
Pengaruh Indeks Massa Tubuh dan Trigliserida Terhadap Gula Darah dengan Model Regresi Nonparametrik Spline Biprediktor

Dewi Rahma Ente^{1*}, Anna Islamiyati², Raupong³
¹²³ Departemen Statistika, Fakultas MIPA
Universitas Hasanuddin, Makassar, 90245, Indonesia

* Corresponding author, email: Dewi01136@gmail.com

Abstract

The regression approach can be carried out using three approaches namely parametric, nonparametric and semiparametric approaches. Nonparametric regression is a statistical method used to see the relationship between the response variable and the predictor variable when the shape of the data curve is unknown. Diabetes mellitus (DM) or commonly called diabetes is a disease that is found and observed in various parts of the world today. DM is often marked by a significant increase in blood sugar levels. In this study using blood sugar levels as response variables, body mass index and triglycerides as predictor variables. Data were analyzed using truncated linear spline with one, two and three point knots experiments. The best model is obtained based on the minimum generalized cross validation (GCV) value. The results obtained that the best model is linear spline using three point knots.

Keywords: Diabetes Mellitus, GCV, Nonparametric regression, Spline truncated.

Abstrak

Pendekatan regresi dapat dilakukan dengan menggunakan tiga pendekatan yaitu pendekatan parametrik, nonparametrik dan semiparametrik. Regresi Nonparametrik merupakan suatu metode statistik yang digunakan untuk melihat hubungan antara variabel respon dengan variabel prediktor ketika bentuk kurva data tidak diketahui. Diabetes mellitus (DM) atau biasa disebut penyakit kencing manis merupakan penyakit yang banyak ditemukan dan diamati diberbagai dunia saat ini. DM sering ditandai dengan kenaikan kadar gula darah secara signifikan. Pada penelitian ini menggunakan kadar gula darah sebagai variabel respon, indeks massa tubuh dan trigliserida sebagai variabel prediktor. Data dianalisis menggunakan *spline linier truncated* dengan percobaan satu, dua dan tiga titik knot. Model terbaik diperoleh berdasarkan nilai *generalized cross validation (GCV)* yang minimum. Hasil yang didapatkan bahwa model terbaik yaitu spline linier dengan menggunakan tiga titik knot.

Kata Kunci: Diabetes Melitus, GCV, Regresi nonparametrik, Spline truncated.

1. Pendahuluan

Analisis regresi merupakan salah satu teknik statistik yang bertujuan untuk melihat pola hubungan antara variabel respon dengan variabel prediktor dengan mengestimasi pola regresinya. Penentuan estimasi kurva regresi dapat dilakukan dengan tiga pendekatan yaitu regresi parametrik, regresi nonparametrik dan regresi semiparametrik [1]. Pendekatan regresi parametrik digunakan ketika pola kurva regresi diketahui,

pendekatan regresi nonparametrik digunakan ketika pola kurva regresi tidak diketahui, dan apabila sebagian bentuk pola kurva regresi diketahui dan sebagian lagi tidak diketahui maka pendekatan yang digunakan adalah pendekatan regresi semiparametrik.

Pendekatan nonparametrik memiliki fleksibilitas yang tinggi karena dapat mengestimasi bentuk kurvanya berdasarkan sebaran data dengan memperhatikan kemulusan kurva tanpa dipengaruhi faktor subyektifitas peneliti [2]. Regresi nonparametrik terdiri atas beberapa metode pendekatan salah satunya adalah spline. Spline merupakan fungsi yang diperoleh dengan meminimumkan *Ordinary Least Square* (OLS) yaitu kriteria estimasi yang menggabungkan *goodness of fit* dengan kemulusan kurva [3]. Salah satu kelebihan pendekatan Spline adalah model cenderung mencari sendiri estimasi data kemanapun pola data tersebut bergerak. Kelebihan ini terjadi karena dalam Spline terdapat titik-titik knot, yaitu titik-perpaduan bersama yang menunjukkan terjadinya perubahan pola perilaku data [1]. Pengembangan regresi spline telah banyak dilakukan oleh peneliti diantaranya spline kubik [4], spline penalized [5], dan spline kuantil [6-7].

Diabetes Melitus (DM) merupakan penyakit yang banyak ditemukan dan diamati diberbagai dunia saat ini. Penyakit ini merupakan salah satu masalah kesehatan yang besar. WHO (2016) mencatat bahwa 70% total kematian di dunia diakibatkan karena penyakit tidak menular salah satunya DM. 90 – 95% dari kasus diabetes adalah diabetes tipe II yang sebagian besar dapat dicegah karena disebabkan oleh gaya hidup tidak sehat. Indonesia juga menghadapi situasi ancaman diabetes serupa dengan dunia. *International Diabetes Federation* (IDF) Atlas 2017 melaporkan bahwa epidemi Diabetes di Indonesia masih menunjukkan kecenderungan meningkat. Indonesia adalah negara peringkat keenam di dunia setelah Tiongkok, India, Amerika Serikat, Brazil dan Meksiko dengan jumlah penyandang Diabetes usia 20-79 tahun sekitar 10,3 juta orang [8]. Sejalan dengan hal tersebut, Riset Kesehatan Dasar (Riskesdas) memperlihatkan peningkatan angka prevalensi Diabetes yang cukup signifikan, yaitu dari 6,9% di tahun 2013 menjadi 8,5% di tahun 2018, sehingga estimasi jumlah penderita di Indonesia mencapai lebih dari 16 juta orang. Karena hal tersebut diharapkan adanya upaya untuk mengetahui faktor yang mempengaruhi dan dapat menekan pertambahan jumlah penderita DM di Indonesia.

Penelitian ini menggunakan data penderita diabetes mellitus II dan dimodelkan dengan menggunakan regresi nonparametrik *spline truncated*. *Truncated* merupakan sebuah fungsi yang dapat diartikan sebagai fungsi potongan. Metode tersebut dipilih karena berdasarkan plot pada studi awal yang telah dilakukan, diketahui bahwa plot variabel-variabel yang mempengaruhi gula darah pasien penderita DM di Rumah Sakit Pendidikan Universitas Hasanuddin tidak diketahui bentuk polanya

2. Material dan Metode

Data yang digunakan pada penelitian ini adalah data hasil rekam medis penderita penyakit diabetes melitus tipe II (DM II) tahun 2014-2018 di Rumah Sakit Pendidikan Universitas Hasanuddin. Jumlah data yang digunakan sebanyak 85 sampel. Penelitian ini menggunakan satu variabel respon (y) dan dua variabel prediktor (x_1 dan x_2). Faktor gula darah yang dinyatakan dalam mg/dL sebagai variabel respon (y), indeks massa tubuh (x_1) dinyatakan dalam Kg/m dan trigliserida (x_2) dinyatakan dalam mg/dL sebagai variabel prediktor. Data dianalisis menggunakan regresi nonparametrik biprediktor dengan menggunakan estimator spline truncated.

Misalkan diberikan data berpasangan $(x_{1i}, x_{2i}, \dots, x_{mi}, y_i)$, $i = 1, 2, \dots, n$ mengikuti model regresi nonparametrik multiprediktor sebagai berikut

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{mi}) + \varepsilon_{ji} \quad (1)$$

Bila persamaan (1) diasumsikan bersifat adiktif dan dihipotesiskan dengan estimator spline maka diperoleh model regresi nonparametrik spline dengan beberapa prediktor sebagai berikut

$$y_i = \beta_0 + \sum_{j=1}^m \left(\sum_{h=1}^p \beta_{jh} x_{ji}^h + \sum_{l=1}^r \beta_{jp+l} (x_{ji} - \kappa_{jl})_+^p \right) + \varepsilon_{ji} \quad (2)$$

Estimasi parameter β dapat dilakukan dengan meminimumkan jumlah kuadrat galat dengan menggunakan metode *least square*.

$$\varepsilon' \varepsilon = (\mathbf{y} - \mathbf{X}\beta)' (\mathbf{y} - \mathbf{X}\beta) \quad (3)$$

Misalkan $\varepsilon' \varepsilon = Q$ maka selanjutnya diturunkan terhadap vektor β dan disamakan dengan nol, maka diperoleh hasil estimasi sebagai berikut

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \quad (4)$$

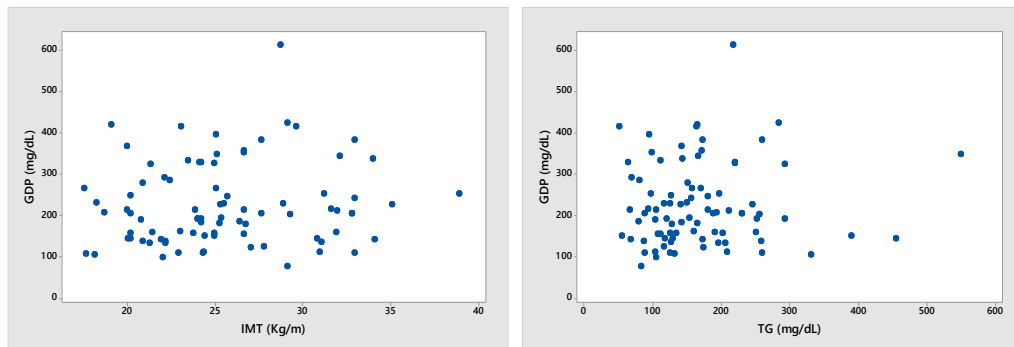
Dari estimasi model regresi nonparametrik spline dengan beberapa prediktor yang diperoleh pada persamaan (4) selanjutnya akan diterapkan dengan data penderita DM II dengan model regresi nonparametrik spline biprediktor sebagai berikut

$$y_i = \beta_0 + \sum_{j=1}^2 \left(\sum_{h=1}^p \beta_{jh} x_{ji}^h + \sum_{l=1}^r \beta_{jp+l} (x_{ji} - \kappa_{jl})_+^p \right) + \varepsilon_{ji} \quad (5)$$

Langkah awal pada penelitian ini adalah membuat scatterplot untuk melihat pola hubungan antara indeks massa tubuh dan trigliserida terhadap gula darah pasien DM II. Selanjutnya memodelkan gula darah pasien DM II dengan menggunakan 1, 2 dan 3 titik knot secara linier. Setelah itu menentukan model terbaik berdasarkan nilai GCV minimum, kemudian menginterpretasikan model yang diperoleh.

3. Hasil dan Diskusi

Identifikasi pola hubungan antara variabel respon gula darah (y) dengan variabel prediktor indeks massa tubuh (x_1) dan trigliserida (x_2) dapat dilihat pada hasil *scatter plot* yang ditunjukkan pada Gambar 1.



Gambar 1. *Scatter Plot* variabel respon dan variabel prediktor data pasien DM II di rumah sakit pendidikan unhas

Berdasarkan Gambar 1 dapat dilihat bahwa pola hubungan gula darah pasien DM II dengan variabel prediktor yaitu Indeks Massa Tubuh (x_1) dan Trigliserida cenderung (x_2) tidak mengikuti pola tertentu. Oleh karena itu data gula darah pasien DM II dengan variabel prediktor akan dimodelkan menggunakan regresi nonparametrik estimator spline truncated.

3.1 Pemilihan Titik Knot Optimal Gula Darah Pasien DM II (y) Terhadap Indeks Massa Tubuh (x_1)

Pemodelan spline linier dimulai dengan pemilihan titik knot optimal dari satu, dua dan tiga titik knot. Pemilihan titik knot menggunakan metode trial and error pada nilai indeks massa tubuh sampai ditemukan nilai GCV yang minimum. Nilai GCV indeks massa tubuh untuk satu, dua dan tiga titik knot pada spline linier terdapat pada Tabel 1.

Tabel 1. Nilai GCV Indeks Massa Tubuh (x_1) Model Spline Linier Satu, Dua dan Tiga Titik Knot

No	Titik Knot			GCV
	K1	K2	K3	
1	26.4	-	-	10418.05
	27.05	-	-	10406.25
	28.76	-	-	10336.01
	29.13	-	-	10350.05
2	27.68	28.76	-	10492.92
	28.76	30.86	-	10377.53
	28.76	31.11	-	10389.59
	28.76	31.63	-	10446.53
3	26.66	28.76	30.86	10186.5
	27.68	28.76	30.86	10047.24
	27.81	28.76	30.86	10000.78
	28.76	30.86	31.11	10613.2

Berdasarkan Tabel 1 untuk indeks massa tubuh (x_1) didapatkan titik knot optimal berada pada tiga titik knot dengan nilai GCV minimum sebesar 10000,78 pada titik knot $K1 = 27,81$, $K2 = 28,76$ dan $K3 = 30,86$

3.2 Pemilihan Titik Knot Optimal Gula Darah Pasien DM II (y) Terhadap Trigliserida (x_2)

Nilai GCV trigliserida untuk satu, dua dan tiga titik knot pada spline linier terdapat pada Tabel 2.

Tabel 2. Nilai GCV Trigliserida (x_2) Model Spline Linier Satu, Dua dan Tiga Titik Knot

No	Titik Knot			GCV
	K1	K2	K3	
1	151	-	-	10517.23
	157	-	-	10504.01
	165	-	-	10484.03
	167	-	-	10482.35
	171	-	-	10484.62
	172	-	-	10485.44
2	118	167	-	10153.9
	119	167	-	10148.92
	121	167	-	10146.48
	126	167	-	10149.55
	127	167	-	10158.15
	128	167	-	10165.02
3	106	121	167	10399.96
	105	121	167	10400.42
	121	165	167	10176.53
	121	167	171	10002.07
	121	167	175	9871.054
	121	167	181	9906.002

Berdasarkan Tabel 2 untuk trigliserida (x_2) didapatkan titik knot optimal berada pada tiga titik knot dengan nilai GCV minimum sebesar 9871,054 pada titik knot $K1 = 121$, $K2 = 167$ dan $K3 = 175$.

3.3 Pemilihan Titik Knot Optimal Regresi Nonparametrik Biprediktor Spline Linier

Pada tahap ini dicari titik knot optimal pada data gula darah pasien DM II (y) dengan dua variabel prediktor yang diduga mempengaruhi yaitu indeks massa tubuh (x_1) dan

trigliserida (x_2). Nilai GCV Biprediktor untuk satu, dua dan tiga titik knot pada spline linier terdapat pada Tabel 3.

Tabel 3. Nilai GCV Spline linier Biprediktor

	Titik Knot			GCV
	K1	K2	K3	
X ₁	27.81	28.76	30.86	9458.2
X ₂	121	167	175	
X ₁	27.68	28.76	30.86	9491.048
X ₂	121	167	176	
X ₁	27.68	28.76	30.86	9487.068
X ₂	121	167	175	
X ₁	27.68	28.76	30.86	9559.422
X ₂	121	167	181	
X ₁	27.81	28.76	30.86	9529.943
X ₂	121	167	181	

Berdasarkan Tabel 3 spline linier biprediktor didapatkan titik knot optimal berada nilai GCV minimum sebesar 9458,2. Hasil ini digunakan dalam pemodelan gula darah pasien penderita DM tipe II di Rumah Sakit Pendidikan Universitas Hasanuddin. Hasil estimasi parameter disajikan dalam Tabel 4.

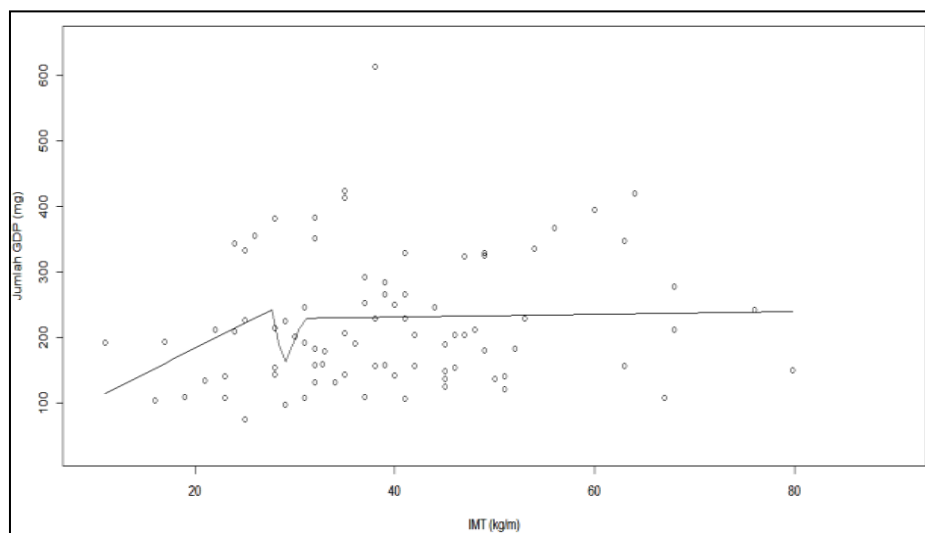
Tabel 4. Estimasi Spline Linier Biprediktor dengan Tiga Titik Knot

Parameter	Estimasi
β_0	209.64
β_1	4.91
β_2	132.88
β_3	-224.15
β_4	97.3
β_5	-1.39
β_6	4.94
β_7	-17.09
β_8	13.62

Berdasarkan estimasi parameter pada Tabel 4 diperoleh estimasi model spline kubik terboboti dengan knot optimal adalah:

$$\hat{y} = 209,64 + 4,91x_1 + 132,88(x_1 - 27,81) - 224,15(x_1 - 28,76) + 97,3(x_1 - 30,86) - 1,39x_2 + 4,94(x_2 - 121) - 17,09(x_2 - 167) + 13,62(x_2 - 175)$$

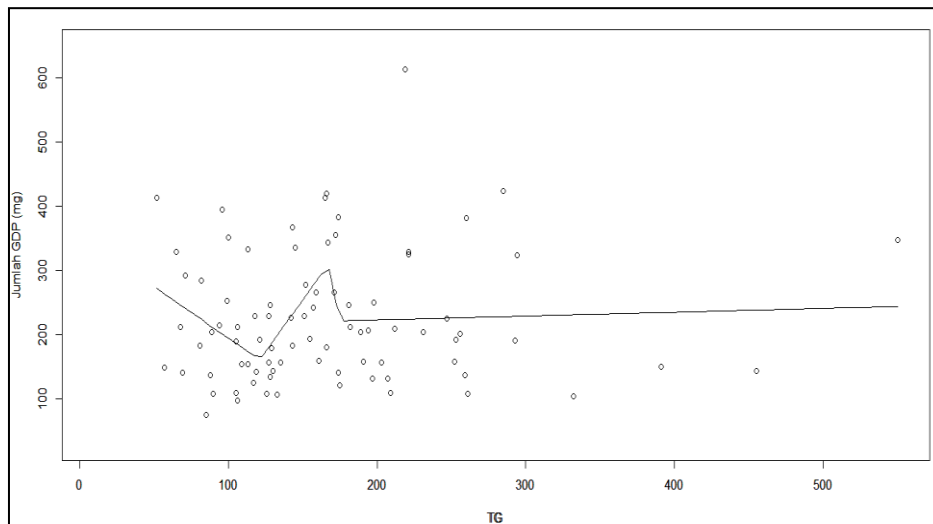
Berdasarkan hasil estimasi model spline linier biprediktor pada Gambar 2 hubungan gula darah penderita DM II dengan Indeks Massa Tubuh (IMT), optimal pada penggunaan tiga titik knot yaitu 27,81 Kg/m, 28,76 Kg/m, dan 30,76 Kg/m. Hasil ini menunjukkan bahwa terdapat empat pola perubahan gula darah berdasarkan indeks massa tubuh (x_1). Pertama, gula darah cenderung naik hingga IMT sebesar 27,81Kg/m. Kedua gula darah perlahan turun pada IMT pada rentang 27,81 Kg/m - 28,76 Kg/m. Pola ketiga gula darah cenderung naik secara signifikan pada antara IMT 28,76 Kg/m - 30,76 Kg/m, dan keempat gula darah naik secara lambat pada IMT besar dari 30,76 Kg/m. Ini berarti bahwa kenaikan IMT mempengaruhi gula darah pasien penderita DM II. Pada pola pertama berarti kenaikan gula darah pada IMT 27,81 Kg/m berada pada tahap pra obesitas. Pola ketiga dan keempat kenaikan gula darah pasien penderita DM II berada pada tahap obesitas tingkat 1 dan obesitas tingkat 2. Penimbunan lemak yang berlebihan di dalam tubuh dapat mengakibatkan resistensi insulin yang berpengaruh terhadap kadar glukosa darah. Ini berarti bahwa pihak tenaga medis diharapkan melakukan pengontrolan yang lebih ketika pasien DM II memiliki Indeks Massa Tubuh berlebih.



Gambar 2. Grafik estimasi spline linier biprediktor dengan knot optimal untuk faktor IMT

Untuk variabel x_2 yaitu trigliserida, optimal pada tiga titik knot yaitu pada nilai 121 mg/dL, 167 mg/dL dan 175 mg/dL. Hasil ini menunjukkan bahwa terdapat empat pola perubahan gula darah terhadap trigliserida. Pada pola perubahan yang pertama, gula darah penderita DM II masih berada diatas kenormalan hingga 121 mg/dL. Pada pola perubahan yang kedua yaitu gula darah cenderung naik saat trigliserida berada pada 121 mg/dL hingga 167 mg/dL. Pola ketiga, terjadi penurunan gula darah saat trigliserida berada pada 167 mg/dL hingga 175 mg/dL. Pola keempat terjadi peningkatan gula darah cenderung lambat namun sangat berisiko pada trigliserida besar dari 175 mg/dL. karena

Kadar trigliserida yang tinggi memberikan sinyal adanya resistansi insulin (sebuah kondisi dimana tubuh tidak berespon terhadap insulin) akibatnya kadar gula darah dan trigliserida terbaca tinggi. Ini berarti bahwa pihak tenaga medis juga diharapkan melakukan pengontrolan yang lebih ketat ketika pasien DM II memiliki kadar Trigliserida besar dari 167 mg/dL karena pada kondisi tersebut dikhawatirkan mengakibatkan komplikasi kardiovaskuler.



Gambar 3. Grafik estimasi spline linier biprediktor dengan knot optimal untuk faktor Trigliserida

4. Kesimpulan

Berdasarkan hasil yang diperoleh dapat disimpulkan bahwa variabel indeks massa tubuh dan trigliserida berpengaruh terhadap kenaikan dan penurunan gula darah pasien penderita DM II. Model spline linier biprediktor yang terbentuk optimal pada penggunaan tiga titik knot sehingga menghasilkan model berikut

$$\hat{y} = 209,64 + 4,91x_1 + 132,88(x_1 - 27,81) - 224,15(x_1 - 28,76) + 97,3(x_1 - 30,86) - 1,39x_2 + 4,94(x_2 - 121) - 17,09(x_2 - 167) + 13,62(x_2 - 175)$$

Pada penelitian ini menggunakan estimator spline truncated, sehingga untuk penelitian berikutnya dapat menggunakan estimator spline lainnya seperti spline smooting dan spline penalized serta dapat menggunakan multivariabel.

Daftar Pustaka

- [1] Budiantara, I.N. *Spline dalam Regresi Nonparametrik dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang*. Pidato Pengukuhan Untuk Jabatan Guru Besar pada Jurusan Statistika, Surabaya: Institut Teknologi Sepuluh Nopember (ITS), 2009.

- [2] Eubank, R.L. *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York. 1988.
- [3] Wahba G. *Spline Models for Observational Data*. SIAM Pennsylvania. 1990.
- [4] Ramdhani, Z.A., Islamiyati, A., dan Raupong. Hubungan Faktor Kolesterol Terhadap Gula Darah Diabetes dengan Spline Kubik Terbobot. *ESTIMASI: Journal of Statistics and Its Application*, 1 (1) : 32-39, 2020.
- [5] Islamiyati, A., Fatmawati, and Chamidah, N. Penalized Spline Estimator with Multi Smoothing Parameters in Bi-Response Multi-Predictor Nonparametric Regression Model For Longitudinal Data . *Songklanakarinn Journal of Sciences and Technology*, 42(4): 897-909, 2020.
- [6] Putri, W.N.A., Islamiyati, A., dan Anisa. Penggunaan Regresi Kuantil Spline Multivariat pada Perubahan Trombosit Pasien Demam Berdarah Dengue. *ESTIMASI: Journal of Statistics and Its Application*, 1 (1) : 1-9, 2020.
- [7] Aprilia, B., Islamiyati, A., Anisa, dan Ilyas, N. Estimasi Model Regresi Kuantil Spline Kuadratik pada Data Trombosit dan Hematokrit Pasien DBD. *ESTIMASI: Journal of Statistics and Its Application*, 1 (2) : 58-64, 2020.
- [8] Infodatim. Pusat Data dan Informasi Kementerian Kesehatan RI : Hari Diabetes Sedunia, ISSN: 2442-7659, 2019.