

**PEMODELAN JUMLAH KASUS KEMATIAN COVID-19 DI
JAWA TIMUR DENGAN PENDEKATAN REGRESI
BINOMIAL NEGATIF BERDASARKAN ESTIMATOR
*PENALIZED SPLINE***

SKRIPSI



MELLYA PUJI LESTARI

**PROGRAM STUDI S-1 STATISTIKA
DEPARTEMEN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS AIRLANGGA**

2022

**PEMODELAN JUMLAH KASUS KEMATIAN COVID-19 DI
JAWA TIMUR DENGAN PENDEKATAN REGRESI
BINOMIAL NEGATIF BERDASARKAN ESTIMATOR
*PENALIZED SPLINE***

SKRIPSI



MELLYA PUJI LESTARI

**PROGRAM STUDI S-1 STATISTIKA
DEPARTEMEN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS AIRLANGGA**

2022

**PEMODELAN JUMLAH KASUS KEMATIAN COVID-19 DI
JAWA TIMUR DENGAN PENDEKATAN REGRESI
BINOMIAL NEGATIF BERDASARKAN ESTIMATOR
*PENALIZED SPLINE***

SKRIPSI

Sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Statistika
pada Fakultas Sains dan Teknologi
Universitas Airlangga

Mellya Puji Lestari
NIM 081811833002

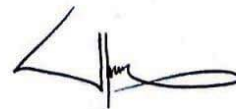
Disetujui oleh :

Pembimbing I,



Dr. Nur Chamidah, S.Si., M.Si.
NIP. 197206021998022001

Pembimbing II,



Dr. Toha Saifudin, S.Si., M.Si.
NIP. 197501061999031002

LEMBAR PENGESAHAN NASKAH SKRIPSI

Judul : Pemodelan Jumlah Kasus Kematian Covid-19 di Jawa Timur dengan Pendekatan Regresi Binomial Negatif Berdasarkan Estimator *Penalized Spline*

Penyusun : Mellya Puji Lestari

NIM : 081811833002

Pembimbing I : Dr. Nur Chamidah, S.Si., M.Si.

Pembimbing II : Dr. Toha Saifudin, S.Si., M.Si.

Tanggal Seminar : 20 Juni 2022

Disetujui oleh:

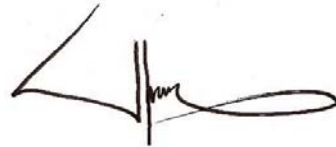
Pembimbing I



Dr. Nur Chamidah, S.Si., M.Si.

NIP. 197206021998022001

Pembimbing II



Dr. Toha Saifudin, S.Si., M.Si.

NIP. 197501061999031002

Mengetahui,

Ketua Departemen Matematika
Fakultas Sains dan Teknologi
Universitas Airlangga



Dr. Herry Suprajitno, S.Si., M.Si.

NIP. 196804041994031020

Koordinator Prodi S-1 Statistika
Fakultas Sains dan Teknologi
Universitas Airlangga



Drs. H. Sediono, M.Si.

196107121987011001

PEDOMAN PENGGUNAAN SKRIPSI

Skripsi ini tidak dipublikasikan, namun tersedia di perpustakaan dalam lingkungan Universitas Airlangga, diperkenankan untuk dipakai sebagai referensi kepustakaan, tetapi pengutipan harus seizin penyusun dan harus menyebutkan sumbernya sesuai kebiasaan ilmiah. Dokumen skripsi ini merupakan hak milik Universitas Airlangga.

SURAT PERNYATAAN TENTANG ORISINALITAS

Yang bertanda tangan di bawah ini, saya:

Nama : Melly Puji Lestari

NIM : 081811833002

Program Studi : Statistika

Fakultas : Sains dan Teknologi Universitas Airlangga

Jenjang : Sarjana (S1)

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul :

Pemodelan Jumlah Kasus Kematian Covid-19 di Jawa Timur dengan Pendekatan Regresi Binomial Negatif Berdasarkan Estimator *Penalized Spline*

Apabila suatu saat nanti melakukan tindakan plagiat, maka saya akan menerima sanksi yang telah ditetapkan. Demikian surat pernyataan ini saya buat dengan sebenar-benarnya.

Surabaya, 01 Juli 2022



Melly Puji Lestari

NIM. 081811833002

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Allah SWT karena telah memberikan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Pemodelan Jumlah Kasus Kematian Covid-19 di Jawa Timur dengan Pendekatan Regresi Binomial Negatif Berdasarkan Estimator *Penalized Spline*”. Pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Prof. Dr. Mohammad Nasih, S.E., M.T., Ak., CMA. Selaku Rektor Universitas Airlangga, Prof. Dr. Moh Yasin, M. Si. Selaku Dekan Fakultas Sains dan Teknologi Universitas Airlangga, dan Drs. H. Sediono, M.Si. selaku Ketua Program Studi S-1 Statistika Universitas Airlangga yang telah memberikan kesempatan pada penulis untuk menimba ilmu di Program Studi Statistika Universitas Airlangga.
2. Kemendikbud atas beasiswa pendidikan (Bidikmisi) yang penulis dapatkan.
3. Ibu Dr. Nur Chamidah, S.Si., M.Si. selaku dosen pembimbing I dan Bapak Dr. Toha Saifudin, S.Si., M.Si. selaku dosen pembimbing II yang senantiasa memberikan nasehat, bimbingan serta dukungan.
4. Seluruh dosen Statistika Universitas Airlangga yang telah memberikan ilmu dan pengalaman yang bermanfaat selama perkuliahan.
5. Keluarga tercinta yang telah memberikan dukungan dan doa untuk penulis.
6. Teman-teman seperjuangan di Universitas Airlangga.
7. Ratih R.S, Mindarti Arifa, Aisah Nur K., Putri Fardha, Sakinah Priandi, Oky Putri R selaku sahabat penulis yang telah memberikan dukungan dan kebersamai penulis dalam penyusunan skripsi.

Penulis menyadari bahwa masih banyak kekurangan dalam penyusunan skripsi ini sehingga penulis mengharapkan kritik serta saran yang membangun dari semua pihak guna menyempurnakan skripsi ini.

Surabaya, 01 Juli 2022



Mellya Puji Lestari

Mellya Puji Lestari, 2022. **Pemodelan Jumlah Kasus Kematian Covid-19 di Jawa Timur dengan Pendekatan Regresi Binomial Negatif Berdasarkan Estimator *Penalized Spline***. Skripsi di bawah bimbingan Dr. Nur Chamidah, S.Si., M.Si. dan Dr. Toha Saifudin, S.Si., M.Si., Program Studi S-1 Statistika, Departemen Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga, Surabaya.

ABSTRAK

Jumlah kematian Covid-19 masih terus meningkat seiring berjalannya waktu pada tahun 2020. Jawa Timur pernah menjadi provinsi dengan jumlah kematian tertinggi. Faktor yang diduga menjadi penyebab bertambahnya kasus konfirmasi dan kasus kematian Covid-19 antara lain diabetes melitus, hipertensi, dan *tuberculosis*. Penelitian ini bertujuan untuk memodelkan jumlah kasus kematian Covid-19 di Jawa Timur yang dipengaruhi faktor penyakit penyerta, yaitu persentase penderita diabetes melitus, persentase penderita hipertensi di atas 15 tahun, dan persentase penderita *tuberculosis*. Untuk memodelkan jumlah kasus kematian Covid-19 di Jawa Timur yang berupa data count, maka dapat menggunakan regresi poisson. Namun pada kasus sehari-hari ditemukan adanya kasus overdispersi. Regresi negatif binomial dapat mengatasi kasus overdispersi. Dalam penelitian ini, analisis yang digunakan adalah pendekatan regresi binomial negatif nonparametrik menggunakan algoritma *localscore* dan *backfitting* berdasarkan estimator *penalized spline*. Estimator *spline* digunakan karena memiliki fleksibilitas lebih yang dapat digunakan dalam perubahan pola perilaku secara lokal. Dari ketiga faktor penyakit penyerta yang mempunyai pengaruh terbesar terhadap jumlah kasus kematian Covid-19 di Jawa Timur adalah *tuberculosis*. Saat persentase penderita *tuberculosis* lebih dari 0,1 persen, maka setiap kenaikan satu persen penderita *tuberculosis* maka jumlah kematian Covid-19 akan mengalami perubahan sebesar 4,8 kali jumlah kasus kematian semula. Dari hasil penelitian, estimasi jumlah kematian Covid-19 di Jawa Timur dengan model regresi binomial negatif nonparametrik sudah sesuai dengan nilai deviance $0,056 < \chi_{(20;0,05)}^2 = 31,414$ dan diperoleh MAPE sebesar 22,099% termasuk dalam kategori cukup baik. Kedua nilai tersebut lebih kecil dibandingkan dengan nilai deviance dan MAPE pada model regresi binomial negatif parametrik yang masing-masing sebesar 11,137 dan 23,72%. Hal ini membuktikan bahwa pemodelan kasus kematian Covid-19 di Jawa Timur lebih baik menggunakan regresi binomial negatif dengan pendekatan nonparametrik berdasarkan estimator *penalized spline* daripada menggunakan regresi binomial negatif dengan pendekatan parametrik.

Kata Kunci : Kasus Kematian Covid-19, Regresi Binomial Negatif, Overdispersi, Nonparametrik, Penalized Spline

Mellya Puji Lestari, 2022. **Modeling the Number of Covid-19 Death Cases in East Java with a Negative Binomial Regression Approach Based on the Penalized Spline Estimator**. The undergraduate thesis is guided by Dr. Nur Chamidah, S.Si., M.Si., and Dr. Toha Saifudin, S.Si., M.Si., Statistics Study Program, Mathematics Department, Faculty of Science and Technology, Airlangga University, Surabaya.

ABSTRACT

The number of Covid-19 deaths continues to increase over time in 2020. East Java was once the province with the highest number of deaths. Factors suspected to be the cause of the increase in confirmed cases and death cases of Covid-19 include diabetes mellitus, hypertension, and tuberculosis. This study aims to model the number of Covid-19 deaths in East Java that are influenced by comorbid factors, namely the percentage of people with diabetes mellitus, the percentage of people with hypertension over 15 years, and the percentage of people with tuberculosis. To model the number of Covid-19 deaths in East Java in the form of count data, you can use Poisson regression. However, in everyday cases, there are cases of overdispersion. Binomial negative regression can solve the case of overdispersion. In this study, the analysis used is a nonparametric negative negative binomial regression approach using a local scoring and backfitting algorithm based on a penalized spline estimator. The spline estimator is used because it has more flexibility to be used in changing behavior patterns locally. Of the three comorbidities that have the greatest influence on the number of Covid-19 deaths in East Java, tuberculosis is. When the percentage of tuberculosis patients is more than 0.1 percent, then for every one percent increase in tuberculosis patients, the number of Covid-19 deaths will experience a change of 4.8 times the original number of deaths. From the results of the study, the estimated number of Covid-19 deaths in East Java using a nonparametric negative binomial regression model was in accordance with a deviance value of $0,056 < \chi^2_{(20;0,05)} = 31,414$ and obtained an MAPE of 22,099% which was included in the fairly good category. Both values are smaller than the deviance and MAPE values in the negative parametric binomial regression model which are 11,137 and 23,72%, respectively. This proves that it is better to model Covid-19 death cases in East Java using negative binomial regression with a nonparametric approach based on a penalized spline estimator than using negative binomial regression with a parametric approach.

Keyword : *Number of Covid-19 Death Cases, Negative Binomial Regression, Overdispersion, Nonparametric, Penalized Spline Estimator*

DAFTAR ISI

	Halaman
LEMBAR JUDUL	i
LEMBAR PENGESAHAN NASKAH SKRIPSI.....	iii
PEDOMAN PENGGUNAAN SKRIPSI	iv
SURAT PERNYATAAN TENTANG ORISINALITAS	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	5
1.5 Batasan Masalah	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Virus Corona 2019 (Covid-19)	6
2.2 Kematian/Mortalitas	7
2.3 Faktor-Faktor Terkait Jumlah Kasus Kematian Covid-19.....	7
2.4 Regresi Nonparametrik	8
2.5 Overdispersi dan Distribusi Binomial Negatif.....	9
2.6 Regresi Binomial Negatif	10
2.7 <i>Generalized Additive Model</i> (GAM)	11
2.8 <i>Estimator Penalized Spline</i>	12
2.9 Pemilihan Parameter Penghalus λ Optimal.....	16
2.10 <i>Deviance</i>	17
2.11 Algoritma Local Scoring.....	18
2.12 Ketepatan Estimasi Jumlah Kasus Kematian Covid-19.....	19

2.13	<i>Open Source Software-R (OSS-R)</i>	20
BAB III METODE PENELITIAN.....		21
3.1	Data dan Sumber Data	21
3.2	Variabel Penelitian.....	21
3.3	Langkah-langkah Analisis	22
BAB IV HASIL DAN PEMBAHASAN		228
4.1	Karakteristik Data	228
4.2	Algoritma untuk Mengestimasi Model Regresi Binomial Negatif berdasarkan Estimator Penalized Spline	29
4.3	Estimasi dan Analisis Model Jumlah Kasus Kematian Covid-19 di Jawa Timur.....	31
BAB V PENUTUP.....		41
5.1	Kesimpulan	41
5.2	Saran	42
DAFTAR PUSTAKA		423
LAMPIRAN.....		447

DAFTAR TABEL

Nomor	Judul Tabel	Halaman
2.1	Tingkat Akurasi Prediksi Berdasarkan Nilai MAPE	19
3.1	Variabel Penelitian	22
4.1	Deskripsi Jumlah Kasus Kematian Covid-19 di Jawa Timur dengan 3 Variabel Prediktor	28
4.2	Statistik Uji <i>Goodness of Fit</i> Model Regresi Poisson	31
4.3	Nilai Parameter Penghalus Optimal Berdasarkan GCV Minimum	34
4.4	Validasi Menggunakan MAPE	40

DAFTAR GAMBAR

Nomor	Judul Gambar	Halaman
3.1	<i>Flowchart</i> untuk Menentukan Orde Polinomial, Parameter Penghalus, dan Jumlah Knot	26
3.2	<i>Flowchart</i> Algoritma <i>Localscoring</i>	27
4.1	<i>Scatter Plot</i> Persentase Penderita <i>Diabetes Melitus</i> terhadap Jumlah Kasus Kematian Covid-19	32
4.2	<i>Scatter Plot</i> Persentase Penderita Hipertensi diatas 15 Tahun terhadap Jumlah Kasus Kematian Covid-19	33
4.3	<i>Scatter Plot</i> Persentase Penderita <i>Tuberculosis</i> terhadap Jumlah Kasus Kematian Covid-19	33
4.4	Plot estimasi Jumlah Kasus Kematian Covid-19 <i>insample</i>	40

DAFTAR LAMPIRAN

No	Judul
1	Data
2	<i>Output</i> Statistika Deskriptif Data Jumlah Kasus Kematian Covid-19 dan Variabel yang diduga Berpengaruh di Provinsi Jawa Timur
3	<i>Script</i> Program R untuk Mengecek Gejala <i>Overdispersion</i>
4	<i>Script</i> Program R untuk Melakukan Pemodelan Parametrik Menggunakan Regresi Binomial Negatif
5	Program Pemilihan Orde, Jumlah Knot, Titik Knot, dan Parameter Penghalus Optimal dengan Kriteria GCV Minimum untuk Masing-Masing Prediktor Berdasarkan Estimator <i>Penalized Spline</i>
6	Program <i>Localscoring</i> Untuk Mengestimasi Model Regresi Binomial Negatif Nonparametrik Berdasarkan Estimator <i>Penalized Spline</i> .
7	<i>Output</i> Uji <i>Overdispersion</i>
8	<i>Output</i> Program Pemilihan Orde, Jumlah Knot, Titik Knot, dan Parameter Penghalus Optimal dengan Kriteria GCV Minimum untuk Masing-Masing Prediktor Berdasarkan Estimator <i>Penalized Spline</i> .
9	<i>Output</i> Program <i>Localscoring</i> Untuk Mengestimasi Model Regresi Binomial Negatif Nonparametrik Berdasarkan Estimator <i>Penalized Spline</i>
10	Prediksi Data Menggunakan Model Regresi Binomial Negatif Nonparametrik Berdasarkan Estimator <i>Penalized Spline</i>
11	<i>Output</i> Program untuk Mengestimasi Model Regresi Binomial Negatif dengan Pendekatan Parametrik
12	Prediksi Data Menggunakan Model Regresi Binomial Negatif dengan Pendekatan Parametrik