

Comparison of Functional Regression and Functional Principal Component Regression for Estimating Non-Invasive Blood Glucose Level*

Perbandingan Metode Regresi Fungsional dan Regresi Komponen Utama Fungsional untuk Menduga Kadar Glukosa Darah pada Alat Non-Invasif

Nurul Fadhilah¹, Erfiani^{2‡}, and Indahwati^{3‡}

¹Department of Statistics, IPB University, Indonesia

²Department of Statistics, IPB University, Indonesia

³Department of Statistics, IPB University, Indonesia

‡corresponding author: erfiani_ipb@yahoo.com

Copyright © 2021 Nurul Fadhilah, Erfiani, and Indahwati. 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

The calibration method is an alternative method that can be used to analyze the relationship between invasive and non-invasive blood glucose levels. Calibration modeling generally has a large dimension and contains multicollinearities because usually in functional data the number of independent variables (p) is greater than the number of observations ($p > n$). Both problems can be overcome using Functional Regression (FR) and Functional Principal Component Regression (FPCR). FPCR is based on Principal Component Analysis (PCA). In FPCR, the data is transformed using a polynomial basis before data reduction. This research tried to model the equations of spectral calibration of voltage value excreted by non-invasive blood glucose level monitoring devices to predict blood glucose using FR and FPCR. This study aimed to determine the best calibration model for measuring non-invasive blood glucose levels with the FR and FPCR. The results of this research showed that the FR model had a bigger coefficient determination (R^2) value and lower Root Mean Square Error (RMSE) and Root Mean Square Error Prediction (RMSEP) value than the FPCR model, which was 12.9%, 5.417, and 5.727 respectively. Overall, the calibration modeling with the FR model is the best model for estimate blood glucose level compared to the FPCR model.

Keywords: calibration, functional principal components regression, functional regression, non-invasive.

* Received: Aug 2020; Reviewed: Oct 2020; Published: Mar 2021

1. Pendahuluan

Pada tahapan analisis dalam pemodelan, permasalahan yang sering terjadi adalah jumlah peubah bebas lebih besar daripada banyaknya pengamatan. Jika jumlah peubah bebas lebih besar daripada banyaknya pengamatan, estimasi model atau parameter tidak dapat diselesaikan dengan mudah dan memungkinkan adanya multikolinieritas atau hubungan antara peubah (Naes et al., 2002). Multikolinieritas dapat menyebabkan model yang akan diprediksi menjadi kurang akurat karena ketelitian dari dugaan parameter yang sangat rendah (Notodiputro, 2003). Metode untuk mengatasi hal ini adalah dengan melakukan transformasi pada data. Regresi Fungsional (RF) merupakan salah satu metode yang dapat digunakan. Konsep dari Regresi Fungsional adalah dengan mengestimasi parameter beta menggunakan ekspansi linier basis fungsi antara lain basis polinomial, basis B-spline dan basis Fourier. Peubah bebas yang memiliki dimensi tinggi akan di ekspansi menggunakan basis fungsi sehingga memiliki dimensi yang lebih kecil namun tidak menghilangkan informasi yang ada dalam data.

Metode lain yang dapat digunakan adalah Regresi Komponen Utama Fungsional (RKUF). Pada prinsipnya, RKUF digunakan untuk menganalisis data fungsional dengan cara kerja yang sama dengan Regresi Komponen Utama (RKU). Pada RKU komponen utama yang dihasilkan merupakan kombinasi linier dari seluruh peubah yang diamati dan saling bersifat orthogonal, sedangkan pada RKUF komponen utama yang dihasilkan merupakan kombinasi linier dari fungsi dalam himpunan data (Berrendero et al., 2011). RKU masih memiliki banyak kekurangan yaitu dengan jumlah peubah bebas yang besar dan dimensi yang besar maka proses perhitungan akan mengalami kendala dalam menghitung matriks ragam peragam (Bellman, 1961). Penelitian sebelumnya yang telah menerapkan RKUF yaitu (Lestari, 2014) menerapkan Analisis Komponen Utama Fungsional (AKUF) dan Regresi Komponen Utama Fungsional (RKUF) untuk memprediksi curah hujan dengan menggunakan data GCM. Pada hasil penelitian tersebut secara umum metode RKUF lebih baik dan konsisten dibandingkan dengan metode RKU dan RKU-FT dalam memprediksi curah hujan. (Shang, 2014) melakukan kajian penggunaan metode RKUF. Risqa, (2016) melakukan pemodelan Regresi Komponen Utama Fungsional (RKUF) pada data spektroskopi. Pada hasil penelitian tersebut antara model RKU dan RKUF dengan basis polinomial maka model RKUF polinomial merupakan model yang lebih baik.

Pada penelitian ini, data yang digunakan memiliki peubah bebas yang lebih banyak dibandingkan dengan banyaknya pengamatan. Data penelitian ini merupakan hasil keluaran spektroskopi berupa nilai residu intensitas yang kontinu. Data yang berbentuk kurva atau fungsi atau biasa disebut data fungsional memerlukan penanganan yang berbeda dengan data pada umumnya. Diperlukan beberapa pendekatan dalam peringkasan data dugaan kadar glukosa darah, sehingga nilai dugaan yang diperoleh dapat sedekat mungkin dengan kadar glukosa darah hasil pengukuran dengan metode invasif (gold standard). Data residu intensitas akan ditransformasi terlebih dahulu menggunakan basis fungsi polinomial. Dengan metode RF dan RKUF diharapkan dapat diperoleh hasil prediksi peubah respon dengan nilai error yang relatif kecil jika terdapat banyak peubah bebas dengan jumlah pengamatan yang terbatas.

Berdasarkan uraian diatas, tujuan dari penulisan ini adalah menduga kadar glukosa darah pada model kalibrasi alat pengukur kadar glukosa darah non-invasif dengan menggunakan Regresi Fungsional dan Regresi Komponen Utama Fungsional.

2. Metodologi

2.1 Data

Data yang digunakan pada penelitian ini adalah data primer yang merupakan bagian dari penelitian pengembangan dan uji klinis purwarupa alat pemantauan kadar glukosa darah secara non-invasif. Penelitian ini didanai oleh Program Insentif Riset Sistem Inovasi Nasional, Kementerian Riset, Teknologi dan Pendidikan Tinggi. Penelitian dilaksanakan pada bulan April 2016 sampai Januari 2017 oleh Tim non-invasif biomarking dari Departemen Statistika yang bekerjasama dengan Departemen Fisika, Gizi Masyarakat, Biologi, dan Ilmu Komputer Institut Pertanian Bogor (IPB). Pelaksanaan penelitian bertempat di Laboratorium Biokimia, Departemen Gizi Masyarakat, IPB. Penelitian dilakukan dengan melibatkan 118 responden yang merupakan mahasiswa IPB. Data yang dianalisis yaitu data kadar glukosa darah invasif hasil pengukuran laboratorium Prodia yang dilakukan dengan cara responden diambil darahnya sebanyak kurang lebih 4 ml pada pembuluh darah vena di lengan. Data kadar glukosa darah invasif merupakan peubah respon (Y). Sedangkan peubah bebas (X) merupakan hasil pengukuran alat pemantau kadar glukosa darah non-invasif berupa residu intensitas cahaya yang melewati jari tangan responden dengan alat yang dirangkai menggunakan lampu 1600 nm. Data spektrum residu intensitas terdiri atas 5 ulangan, setiap ulangan terdiri atas 5 modulasi, setiap modulasi terdapat 2 data puncak dan setiap data puncak terdiri atas 3 data sehingga diperoleh 150 peubah residu intensitas.

2.2 Metode

Tahapan analisis pada penelitian ini adalah:

- A. Melakukan eksplorasi pada data hasil pengukuran alat pemantau kadar glukosa darah invasif.
- B. Melakukan prapemrosesan pada data hasil pengukuran alat pemantau kadar glukosa darah non-invasif.
 1. Melihat pola yang dibentuk residu intensitas terhadap *time domain* pada 10 modulasi yaitu modulasi 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% dan 90%. Modulasi merupakan pengaturan lebar intensitas inframerah dalam suatu periode untuk mendapatkan tegangan rata-rata yang berbeda. Setiap modulasi diulang sebanyak dua kali. Modulasi yang dipilih adalah modulasi yang membentuk puncak.
 2. Memisahkan data puncak residu intensitas pada modulasi yang terpilih.
 3. Membuang data awal dan data akhir dari setiap puncak spektrum yang telah dipilih. Hal ini dilakukan karena dikhawatirkan data tersebut merupakan *noise* dari hasil alat yang belum stabil saat memulai dan mengakhiri pengukuran.
 4. Meringkas data residu intensitas pada setiap puncak. Peringkasan dilakukan agar setiap responden memiliki jumlah peubah X yang sama. Peringkasan dilakukan dengan pemilihan tiga data yakni data awal, data tengah dan data akhir. Hal ini dilakukan untuk mempertahankan pola pada data, sehingga tidak banyak kehilangan informasi.
 5. Menyusun matriks data residu intensitas yang diperoleh pada tahapan nomor 4 yang selanjutnya akan dianalisis dan merupakan nilai-nilai peubah X.
- C. Membagi data menjadi dua bagian secara acak yaitu data *training* (data untuk menyusun model) sebesar 80% dan data *testing* (data untuk validasi model) sebesar 20%.

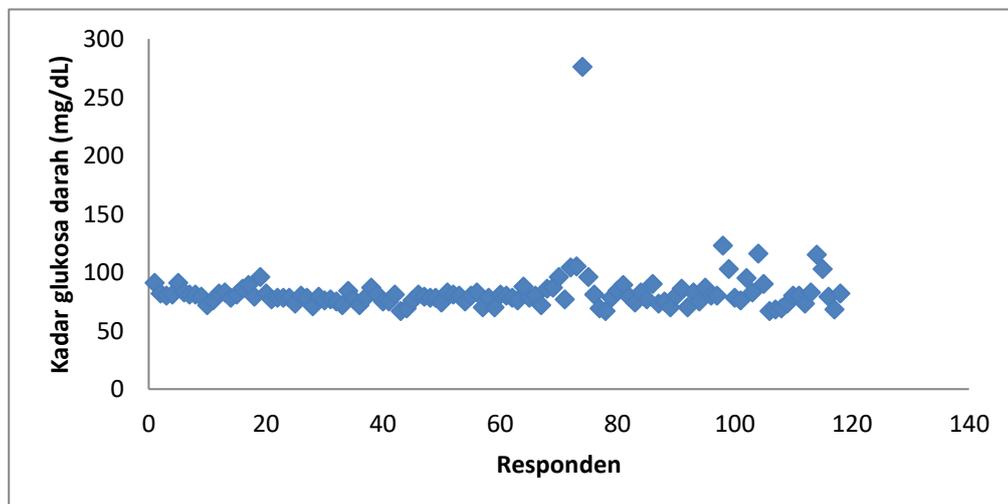
- D. Melakukan analisis menggunakan semua data dan data tanpa pencilan (jika terdapat pencilan) hasil pengukuran invasif dengan tahapan yaitu:
1. Pemeriksaan multikolinearitas menggunakan matrik korelasi.
 2. Pemodelan dengan pendekatan RF dan RKUF.
Model Regresi Fungsional (RF) sebagai berikut (Saeys et al., 2008):

$$y_i = a + \int_{\lambda_1}^{\lambda_m} x_i(s)\beta(s) ds + \varepsilon_i$$
 di mana λ_1 dan λ_m adalah titik awal dan akhir dari fungsi.
 3. Melakukan perulangan sebanyak 100 kali pada pemodelan kedua metode kemudian menghitung rata-rata dan ragam pada R^2 , RMSE dan RMSEP.
- E. Mengevaluasi model terbaik dengan membandingkan keakuratan model dari nilai R^2 , RMSE dan RMSEP.

3. Hasil dan Pembahasan

3.1 Deskripsi Data Kadar Glukosa Darah Invasif

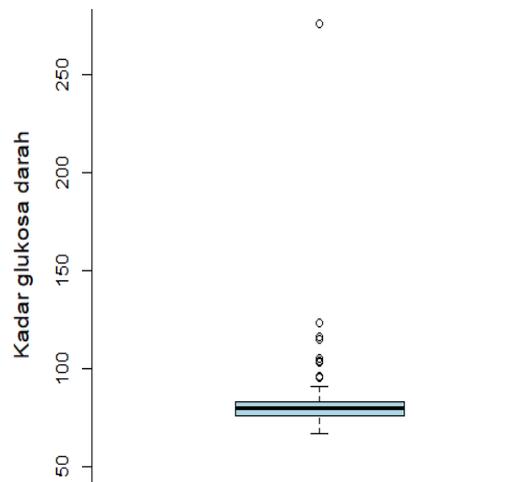
Data hasil pengukuran kadar glukosa darah invasif yang digunakan adalah hasil keluaran pengukuran oleh Laboratorim Prodia. Data tersebut diperoleh dari 118 responden yang merupakan mahasiswa IPB. Gambar 1 merupakan representasi visual dari hasil pengukuran kadar glukosa darah invasif. Dapat dilihat, plot kadar glukosa darah invasif memiliki besaran data yang cenderung sama. Hasil pengukuran kadar glukosa darah memiliki rata-rata sebesar 82.64 mg/dL dengan median 80 mg/dL. Data hasil pengukuran kadar glukosa darah tertinggi adalah sebesar 276 mg/dL sedangkan kadar glukosa darah invasif terendah adalah 67 mg/dL.



Gambar 1: Plot kadar glukosa darah invasive

Hasil pengukuran kadar glukosa darah secara invasif sebagian besar menunjukkan kurang dari 108 mg/dL atau glukosa darah normal. Terdapat sekitar 106 responden yang memiliki kadar glukosa darah normal. Terdapat sekitar 8 responden memiliki kadar glukosa darah rendah yaitu kadar glukosa darah kurang dari 70 mg/dL. Terdapat 3 responden yang termasuk dalam kategori prediabetes yaitu kadar glukosa darah yang berada diantara 108 mg/dL sampai 125 mg/dL. Kadar glukosa darah yang termasuk dalam kategori prediabetes tersebut yaitu 115 mg/dL, 116 mg/dL dan 123

mg/dL. Terdapat 1 responden yang termasuk dalam kategori diabetes yaitu sebesar 276 mg/dL atau lebih besar dari 126 mg/dL. Hasil pengukuran 118 responden juga disajikan dalam bentuk diagram kotak garis yang terdapat pada Gambar 2.



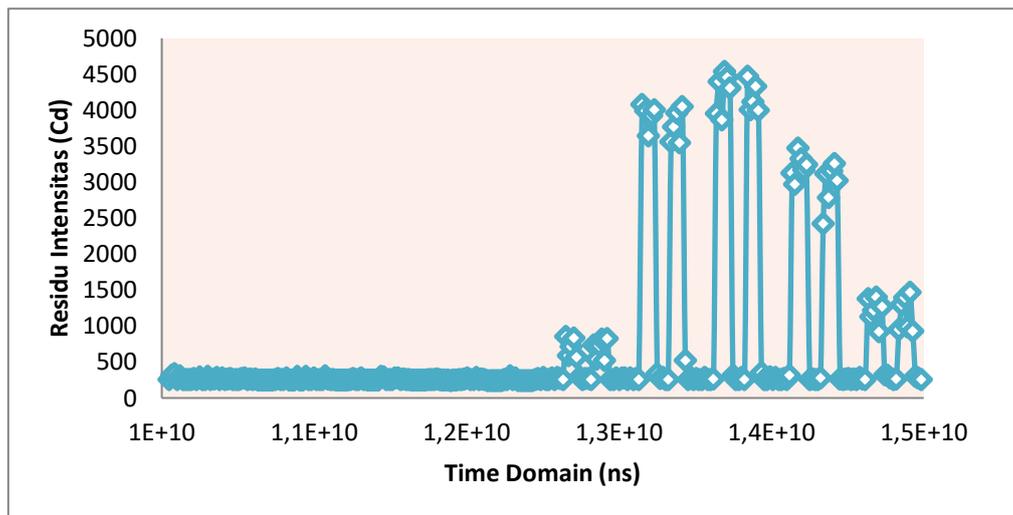
Gambar 2: Boxplot kadar glukosa darah invasif

Berdasarkan Gambar 2, data kadar glukosa darah invasif memiliki nilai keragaman yang kecil. Terdapat beberapa responden yang menjadi pencilan atas. Pencilan ini terjadi dikarenakan beberapa responden memiliki kadar glukosa darah yang bernilai jauh lebih tinggi dibandingkan dengan responden lainnya. Keberadaan data pencilan akan mengganggu dalam proses analisis data. Pencilan dapat memberikan pengaruh yang signifikan terhadap hasil pendugaan koefisien regresi (Hatch & Prihoda, 1992) dan menyebabkan nilai residual dari model yang terbentuk menjadi besar. Salah satu upaya yang dapat dilakukan untuk mengatasi pencilan tersebut adalah dengan cara menghilangkan atau menghapus data pencilan. Mengatasi pencilan tersebut sebagai upaya yang dapat dilakukan agar model yang diperoleh semakin baik.

Pencilan yang terdapat pada hasil pengukuran kadar glukosa darah secara invasif ada 12 responden. Kadar glukosa darah tersebut yaitu 95mg/dL, 96mg/dL (3 responden), 103mg/dL (2 responden), 104mg/dL, 105mg/dL, 115mg/dL, 116mg/dL, 123mg/dL dan 276mg/dL. Analisis dalam penelitian ini akan dilakukan dengan dua kelompok data, yaitu menggunakan semua data hasil pengukuran invasif (118 responden) dan data tanpa pencilan (106 responden).

3.2 Prapemrosesan Data Kadar Glukosa Darah Non Invasif

Alat pengukur kadar glukosa darah non-invasif pada penelitian ini menggunakan lampu 1600 nm. Hasil keluaran alat pengukur glukosa darah non-invasif berupa spektrum residu intensitas terhadap time domain. Hasil spektrum pengukuran kadar glukosa non-invasif dengan satu jari tangan dari satu responden diulang sebanyak lima kali ulangan. Pola spektrum yang dihasilkan merupakan intensitas dari cahaya inframerah yang dilewatkan terhadap time domain. Alat ini bekerja dengan dua kondisi yang terjadi yaitu saat lampu mati dan sensor menyala serta saat lampu dan sensor menyala bersamaan. Hasil spektrum tersebut ditunjukkan pada Gambar 3.

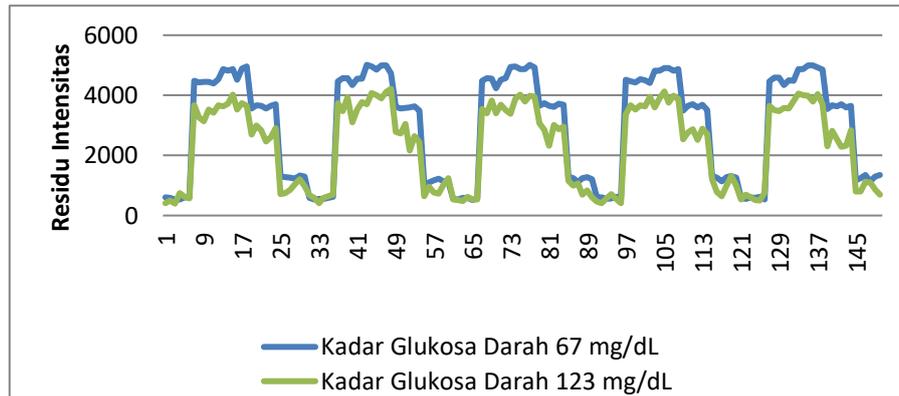


Gambar 3: Spektrum kadar glukosa darah non-invasif

Pengaturan tingkat pencahayaan dari lampu inframerah dikenal dengan modulasi. Spektrum kadar glukosa darah memiliki sepuluh modulasi yaitu 0%, 10%, 20%, sampai dengan modulasi 90%. Masing-masing modulasi diulang sebanyak dua kali untuk setiap pengukuran. Kondisi yang terjadi pada modulasi yang membentuk puncak adalah ketika lampu dan sensor menyala bersamaan. Intensitas yang dipancarkan oleh sinar inframerah dapat terbaca dengan baik oleh sensor pada kelima modulasi terakhir yaitu modulasi 50%, 60%, 70%, 80% dan 90%. Sehingga, penelitian ini hanya menggunakan modulasi 50% sampai dengan modulasi 90%.

Setiap modulasi terdiri atas dua kali pengukuran, sehingga terdapat sepuluh puncak yang terbentuk. Data puncak spektrum kadar glukosa darah umumnya memiliki data residu intensitas paling sedikit lima data dan paling banyak tujuh data. Data awal dan data akhir pada setiap puncak dibuang karena dikhawatirkan data tersebut merupakan noise dari alat non-invasif saat memulai dan mengakhiri pengukuran. Jumlah data residu intensitas yang dimiliki setiap responden berbeda-beda sehingga diperlukan penyeragaman data. Penyeragaman dilakukan dengan cara pengambilan tiga angka dengan tujuan untuk mempertahankan pola trend, sehingga tidak kehilangan informasi pada data. Tiga angka yang diambil adalah data awal, data tengah dan data akhir. Secara keseluruhan data spektrum residu intensitas terdiri atas 5 ulangan, setiap ulangan terdiri atas 5 modulasi, setiap modulasi terdapat 2 data puncak dan setiap data puncak terdiri atas 3 data sehingga diperoleh 150 peubah residu intensitas. Nilai tersebut yang akan menjadi nilai pada peubah X. Hasil spektrum gabungan kadar glukosa darah berbeda setelah dilakukan penyeragaman data yang disajikan pada Gambar 4.

Gambar 4 menunjukkan perbandingan spektrum yang dihasilkan antara kadar glukosa darah tinggi (123 mg/dL) dan kadar glukosa darah rendah (67 mg/dL). Spektrum pada kadar glukosa darah rendah memiliki nilai residu intensitas lebih tinggi daripada spektrum kadar glukosa darah tinggi. Hal ini dikarenakan, pada saat cahaya inframerah yang ditembakkan oleh alat pengukur kadar glukosa non-invasif menghasilkan intensitas cahaya yang dilewatkan lebih besar daripada yang diserap tubuh. Hal ini berarti, semakin tinggi kadar glukosa darah yang diukur maka intensitas cahaya yang dilewatkan semakin sedikit, sehingga nilai residu intensitas cahaya semakin rendah.



Gambar 4: Spektrum gabungan kadar glukosa darah 67mg/dL dan 123mg/dL

3.3 Analisis Semua Data

Penelitian ini menggunakan data hasil pengukuran secara invasif dan non invasif sebanyak 118 responden. Hasil pengukuran kadar glukosa darah invasif yang diperoleh akan digunakan sebagai peubah Y yang merupakan vektor berukuran 118×1 . Adapun peubah X yang digunakan merupakan hasil pengukuran kadar glukosa darah non-invasif berupa matriks yang berukuran 118×150 . Data yang digunakan akan dibagi menjadi dua bagian yaitu data training dan data testing. Proporsi yang digunakan dalam pengambilan sampel adalah 80% data training dan 20% data testing. Pengambilan sampel diulang sebanyak 100 kali. Sebelum melakukan pemodelan, masalah yang dapat muncul dalam analisis regresi yaitu multikolinearitas. Menurut Montgomery & Peck (1982), beberapa cara yang dapat dilakukan untuk pemeriksaan multikolinearitas yaitu dengan menggunakan nilai matrik korelasi X, nilai Variance Inflation Factor (VIF), nilai eigen dari matriks $X'X$ dan determinan dari matriks $X'X$. Cara mendeteksi adanya multikolinearitas yang paling sederhana adalah dengan melakukan pemeriksaan nilai korelasi pada matriks korelasi X (Ghozali, 2013). Jika antar peubah X terdapat korelasi yang tinggi umumnya mendekati 1 dan -1, maka hal ini merupakan indikasi adanya multikolinearitas. Tabel 1 merupakan matriks korelasi antara peubah X hasil pengukuran glukosa darah non-invasif yang berupa data spektrum residu intensitas.

Tabel 1: Matrik korelasi antar peubah X

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	...	X_{150}
X_1	1.00	0.91	0.83	0.83	0.88	0.87	0.73	0.72	0.74	...	0.78
X_2	0.91	1.00	0.94	0.84	0.88	0.90	0.74	0.77	0.75	...	0.84
X_3	0.83	0.94	1.00	0.79	0.82	0.83	0.76	0.80	0.78	...	0.80
X_4	0.83	0.84	0.79	1.00	0.89	0.86	0.73	0.74	0.72	...	0.80
X_5	0.88	0.88	0.82	0.89	1.00	0.88	0.74	0.72	0.70	...	0.78
X_6	0.87	0.90	0.83	0.86	0.88	1.00	0.74	0.73	0.72	...	0.79
X_7	0.73	0.74	0.76	0.73	0.74	0.74	1.00	0.96	0.93	...	0.77
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
X_{148}	0.76	0.81	0.78	0.80	0.78	0.79	0.82	0.84	0.84	...	0.91
X_{149}	0.79	0.85	0.83	0.78	0.79	0.80	0.81	0.83	0.84	...	0.94
X_{150}	0.78	0.84	0.80	0.80	0.78	0.79	0.77	0.80	0.82	...	1.00

Tabel 1 menunjukkan bahwa korelasi antara peubah X memiliki nilai korelasi antara $0.62 \leq r \leq 0.96$. Nilai korelasi cukup tinggi karena mendekati nilai 1. Pendekatan model kalibrasi yang digunakan dalam penelitian ini adalah Regresi Fungsional dan Regresi Komponen Utama Fungsional.

3.3.1. Regresi Fungsional (RF)

Sebelum melakukan pemodelan data menggunakan regresi fungsional, perlu dilakukan transformasi data dengan basis fungsi polinomial. Pada penelitian ini digunakan data hasil pengukuran alat pemantau kadar glukosa darah non-invasif berupa residu intensitas cahaya yang melewati jari tangan responden dengan alat yang dirangkai menggunakan lampu 1600 nm sebagai peubah prediktor dan data hasil pengukuran kadar glukosa darah invasif sebagai peubah respon.

Dengan menggunakan data hasil transformasi, pada model regresi fungsional dengan basis polinomial terlebih dahulu data dibagi menjadi dua bagian secara acak yaitu data training (data untuk menyusun model) sebesar 80% dan data testing (data untuk validasi model) sebesar 20%. Pengambilan sampel pada data *training* dan *testing* diulang sebanyak 100 kali ulangan.

Tabel 2: Penduga parameter dan statistik uji untuk analisis semua data

Variabel	Estimate	Std. Error	T value	P value
(Intercept)	83.360	2.345	35.550	$<2 \times 10^{-16}$ ***
xtrain.monomial0	8.014×10^{-3}	6.143×10^{-3}	1.305	0.195
xtrain.monomial1	-9.428×10^{-4}	8.601×10^{-4}	-1.096	0.276
xtrain.monomial2	3.080×10^{-5}	3.418×10^{-5}	0.901	0.370
xtrain.monomial3	-4.214×10^{-7}	5.561×10^{-7}	-0.758	0.451
xtrain.monomial4	2.584×10^{-9}	3.951×10^{-9}	0.654	0.515
xtrain.monomial5	-5.867×10^{-12}	1.016×10^{-11}	-0.577	0.565

Tabel 2 menunjukkan bahwa tidak ada satupun variabel bebas yang berpengaruh terhadap hasil pengukuran kadar glukosa darah invasif kecuali intercept dari model yang berpengaruh signifikan. Hal ini juga dijelaskan dengan hasil nilai kebaikan model yang hanya 5.48% yang berarti bahwa peubah xtrain.monomial0 sampai xtrain.monomial5 hanya mampu menerangkan total keragaman hasil pengukuran kadar glukosa darah invasif sebesar 5.48%.

3.3.2. Regresi Komponen Utama Fungsional (RKUF) Basis Polynomial

Peubah prediktor yang digunakan pada metode RKUF ialah data yang telah di transformasi menjadi data fungsional dengan menggunakan basis polinomial. Sedangkan data hasil pengukuran kadar glukosa darah invasif digunakan sebagai peubah respon. Hasil analisis menggunakan metode RKUF dengan transformasi polinomial pada Tabel 3, menunjukkan bahwa komponen utama fungsional pertama basis polinomial (kfp1) mampu menjelaskan sebanyak 97.70% dari keragaman yang ada dan komponen utama fungsional kedua basis polinomial (kfp2) mampu menjelaskan 1.34% dari keragaman yang ada dan komponen utama fungsional ketiga basis polinomial (kfp3) mampu menjelaskan 0.56% dari keragaman dan komponen utama fungsional keempat basis polinomial (kfp4) mampu menjelaskan 0.24% dari keragaman yang ada.

Tabel 3: Proporsi keragaman setiap komponen

	Komponen			
	Kfp1	Kfp2	Kfp3	Kfp4
Proporsi Keragaman	97.70	1.34	0.56	0.24
Proporsi Komulatif	97.70	99.04	99.60	99.82

Tabel 3 menunjukkan komponen utama fungsional pertama basis polinomial (kfp1) mampu menjelaskan sebanyak 97.70% yang artinya memenuhi kriteria > 90%. Sehingga untuk skor komponen utama fungsional yang digunakan pada RKUF sebanyak satu yaitu kfp1 dalam membentuk model regresi.

Tabel 4: Penduga parameter dan statistik ujiinya

Parameter	Penduga	Galat Baku	t-Hitung	Nilai-p
Intersep	83.362	2.314	36.017	$<2 \times 10^{-16}$ ***
Kfp1	-0.0001	0.0004	-0.507	0.613

Tabel 4 menunjukkan bahwa tidak ada satupun variabel komponen yang berpengaruh terhadap hasil pengukuran kadar glukosa darah invasif kecuali intersep dari model yang berpengaruh signifikan. Hal ini juga dijelaskan dengan hasil nilai kebaikan model R^2 yang hanya 0.2% yang berarti bahwa peubah komponen Kfp1 hanya mampu menerangkan total keragaman hasil pengukuran kadar glukosa darah invasif sebesar 0.2%. Model akhir dari metode RKUF sebagai berikut :

$$y = 83.362 - 0.0001Kfp1$$

3.4 Analisis Data Tanpa Pencilan

Penelitian ini menggunakan data sebanyak 106 responden. Hasil pengukuran kadar glukosa darah invasif yang diperoleh akan digunakan sebagai peubah Y yang merupakan vektor berukuran 106×1 . Adapun peubah X yang digunakan merupakan hasil pengukuran kadar glukosa darah non-invasif berupa matriks yang berukuran 106×150 . Data yang digunakan akan dibagi menjadi dua bagian yaitu data training dan data testing. Proporsi yang digunakan dalam pengambilan sampel adalah 80% data training dan 20% data testing. Pengambilan sampel diulang sebanyak 100 kali. Proses analisis sama halnya dengan analisis menggunakan semua data yaitu pemodelan dengan metode Regresi Fungsional dan Regresi Komponen Utama Fungsional.

3.5 Regresi Fungsional (RF)

Sebelum melakukan pemodelan data menggunakan regresi fungsional, perlu dilakukan transformasi data dengan basis fungsi polinomial. Dengan menggunakan data hasil transformasi, pada model regresi fungsional dengan basis polinomial terlebih dahulu data dibagi menjadi dua bagian secara acak yaitu data *training* (data untuk menyusun model) sebesar 80% dan data *testing* (data untuk validasi model) sebesar 20%. Pengambilan sampel pada data *training* dan *testing* diulang sebanyak 100 kali ulangan.

Tabel 5 menunjukkan bahwa tidak ada satupun variabel bebas yang berpengaruh terhadap hasil pengukuran kadar glukosa darah invasif pada data tanpa pencilan kecuali intercept dari model yang berpengaruh signifikan. Hal ini juga dijelaskan dengan hasil nilai kebaikan model yang hanya 12.98% yang berarti bahwa peubah

xtrain.monomial0 sampai xtrain.monomial5 hanya mampu menerangkan total keragaman hasil pengukuran kadar glukosa darah invasif sebesar 12.98%.

Tabel 5: Penduga parameter dan statistik uji untuk analisis data tanpa pencilan

Variabel	Estimate	Std. Error	T value	P value
(Intercept)	78.75	0.606	129.913	$<2 \times 10^{-16}$ ***
xtrain.monomial0	2.603×10^{-3}	1.563×10^{-3}	1.665	0.099
xtrain.monomial1	-3.550×10^{-4}	2.170×10^{-4}	-1.636	0.106
xtrain.monomial2	1.378×10^{-5}	8.576×10^{-6}	1.606	0.112
xtrain.monomial3	-2.154×10^{-7}	1.395×10^{-7}	-1.544	0.126
xtrain.monomial4	1.459×10^{-9}	9.922×10^{-10}	1.471	0.145
xtrain.monomial5	-3.579×10^{-12}	2.560×10^{-12}	-1.398	0.166

3.5.1. Regresi Komponen Utama Fungsional (RKUF) Basis Polynomial

Proses pada analisis ini sama halnya dengan proses pemodelan RKUF menggunakan semua data. Hasil analisis menggunakan metode RKUF data tanpa pencilan dengan transformasi polinomial pada Tabel 6, menunjukkan bahwa komponen utama fungsional pertama basis polinomial (kfp1) mampu menjelaskan sebanyak 97.57% dari keragaman yang ada dan komponen utama fungsional kedua basis polinomial (kfp2) mampu menjelaskan 1.56% dari keragaman yang ada dan komponen utama fungsional ketiga basis polinomial (kfp3) mampu menjelaskan 0.50% dari keragaman dan komponen utama fungsional keempat basis polinomial (kfp4) mampu menjelaskan 0.22% dari keragaman yang ada.

Tabel 6: Proporsi keragaman setiap komponen

	Komponen			
	Kfp1	Kfp2	Kfp3	Kfp4
Proporsi Keragaman	97.57	1.56	0.50	0.22
Proporsi Kumulatif	97.57	99.03	99.53	99.75

Tabel 6 menunjukkan komponen utama fungsional pertama basis polinomial (kfp1) mampu menjelaskan sebanyak 97.57% yang artinya memenuhi kriteria $> 90\%$. Sehingga untuk skor komponen utama fungsional yang digunakan pada RKUF sebanyak satu yaitu kfp1 dalam membentuk model regresi.

Tabel 7: Penduga parameter dan statistik ujinya

Parameter	Penduga	Galat Baku	t-Hitung	Nilai-p
Intersep	78.750	0.602	130.725	2×10^{-16}
Kfp1	-0.0002	0.0001	-2.786	0.006

Tabel 7 merupakan tabel penduga parameter untuk metode RKUF basis polinomial pada data tanpa pencilan dengan menggunakan satu skor komponen utama fungsional. Hasil nilai kebaikan model R^2 sebesar 0.85% yang berarti bahwa peubah komponen Kfp1 hanya mampu menerangkan total keragaman hasil pengukuran kadar glukosa darah invasif sebesar 0.85%. Model akhir dari metode RKUF pada data tanpa pencilan sebagai berikut :

$$y = 78.75 - 0.0002Kfp1$$

3.6 Perbandingan Ukuran Kebaikan Model

Evaluasi hasil peramalan menggunakan data testing dimaksudkan untuk mengetahui akurasi peramalan yang dinyatakan dengan *Root Mean Square Error Prediction* (RMSEP). Dilakukan pengulangan pemodelan data mulai dari pembagian data (*training* dan *testing*) sampai evaluasi model sebanyak $m = 100$. Dari 100 model tersebut akan dihitung nilai rata-rata dan ragam dari R^2 , *Root Mean Square Error* (RMSE), dan *Root Mean Square Error Prediction* (RMSEP). Semakin tinggi nilai R^2 menunjukkan model yang diperoleh semakin baik. Model yang terbaik adalah model yang memiliki RMSE terkecil dan R^2 terbesar. Adapun hasil kebaikan model masing-masing pendekatan disajikan pada Tabel 8.

Tabel 8: Ukuran kebaikan model untuk analisis semua data dan data tanpa pencilan

Model	Ringkasan Statistik	Analisis Semua Data			Analisis Data Tanpa Pencilan		
		Data Training		Data Testing	Data Training		Data Testing
		R^2	RMSE	RMSEP	R^2	RMSE	RMSEP
RF	Rata-rata	0.041	19.453	18.303	0.129	5.417	5.727
	Ragam	0.000	31.328	172.683	0.001	0.030	0.464
RKUF	Rata-rata	0.005	19.255	17.684	0.067	5.435	5.594
	Ragam	0.000	30.207	184.038	0.000	0.029	0.447

Berdasarkan Tabel 8 nilai ukuran kebaikan model yang diperoleh antara model RF dengan RKUF tidak terpaut jauh. Hasil analisis data tanpa pencilan atau pada rentang data yang lebih sempit menunjukkan bahwa model RF memiliki nilai R^2 terbesar yaitu 12.9%. Artinya, sebesar 12.9% keragaman dari hasil pengukuran glukosa darah invasif yang dapat dijelaskan oleh model, sedangkan sisanya 87.1% tidak dapat dijelaskan oleh model. Berdasarkan nilai RMSE dan RMSEP pada model RF juga memiliki nilai yang lebih kecil dibandingkan dengan model RKUF. Dengan demikian, model yang diterapkan pada alat pengukuran kadar glukosa darah non-invasif menggunakan analisis data tanpa pencilan atau pada rentang data yang lebih sempit.

4. Simpulan dan Saran

Ukuran kebaikan model regresi fungsional dan regresi komponen utama fungsional memperoleh hasil yang tidak terpaut jauh. Model regresi fungsional memiliki nilai R^2 terbesar, sedangkan nilai RMSE dan RMSEP yang diperoleh lebih kecil dibandingkan model regresi komponen utama fungsional. Secara keseluruhan model kalibrasi dengan pendekatan regresi fungsional merupakan metode yang baik dalam menduga kadar glukosa darah non-invasif dibandingkan dengan regresi komponen utama fungsional pada rentang data yang lebih sempit.

Saran untuk penelitian selanjutnya dalam pengambilan sampel pada penelitian kadar glukosa darah non-invasif lebih diperbanyak dengan memperhatikan proporsi kadar glukosa darah normal, pradiabetes, dan diabetes. Saran kedua dalam menduga kadar glukosa darah non-invasif dapat menggunakan metode yang dapat mengatasi pencilan pada data.

Daftar Pustaka

- Bellman, R. E. (1961). *Adaptive Control Process: a Guided Tour*. New Jersey: Princeton.
- Berrendero, J. R., Juste, A., & Svarc, M. (2011). Principal Components for Multivariate Functional Data. *J Elsevier Science Direct Computational Statistics and Data Analysis*, 55: 2619–2634. <https://doi.org/10.1016/j.csda.2011.03.011>
- Montgomery, D. C., & Peck, E. A. (1982). *Introduction to Linear Regression Analysis*. New York (US): John Wiley & Sons.
- Ghozali, I. (2013). *Aplikasi Analisis Multivariate dengan Program SPSS (7th ed.)*. Semarang (ID): Badan Penerbit Universitas Diponegoro.
- Hatch, J. P., & Prihoda, T. J. (1992). The Effect of Influential Outliers on Parameter Estimation in Regression Analysis. *Biofeedback and Self-Regulation*, (17(2)): 153–156.
- Lestari, D. N. (2014). *Pemodelan Statistical Downscaling dengan Analisis Komponen Utama Fungsional untuk Prediksi Curah Hujan (Studi Kasus: Stasiun Klimatologi Bondan Kabupaten Indramayu)*. Bogor(ID): Institut Pertanian Bogor.
- Naes, T., Issakson, T., Fearn, T., & Davies, T. (2002). *Multivariate Calibration and Classification*. United Kingdom (GB): NIR Publications Chichester.
- Notodiputro, K. A. (2003). *Pendekatan Statistika dalam Kalibrasi*. Bandung.
- Risqa, A. (2016). *Pemodelan Regresi Komponen Utama Fungsional pada Data Spektroskopi*. Bogor(ID): Institut Pertanian Bogor.
- Saeys, W., Ketelaere, B. D., & Darius, P. (2008). Potential Applications of Functional Data Analysis in Chemometrics. *Journal of Chemometrics*, (22): 335–344.
- Shang, H. L. (2014). A survey of functional principal component analysis. *AStA Advances in Statistical Analysis*, 98(2): 121–142.