor (ID): Institut Pertanian Bogor
- [IDF] International Diabetes Federation. 2019. IDF Diabetes Atlas Ninth Edition [internet]. [diunduh 2020 Jan 5]. <https://diabetesatlas.org/>.
- Johnson AR, Wichern DW. 2007. *Applied Multivariate Statistical Analysis 6th Edition*. New Jersey(US): Wiley
- Kania A. 2020. Pendugaan kadar glukosa darah *Non-Invasive* menggunakan regresi kuadrat terkecil parsial dengan beberapa pendekatan peringkasan [skripsi]. Bogor(ID): Institut Pertanian Bogor
- Kermani SK, Khatony A, Jalalil R, Rezaei M, Abdi A. 2017. Accuracy and precision of measured blood sugar values by three glucometers compared to the standard technique. *Journal of Clinical and Diagnostic Research*. 11(4): OC05-OC08. DOI: 10.7860/JCDR/2017/23926.9613
- Naes T, Issakson T, Fearn T, Davies T. 2002. A User-friendly Guide to Multivariate Calibration and Classification. United Kingdom: NIR Publication Chichester.
- Raschka S, Mirjalili V. 2017. Python Machine Learning Second Edition. *Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow*. Birmingham(UK): Packt Publishing, Ltd.
- Rousseeuw PJ, Van Den Bossche W. 2018. *Detecting Deviating Data Cells*. Technometrics. 60(2):135-145. doi:10.1080/00401706.2017.1340909
- Setiawan. 2007. Pendekatan Regresi Kontinum dalam Model Kalibrasi. [Tesis]. Bogor(ID): IPB University
- Shao X, Yadong Z. 2004. Determination of Chlorogenic Acid in Plant Samples by Using Near-Infrared Spectrum with Wavelet Transform Preprocessing. *Analytical Sciences*. China: The Japan Society for Analytical Chemistry.
- Soemartini. 2007. Pencilan(Outlier). Jatinangor(ID): Universitas Padjajaran

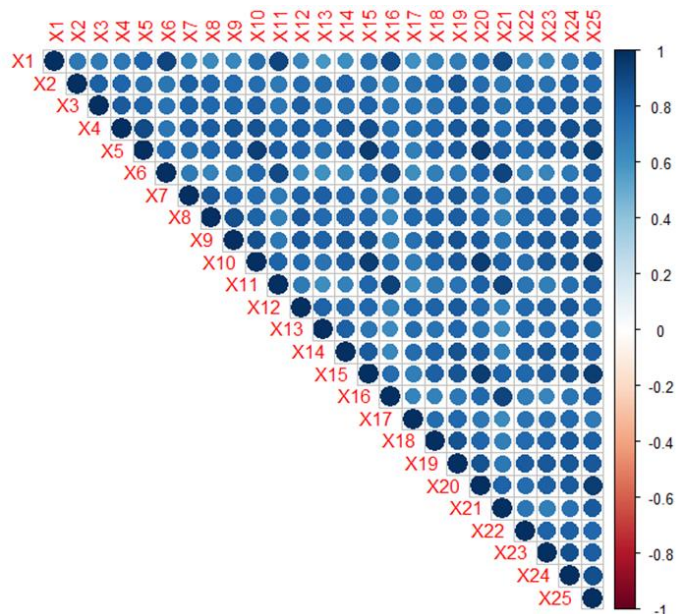
- Stuart B. 2004. *Infrared Spectroscopy: Fundamentals and Applications*. New Jersey (US): John Wiley and Sons, Ltd.
- Sundberg R. 2000. Shrinkage Regression. *Encyclopedia of Environmetrics*. 4:1994-1998
- Suri NNRR, Murty MN, Athitan G. *Outlier Detection: Techniques and Applications*. Switzerland: Springer
- Umar U, Amin I. 2019. Monitoring Kadar Glukosa Darah *Non-Invasive* Menggunakan Sensor Photoacoustic. *Celebes Health Journal*. 1(2): 99-111. Sulawesi (ID): Lembaga Layanan Pendidikan Tinggi (LL-DIKTI) Wilayah IX Sulawesi
- Willems G, Aelst SV. 2005. Multivariate Regression S-Estimators for Robust Estimation and Inference. *Statistica Sinica* 15:981-1001
- [WHO] World Health Organization. 2006. *Guidelines For The Prevention Management and Care of Diabetes melitus*. Kaira: EMRO Technical Publicationtns Series 32.
- Zou H, Hastie T. 2005. Regularization and variable selection via the *elastic-net*. *Journal of the Royal Statistical Society Series B*. 67(2):301-320.

LAMPIRAN

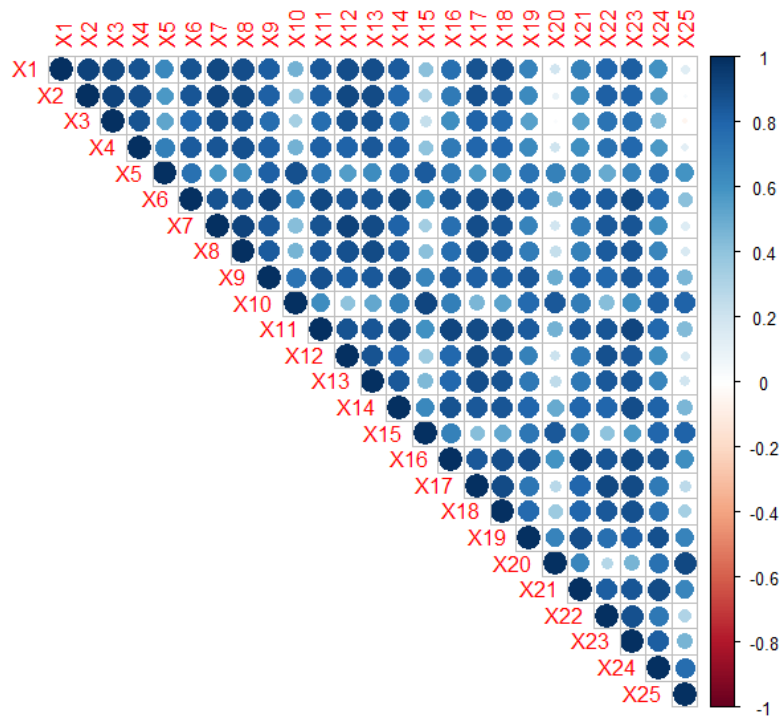
Lampiran 1 *Corrplot* matriks korelasi antar peubah data 2017 peringkasan 3 angka



Lampiran 2 *Corrplot* matriks korelasi antar peubah data tahun 2017 peringkasan luas



Lampiran 3 *Corrplot* matriks korelasi antar peubah data tahun 2019 peringkasan luas



Lampiran 4 Nilai alpha, lambda, dan RMSEP data 2017 dengan peringkasan 3 angka

Data 2017 Peringkasan 3 Angka			
Alpha	Lambda	RMSEP	
1	0.57	43.91	
0.6	8.60	11.06	
0.5	9.58	10.04	
0.6	9.22	9.59	
0.9	8.25	7.04	
0.7	9.51	11.70	
0.3	1.91	43.27	
0.7	8.63	9.54	
0.5	9.63	13.73	
0.6	9.05	9.28	
1	2.89	42.25	
0.5	8.94	8.60	
0.5	10.04	11.48	
1	2.60	42.63	
1	7.45	11.48	
0.5	8.71	7.97	
0.6	0.81	43.73	
1	0.57	43.99	

0.5	10.23	11.75
0.9	0.75	44.01
0.4	0.92	43.37
0.6	8.40	7.77
0.5	9.40	9.94
0.8	0.95	42.96
0.5	8.61	9.32
0.6	8.99	9.15
0.5	9.19	11.84
1	8.79	8.04
0.6	9.30	8.77
0.6	9.16	10.29
0.7	9.11	7.98
0.8	0.68	42.93
0.3	2.97	41.89
0.6	9.97	12.00
1	8.75	8.63
0.7	8.10	10.39
0.6	8.68	9.86
0.6	8.98	11.14
0.6	8.85	7.87
0.8	3.73	7.71
0.6	7.64	8.41
0.9	9.15	9.37
0.8	0.85	44.11
1	8.27	10.31
0.7	8.73	8.09
0.6	8.45	9.55
0.5	8.35	9.32
0.5	8.75	7.35
1	6.16	6.40
0.5	8.88	9.77
1	0.56	43.82
0.5	9.68	14.04
0.7	8.41	10.87
1	5.97	11.74
1	0.97	41.93
0.5	8.56	9.06
0.5	9.20	6.91
1	8.84	7.49
0.6	8.02	10.26
0.5	8.24	10.52
0.9	8.32	12.45
0.9	6.13	10.83

0.7	8.76	8.10
0.3	2.82	42.54
0.6	9.82	10.15
0.6	8.57	7.59
1	7.50	10.18
0.5	9.59	8.01
0.6	8.43	8.50
0.4	9.24	8.62
0.6	8.76	8.16
0.2	4.08	42.00
0.5	9.94	13.78
0.6	8.50	10.05
0.5	8.62	11.85
0.5	8.25	11.01
0.6	8.89	8.37
0.6	9.42	11.07
1	8.44	9.47
0.6	7.94	10.63
1	9.56	9.91
1	0.12	44.84
0.6	8.72	10.76
0.7	8.61	8.90
0.6	1.09	43.28
0.4	1.96	42.39
0.6	8.80	6.72
0.6	9.25	11.96
1	8.71	6.04
0.9	1.54	42.14
1	3.01	42.33
0.6	8.53	10.13
0.5	8.67	6.35
1	0.35	45.27
0.6	8.81	11.62
0.6	8.48	8.90
0.6	9.60	10.48
0.6	8.94	8.62
0.5	8.90	7.94
0.9	6.12	12.06

Lampiran 5. Nilai alpha, lambda, dan RMSEP data 2017 dengan Peringkasan Luas

Data 2017 Peringkasan Luas		
Alpha	Lambda	RMSEP

0.5	3.97	9.57
0.4	4.74	10.68
0.6	1.82	10.65
1	0.70	8.89
0.9	4.04	9.60
0.5	4.51	11.34
0.4	4.83	10.89
0.4	4.40	14.36
0.4	4.07	7.02
1	0.66	7.51
0.4	4.14	9.65
0.3	4.51	11.17
0.9	2.10	8.18
0.3	4.52	11.04
0.5	3.72	8.57
0.3	4.94	10.42
0.4	3.62	9.89
0.4	4.23	8.56
0.2	4.54	41.91
1	0.67	8.06
0.5	1.18	6.41
0.9	4.49	8.06
0.4	4.44	8.17
0.4	4.13	12.36
0.4	4.52	6.56
0.4	4.49	7.51
0.3	4.21	7.90
0.9	0.73	7.93
0.3	4.21	8.36
0.9	3.64	6.16
0.4	4.01	8.82
0.4	4.31	42.44
0.7	5.03	8.97
0.9	4.92	10.35
0.7	4.25	8.34
0.4	4.04	41.76
0.9	4.96	7.60
0.4	4.66	13.31
0.3	4.13	10.39
0.4	4.13	41.69
0.3	4.53	7.14
0.5	1.89	7.10
0.4	4.09	10.37
0.5	4.27	7.14

0.6	4.12	11.23
0.4	4.09	12.91
1	4.15	8.30
1	4.94	11.79
0.3	1.54	41.92
1	3.37	13.77
0.3	4.16	41.98
0.4	4.03	10.99
0.4	3.89	9.61
0.7	2.26	8.81
1	3.82	6.36
0.4	3.98	11.62
0.5	4.31	42.51
0.3	4.20	9.26
0.4	4.00	13.15
0.4	4.15	8.91
0.5	3.88	41.67
1	0.95	8.92
0.9	4.54	10.44
0.2	4.00	41.59
0.8	4.30	10.82
0.9	0.71	7.90
0.5	3.79	42.41
0.3	5.20	42.53
0.3	4.15	10.50
0.8	1.57	42.83
0.4	3.63	42.12
0.4	4.06	7.17
0.5	4.14	10.00
0.4	4.20	42.53
0.4	5.35	9.22
1	2.02	8.97
0.4	4.16	11.63
0.3	4.67	8.38
0.4	1.17	8.51
0.7	5.10	10.20
0.9	0.85	7.90
0.6	3.83	42.29
0.4	4.71	41.16
0.3	4.26	12.19
0.4	3.94	8.23
0.9	4.56	9.56
0.7	0.96	9.09
0.5	4.15	10.94

0.2	4.50	8.03
0.5	4.05	7.31
0.5	1.41	7.90
0.3	4.57	9.73
0.3	4.88	42.89
0.7	1.02	8.86
1	3.28	7.63
0.9	1.89	9.13
0.7	0.82	8.98
0.7	0.94	6.29
0.5	4.25	5.62
0.9	0.81	9.47

Lampiran 6. Nilai alpha, lambda, dan RMSEP data 2019 dengan Peringkasan Luas

Data 2019 Peringkasan Luas		
Alpha	Lambda	RMSEP
1	17.92	71.90
1	14.32	148.78
1	12.95	55.06
1	20.32	71.02
1	23.02	61.25
1	9.83	143.10
1	13.29	147.88
1	12.82	78.77
1	18.95	52.46
1	14.77	146.51
1	17.20	70.28
1	12.30	78.41
1	13.04	58.30
1	12.33	83.95
1	16.91	76.36
1	19.61	72.94
1	16.59	83.83
1	11.00	59.31
1	19.52	65.00
1	12.56	64.15
1	13.55	77.93
1	9.98	146.69
1	12.32	62.77
1	12.09	72.76
1	14.87	59.86
1	18.31	64.80

1	11.46	73.91
1	11.46	146.80
1	12.97	56.90
1	15.71	71.46
1	11.20	74.45
1	12.77	64.18
1	11.91	57.12
1	14.62	70.07
1	12.85	56.46
1	12.30	55.80
1	12.91	59.58
1	17.54	69.10
1	11.34	75.04
1	14.66	68.27
1	18.62	69.52
0.9	16.73	82.03
0.7	15.08	80.33
1	14.82	63.10
1	12.42	150.00
1	14.63	68.50
1	14.46	85.42
1	16.67	61.79
1	15.62	66.51
1	12.60	59.04
1	14.99	66.41
1	16.88	58.07
1	9.91	75.22
1	13.42	57.81
0.3	17.77	93.80
1	14.59	68.39
1	13.22	57.95
1	14.07	66.05
1	16.20	78.55
1	17.94	72.52
1	8.97	150.02
1	15.92	85.46
1	14.52	59.40
1	11.95	78.62
1	10.87	73.12
1	13.66	59.95
0.9	13.59	66.54
1	17.49	75.65
1	15.13	81.46
1	9.88	76.80

1	10.39	143.16
1	10.94	71.33
1	12.60	72.70
1	10.20	73.66
0.9	14.20	74.77
1	14.05	61.85
1	13.49	61.06
1	15.56	139.62
1	11.50	53.07
1	12.72	63.96
1	12.66	65.37
1	12.76	67.08
1	13.41	55.24
1	12.17	68.34
1	12.08	61.90
1	16.25	60.98
1	16.82	64.47
0.6	18.45	87.09
1	11.86	140.12
1	13.72	64.75
1	14.86	68.12
1	13.06	65.30
1	16.06	51.59
1	11.50	55.19
1	15.71	72.66
1	12.65	144.74
1	11.02	70.64
1	12.79	52.41
1	13.14	65.33
1	9.65	140.76
