

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

Fariz Mufti Rusdana¹, Itasia Dina Sulvianti^{2‡}, Erfiani³

^{1,2,3}Department of Statistics, IPB University, Indonesia
[‡]corresponding author: itasiasu@apps.ipb.ac.id

Copyright © 2022 Fariz Mufti Rusdana, Itasia Dina Sulvianti, and Erfiani. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

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.

* Received: Aug 2021; Reviewed: Sep 2021; Published: Jan 2022

1. Pendahuluan

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. 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 peubah ganda adalah sebuah fungsi matematis untuk menduga peubah respons Y berdasarkan informasi pada peubah penjelas X yang tersedia dengan tujuan untuk menemukan hubungan antara satuan pengukuran yang memiliki proses yang relatif lebih mudah dan murah dengan satuan pengukuran yang memiliki proses yang lama dan lebih mahal sehingga pengukuran yang mahal dapat diprediksi dengan cepat, murah serta akurasi yang tinggi (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.

2. Metodologi

2.1 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*.

2.2 Metode Penelitian

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:

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

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*

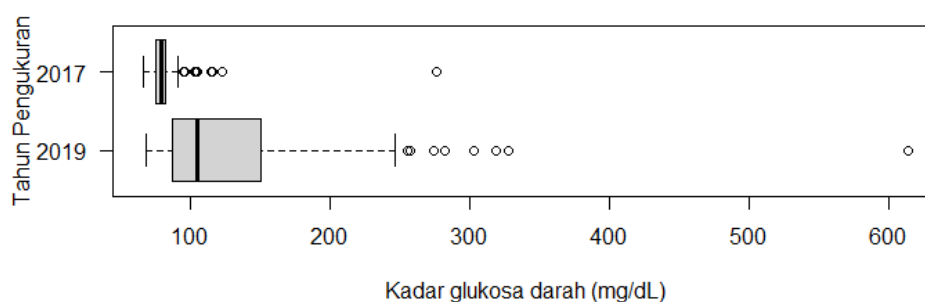
- b. Melakukan peringkasan data 3 angka setiap puncak data spektrum residu intensitas masing-masing *time-domain* pada data hasil pengukuran produk tahun 2017.
3. Pemeriksaan multikolinearitas pada peubah penjelas X yang sudah didapat pada langkah (2b dan 2c) menggunakan matriks korelasi.
4. Melakukan 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.

3. Hasil dan Pembahasan

3.1 Eksplorasi Data Kadar Glukosa Darah *Invasive*

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.

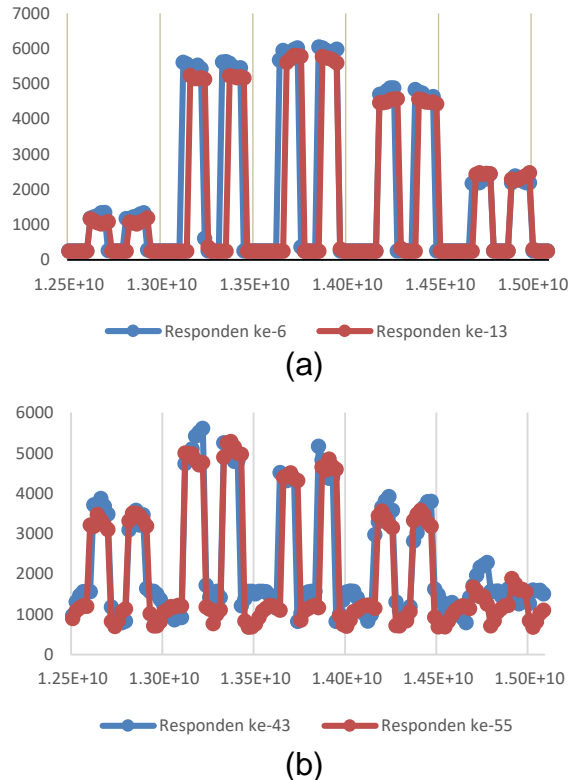


Gambar 1 Diagram kotak garis hasil pengukuran kadar glukosa darah secara *invasive* dengan produk tahun 2017 dan 2019

3.2 Eksplorasi Data Kadar Glukosa Darah *Non-Invasive*

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. Pada data desain alat tahun 2019 terlihat bahwa kedua kadar glukosa darah yang sama tidak

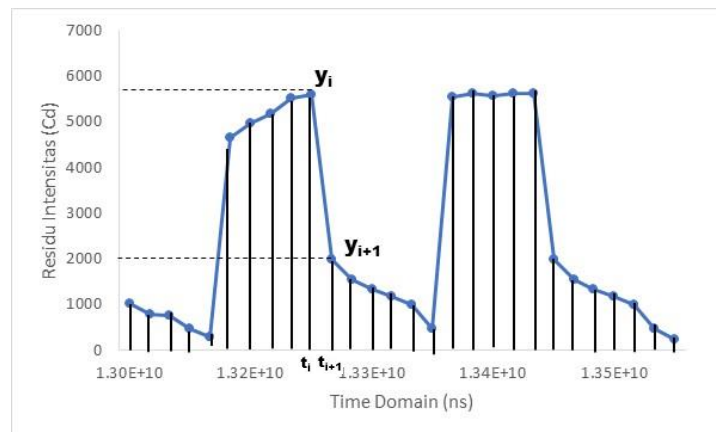
menghasilkan grafik spektrum residu intensitas yang sama. Aurelia (2020) mengatakan bahwa, pada data desain alat tahun 2019 peringkasan berdasarkan periode lebih tepat untuk digunakan. Oleh karena itu, pendekatan peringkasan luas yang akan digunakan pada penelitian ini adalah peringkasan luas berdasarkan periode.



Gambar 2 Plot *time-domain* terhadap residu intensitas yang memiliki kadar glukosa darah yang sama hasil pengukuran alat (a) tahun 2017, dan (b) tahun 2019

3.3 Praproses Data Residu Intensitas

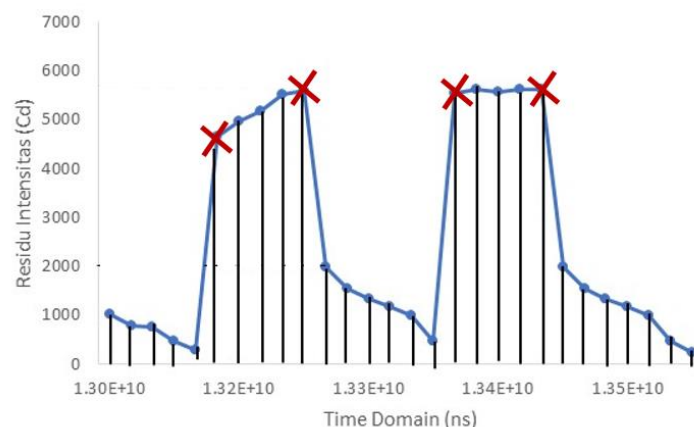
Setiap data puncak residu intensitas memiliki banyaknya data residu intensitas antara 5 sampai 7 data. Banyaknya data residu intensitas setiap puncaknya berbeda-beda, hal ini dapat menyebabkan banyaknya amatan setiap peubah X berbeda-beda. Oleh karena itu akan dilakukan peringkasan pada kedua data hasil 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 3 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 4 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.

3.4 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, 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.

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 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
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
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

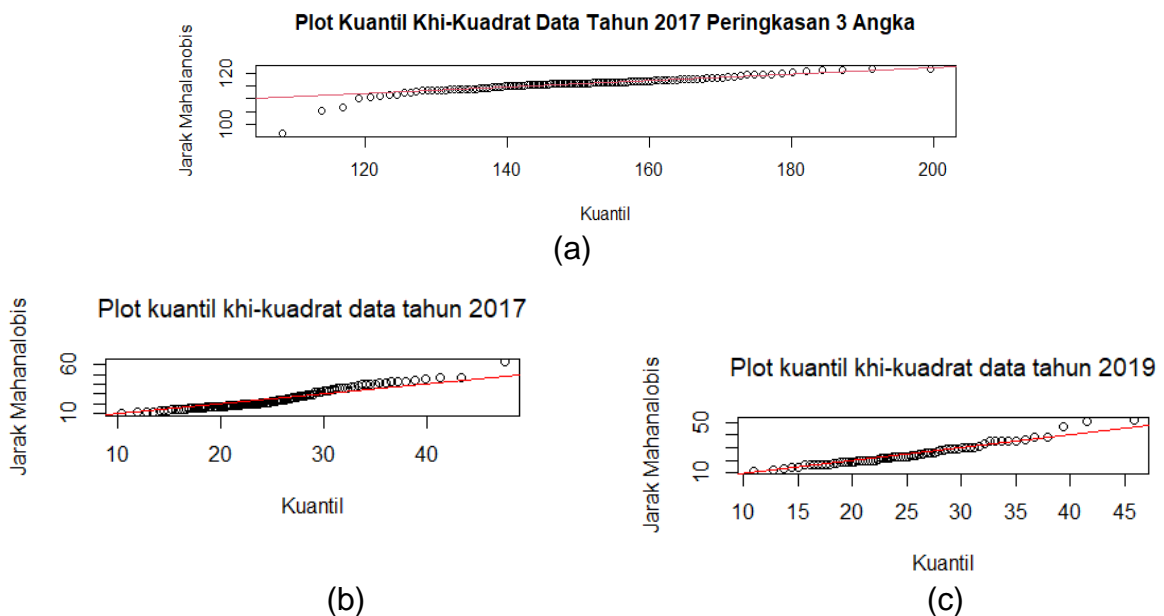
Tabel 3 Matriks korelasi antar peubah X data hasil pengukuran produk tahun 2019 peringkasan luas

	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
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
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

3.5 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.

Gambar 5 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 $\chi_{0,5}^2$ berderajat bebas p pada masing-masing dataset 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.

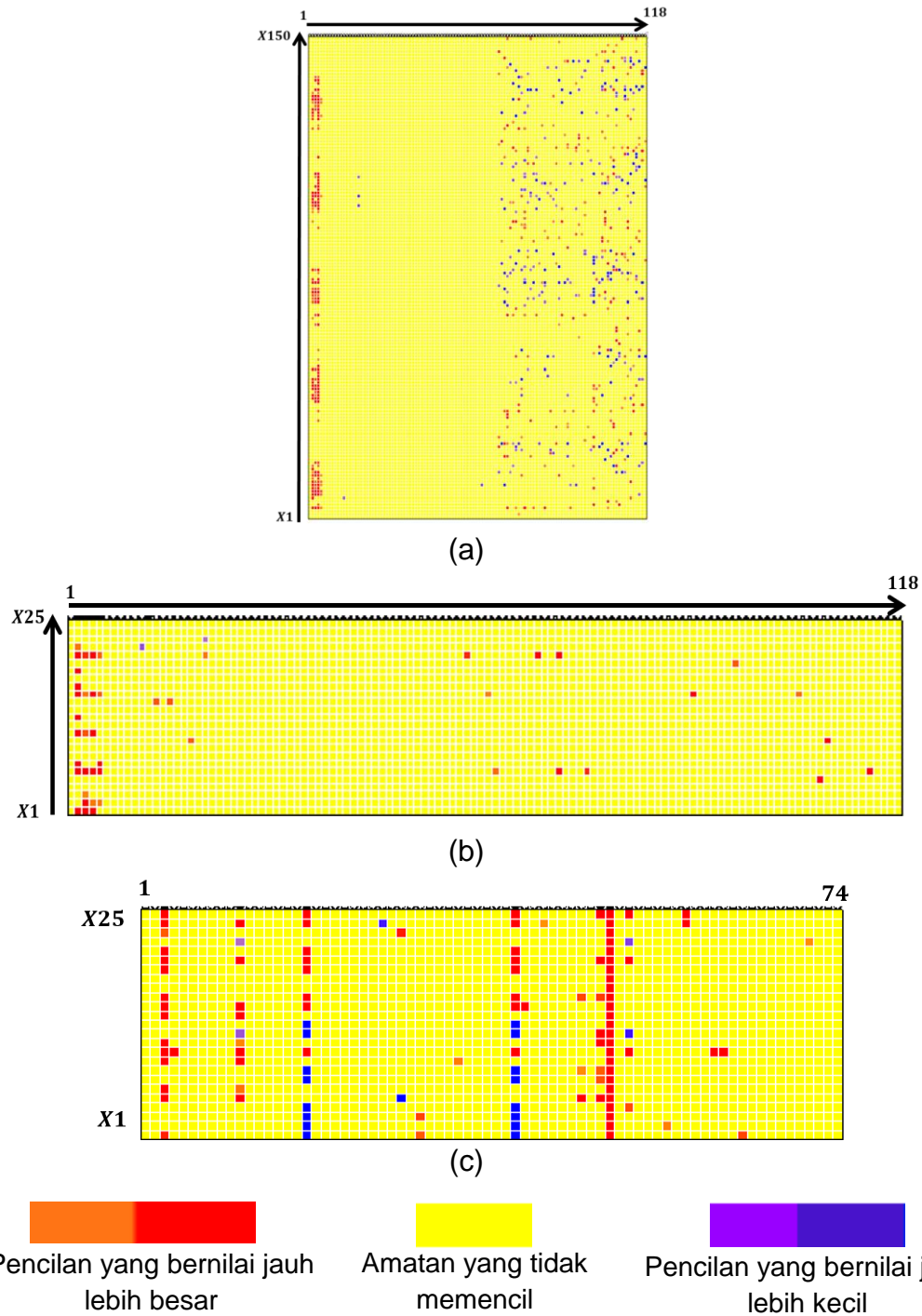


Gambar 5 Plot kuantil khi-kuadrat data desain alat (a) tahun 2017 peringkasan 3 angka, (b) tahun 2017 peringkasan luas, dan (c) tahun 2019 peringkasan luas

3.6 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*.

Gambar 6 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.



Gambar 6 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

3.7 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 pemodelan regresi *elastic net*

Peubah yang terseleksi	Koefisien Regresi <i>Elastic Net</i>		
	Data Desain Alat Tahun 2017 Peringkasan 3 Angka	Data Desain Alat Tahun 2017 Peringkasan Luas	Data Desain Alat Tahun 2019 Peringkasan Luas
Intersept	80,11	74,20	145,44
X_3	0	0	0
X_4	$9,1 \times 10^{-3}$	0	0
X_{10}	0	0	0
X_{13}	0	$1,5 \times 10^{-11}$	0
X_{15}	0	0	0
X_{17}	$-2,4 \times 10^{-3}$	$-1,3 \times 10^{-12}$	-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:

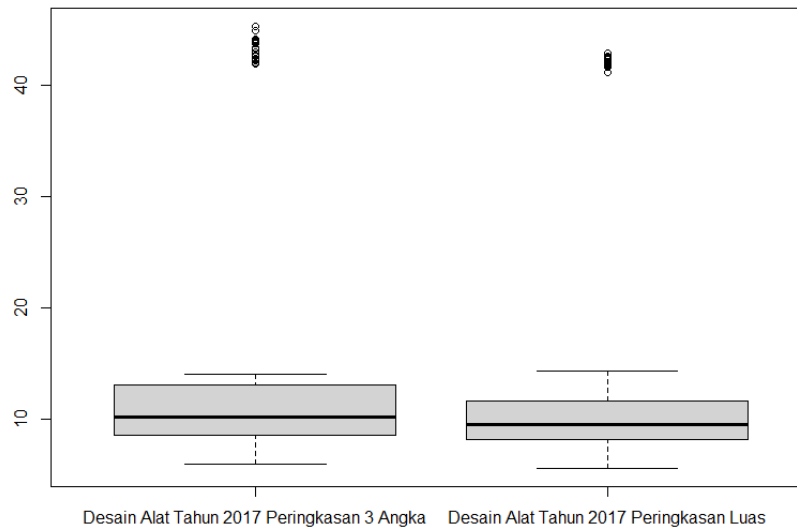
$$Y = 79,91 + 0,005X_{33} - 0,00063 X_{136}$$

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

$$Y = 72,45 - 2,52 \times 10^{-19} X_2 + 6,42 \times 10^{-11} X_6 - 6,18 \times 10^{-12} X_{14}$$

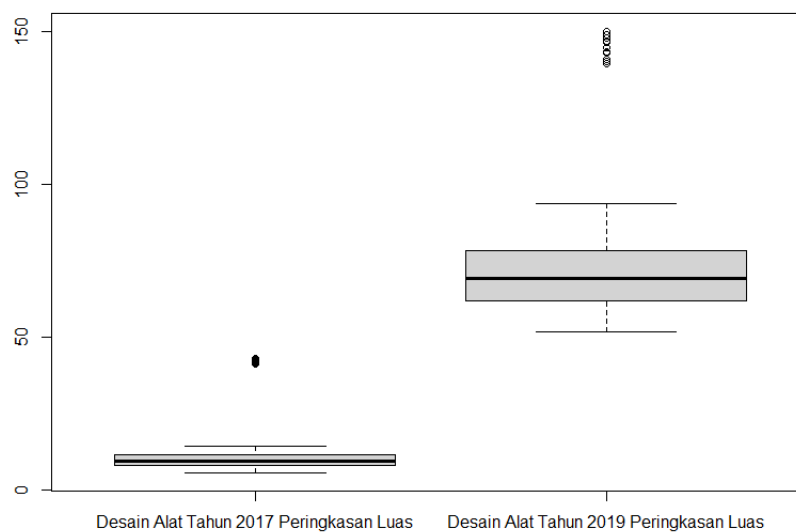
Persamaan regresi *elastic net* dengan algoritma DDC pada data desain alat tahun 2019 peringkasan luas:

$$Y = 147,1 - 2,57X_{11} - 1,62 X_{17}$$



Gambar 7 Diagram kotak garis RMSEP hasil pemodelan desain alat tahun 2017 peringkasan 3 angka dan peringkasan luas

Tingkat akurasi yang akan diukur pada pemodelan ini menggunakan nilai *Root Mean Square Error Prediction* (RMSEP) setiap set data. Diagram kotak garis untuk RMSEP hasil pemodelan pada desain alat 2017 dengan peringkasan 3 angka dan peringkasan luas menunjukkan bahwa adanya perbedaan nilai RMSEP pada kedua metode peringkasan. Pada metode peringkasan 3 angka memiliki nilai RMSEP yang lebih besar daripada peringkasan luas. Nilai median pada peringkasan 3 angka adalah sebesar 10,27 dan 9,58 pada peringkasan luas. Metode peringkasan 3 angka juga memiliki lebar kotak yang lebih lebar dibandingkan dengan peringkasan luas. Oleh karena itu, dapat dikatakan bahwa metode peringkasan luas lebih akurat dibandingkan metode peringkasan 3 angka.



Gambar 8 Diagram kotak garis RMSEP hasil pemodelan dengan peringkasan luas untuk desain alat tahun 2017 dan 2019

Membandingkan purwarupa alat tahun 2017 dan 2019 akan dilakukan dengan membandingkan nilai RMSEP dari metode peringkasan terbaik pada desain alat tahun 2017 yaitu metode peringkasan luas dengan desain alat tahun 2019 peringkasan luas. Berdasarkan diagram kotak garis RMSEP hasil pemodelan desain alat tahun 2017 dan 2019 dengan peringkasan luas, dapat dilihat bahwa kedua desain alat memiliki nilai RMSEP yang berbeda. Desain alat tahun 2019 memiliki nilai RMSEP yang jauh lebih besar dari desain alat tahun 2017, dapat dilihat dari letak kotak pada desain alat tahun 2019 lebih tinggi dibandingkan desain alat tahun 2017. Nilai median dari RMSEP desain alat tahun 2017 juga memiliki nilai yang jauh lebih besar yaitu sebesar 69,31. Produk tahun 2017 memiliki hasil yang lebih stabil dibandingkan dengan tahun 2019, dapat dilihat dari lebar kotaknya yang lebih kecil dibandingkan dengan lebar kotak dari produk tahun 2019.

4. Simpulan dan Saran

Metode peringkasan 3 angka dan metode peringkasan luas memberikan hasil yang berbeda. Metode peringkasan luas memiliki hasil yang lebih akurat serta lebih stabil dibandingkan dengan metode peringkasan 3 angka. Selain itu, 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. Oleh karena itu, metode peringkasan terbaik adalah metode peringkasan luas.

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.

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, Heriwiyantri, 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*. Kairo: EMRO Technical Publications 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.