

**ESTIMASI DAN PERAMALAN JUMLAH KASUS
COVID-19 DI SULAWESI SELATAN PADA DATA
GOOGLE TRENDS DAN GOOGLE MOBILITY
MENGUNAKAN MODEL ADITIF UMUM**

SKRIPSI



RIFKA YULIA SARI IFADAH LATIF

H051181506

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
OKTOBER 2022**

**ESTIMASI DAN PERAMALAN JUMLAH KASUS COVID-19
DI SULAWESI SELATAN PADA DATA GOOGLE TRENDS
DAN GOOGLE MOBILITY MENGGUNAKAN
MODEL ADITIF UMUM**

SKRIPSI

**Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Sains pada
Program Studi Statistika Departemen Statistika Fakultas Matematika dan
Ilmu Pengetahuan Alam Universitas Hasanuddin**

RIFKA YULIA SARI IFADAH LATIF

H051181506

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
OKTOBER 2022**

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

Estimasi dan Peramalan Jumlah Kasus COVID-19 Di Sulawesi Selatan Pada Data Google Trends dan Google Mobility Menggunakan Model Aditif Umum

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun

Makassar, 21 Oktober 2022



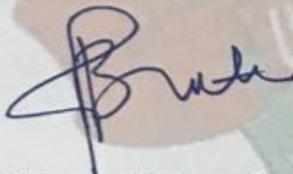
Rifka Yulia Sari Ifadah Latif

NIM H051181506

**ESTIMASI DAN PERAMALAN JUMLAH KASUS COVID-19
DI SULAWESI SELATAN PADA DATA GOOGLE TRENDS
DAN GOOGLE MOBILITY MENGGUNAKAN
MODEL ADITIF UMUM**

Disetujui Oleh:

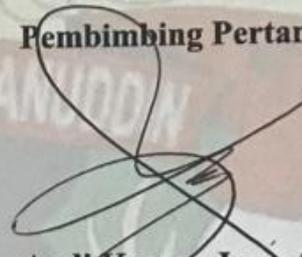
Pembimbing Utama



Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.

NIP. 19740713 199903 2 001

Pembimbing Pertama



Andi Kresna Jaya, S.Si, M.Si.

NIP. 19731228 200003 1 001

Ketua Program Studi



Dr. Nurtiti Sunusi, S.Si., M.Si.

NIP. 19720117 199703 2 002

Pada 21 Oktober 2022

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Rifka Yulia Sari Ifadah Latif
NIM : H051181506
Program Studi : Statistika
Judul Skripsi : Estimasi dan Peramalan Jumlah Kasus COVID-19 Di Sulawesi Selatan Pada Data Google Trends dan Google Mobility Menggunakan Model Aditif Umum

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

1. Ketua : Sri Astuti Thamrin, S.Si., M.Stat., Ph.D. (.....)
2. Sekretaris : Andi Kresna Jaya, S.Si., M.Si. (.....)
3. Anggota : Dr. Nurtiti Sunusi, S.Si., M.Si. (.....)
4. Anggota : Sitti Sahriman, S.Si., M.Si. (.....)

Ditetapkan di : Makassar

Tanggal : 21 Oktober 2022

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji hanya milik Allah *Subhanallahu Wa Ta'ala* atas segala limpahan rahmat dan hidayah-Nya yang telah diberikan kepada penulis sampai saat ini. Shalawat serta salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa sallam. Alhamdulillahirobbil'aalamiin*, berkat rahmat dan kemudahan yang diberikan oleh Allah *Subhanallahu Wa Ta'ala*, penulis dapat menyelesaikan tugas akhir ini.

Penulis tidak akan sampai pada titik ini tanpa dukungan dan bantuan dari pihak yang selalu ada, peduli dan menyayangi penulis. Oleh karena itu, penulis haturkan rasa terima kasih yang setulus-tulusnya serta penghargaan yang setinggi-tingginya kepada keluarga terutama orang tua penulis, Ayahanda **Latif** dan Ibunda **Masnah** yang telah memberikan dukungan penuh, pengorbanan, kesabaran hati, cinta dan kasih sayang, serta dengan ikhlas telah mengiringi setiap langkah penulis dengan doa dan restunya. Teruntuk adikku **Muhammad Rizqi** dan **Salsabila**, terima kasih telah menjadi adik yang sangat baik dan selalu ada.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada:

1. **Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.**, selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Nurtiti Sunusi S.Si., M.Si.**, selaku Ketua Departemen Statistika, segenap Dosen Pengajar dan Staf yang telah membekali ilmu dan kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.
4. **Ibu Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.**, selaku Pembimbing Utama dan **Bapak Andi Kresna Jaya S.Si., M.Si.**, selaku Pembimbing Pendamping yang dengan penuh kesabaran telah meluangkan waktu dan pemikirannya di tengah berbagai kesibukan dan prioritasnya untuk senantiasa memberikan

arahan, dorongan, dan motivasi kepada penulis mulai dari awal hingga selesainya penulisan tugas akhir ini.

5. **Ibu Dr. Nurtiti Sunusi, S.Si., M.Si., dan Ibu Sitti Sahrinan, S.Si., M.Si.,** selaku Tim Penguji yang telah memberikan kritikan yang membangun dalam penyempurnaan penyusunan tugas akhir ini serta waktu yang telah diberikan kepada penulis.
6. **Ibu Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.,** selaku Penasehat Akademik penulis. Terima kasih atas segala bantuan, nasehat serta motivasi yang selalu diberikan kepada Penulis selama menjalani pendidikan di Departemen Statistika.
7. Ibunda **Almh. Sitti Dg Layu dan Almh. Hj. Andi Upe** yang senantiasa memberikan dukungan, doa tulus dan kasih sayang, hingga di setengah perjalanan penyusunan tugas akhir ini.
8. Teman-teman **Statistika 2018**, terima kasih untuk segala kebersamaan yang tak terlupakan selama 4 tahun, kerja sama, suka dan duka serta dukungan yang telah diberikan kepada penulis setiap kali penulis menghadapi hambatan dalam menyelesaikan tugas akhir ini.
9. Teman-teman **Netijen, Fani Fahira, Claudian T Tangdilomban, Nurul Hijrah, Ika Pratiwi Haya, Reni Roihanah, Emi Astuti, Putri Aprilia S,** dan **Hafis Dwi Fernando**, yang selalu membersamai sejak awal perkuliahan, dan banyak membantu segala urusan perkuliahan, serta senantiasa menjadi tempat bertukar pikiran. Serta teman-teman **Koperasi**.
10. Teman-teman **KKN Gelombang 106 Posko Bulukumba 2** terima kasih untuk segala dukungan dan kekeluargaannya.
11. Teman-teman **Ugralieza**, yang selalu ada setiap kali penulis butuh bantuan.
12. Diri sendiri, yang telah berjuang dan selalu kembali bangkit setelah menghadapi banyak kesulitan dan tantangan hidup, terima kasih karena tidak menyerah, *u did it*.
13. Semua pihak yang tidak dapat penulis sebutkan satu-persatu, semoga segala dukungan dan partisipasi yang diberikan kepada penulis bernilai ibadah disisi Allah *Subhanallahu Wa Ta'ala*.

Penulis menyadari bahwa masih banyak kekurangan dalam tugas akhir ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini memberikan manfaat untuk pembaca.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Makassar, 21 Oktober 2022

Rifka Yulia Sari Ifadah Latif

**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMIK**

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Rifka Yulia Sari Ifadah Latif
NIM : H051181506
Program Studi : Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-eksklusif** (*Non-exclusive Royalty- Free Right*) atas tugas akhir saya yang berjudul:

**“Estimasi dan Peramalan Jumlah Kasus COVID-19 Di Sulawesi Selatan
Pada Data Google Trends dan Google Mobility
Menggunakan Model Aditif Umum”**

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada tanggal, 21 Oktober 2022

Yang menyatakan

(Rifka Yulia Sari Ifadah Latif)

ABSTRAK

Pemantauan kejadian COVID-19 dapat dilakukan dengan memanfaatkan data jejaring sosial misalnya, Google Trends dan COVID 19 - Google Mobility Reports. Data tersebut sebagai variabel prediktor dapat diintegrasikan dengan data klinis kasus positif COVID-19 di Sulawesi Selatan sebagai variabel respon, sehingga menghasilkan model yang akurat dalam meramalkan banyaknya kejadian penyakit. Jumlah kasus positif COVID-19 di Sulawesi Selatan merupakan data diskrit yang mengalami overdispersi serta tidak berdistribusi normal. Oleh karena itu, analisis yang tepat untuk digunakan adalah regresi nonparametrik yaitu model aditif umum, dengan estimator *spline* kubik dan variabel respon berdistribusi Binomial Negatif. Metode estimasi parameter dilakukan dengan menggunakan metode maksimum *likelihood*. Ukuran kebaikan model ditinjau berdasarkan nilai *Akaike's Information Criterion* (AIC) dan nilai koefisien determinasi. Untuk ukuran kebaikan peramalan ditinjau berdasarkan nilai *Root Mean Square Error of Prediction* (RMSEP) dan *Mean Absolute Percentage Error* (MAPE). Model terbaik terbentuk dari variabel Google Trends dan Google Mobility serta variabel struktur temporal dengan nilai AIC terkecil 6596,145 dan nilai koefisien determinasi terbesar yaitu 73,4%. Model tersebut juga menghasilkan ukuran kebaikan peramalan MAPE yaitu 47,9% serta nilai RMSEP yaitu 121,038.

Kata Kunci: Jumlah Kasus Positif COVID-19, Google Trends, Google Mobility Reports, Model Aditif Umum, Distribusi Binomial Negatif, *Spline* Kubik.

ABSTRACT

Monitoring of COVID-19 cases can be done by utilizing social network data, for the example, Google Trends and COVID 19 - Google Mobility Reports. These data as predictor variables can be integrated with positive cases of COVID-19 clinical data in South Sulawesi as a response variable, this will produce an accurate model to forecast the number of cases of the disease. The number of positive COVID-19 cases in South Sulawesi is a discrete data that is overdispersed and not normally distributed. Therefore, nonparametric regression analysis that is general additive model with a cubic spline estimator and the response variable with a negative binomial distribution, is right to use. The parameter estimation is solved by using the maximum likelihood method. The measure of the goodness of fit of the model is reviewed based on the value of Akaike's Information Criterion (AIC) and the value of the coefficient of determination. Then for the measure of the goodness of forecasting, it is reviewed based on the value of Root Mean Square Error of Prediction (RMSEP) and the value of Mean Absolute Percentage Error (MAPE). The best model is formed from several variables, that is Google Trends, Google Mobility and temporal structure. The model produces the smallest AIC value that is 6596,145 and the largest coefficient of determination, that is 73,4%. The model also produces value of MAPE, that is 47,9% and value of RMSEP, that is 121,038.

Keywords: Total of Positive Cases of COVID-19, Google Trends, Google Mobility Reports, Generalized Additive Model, Negative Binomial Distribution, Cubic Spline.

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL.....	ii
LEMBAR PERNYATAAN KEOTENTIKAN.....	iii
HALAMAN PERSETUJUAN PEMBIMBING	iv
HALAMAN PENGESAHAN.....	v
KATA PENGANTAR	vi
ABSTRAK	x
ABSTRACT	xi
DAFTAR ISI.....	xii
DAFTAR GAMBAR	xiv
DAFTAR TABEL.....	xv
DAFTAR LAMPIRAN.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Analisis Regresi.....	5
2.2 Regresi Linear	6
2.3 Korelasi <i>Pearson Product Moment</i>	6
2.4 Distribusi Binomial Negatif	7
2.5 Model Linear Umum	10
2.6 Model Aditif Umum.....	11
2.7 Metode <i>Maximum Likelihood Estimation</i>	13
2.8 Evaluasi Model.....	14
2.8.1 <i>Akaike's Information Criterion</i>	14
2.8.2 Nilai Koefisien Determinasi	14
2.9 Validasi Model	15
2.9.1 <i>Mean Absolute Percentage Error</i>	15
2.9.2 <i>Root Mean Square Error of Prediction</i>	16
2.10 Kasus Positif COVID-19.....	16

2.11	Google Trends	17
2.12	Google Mobility	17
BAB III METODOLOGI PENELITIAN.....		18
3.1	Sumber Data	18
3.2	Variabel Penelitian dan Definisi Operasional Variabel.....	18
3.3	Metode dan Alat Analisis Statistika	20
3.4	Tahapan Analisis Data.....	20
BAB IV HASIL DAN PEMBAHASAN		23
4.1	Statistik Deskriptif Data	23
4.2	Korelasi Variabel Prediktor Terhadap Variabel Respon	26
4.3	Estimasi Parameter Model Aditif Umum Pada Jumlah Kasus Positif COVID-19 di Sulawesi Selatan.....	28
4.4	Peramalan Jumlah Kasus Positif COVID-19 di Sulawesi Selatan	37
BAB V PENUTUP.....		40
5.1	Kesimpulan.....	40
5.2	Saran	40
DAFTAR PUSTAKA		41
LAMPIRAN		43

DAFTAR GAMBAR

Gambar 4.1 Statistik Deskriptif Data Variabel y	23
Gambar 4.2 Statistik Deskriptif Data Variabel x_1 dan x_2	24
Gambar 4.3 Statistik Deskriptif Data Variabel x_3, x_4, x_5 dan x_6	24
Gambar 4.4 Statistik Deskriptif Data Variabel x_7 dan x_8	26
Gambar 4.5 Pengaruh x_1 Terhadap y	32
Gambar 4.6. Pengaruh x_2 Terhadap y	32
Gambar 4.7 Pengaruh x_3 Terhadap y	33
Gambar 4.8 Pengaruh x_4 Terhadap y	33
Gambar 4.9 Pengaruh x_5 Terhadap y	34
Gambar 4.10 Pengaruh x_6 Terhadap y	34
Gambar 4.11 Pengaruh x_7 Terhadap y	35
Gambar 4.12 Nilai Peramalan dan Nilai Observasi Kasus Positif COVID-19 di Sulawesi Selatan	37

DAFTAR TABEL

Tabel 2.1 <i>Range</i> Nilai MAPE.....	16
Tabel 3.1 Variabel Respon dan Variabel Prediktor.....	19
Tabel 4.1 Nilai Korelasi Variabel Respon dan Variabel Prediktor	26
Tabel 4.2 Parameter Model Aditif Umum Pada Jumlah Kasus Positif COVID-19 di Sulawesi Selatan.....	31
Tabel 4.3 Nilai Evaluasi Model.....	36
Tabel 4.4 Nilai Validasi Model	38

DAFTAR LAMPIRAN

Lampiran 1. Data Variabel Respon dan Variabel Prediktor.....	44
Lampiran 2. Pengujian Overdispersi	45
Lampiran 3. Pengujian Multikolinearitas.....	46
Lampiran 4. Plot Sebaran Data Variabel Prediktor Terhadap Variabel Respon .	47
Lampiran 5. Plot <i>Time Series</i> Perbandingan <i>Trend</i> antara Variabel Respon dan Variabel Prediktor.....	48
Lampiran 6. Simulasi Perhitungan Estimasi Parameter Model.....	50
Lampiran 7. Plot <i>Time Series</i> Perbandingan Nilai Observasi dan Nilai Peramalan Model <i>Time Lag</i>	56
Lampiran 8. Syntax R Studio Model Aditif Umum	60
Lampiran 9. Plot <i>Time Series</i> Jumlah Kasus COVID-19 Berdasarkan Peristiwa Penting	61

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sejak pertama kali teridentifikasi di Indonesia pada awal Maret 2020, *Coronavirus Disease-2019* (COVID-19) telah menyebar dan menginfeksi ke seluruh provinsi di Indonesia, termasuk Provinsi Sulawesi Selatan (Gugus Tugas Percepatan Penanganan COVID-19, 2020). Untuk mengurangi dampak negatif yang ditimbulkan akibat penyakit menular COVID-19 perlu dilakukan upaya preventif, yaitu dengan pemantauan yang akurat terkait resiko dan kejadian penyakit. Langkah tersebut sebagai alternatif pengamatan epidemiologi COVID-19 serta sebagai deteksi dini aktivitas penyakit untuk memberikan respon yang cepat dalam penanganannya (Marques-Toledo *et al.*, 2017).

Pemantauan kejadian COVID-19 dapat dilakukan dengan memanfaatkan data jejaring sosial. Preis dan Moat (2014) menyatakan saat ini perkembangan bidang ilmu sosial komputasi bertujuan untuk mengukur fenomena sosial menggunakan kumpulan data besar atau '*big data*', yang dapat diperoleh dari jejaring sosial (Marques-Toledo *et al.*, 2017). Data pada jejaring sosial tertentu mampu melacak aktivitas penyakit. Menurut Generous (2014), pemanfaatan data jejaring sosial merupakan pendekatan alternatif untuk surveilans penyakit menular (Marques-Toledo *et al.*, 2017). Beberapa studi juga menunjukkan data jejaring sosial mampu menunjukkan *trend* epidemiologi nasional.

Jejaring sosial yang dimaksud misalnya Google Trends, yaitu platform untuk mengetahui *trend* suatu kata yang dicari di dalam Google. Platform ini merupakan sumber data potensial yang menunjukkan respon (*sentiment*) masyarakat atas pandemi COVID-19. Selain itu, terdapat sumber data teragregasi dari Google (COVID 19 - Google Mobility Reports) yang memuat *big data* mobilitas manusia. Mobilitas terbukti berpengaruh terhadap tingkat penyebaran, sehingga mobilitas sering diangkat sebagai faktor penting dalam upaya pengendalian penyebaran penyakit. Seperti yang telah dilakukan di China, pola penularan COVID-19 dikaji berdasarkan data mobilitas penduduk di Wuhan (Kraemer *et al.*, 2020).

Big data yang diperoleh dari Google Trends dan COVID 19 - Google Mobility Reports dapat diintegrasikan dengan data klinis kasus positif COVID-19 di Sulawesi Selatan, dari data tersebut dapat dihasilkan model yang akurat dalam meramalkan banyaknya kejadian penyakit. Untuk membentuk suatu model yang akurat, dapat dilakukan dengan analisis regresi, yaitu suatu cara untuk memodelkan hubungan antara variabel respon y dengan variabel prediktor x . Analisis regresi bertujuan untuk mengestimasi parameter model regresi pada jumlah kasus positif COVID-19 di Sulawesi Selatan.

Jumlah kasus positif COVID-19 di Sulawesi Selatan merupakan data diskrit, dan hal ini dapat digolongkan dalam data yang berdistribusi Poisson. Model regresi yang dapat digunakan untuk menganalisis hubungan antara variabel prediktor dengan variabel respon adalah model regresi Poisson. Pemodelan regresi Poisson mensyaratkan equidispersi, yaitu kondisi dengan nilai rata-rata dan ragam variabel respon bernilai sama. Namun, pada data Jumlah kasus positif COVID-19 di Sulawesi Selatan terjadi overdispersi, yaitu ragam dari variabel respon lebih besar dari nilai rata-rata variabel respon (Ulfa *et al.*, 2021).

Salah satu penanganan overdispersi pada regresi Poisson adalah menggunakan regresi Binomial Negatif. Fitriani dan Fatikhurizqi (2020) memodelkan jumlah kasus COVID-19 dengan regresi Poisson dan regresi Binomial Negatif, hasil penelitian tersebut menunjukkan model regresi Binomial Negatif lebih baik untuk digunakan. Namun, kedua model tersebut merupakan model regresi parametrik, yang digunakan untuk data dengan asumsi linearitas, hal ini menjadikan penggunaan regresi parametrik kurang tepat digunakan, karena pada data jumlah kasus positif COVID-19 di Sulawesi Selatan terdapat pola data yang tidak beraturan, yaitu hubungan antara variabel respon dan variabel prediktor tidak mengikuti asumsi linearitas serta variabel respon tidak berdistribusi normal. Oleh karena itu, analisis yang tepat untuk digunakan adalah regresi nonparametrik yaitu model aditif umum.

Model aditif umum merupakan pengembangan dari model linear umum. Model ini mengganti fungsi linear pada model linear umum dengan fungsi aditif. Fungsi aditif pada model tidak memiliki bentuk yang kaku, sehingga model ini dapat digunakan meskipun hubungan variabel respon dan variabel prediktor tidak

linear. Distribusi variabel respon pada model aditif umum menghendaki distribusi yang termasuk dalam keluarga eksponensial. Model aditif umum banyak digunakan dalam penelitian terdahulu, penelitian tersebut mampu menghasilkan model dengan ukuran kebaikan yang cukup tinggi.

Marques-Toledo, dkk (2017) meneliti kasus demam berdarah di Brazil pada tahun 2012-2016 yang datanya mengalami overdispersi menggunakan model aditif umum dengan distribusi Binomial Negatif, penelitian ini menghasilkan model dengan koefisien determinasi sebesar 94%. Thurston, dkk (2000) meneliti kasus kanker paru-paru menggunakan model aditif umum Binomial Negatif, dan menarik kesimpulan bahwa model yang dihasilkan merupakan penanganan yang lebih baik pada data overdispersi. Nainggolan, dkk (2017) menerapkan model aditif umum pada kasus rata-rata lama sekolah di Jawa Tengah yang menghasilkan nilai koefisien determinasi sebesar 85,73%, hal ini dinilai lebih baik dalam memodelkan data dengan kenonlinearan. Oleh karena itu, pada penelitian ini jumlah kasus positif COVID-19 di Provinsi Sulawesi Selatan diestimasi dan diramalkan menggunakan model aditif umum.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang, rumusan masalah yang dikaji dalam penelitian ini, yaitu:

1. Bagaimana estimasi parameter model aditif umum dari jumlah kasus positif COVID-19 di Sulawesi Selatan pada data Google Trends dan data Google Mobility Reports?
2. Bagaimana peramalan jumlah kasus positif COVID-19 di Sulawesi Selatan dengan menggunakan model aditif umum pada data Google Trends dan data Google Mobility Reports?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini akan mengarahkan pembahasan sehingga tidak terlalu luas. Adapun batasan masalahnya dirincikan sebagai berikut:

1. Data yang digunakan adalah data sekunder jumlah kasus harian terkonfirmasi positif COVID-19 di Provinsi Sulawesi Selatan, data Google Trends dan data Google Mobility, pada interval waktu 19 Maret 2020 sampai 15 Januari 2022.

2. Metode analisis yang digunakan adalah model aditif umum dengan variabel respon berdistribusi Binomial Negatif.
3. Ukuran kebaikan model ditinjau berdasarkan nilai *Akaike's Information Criterion* (AIC) dan nilai koefisien determinasi. Untuk ukuran kebaikan peramalan ditinjau berdasarkan nilai *Root Mean Square Error of Prediction* (RMSEP) dan *Mean Absolute Percentage Error* (MAPE).

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan dari penelitian ini, yaitu:

1. Memperoleh hasil estimasi parameter model aditif umum dari jumlah kasus harian positif COVID-19 di Sulawesi Selatan pada data Google Trends dan data Google Mobility Reports.
2. Memperoleh hasil peramalan jumlah kasus positif COVID-19 di Sulawesi Selatan dengan model aditif umum pada data Google Trends dan data Google Mobility Reports.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini diuraikan sebagai berikut:

1. Penelitian ini berkontribusi memberikan penjelasan sesuai dengan ilmu statistika tentang spesifikasi data dan pengolahannya, kemudian dapat menjadi acuan dalam melakukan analisis peningkatan kasus COVID-19 oleh banyak pihak, sehingga semakin banyak daerah yang dapat menganalisis kondisinya sendiri dengan lebih mudah dan efisien.
2. Memberikan gambaran bagaimana data pada jejaring sosial dapat dimanfaatkan untuk meramalkan jumlah aktivitas penyakit menular khususnya COVID-19, sebagai suatu upaya surveilans dan deteksi dini jumlah kasus atau kejadian penyakit di masa yang akan datang.
3. Dapat menjadi acuan bagi pemerintah dan tenaga medis terkait kondisi pandemi di masa yang akan datang, sehingga mendorong percepatan kebijakan yang dapat menekan dampak negatif COVID-19, serta sebagai bentuk persiapan dalam penanganannya.
4. Memberikan gambaran kepada masyarakat luas mengenai aplikasi dari ilmu statistika dalam kehidupan nyata khususnya dalam bidang kesehatan.

BAB II

TINJAUAN PUSTAKA

2.1 Analisis Regresi

Menurut Budiantara (2000) analisis regresi merupakan metode yang digunakan untuk menentukan bentuk sebuah model atau hubungan antar variabel pada suatu kumpulan data yang menjelaskan hubungan sebab akibat atau keterkaitan antarkejadian. Dalam hal ini, analisis regresi digunakan untuk memodelkan atau mencari pola hubungan antara variabel respon dengan variabel prediktor (Sanusi *et al.*, 2019). Jika variabel respon adalah y_i dan variabel prediktor adalah x_i dengan $i = 1, 2, \dots, n$, maka model hubungan fungsionalnya dapat dituliskan sebagai

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (2.1)$$

dengan:

$f(x_i)$ = kurva regresi

ε_i = error acak yang diasumsikan identik, independen, dan berdistribusi normal dengan mean nol dan varian σ^2 (Eubank, 1998).

Silverman (1985) menyatakan tujuan utama analisis regresi ada dua, yaitu untuk memberikan cara mengeksplorasi hubungan antara variabel respon dan variabel prediktor, serta untuk tujuan peramalan (Jao, 2020).

Analisis regresi terdiri atas beberapa bentuk, yaitu pendekatan regresi parametrik, regresi nonparametrik, dan regresi semiparametrik. Untuk analisis regresi parametrik terdapat asumsi yang disyaratkan, asumsi tersebut sangat kuat dan kaku. Selain itu diperlukan pengetahuan masa lalu terkait karakteristik datanya. Tujuan utama analisis regresi yaitu mencari bentuk estimasi kurva regresi. Dalam regresi parametrik estimasi kurva regresi ekuivalen dengan estimasi terhadap parameter-parameter modelnya. Berbeda dengan regresi parametrik, dalam regresi nonparametrik bentuk kurva regresi diasumsikan tidak diketahui, dan hanya diasumsikan sebagai kurva halus (*smooth*), hal ini mengindikasikan kurva tersebut termuat dalam suatu ruang fungsi tertentu (Hidayat *et al.*, 2018).

Analisis regresi digunakan untuk menyelidiki pola hubungan fungsional antara variabel respon dan variabel prediktor. Berdasarkan pola hubungan, analisis regresi dibedakan atas regresi linear dan regresi non-linear. Analisis regresi linear memodelkan hubungan antara variabel respon dan variabel prediktor dengan pola linear, sedangkan analisis regresi non-linear adalah suatu metode untuk membentuk model yang menggambarkan hubungan antara variabel respon dan variabel prediktor yang tidak linear.

2.2 Regresi Linear

Regresi linear mempelajari terkait hubungan kausal antara dua buah variabel yang direpresentasikan dengan garis lurus. Analisis regresi juga dapat digunakan untuk menentukan hubungan kausal antara variabel X dan variabel Y sesuai dengan keadaan yang sebenarnya (Paiman, 2019).

Persamaan yang dapat digunakan untuk memperoleh garis regresi yang terdapat pada *scatter plot* disebut dengan persamaan regresi. Untuk menempatkan garis regresi yang terdapat pada data dapat dilakukan dengan menggunakan metode kuadrat terkecil. Oleh karena itu bentuk persamaan regresinya dapat dituliskan sebagai berikut.

$$Y = \alpha + \beta X, \quad (2.2)$$

dengan

Y = nilai taksiran untuk variabel respon

α = titik potong garis regresi pada sumbu Y

β = gradien garis regresi

X = nilai variabel prediktor

2.3 Korelasi *Pearson Product Moment*

Korelasi *pearson product moment* adalah metode statistika yang digunakan untuk menentukan suatu besaran yang menyatakan kuat hubungan dua variabel. Korelasi *pearson* dapat digunakan untuk data yang berskala interval atau rasio. Nilai koefisien korelasi berada di antara $-1 < r < 1$, yaitu apabila $r = -1$ korelasi negatif sempurna, dan apabila $r = 1$ korelasi positif sempurna. Jika koefisien korelasi menunjukkan angka 0, maka tidak terdapat hubungan antara

dua variabel yang dikaji. Nilai korelasi pearson dapat dihitung dengan persamaan berikut ini (Jamun *et al.*, 2020).

$$r_{xy} = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{\{n \sum X_i^2 - (\sum X_i)^2\} \{n \sum Y_i^2 - (\sum Y_i)^2\}}} \quad (2.3)$$

dengan

r_{xy} = korelasi antara x dan y

X_i = nilai variabel prediktor X ke i

Y_i = nilai variabel respon Y ke i

n = jumlah observasi

2.4 Distribusi Binomial Negatif

Distribusi Binomial Negatif adalah distribusi probabilitas diskrit yang memodelkan jumlah barisan percobaan Bernoulli yang dibutuhkan sampai terjadi k buah sukses. Probabilitas sukses setiap percobaan yaitu θ , dan probabilitas gagal yaitu $1 - \theta$. Distribusi peubah acak X yang menyatakan jumlah percobaan sampai k sukses, fungsi probabilitasnya dapat dituliskan sebagai berikut.

$$\Pr(X = x) = \begin{cases} \binom{x-1}{k-1} \theta^k (1-\theta)^{x-k}, & x \geq k, k+1, \dots \text{ dan } 0 \leq \theta \leq 1 \\ 0, & x \text{ lain} \end{cases} \quad (2.4)$$

dengan $\Pr(X = x)$ adalah probabilitas terjadi sukses ke- k pada percobaan ke- x .

Distribusi probabilitas dari peubah acak X , dapat dinotasikan menjadi bentuk lain, dengan menggunakan transformasi $Y = X - k$, dengan Y menyatakan jumlah kegagalan sebelum terjadi k buah sukses. Distribusi probabilitas dari peubah acak Y dapat dinyatakan sebagai berikut (Shafira, 2011).

$$\Pr(Y = y) = \begin{cases} \binom{y+k-1}{k-1} \theta^k (1-\theta)^y, & y = 0, 1, 2, \dots \text{ dan } 0 \leq \theta \leq 1 \\ 0, & y \text{ lain} \end{cases} \quad (2.5)$$

dengan $\Pr(Y = y)$ adalah probabilitas terjadi sukses ke- k setelah y kegagalan.

Distribusi Binomial Negatif dapat didefinisikan untuk setiap nilai positif dari k dengan menggunakan fungsi Gamma sebagai pengganti dari faktorial atau kombinasi, yaitu:

$$\Pr(Y = y) = \begin{cases} \frac{\Gamma(y+k)}{\Gamma(k)y!} \theta^k (1-\theta)^y, & y = 0, 1, 2 \dots \text{ dan } 0 \leq \theta \leq 1 \\ 0 & , y \text{ lain} \end{cases} \quad (2.6)$$

Fungsi probabilitas Binomial Negatif pada Persamaan (2.6) memiliki bentuk yang similar dengan bentuk fungsi probabilitas Binomial Negatif penurunan distribusi campuran Poisson-Gamma berikut ini (Shafira, 2011).

$$\Pr(Y = y) = \frac{\Gamma\left(y + \frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right) y!} \left(\frac{\mu}{\mu + \frac{1}{\alpha}}\right)^y \left(\frac{\frac{1}{\alpha}}{\mu + \frac{1}{\alpha}}\right)^{\frac{1}{\alpha}}, \quad (2.7)$$

dengan parameter pada Persamaan (2.6) $k = \frac{1}{\alpha}$ dan $\theta = \frac{1}{\alpha\mu + 1}$.

Regresi Binomial Negatif digunakan untuk mengatasi permasalahan dalam pemodelan regresi Poisson. Distribusi Binomial Negatif sangat umum digunakan untuk data *count* dalam membantu mengatasi masalah overdispersi (Bouk, 2016).

Jika Y_1, \dots, Y_p merupakan himpunan data *count*. Model Binomial Negatifnya didefinisikan sebagai

$$f(y; \mu, \alpha) = \frac{\Gamma\left(y + \frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right) y!} \left(\frac{\alpha\mu}{1 + \alpha\mu}\right)^y \left(\frac{1}{1 + \alpha\mu}\right)^{\frac{1}{\alpha}}, \quad (2.8)$$

dengan $\mu = E(Y)$ dan α adalah parameter dispersi. Berdasarkan model tersebut rata-rata $E(y) = \mu$, dan variansi adalah $Var(y) = \mu + \alpha\mu^2$.

Jika dituliskan dalam format keluarga eksponensial, maka bentuk *log-likelihoodnya* adalah

$$\ell(\mu, \alpha; y) = \sum_{i=1}^n \left\{ \left(\sum_{r=0}^{y_i-1} \ln(1 + \alpha r) \right) - \ln(y_i!) + y_i \ln \mu - \left(y_i + \frac{1}{\alpha} \right) \ln(1 + \alpha\mu) \right\}. \quad (2.9)$$

Dapat diketahui bahwa *link* kanoniknya adalah

$$\eta_i = \ln\left(\frac{\alpha\mu}{1 + \alpha\mu}\right).$$

Dari fungsi penghubung tersebut akan menghasilkan bentuk yang cukup rumit sehingga interpretasi dari parameter model akan menjadi lebih sulit. Hilbe (2011) menyatakan bahwa fungsi penghubung untuk model Binomial Negatif yang umum digunakan adalah logaritma atau *log link* yaitu (Wahyuni, 2011)

$$g(\mu) = \log(\mu).$$

Distribusi Binomial Negatif merupakan salah satu keluarga distribusi eksponensial. Hilbe (2011) menyatakan distribusi dari y merupakan keluarga eksponensial jika fungsi probabilitasnya berbentuk:

$$f(y; \theta; \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\}, \quad (2.10)$$

dengan θ merupakan parameter kanonik dan ϕ merupakan parameter dispersi. Fungsi peluang yang dapat ditulis sesuai Persamaan (2.10) dikatakan anggota keluarga eksponensial (Wahyuni, 2011).

Oleh karena itu,

$$\begin{aligned} f(y; \mu, \alpha) &= \frac{\Gamma\left(y + \frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right) y!} \left(\frac{\alpha\mu}{1 + \alpha\mu}\right)^y \left(\frac{1}{1 + \alpha\mu}\right)^{\frac{1}{\alpha}} \\ &= \exp\left(\ln\left(\frac{\Gamma\left(y + \frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right) y!} \left(\frac{\alpha\mu}{1 + \alpha\mu}\right)^y \left(\frac{1}{1 + \alpha\mu}\right)^{\frac{1}{\alpha}}\right)\right) \\ &= \exp\left(\ln\left(\frac{\Gamma\left(y + \frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right) y!}\right) + y \ln\left(\frac{\alpha\mu}{1 + \alpha\mu}\right) + \frac{1}{\alpha} \ln\left(\frac{1}{1 + \alpha\mu}\right)\right) \end{aligned} \quad (2.11)$$

$$= \exp \left(y \ln \left(\frac{\alpha \mu}{1 + \alpha \mu} \right) + \ln \left(\frac{\Gamma \left(y + \frac{1}{\alpha} \right)}{\Gamma \left(\frac{1}{\alpha} \right) y!} \right) + \frac{1}{\alpha} \ln \left(\frac{1}{1 + \alpha \mu} \right) \right),$$

dengan $c(y, \phi) = \ln \left(\frac{\Gamma \left(y + \frac{1}{\alpha} \right)}{\Gamma \left(\frac{1}{\alpha} \right) y!} \right)$, $\alpha(\phi) = 1$, $b(\theta) = -\frac{1}{\alpha} \ln \left(\frac{1}{1 + \alpha \mu} \right)$, dan $\theta = \ln \left(\frac{\alpha \mu}{1 + \alpha \mu} \right)$, sehingga terbukti bahwa distribusi Binomial Negatif merupakan anggota dari keluarga eksponensial.

2.5 Model Linear Umum

Model linear umum merupakan perluasan dari model regresi linear dalam mengatasi variabel respon yang tidak berdistribusi normal. Model ini memungkinkan variabel respon berdistribusi keluarga eksponensial. Secara umum model linear umum dapat dituliskan sebagai berikut.

$$g(\mu_i) = X_i \beta, \tag{2.12}$$

dengan $\mu_i = E(Y_i)$ dan Y_i merupakan keluarga distribusi eksponensial. Adapun komponen-komponen dalam model linear umum adalah sebagai berikut.

1. Komponen acak yaitu variabel respon berdistribusi keluarga eksponensial.
2. Komponen tetap yaitu variabel prediktor linear, untuk variabel respon Y_1, \dots, Y_n komponen tetapnya didefinisikan sebagai berikut.

$$\boldsymbol{\eta} = \mathbf{X}' \boldsymbol{\beta}, \tag{2.13}$$

dengan

$\boldsymbol{\eta}$ = vektor nilai variabel respon berukuran $n \times 1$

\mathbf{X} = matriks nilai variabel prediktor berukuran $n \times p$

$\boldsymbol{\beta}$ = vektor koefisien regresi berukuran $p \times 1$

3. Fungsi penghubung yaitu komponen yang menjelaskan hubungan antara μ dengan prediktor linear η , sehingga diperoleh persamaan berikut.

$$g(\mu) = \eta. \tag{2.14}$$

Fungsi penghubung untuk variabel respon yang berdistribusi normal pada model regresi linear dapat digunakan fungsi penghubung *identity*, sehingga persamaannya dapat dituliskan sebagai berikut (Valentina, 2019).

$$E(y) = \mu = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p = \beta_0 + \sum_{i=1}^p \beta_i X_i. \quad (2.15)$$

2.6 Model Aditif Umum

Model aditif umum pertama kali dikembangkan oleh Hestie dan Tibshirani pada 1986. Model aditif umum merupakan pengembangan dari model linear umum, karena model ini mengembangkan model linear umum dengan cara mengganti prediktor linear $\sum_{i=1}^p \beta_i X_i$ dengan prediktor aditif $\sum_{i=1}^p f_i(X_i)$. Model aditif umum juga dapat disebut pengembangan model aditif ke dalam bentuk distribusi eksponensial (Fitrihasari *et al.*, 2018).

Seperti pada model linear umum, model aditif umum juga terdiri atas komponen acak, komponen tetap dan fungsi *link* yang menghubungkan kedua komponen tersebut. Komponen acak didefinisikan sebagai variabel respon Y yang dianggap mempunyai distribusi eksponensial dengan bentuk fungsi probabilitas sebagai berikut.

$$f(y; \theta; \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a\phi} + c(y, \phi) \right\}, \quad (2.16)$$

dengan θ adalah parameter kanonik dan ϕ adalah parameter dispersi. Komponen tetap atau prediktor aditif pada model aditif umum adalah kuantitas η dengan bentuk berikut ini:

$$\eta = s_0 + \sum_{i=1}^p s_i(X_i). \quad (2.17)$$

Variabel respon mempunyai fungsi kepadatan peluang (*probability density*) keluarga eksponensial. Hubungan antara variabel respon μ dan η ditunjukkan dengan fungsi *link* $g(\cdot)$, yang dapat dituliskan sebagai:

$$g(\mu) = \eta. \quad (2.18)$$

Fungsi penghubung untuk model nonlinear dapat digunakan fungsi penghubung *log link*, sehingga dapat dituliskan sebagai berikut

$$E(y) = \mu = \exp\left(s_0 + \sum_{i=1}^p s_i(X_i)\right). \quad (2.19)$$

Pada model aditif umum dengan nilai harapan Y yang dapat dituliskan dalam bentuk berikut:

$$E(Y) = f(X_1, \dots, X_p) = s_0 + s_1(X_1) + \dots + s_p(X_p), \quad (2.20)$$

diketahui $s_i(X_i)$ untuk $i = 1, \dots, p$ merupakan fungsi penghalus yang diasumsikan tidak diketahui dan diestimasi dengan pendekatan nonparametrik yaitu fungsi *spline* kubik. Secara umum fungsi *spline* ber-orde m dapat dinyatakan sebagaimana persamaan berikut.

$$s(x_i) = \sum_{r=1}^{m-1} \beta_{i,r} X_i^r + \sum_{k=1}^N \beta_{i,(r+k)} (X_i - K_{ik})^{m-1}, \quad (2.21)$$

$$(X - K_k)^{m-1} = \begin{cases} (X - K_k)^{m-1}, & \text{untuk } x \geq K_n \\ 0, & \text{untuk } x < K_n \end{cases}$$

dengan $i = 1, \dots, p$ variabel prediktor, $k = 1, 2, \dots, n$ banyaknya titik knot, $r = 1, 2, \dots, m$ banyaknya orde dan K_1, K_2, \dots, K_n adalah titik-titik knot. Jika nilai $m = 4$, maka diperoleh *spline* kubik.

Berdasarkan Persamaan (2.21), maka model aditif umum dengan estimator yang didekati dengan fungsi *spline* kubik dapat dinyatakan dalam bentuk matriks sebagai berikut (Christianto, 2019).

$$\mu_i = e^{(X^{(k)}\beta)}, \quad (2.22)$$

dengan,

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}; \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \\ \beta_{m+1} \\ \vdots \\ \beta_{m+k} \end{bmatrix};$$

$$\mathbf{X}(k) = \begin{bmatrix} 1 & x_1 & \cdots & x_1^{m-1} & (x_1 - k_1)^{m-1} & \cdots & (x_1 - k_n)^{m-1} \\ 1 & x_2 & \cdots & x_2^{m-1} & (x_2 - k_1)^{m-1} & \cdots & (x_2 - k_n)^{m-1} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_n & \cdots & x_n^{m-1} & (x_n - k_1)^{m-1} & \cdots & (x_n - k_n)^{m-1} \end{bmatrix}$$

2.7 Metode *Maximum Likelihood Estimation*

Metode *maximum likelihood estimation* (MLE) adalah suatu metode yang digunakan untuk menaksir parameter suatu model yang diketahui fungsi probabilitasnya. MLE mencari titik tertentu untuk memaksimumkan sebuah fungsi. Langkah yang perlu dilakukan dengan metode MLE adalah dengan membuat fungsi *likelihood* distribusi. Kemudian membuat transformasi fungsi tersebut ke dalam bentuk *ln*. Selanjutnya hasil yang diperoleh diturunkan secara parsial terhadap parameter dan menyamakannya dengan nol. Dari pendiferensial diperoleh estimasi parameter (Zahro *et al.*, 2018).

Berikut ini didefinisikan fungsi *likelihood* sebagai:

$$L = (\mu, \alpha; y_i) = \prod_{i=1}^n f(\mu, \alpha; y_i). \tag{2.23}$$

Selanjutnya, berdasarkan Persamaan (2.23) dibuat transformasi ke dalam bentuk *ln* atau disebut dengan *log-likelihood* sebagai berikut:

$$\ln(L(\mu, \alpha; y_i)) = l(\mu, \alpha; y_i) = \sum_{i=1}^n \ln f(\mu, \alpha; y_i). \tag{2.24}$$

Kemudian distribusi keluarga eksponensial dalam hal ini distribusi Binomial Negatif dari Persamaan (2.24) disubstitusi ke dalam fungsi *log-likelihood* sebagai berikut (Wahyuni, 2011):

$$l(\mu, \alpha) = \sum_{i=1}^n \left\{ \left(\sum_{r=0}^{y_i-1} \ln(1 + \alpha r) \right) - \ln(y_i!) + y_i \ln \mu - \left(y_i + \frac{1}{\alpha} \right) \ln(1 + \alpha \mu) \right\}. \tag{2.25}$$

2.8 Evaluasi Model

Pemodelan suatu data dianggap sebagai suatu proses untuk mengganti satu set data observasi dengan satu set nilai ramalannya. Nilai peramalan pada pemodelan secara umum tidak akan tepat sama dengan nilai-nilai y nya. Nilai ketidaktepatan tersebut dapat diukur, suatu ukuran (kuantitas) untuk uji ketepatan model disebut juga dengan uji *goodness of fit*. Ketika ketidaktepatan kecil, maka hal tersebut dapat ditoleransi. Namun sebaliknya, jika ketidaktepatannya besar, maka tidak bisa ditoleransi. Berikut adalah nilai yang dapat digunakan untuk memilih model terbaik serta untuk mengevaluasi model.

2.8.1 Akaike's Information Criterion

Akaike's Information Criterion (AIC) adalah salah satu nilai yang dapat digunakan untuk memilih model terbaik dari beberapa model yang disajikan. AIC merupakan kriteria yang menyeimbangkan *goodness of fit* dari suatu model berdasarkan nilai *likelihood* dengan banyaknya parameter dari suatu model. Oleh karena itu untuk menentukan model yang terbaik dengan AIC didasarkan pada nilai AIC terkecil. Untuk memperoleh nilai AIC dapat digunakan persamaan berikut (Zahro *et al.*, 2018):

$$AIC = -2 \log L_{fit} + 2p, \quad (2.26)$$

dengan

$\log L_{fit}$ = nilai *log-likelihood* model

p = banyaknya parameter model

2.8.2 Nilai Koefisien Determinasi

Koefisien determinasi (*Adjusted R²*) merupakan proporsi variabilitas dalam suatu data yang dihitung berdasarkan pada model statistik. Koefisien determinasi (*Adjusted R²*) secara umum digunakan untuk mengukur seberapa jauh kemampuan model dalam menerangkan variasi variabel respon. Nilai dari *Adjusted R²* dapat dihitung dengan menggunakan rumus (Ernawatiningsih, 2019):

$$R_{adj}^2 = 1 - \left(\frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2} \right) \left(\frac{n-1}{n-p-1} \right), \quad (2.27)$$

Dengan

- y_i = nilai observasi respon ke-i
- \hat{y}_i = nilai peramalan respon ke-i
- \bar{y}_i = nilai rata-rata respon ke-i
- n = jumlah observasi
- p = jumlah parameter dalam model

2.9 Validasi Model

Validasi model merupakan tolak ukur seberapa besar model tersebut dapat dipercaya untuk kemudian digunakan dalam proses pengambilan keputusan. Validasi model dilakukan dengan membandingkan *output* sebagai kinerja dari model dengan nilai observasi. Berikut adalah beberapa ukuran untuk menilai validasi model.

2.9.1 Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) adalah Nilai rata-rata perbedaan absolut antara nilai peramalan dan nilai observasi yang dituliskan sebagai hasil persentasi dari nilai observasi. Nilai MAPE dapat dihitung menggunakan persamaan berikut (Maricar, 2019).

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^T \frac{|y_t - \hat{y}_t|}{y_t} \times 100\%, \quad (2.28)$$

dengan

- y_t = nilai observasi
- \hat{y}_t = nilai peramalan
- T = jumlah data

Semakin rendah nilai MAPE, kemampuan dari model peramalan yang digunakan dapat dikatakan baik. Pada MAPE terdapat *range* nilai yang dapat dijadikan sebagai tolak ukur kemampuan peramalan, *range* tersebut disajikan pada tabel berikut.

Tabel 2. 1 Range Nilai MAPE

<i>Range MAPE</i>	Kemampuan Peramalan Model
< 10%	Sangat Baik
10% – 20%	Baik
21% – 50%	Layak
> 50%	Buruk

2.9.2 *Root Mean Square Error of Prediction*

Salah satu cara yang dapat digunakan untuk melakukan validasi model yaitu menurut Brockwell & Davis (2002), dengan menghitung nilai *Root Mean Square Error of Prediction* (RMSEP). Formula yang digunakan, yaitu:

$$\text{RMSEP} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (2.29)$$

dengan

y = nilai observasi

\hat{y} = nilai peramalan

n = jumlah data

Semakin kecil nilai *Root Mean Square Error of Prediction* (RMSEP), maka semakin kecil perbedaan antara nilai peramalan dengan nilai observasi. Hal ini menunjukkan bahwa model yang dibentuk menghasilkan nilai peramalan yang semakin akurat (Handayani *et al.*, 2019).

2.10 Kasus Positif COVID-19

Pandemi COVID-19 yang berasal dari Wuhan, Tiongkok setiap hari masih menunjukkan adanya aktivitas dari penyakit ini. Penyebaran COVID-19 sangat cepat, saat ini berdasarkan data yang dihimpun Gugus Tugas Percepatan Penanganan COVID-19, sampai pada tanggal 14 Februari 2022 jumlah kasus terkonfirmasi positif COVID-19 Sulawesi Selatan adalah 114,146 (2.3% dari jumlah terkonfirmasi nasional). Seseorang dinyatakan terkonfirmasi positif COVID-19, baik memiliki gejala atau tidak bergejala, dibuktikan dengan pemeriksaan laboratorium RT-PCR/RDT-Antigen (Kemenkes RI, 2021).

2.11 Google Trends

Google Trends merupakan situs web yang dimiliki Google.Inc yang berisikan *trend* penggunaan kata kunci di website mesin pencari Google dan berita yang sedang *trend*. Google Trends dapat dimanfaatkan untuk *research* (riset). Mesin pencari Google merekam setiap kata yang diinput oleh pengguna Google sejak tahun 2004. Kata kunci dapat dicari oleh pengguna dengan 5 kata sekaligus. Google Trends menyajikan grafik statistik popularitas pencarian kata dengan data persentase pada kurun waktu tertentu. Data dapat ditampilkan berdasarkan wilayah baik provinsi maupun kota, serta dalam berbagai bahasa (Riyanto, 2014).

2.12 Google Mobility

Set data mobilitas masyarakat diperoleh dari COVID 19 – Google Mobility Reports yang disusun oleh Google. Data ini merupakan rekaman pergerakan manusia di setiap wilayah geografis, yang diperoleh dari jejak pergerakan telepon seluler berbasis Android. Data pada COVID 19 - Google Mobility Reports dikelompokkan secara harian. Data perhari menunjukkan *trend* naik atau *trend* turun mobilitas di area-area yang telah ditentukan yaitu, pertokoan dan tempat rekreasi, tempat perbelanjaan dan apotek, taman, tempat transit transportasi, lingkungan perumahan, dan area perkantoran. Tempat-tempat ini dipilih mewakili tempat-tempat yang biasanya menjadi lokasi terjadinya interaksi sosial, yang kemudian dibandingkan dengan *baseline* tertentu. Google merekam data mobilitas ini sejak tanggal 2 Februari 2020 (Nugroho & Rakhman, 2021).