

PENERAPAN ANALISIS LASSO DALAM MENENTUKAN PENYEBAB UTAMA FAKTOR KEMATIAN PASIEN KOMORBID YANG TERINFEKSI COVID-19

Agung Lesmana Siregar¹, Rina Filia Sari², Rina Widyasari³

Program Studi Matematika Fakultas Sains Dan Teknologi, Universitas Islam Negeri Sumatera Utara Medan^{1,2,3}

Email: agunglesmanasiregar@gmail.com¹, rinafiliasari@uinsu.ac.id²,
rinawidyasari@uinsu.ac.id³

Corresponding Author: Agung Lesmana Siregar email: agunglesmanasiregar@gmail.com

Abstrak. Tingkat kematian akibat COVID-19 saat ini sangatlah mengkhawatirkan. Pada pasien dengan komorbid (penyakit pembawa) tingkat kematian semakin tinggi. Tujuan dari penelitian ini adalah untuk mengetahui komorbid apa saja yang berisiko kematian akibat COVID-19 berdasarkan ketepatan datang pengobatan di RSU Haji Kota Medan, dengan menggunakan metode LASSO (Least Absolute Shrinkage and Selection Operator). Keunggulan metode ini adalah mendapatkan hasil koefisien nol atau mendekati nol, yang menyebabkan terseleksi variabel independen dengan diketahui koefisien dari usia, kadar gula darah, tekanan darah, hemoglobin, kreatinin dan saturasi oksigen. Berdasarkan koefisien hemoglobin memiliki nilai tertinggi dan menjadi variabel independen yang berpengaruh terhadap kematian akibat COVID-19. Hemoglobin dapat mempengaruhi tingkat oksigen di dalam tubuh karena saturasi oksigen adalah presentase hemoglobin yang mengikat oksigen. Jika tekanan hemoglobin turun menyebabkan sesak nafas sampai menyebabkan sesak nafas sampai menyebabkan kematian. Sehingga hemoglobin menjadi variabel independen yang berpengaruh terhadap kematian akibat COVID -19.

Kata Kunci: CoViD-19, LASSO (Least Absolute Shrinkage and Selection Operator), Hemoglobin.

Abstract. The current death rate due to COVID-19 is very worrying. In patients with comorbidities (carrier diseases) the mortality rate is higher. The aim of this research is to find out what comorbidities are at risk of death due to COVID-19 based on the accuracy of arriving for treatment at the Haji RSU Medan City, using the LASSO (Least Absolute Shrinkage and Selection Operator) method. The advantage of this method is that it produces a coefficient of zero or close to zero, which results in the selection of independent variables with known coefficients of age, blood sugar levels, blood pressure, hemoglobin, creatinine and oxygen saturation. Based on the hemoglobin coefficient, it has the highest value and is an independent variable that influences deaths due to COVID-19. Hemoglobin can affect oxygen levels in the body because oxygen saturation is the percentage of hemoglobin that binds oxygen. If the hemoglobin pressure drops, it causes shortness of breath, causing shortness of breath and causing death. So hemoglobin becomes an independent variable that influences death due to COVID-19.

Keywords: CoViD-19, LASSO (Least Absolute Shrinkage and Selection Operator), Hemoglobin.

A. Pendahuluan

Penyakit merupakan tidak terjadinya mekanisme adaptasi suatu organisme untuk bereaksi secara tepat terhadap rangsangan sehingga menimbulkan gangguan pada fungsi dan struktur dari bagian sistem terhadap tubuh. Penyakit menular merupakan suatu penyakit yang dapat berpindah gangguan pada fungsi bagian sistem tubuh secara langsung maupun melalui perantara. Sebagai contoh, dilaporkan pada 31 Desember 2019, WHO kantor negara China menyatakan adanya penyakit pneumonia yang tidak diketahui asalnya di Kota Wuhan, Provinsi Hubei, Cina. Coronavirus Disease 2019 (COVID-19) adalah penyakit menular yang disebabkan oleh Severe Acute Respiratory Syndrome Coronavirus 2 (SARSCoV-2). COVID-19 masuk ke



Indonesia khususnya di Sumatera Utara pada tanggal 18 Maret 2020. Diketahui kasus yang terkonfirmasi sebanyak 106.519 pasien positif COVID-19. Menurut data dari Dinas Kesehatan Kota Medan, jumlah kematian akibat COVID-19 di Kota Medan yang terkonfirmasi hingga tanggal 04 Juli 2021 berjumlah 643 orang.

Terdapat beberapa penelitian terdahulu yang relevan dengan penelitian ini. Penelitian Galvao dan Roncalli, (2021) dengan metode Survival Analysis Model Cox PH juga menyatakan pasien COVID-19 yang memiliki komorbiditas dapat meningkatkan risiko kematian pasien COVID-19. Metode yang digunakan pada penelitian ini adalah LASSO. LASSO adalah pengembangan regresi ridge yang menyelesaikan bentuk regresi yang mempunyai multikolinearitas. LASSO ialah metode regresi yang dapat menurunkan koefisien regresi menjadi nol. Perbedaan antara regresi LASSO dan regresi ridge terletak pada kendala perkiraan koefisien regresi. Perbedaan masalah antara kedua metode tersebut membuat koefisien regresi yang diperoleh dari regresi LASSO membuat lebih kecil dibanding dengan koefisien regresi ridge.

Sehingga koefisien regresi ridge hanya diturunkan mendekati nol, sedangkan koefisien regresi LASSO memungkinkan untuk menurun sampai tepat nol. Ini membuat kelebihan pada regresi LASSO yaitu dipakai sebagai menyeleksi variabel independen pada bentuk, sehingga hanya variabel-variabel terbaik yang masuk kedalam bentuk. Hal ini juga bertujuan memudahkan dalam menginterpretasikan bentuk regresi. Oleh itu, LASSO adalah penyelesaian dari kendala pada interpretasi dan multikolinearitas yang sering terjadi pada analisis regresi berganda. Oleh karena itu, LASSO dapat menyusutkan koefisien menjadi nol. Hal ini mengakibatkan LASSO dapat membuat bentuk dengan peubah penjelas yang lebih sedikit dan ringkas.

B. Metode Penelitian

a Jenis Penelitian

Penelitian ini adalah penelitian terapan melalui pendekatan kuantitatif. Penelitian terapan berhubungan dengan kenyataan-kenyataan praktis, penerapan serta pengembangan ilmu pengetahuan yang dihasilkan oleh penelitian mendasar dalam kehidupan.

b Sumber Data

Sumber data yang diperoleh peneliti berasal dari perolehan data sekunder atau berupa dokumentasi tertulis tentang kasus kematian akibat teridentifikasi COVID-19 beserta komorbid pasien yang diperoleh dari RSUD Haji Kota Medan.

c Variabel Penelitian

Variabel Dependen Didefinisikan sebagai waktu pengobatan pasien yang mulai teridentifikasi COVID-19 di RSUD Haji Kota Medan. Variabel Independen dalam penelitian ini adalah komorbid pasien yang mulai teridentifikasi COVID-19 di RSUD Haji Kota Medan yang terdiri dari usia yang mulai teridentifikasi COVID-19, tingkat gula darah rata-rata (mg/dl), tingkat tekanan darah rata-rata (mmHg), tingkat status kadar hemoglobin (g/dL), tingkat kadar kreatinin (U/L) dan saturasi oksigen (%).

d Analisis data

Pendeskripsian data bertujuan untuk menampilkan keseluruhan jumlah data orang yang meninggal akibat COVID-19 dan komorbidnya. Semua variabel maupun berdasarkan masing-masing variabel dan mengetahui persentase data pasien yang meninggal akibat COVID-19 dengan komorbid dan waktu pengobatan. Bentuk dasar persamaan *Least Absolute Shrinkage and Selection Operator* pada persamaan sebagai berikut :

$$LASSO = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$$

Dengan menganggap semua variabel berpengaruh terhadap model, maka semua variabel dimasukkan ke dalam model tersebut.



e Prosedur Penelitian

Prosedur penelitian yang digunakan dalam menganalisis penelitian ini adalah pengumpulan data kematian kasus COVID-19 dan komorbidnya yang diperoleh dari RSUD Haji Kota Medan, mendeskripsikan data kematian kasus COVID-19 dengan variabel independen, melakukan perhitungan regresi antar variabel, melakukan perhitungan Metode Kuadrat Terkecil (*Ordinary Least Square/OLS*), mendeteksi adanya multikolinieritas, melakukan uji *Least Absolute Shrinkage And Selection Operator (LASSO)* dan pemilihan parameter terbaik.

C. Hasil Penelitian Dan Pembahasan

a. Perhitungan Regresi Antar Variabel

Setelah mencari analisis data akan dicari masing-masing setiap A matriks ukuran 7x7, agar mempermudah pengerjaan analisis regresi linear berganda enam variabel. Dengan diketahui nilai n adalah 43, yang diambil dari jumlah pasien sebanyak 43 orang.

$$A = \begin{bmatrix} 43 & 2582 & 7219 & 73,965 & 523,8 & 132,97 & 3485 \\ 2582 & 160650 & 434786 & 4412,42 & 32064,1 & 6694,72 & 209953 \\ 7219 & 434786 & 1294323 & 9762,6 & 68696,5 & 24569,41 & 587327 \\ 73,965 & 4412,42 & 12659,26 & 133,6434 & 889,4775 & 244,9631 & 42282,5 \\ 523,8 & 32064,1 & 87459,2 & 889,4775 & 6707,84 & 25679,44 & 2721660 \\ 132,97 & 6694,72 & 24569,41 & 244,9631 & 25679,44 & 2000,5501 & 158059,4 \\ 3485 & 209953 & 587327 & 42282,5 & 2077134 & 158059,36 & 288415 \end{bmatrix}$$

$$H = \begin{bmatrix} 200 \\ 11902 \\ 33494 \\ 346,37 \\ 2546,5 \\ 492,53 \\ 16224 \end{bmatrix}$$

Adapun hasil determinannya adalah

$$\beta_0 \quad 1,62449E+28$$

$$\det(A) \quad 8,39693E+28$$

$$\det(A1) \quad -2,2725E+26$$

$$\det(A2) \quad -,77014E+25$$

$$\det(A3) \quad -9,3903E+27$$

$$\det(A4) \quad 3,19512E+26$$

$$\det(A5) \quad -3,4359E+27$$

$$\det(A6) \quad 3,07878E+25$$



b. Metode Kuadrat Terkecil

Tujuan dari metode *Ordinary Least Square* adalah meminimumkan Jumlah Kuadrat Galat (JKG). Langkah-langkah dalam meminimumkan JKG adalah mencari Faktor Koreksi (FK), Jumlah Kuadrat Total (JKT), Jumlah Kuadrat Kelompok (JKK), Jumlah Kuadrat Perlakuan (JKP), dan Jumlah Kuadrat Galat (JKG). Pengerjaannya dapat dilihat sebagai berikut.

$$FK = \frac{\sum_{i=1}^n \hat{X}_i^2}{p.k}$$

$$= \frac{196468860,1}{43.6}$$

$$= 761507,2095$$

$$JKT = \sum_{ij} \hat{X}_{ij}^2 - FK$$

$$= 14016,735 - 761507,2095$$

$$= 990722,824$$

$$JKK = \sum_{ij} \frac{X_j^2}{p} - FK$$

$$= \frac{14016,74}{6} - 761507,2095$$

$$= 11109064,17$$

$$JKP = \sum_{ij} \frac{X_i^2}{k} - FK$$

$$= \frac{28033,47}{6} - 761507,2095$$

$$= 32762549,99$$

$$JKG = JKT - JKK - JKP$$

$$= 990722,824 - 11109064,17 - 32762549,99$$

$$= -42880891,33$$

c. Uji F

Tabel 1 Uji F

Model		Sum Squares	of Df	Mean Square	F	Sig.
1	Regression	37362,202	6	6227,034	445,275	,000 ^b
	Residual	517,434	37	13,985		
	Total	37879,636	43			

Berdasarkan *output* di atas, diketahui nilai signifikansi untuk pengaruh Usia, KGD, TD, Hb, Kreatinin dan Saturasi secara simultan terhadap ketepatan waktu pengobatan adalah sebesar $0,000 < 0,05$ dan nilai F hitung $445,275 > F$ tabel 2,36 sehingga dapat disimpulkan bahwa terdapat pengaruh Usia, KGD, TD, Hb, Kreatinin dan Saturasi secara simultan terhadap ketepatan waktu pengobatan.



d. Uji T

Tabel 2 Uji T

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	2,021E-5	,591		,000	1,000
	Usia	-,078	,057	-1,001	-1,361	,182
	KGD	,001	,014	,030	,058	,954
	TD	,857	1,482	,315	,578	,566
	Hb	,487	,228	1,267	2,132	,040
	Kreatinin	-,032	,113	-,022	-,286	,777
	Saturasi	,023	,042	,404	,550	,586

Diketahui nilai signifikansi untuk pengaruh Hb terhadap ketepatan pengobatan adalah sebesar $0,040 < 0,05$ dan nilai t hitung $2,132 > t$ tabel $2,028$. Sehingga dapat disimpulkan bahwa H_1 diterima yang berarti terdapat pengaruh Hb terhadap ketepatan pengobatan.

e. Mendeteksi Adanya Multikolinearitas

Tabel 3 Mendeteksi Adanya Multikolinearitas

Model	Unstandardized Coefficients		Standardized Coefficients		Collinearity Statistics		
	B	Std. Error	Beta	T	Sig.	Tolerance	VIF
1 (Constant)	2,021E-5	,591		,000	1,000		
Usia	-,078	,057	-1,001	-1,361	,182	,001	1465,935
KGD	,001	,014	,030	,058	,954	,001	706,608
TD	,857	1,482	,315	,578	,566	,001	802,589
HB	,487	,228	1,267	2,132	,040	,001	955,490
Kreatinin	-,032	,113	-,022	-,286	,777	,061	16,499
Saturasi	,023	,042	,404	,550	,586	,001	1463,073

f. LASSO

$$\begin{aligned}
 \text{LASSO} &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 \\
 &= (1,62449\text{E}+28) - 0,00270635 X_1 - 0,00044899 X_2 \\
 &\quad - 0,11183018 X_3 + 0,003805105 X_4 - 0,04091849 X_5 + 0,000366655 X_6
 \end{aligned}$$

Tabel 4 LASSO

Variabel independen	Koefisien sebelum penyusutan LASSO	Koefisien LASSO
X_1 (Usia)	0,182	-0,00270635
X_2 (KGD)	0,954	-0,00044899
X_3 (TD)	0,566	-0,11183018
X_4 (Hb)	0,040	0,003805105
X_5 (Kreatinin)	0,777	-0,04091849
X_6 (Saturasi)	0,586	0,000366655



Dapat dilihat berdasarkan tabel di atas koefisien dari variabel independen mengalami penyusutan oleh persamaan LASSO. Sesuai dengan urutan berdasarkan nilai tertinggi dari koefisien LASSO adalah Hb, Saturasi, KGD, Kreatinin, TD dan Usia.

g. Parameter Terbaik

Berdasarkan hasil penelitian dari analisis regresi linear berganda, metode kuadrat terkecil, mendeteksi adanya multikolinearitas sampai uji LASSO, yang memiliki hasil yang signifikan di dalam metode kuadrat terkecil adalah Hb (Hemoglobin) dan memiliki nilai tertinggi berdasarkan koefisien LASSO. Besarnya Hb juga dipengaruhi oleh ketepatan pengobatan dengan rata-rata pengobatan 4 hari dari seluruh jumlah hari pengobatan setiap observasi.

D. Kesimpulan

Berdasarkan penelitian terdahulu yang relevan dengan penelitian ini adalah penelitian Galvao dan Roncalli (2021) dengan metode *Survival Analysis Model Cox PH* juga menyatakan pasien COVID-19 yang memiliki komorbiditas dapat meningkatkan risiko kematian pasien COVID-19. Penelitian ini menggunakan metode *Survival Analysis Model Cox*, yang menghasilkan PH sebagai komorbid yang berpengaruh. Sehingga peneliti mengembangkan penelitian ini dengan metode yang berbeda. Berdasarkan analisis LASSO komorbid pasien yang paling berpengaruh yaitu Hb (Hemoglobin) dengan koefisien 0,003805105 koefisien tersebut paling mendekati nilai nol diantara enam variabel independen yang diteliti. Hb dapat mempengaruhi tingkat kadar tinggi atau rendahnya kadar oksigen di dalam tubuh. Serta berhubungan dengan saturasi oksigen, dimana saturasi oksigen adalah presentase hemoglobin yang mengikat oksigen dibandingkan dengan jumlah total hemoglobin yang ada di dalam darah. Besarnya Hb juga dipengaruhi oleh ketepatan pengobatan dengan rata-rata pengobatan 4 hari dari seluruh jumlah hari pengobatan setiap observasi.

DAFTAR PUSTAKA

- Andana, Aulia, Diah dan Agus. 2017. Model Regresi Menggunakan Least Absolute Shrinkage And Selection Operator (Lasso) Pada Data Banyaknya Gizi Buruk Kabupaten/Kota Di Jawa Tengah. *Jurnal Gaussian: Volume 6 Nomor 1 Halaman 21 – 30*.
- Banguna, Rita Br dan Aida. 2020. Pemetaan Kemiskinan Melalui Pendekatan Geographically Weighted Lasso. *Jurnal Ekonomi Indonesia: Volume 9 Nomor 3 Halaman 233 – 246*.
- Chen, Stephan, Khairil dan Septian. 2020. Penerapan Analisis Lasso Dan Group Lasso Dalam Mengidentifikasi Faktor-Faktor Yang Berhubungan Dengan Tuberkulosis Di Jawa Barat. *Indonesian Journal of Statistics and Its Applications: Volume 4 Nomor 1 Halaman 39 – 54*.
- Fadli Rizal. “Alasan Pria Dinilai Lebih Rentan Mengidap Corona”. *Halo Doc*, 20 April 2020, <https://www.halodoc.com/artikel/alasan-pria-dinilai-lebih-rentan-mengidap-corona>. Diakses pada 19 April 2022
- Fanny, Resty, Anik dan Aam. 2018. Pendugaan Produktivitas Penangkapan Bagan Perahu Dengan Regresi Gulud, Lasso Dan Elastic – Net. *Jurnal Departemen Statistika Institut Pertanian Bogor: Volume 2 Nomor 2 Halaman 7 - 14*.
- Gujarati DN. *Dasar-Dasar Ekonometrika*. Edisi ketiga. Jilid 1. Jakarta: Erlangga; 2006.



- Ghozali, I. 2011. *Aplikasi Analisis Multivariat dengan Program IBM SPSS19*. Semarang: Badan Penerbit Universitas Diponegoro.
- Hasan, Iqbal. 2013. *Analisis data penelitian dengan statistik*. Jakarta: Bumi Aksara.
- Hastie, Trevor, Robert dan Jerome. 2008. *The Elements of Statistical Learning*. California: Springer.
- Heriagus, Alfin. “Memahami Saturasi Oksigen Kritis pada Pasien CoViD- 19”. *Universitas Indonesia*, 16 Juli 2021, <https://www.ui.ac.id/memahami-saturasi-oksigen-kritis-pada-pasien-covid-19/>. Diakses pada 25 November 2021
- Kementerian Kesehatan RI. 2020. *Buku Pedoman P2 CoViD-19 Revisi 5*. Jakarta: Kementerian Kesehatan RI.

