

Penerapan Regresi Nonparametrik Smoothing Spline untuk Data Tersensor dalam Memodelkan Hubungan antara Lamanya Waktu Kesembuhan Rawat Inap pada Pasien Diabetes Melitus Tipe-2 dengan Usia Pasien

Dea Sri Mulyani*, Abdul Kudus

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*deasrim99@gmail.com, abdul.kudus@unisba.ac.id

Abstract. The aim of study is to explain the case. Regression analysis is used to model or look for patterns of relationship between one or more independent variables and one or more response variables. Since the data often does not follow a specific formulation pattern, a more flexible model is required, namely nonparametric regression model approach is an approach used when the shape of the shape of the relationship between the response variable and the independent variable is unknown or information about the shape of the regression function is not available. The nonparametric spline regression model with optimal nodes was applied to patient age and time to recovery after hospitalization in patients with type 2 diabetes mellitus containing right-censored data in a variable containing the right sensor. In this study, the generalized cross-validation (GCV) method was applied to nonparametric smoothing spline regression to determine the optimal smoothing parameter. In order to obtain the GCV value, parameter estimation must be available to form a hat matrix formed from $\hat{f}_\lambda$. Parameter estimation for the smoothing spline from the function $f(\cdot)$ by minimizing PRSS (Penalized Residual Sum of Square). Once the optimal GCV value has been determined, it forms the best estimate of the regression function. Optimal Smoothing parameter by choosing λ with a minimum GCV value. According to the research result, the minimum GCV value is 0.0378 at $\lambda=0.015$ and the GCV value with Kaplan Meier weights is 30.4773.

Keywords: *Censored Data, Diabetes Mellitus, GCV, Kaplan Meier, Nonparametric Regression, Smoothing Spline.*

Abstrak. Penelitian ini bertujuan untuk menjelaskan mengenai kasus Analisis regresi digunakan untuk memodelkan atau mencari pola hubungan antara satu atau lebih variabel bebas dengan satu atau lebih variabel respons. Seringkali data tidak mengikuti pola rumusan tertentu, sehingga diperlukan model yang lebih fleksibel, yaitu regresi nonparametrik. Pendekatan model regresi nonparametrik yaitu merupakan suatu pendekatan yang digunakan apabila bentuk hubungan antara variabel respon dan variabel bebasnya tidak diketahui atau tidak adanya informasi mengenai bentuk fungsi regresinya. Model regresi nonparametrik spline dengan titik knot optimal diaplikasikan pada usia pasien terhadap lamanya waktu kesembuhan rawat inap pada pasien yang menyandang penyakit diabetes melitus tipe-2 yang di dalamnya mengandung data tersensor kanan menjadi variabel yang berisi sensor kanan. Dalam penelitian ini diterapkannya metode GCV (Generalized Cross-Validation) pada regresi nonparametrik smoothing spline untuk menentukan parameter smoothing yang optimal. Untuk mendapatkan nilai GCV, maka perlu adanya nilai estimasi parameter untuk membentuk hat matrix yang dibentuk dari $\hat{f}_\lambda$. Estimasi parameter untuk smoothing spline dari fungsi $f(\cdot)$ dengan meminimumkan PRSS (Penalized Residual Sum of Square). Ketika nilai GCV yang optimal sudah didapatkan maka akan membentuk estimasi fungsi regresi yang terbaik. Parameter penghalus optimal dengan cara memilih λ yang memiliki nilai GCV minimum. Dari hasil penelitian, nilai GCV paling minimum bernilai 0.0378 pada saat $\lambda=0.015$ dan untuk nilai GCV dengan pembobot Kaplan-Meier bernilai 30.4773.

Kata Kunci: *Data Tersensor, Diabetes Melitus, Kaplan Meier, Regresi Nonparametrik, Smoothing Spline.*

A. Pendahuluan

Analisis survival yaitu teknik analisis data waktu pada suatu kejadian atau *event* tertentu. Waktu yang dimaksud yaitu waktu ketahanan yang merupakan data mengukur waktu *event* tertentu seperti kematian, kegagalan, perceraian, dan sebagainya. [1] waktu ketahanan memiliki dua komponen yang harus didefinisikan yaitu titik awal pengamatan dan titik akhir pengamatan yang tercapai ketika *event* sudah terjadi.

Dalam data waktu survival tidak diketahui yang artinya data itu tidak mengikuti sebuah distribusi tertentu yang sudah ada maka, digunakanlah metode nonparametrik. Metode nonparametrik terdapat tiga metode yang terkenal yaitu, metode *Nelson Aalen*, metode *life table*, dan metode *Kaplan- Meier*. Namun pada penelitian ini penulis menggunakan metode *Kaplan-Meier*.

[2] Dalam hal ini, analisis regresi digunakan untuk memodelkan atau mencari pola hubungan antara satu atau lebih variabel bebas dengan satu atau lebih variabel respon. Terdapat dua teknik pendekatan yang dapat digunakan dalam analisis regresi yaitu pendekatan regresi parametrik dan pendekatan regresi nonparametrik. Pada regresi nonparametrik bisa menjadi alternatif karena penggunaan tidak terikat pada asumsi-asumsi yang terdapat pada regresi parametrik. [3] telah mengembangkan pendekatan regresi nonparametrik antara lain *spline*, *kernel*, *polynomial lokal*, *wavelet*, dan *fourier*.

Regresi nonparametrik dapat diaplikasikan ke dalam kehidupan sehari-hari. Pada penelitian ini yang mengaplikasikan regresi nonparametrik dalam bidang kesehatan khususnya pada data pasien penderita diabetes melitus Tipe-2. Dalam penelitian ini akan digunakan regresi nonparametrik *smoothing spline*.

Diabetes melitus atau penyakit gula adalah penyakit kronis yang berlangsung jangka Panjang. Penyakit ini ditandai dengan meningkatnya kadar gula dalam darah (glukosa) hingga di atas normal. Penyakit diabetes yang tidak terkontrol dengan baik dapat menyebabkan konsekuensi serius, yaitu dapat menyebabkan kerusakan pada berbagai organ dan jaringan tubuh contohnya, jantung, ginjal, mata dan saraf. Terdapat dua jenis diabetes, yaitu diabetes tipe1 dan tipe 2 [12].

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut: Bagaimana bentuk estimasi fungsi model regresi nonparametrik *smoothing spline* untuk kasus rawat inap pasien Diabetes Melitus Tipe-2 di Rumah Sakit X pada tahun 2020. Selanjutnya, tujuan dalam penelitian ini untuk mendapatkan estimasi fungsi model regresi nonparametrik *smoothing spline* terbaik pada kasus rawat inap pasien diabetes melitus tipe-2 di Rumah Sakit X pada tahun 2020.

B. Metodologi Penelitian

Peneliti menggunakan metode untuk penerapan regresi nonparametrik *smoothing spline* dalam kasus pasien rawat inap penderita penyakit diabetes melitus tipe-2 di Rumah Sakit X. Populasi dalam penelitian ini adalah seluruh pasien rawat inap penderita penyakit diabetes melitus tipe-2 di Rumah Sakit X.

Bahan yang digunakan merupakan data sekunder yaitu sebanyak 25 data yang diperoleh dari penelitian terdahulu yaitu oleh [4] data yang terdiri dari satu variabel respon dan satu variabel bebas. Kemudian, data yang didapatkan sebanyak 25 rawat inap pasien penderita penyakit diabetes melitus tipe 2. Waktu penelitian yang dilakukan penelitian ini yaitu pada awal bulan Maret sampai dengan akhir bulan Maret 2020. Adapun variabel yang digunakan dalam penelitian ini yaitu sebagai berikut :

1. Variable respon yang digunakan dalam penelitian ini terdiri dari waktu survival (T) yaitu waktu lamanya pasien penderita diabetes melitus tipe 2 menjalani rawat inap dengan satuan waktu hari. Sedangkan, status keputusan (δ) menunjukkan pengkategorian data tersensor, dengan status 0 menyatakan lamanya pasien apabila pasien pulang dalam keadaan belum membaik serta apabila pasien meninggal dunia pada saat rawat inap, sedangkan status 1 menyatakan untuk pasien yang pulang dalam kondisi membaik.

2. Variabel bebas yang digunakan dalam penelitian ini yaitu usia pasien penderita penyakit diabetes melitus tipe-2 yang sedang menjalankan rawat inap, dengan satuan dalam tahun.

Menganalisis data diawali dengan menyusun data. Selanjutnya untuk mendapatkan pemodelan data dengan regresi nonparametrik *smoothing spline* maka dilakukan beberapa tahapan sebagai berikut:

Identifikasi dan Pengumpulan Data

Penulis mengidentifikasi serta mengumpulkan data yang diperlukan dalam penelitian ini

1. *Input Data*
Penulis melakukan penginputan data ke dalam *Microsoft Excel*.
2. Analisis Deskriptif
Penulis melakukan analisis deskriptif untuk melihat gambaran umum mengenai data yang akan digunakan dalam penelitian ini.
3. Estimasi *Kaplan-Meier*
Penulis melakukan estimasi *Kaplan-Meier* untuk mengetahui peluang survival dari pasien rawat inap penderita diabetes melitus tipe 2
4. Analisis *Smoothing Spline*
Penulis melakukan analisis *Smoothing Spline* untuk mengetahui model terbaik. Adapun langkah-langkah berikut ini:
 1. Menginput data berpasangan variabel T dan Z
 2. Menginput titik knot yang merupakan estimasi parameter penghalus
 3. Menghitung nilai *Generalized Cross Validation* ($GCV(\lambda)$)
 4. Membuat matrik $K = Q^t R^{-1} Q$
 5. Menghitung $\hat{f}_\lambda$ pada Persamaan sebagai berikut:

$$\hat{f}_\lambda = (I + \lambda K)^{-1} T = S_\lambda T$$
 6. Menghitung $H_\lambda = (I + \lambda K)^{-1} I$
 7. Membuat matriks W atau sebagai pembobot *Kaplan-Meier* dengan elemen w_i sebagai berikut:

$$w_i = \frac{\delta_{(i)}}{n - i + 1} \prod_{j=1}^{i-1} \left(\frac{n - j}{n - j + 1} \right)^{\delta_{(j)}}$$

8. Menghitung $\hat{f}_\lambda$ pada persamaan berikut :

$$\hat{f}_\lambda = (W + \lambda K)^{-1} W$$
9. Menganalisis dan menginterpretasikan hasil yang telah didapatkan dari estimasi fungsi analisis regresi nonparametrik *smoothing spline*.

C. Hasil Penelitian dan Pembahasan

Penelitian ini menjelaskan gambaran umum serta hasil analisis yang telah penulis olah yaitu dari data pasien rawat inap penderita penyakit diabetes melitus tipe-2 dengan menggunakan analisis regresi nonparametrik *smoothing spline* dengan data yang berhubungan dengan waktu survival.

Berdasarkan hasil analisis deskriptif, dari 25 pasien penderita penyakit diabetes melitus tipe 2 diperoleh bahwa pasien rawat inap penderita penyakit diabetes melitus tipe-2 terdapat 5 pasien yang keadaanya dapat dikatakan belum membaik maupun dalam keadaan meninggal dunia. Berikut merupakan gambaran umum dari variabel yang diteliti.

Tabel 1. Analisis Deskriptif

	N	Minimum	Maximum	Mean	Std. Deviation
USIA	25	43	86	61.28	12.495

Berdasarkan pada **Tabel 1**, diperoleh bahwa pasien rawat inap penderita penyakit diabetes melitus tipe-2 terdapat 25 pasien dengan umur tertinggi pasien yaitu pada umur 86 taun dan umur terendah pada pasien yaitu umur 43 tahun, dan rata-rata pasien penderita diabetes melitus tipe 2 yaitu pada umur 61,28 tahun.

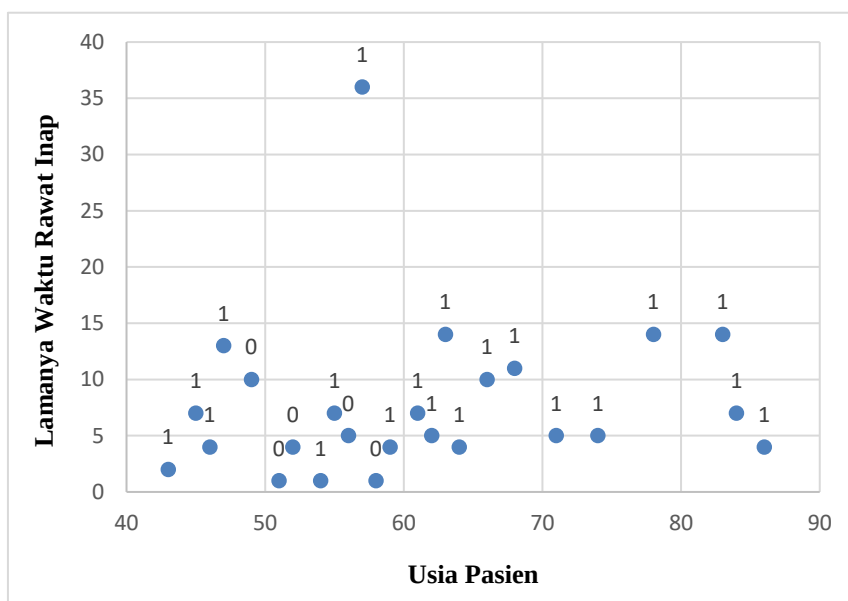
Analisis survival dengan estimasi *Kaplan Meire* digunakan untuk mengetahui peluang survival pasien rawat inap penderita penyakit diabetes melitus tipe-2. Berikut data pasien rawat inap penderita penyakit diabetes melitus tipe-2 yang disajikan pada

Tabel 2. Data Waktu Survival Pasien Penderita Penyakit Diabetes Tipe 2

No	Waktu Rawat Inap (Hari)	Status Kepastian
1	1	0
2	1	1
3	1	0
⋮	⋮	⋮
25	36	1

Pada **Tabel 4.2** di atas merupakan data waktu survival pasien dengan status kepastian yang telah dijelaskan sebelumnya bahwa terdapat dua pengkategorian yaitu 0 dan 1. Status 0 menyatakan lamanya pasien apabila pasien pulang dalam keadaan belum membaik atau pasien meninggal dunia pada saat rawat inap, sedangkan status 1 menyatakan untuk pasien yang pulang dalam kondisi membaik.

Berikut adalah *scatter plot* antara varibel usia (X) dan variabel lama waktu rawat inap (Y).



Gambar 1. Scatter Plot antara varibel usia (X) dan variabel lama waktu rawat inap (Y).

Dari gambar di atas dapat dilihat bahwa data berpecah membentuk suatu fungsi data yang pola hubungannya tidak sederhana dan menunjukkan model untuk fungsi tertentu tidak dapat diwakili suatu fungsi regresi parametrik seperti linear.

Analisis Regresi Nonparametrik**Tabel 3.** Peluang Survival Pasien Penderita Penyakit Diabetes Tipe 2

Lama Waktu Rawat Inap	Usia	n_j	d_j	Status Tersensor	$S(t)$
1	54	25	1	1	0.96
1	51			0	0
1	58			0	0
2	43	22	1	1	0.92
4	46	21	4	1	0.74
4	59			1	0.74
4	64			1	0.74
4	86			1	0.74

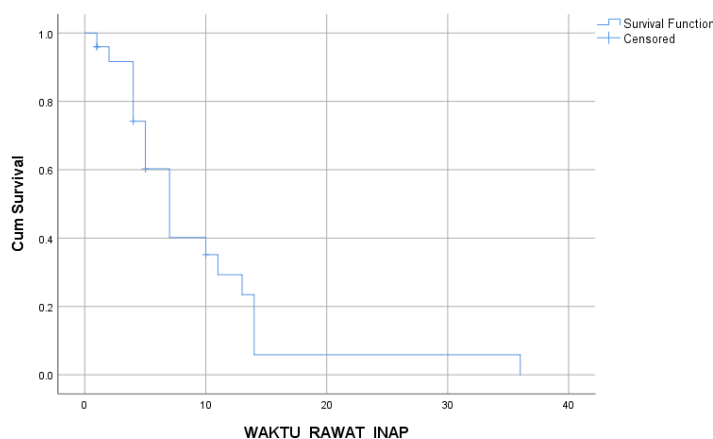
Tabel 4. Peluang Survival Pasien Penderita Penyakit Diabetes Tipe 2 (Lanjutan)

Lama Waktu Rawat Inap	Usia	n_j	d_j	Status Tersensor	$S(t)$
4	52			0	0
5	62	16	3	1	0.6
5	71			1	0.6
5	74			1	0.6
5	56			0	0
7	45	12	4	1	0.4
7	55			1	0.4
7	61			1	0.4
7	84			1	0.4
10	66	8	1	1	0.35
10	49			0	0
11	68	6	1	1	0.29
13	47	5	1	1	0.23
14	63	4	4	1	0.06
14	78			1	0.06
14	83			1	0.06
36	57	1	1	1	0

Berdasarkan hasil perhitungan dari persamaan (2.1) diperoleh bahwa pasien diabetes tipe-2 yang keluar dari rumah sakit dalam keadaan membaik terjadi pada rawat inap hari ke-1 dengan sebanyak satu pasien dari 25 pasien dengan peluang survivalnya adalah $S(1) = 0.96$, hari ke-2 dengann sebanyak satu pasien dari 22 pasien dengan peluang survivalnya adalah $S(2) = 0.92$, hari ke-4 sebanyak 4 pasien dari 21 pasien dengan kemungkinan untuk kondisi pulang membaik sebesar $S(4) = 0.74$, hari ke-5 sebanyak 3 pasien dengan dari 16 pasien kemungkinan untuk pulang dengan kondisi membaik sebesar $S(5) = 0.60$, hari ke-7 sebanyak 4 pasien dengan dari 12 pasien kemungkinan untuk pulang dengan kondisi membaik sebesar $S(7) = 0.40$, hari ke-10 sebanyak 1 pasien dengan dari 8 pasien kemungkinan untuk pulang

dengan kondisi membaik sebesar $S(10) = 0.35$, hari ke-11 sebanyak 1 pasien dengan dari 6 pasien kemungkinan untuk pulang dengan kondisi membaik sebesar $S(11) = 0.29$, hari ke-13 sebanyak 3 pasien dengan dari 5 pasien kemungkinan untuk pulang dengan kondisi membaik sebesar $S(13) = 0.23$, hari ke-14 sebanyak 3 pasien dengan dari 4 pasien kemungkinan untuk pulang dengan kondisi membaik sebesar $S(14) = 0.06$, dan hari ke-36 sebanyak 1 pasien dengan dari 1 pasien kemungkinan untuk pulang dengan kondisi membaik sebesar $S(36) = 0.000$.

Kemudian, kurva Kaplan Meier pasien rawat inap penderita penyakit diabetes melitus tipe 2 yang dihasilkan oleh software sebagai berikut :



Gambar 2. Kurva Kaplan-Meier Pasien Rawat Inap Pasien Penderita Diabetes Melitus Tipe-2

Gambar 4.2 menunjukkan semakin bertambahnya waktu penelitian atau peluang pasien belum pulang maka akan semakin menurun dengan asumsi semakin lamanya pasien di rawat inap maka pasien akan cepat membaik. Peluang survival terbesar yaitu pada rawat inap hari ke-1 sebesar 0.9600 atau 96%, sedangkan peluang survival terkecil yaitu sebesar 0.0000 atau 0% pada hari rawat inap ke-36. Sehingga, bahwa pasien rawat inap diabetes melitus Tipe-2 akan cepat mengalami perbaikan terhadap kondisi kesehatannya pada hari ke-36 dengan nilai peluang survivalnya sebesar 0%. Maka, semakin cepat pulang dengan kondisi keadaan membaik peluang survivalnya akan semakin kecil.

Regresi Smoothing Spline

Estimasi fungsi regresi dengan pendekatan regresi nonparametrik menggunakan persamaan (2.8), menghasilkan :

$$\hat{f} = \begin{pmatrix} 2.003 \\ 6.788 \\ 4.706 \\ 12.345 \\ 10.074 \\ 1.215 \\ \vdots \\ 13.991 \\ 13.988 \\ 7.031 \\ 3.966 \end{pmatrix}$$

Dengan λ optimal yang diperoleh sebesar 0.015. sedangkan nilai GCV yang diperoleh adalah sebesar 0.0378 dan MSE sebesar 0.0353.

Regresi Smoothing Spline dengan Pembobot Kaplan-Meier (K-M)

Estimasi fungsi regresi dengan pendekatan regresi nonparametrik menggunakan persamaan (2.18) dengan pembobot *Kaplan-Meier* menghasilkan :

$$\hat{f} = \begin{pmatrix} 2.002 \\ 6.570 \\ 4.670 \\ 11.478 \\ 9.193 \\ 1.629 \\ \vdots \\ 13.852 \\ 13.811 \\ 7.072 \\ 3.957 \end{pmatrix}$$

Dengan λ optimal yang diperoleh sebesar 0.015. sedangkan nilai GCV yang diperoleh adalah sebesar 44.2746 dan MSE sebesar 30.4373.

D. Kesimpulan

Berdasarkan hasil analisis yang telah diperoleh, maka dalam penelitian ini dapat disimpulkan sebagai berikut.

1. Dari 25 Pasien rawat inap penderita penyakit Diabetes Melitus Tipe-2 usia tertinggi penderita pada usia 86 tahun dan usia terendah pada usia 43 tahun dengan rata-rata penderita penyakit diabetes melitus yaitu pada usia 61 tahun.
2. Dengan bertambahnya waktu penelitian semakin lamanya pasien dirawat maka pasien akan cepat membaik atau akan cepat mengalami perbaikan kondisi kesehatannya. Peluang survival terbesar sebesar 0.96 pada lamanya waktu rawat inap hari ke-1 sedangkan terkecil yaitu 0.00 pada lamanya waktu rawat inap hari ke-36.
3. Penerapan *Smoothing Spline* pada data pasien rawat inap penderita penyakit Diabetes melitus tipe-2 dengan dengan nilai fungsi estimasi diperoleh sebagai berikut :

$$\hat{f} = \begin{pmatrix} 2.003 \\ 6.788 \\ 4.706 \\ 12.345 \\ 10.074 \\ 1.215 \\ \vdots \\ 13.991 \\ 13.988 \\ 7.031 \\ 3.966 \end{pmatrix}$$

4. Berdasarkan estimasi fungsi pada regresi smoothing spline dinyatakan dengan bertambahnya tingginya usia pasien penderita penyakit Diabetes Melitus tipe-2 maka semakin bertambah waktu lamanya waktu rawat inap.
Kemudian, berikut nilai fungsi estimasi diperoleh pembobot estimasi Kaplan-Meier sebagai berikut :

$$\hat{f} = \begin{pmatrix} 2.002 \\ 6.570 \\ 4.670 \\ 11.478 \\ 9.193 \\ 1.629 \\ \vdots \\ 13.852 \\ 13.811 \\ 7.072 \\ 3.957 \end{pmatrix}$$

5. Estimasi smoothing spline terbaik untuk data rawat inap penderita penyakit diabetes melitus tipe-2 dengan adanya pembobotan Kaplan-Meier ataupun tidak memiliki nilai Lambda yang sama yaitu 0.015 dengan nilai CGV tidak sebesar 0.0378 dan GCV dengan pembobot Kaplan-Meier sebesar 30.4773.

Daftar Pustaka

- [1] Etikan. (2017). The Kaplan Meier Estimate in Survival Analysis. Department of Biostatistic, Near East University Faculty of Medicine
- [2] Budiantara, N. I. (2009). Spline dalam Regresi Nonparametrik : Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang.
- [3] Eubank, R. L. (1999). NonParametric Regression and Spline Smoothing. New York: Mercel Dekker Inc
- [4] Departemen Kesehatan RI. (2009). Klasifikasi Umur menurut Kategori. Jakarta: Ditjen Yankes.
- [5] Dursun Aydın, E. Y. (2016). Right-Censored Nonparametric Regression: A Comparative Simulation Study.
- [6] Engle, R. F., Granger, C. W., Rice, J., & Andrew, W. (1986). Semi Parametric Estimasi of the Relation Between Weater and Electrycity Sales.
- [7] Klein, J. P., & Moeschberger, M. L. (2005). Survival Analysis Techniques for Censored and Truncated Data. Springer-Verlag New York.
- [8] Wati, L. (2014). Penentuan Parameter Penghalus Smoothing Spline dalam Regresi Semiparametrik dengan CGV.
- [9] WHO (World Health Organization). (2016). Tentang Populasi Lansia.
- [10] Witten, D., Tibshieani, R., James, G., & Hastie, T. (2021). An Introduction to Statistical Learning.
- [11] Wu, H., & Zhang, J. T. (2006). Nonparametric Regression Methods for Longitudinal Data Analysis. Rochester, New York: A John Wiley & Sons, INC.
- [12] Fatmawati and N. A. K. Rifai, "Klasifikasi Penyakit Diabetes Retinopati Menggunakan Support Vector Machine dengan Algoritma Grid Search Cross-validation," *Jurnal Riset Statistika*, pp. 79–86, Jul. 2023, doi: 10.29313/jrs.v3i1.1945.