

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Demam Tifoid (Tifoid Demam)

Demam tifoid dengan intensitas tinggi dapat berakibat fatal yang dapat menyebabkan kematian, dengan penyebaran demam tifoid tertinggi terdapat di wilayah Asia dan Afrika (Marchello, 2019). Maraknya penyebaran demam tifoid menyebabkan manusia resistensi terhadap vaksin dan antibiotik untuk penanganan Demam Tifoid (Appiah, 2022). Resistensi antibiotik dapat membuat pengelolaan demam tifoid semakin rumit (Masuet-Aumatell, 2020). Resistensi terhadap antibiotik membuat tubuh pengidap demam tifoid menjadi kebal terhadap virus baik untuk melawan penyakit, sehingga dibutuhkan pencegahan lebih awal agar tidak berdampak fatal bagi pengidap demam tifoid.

Demam tifoid akan menjadi fatal karena memicu komplikasi penyakit lain seperti pendarahan organ dalam, perforasi usus dan banyak penyakit dalam lainnya (Rasheed., 2019). Demam tifoid juga menjadi gejala awal dari penyebab kematian dengan kecacatan, dimana dengan penyebaran yang tidak terdeteksi akan mengakibatkan banyak manusia yang terpapar patogen tersebut (Jeffrey D, 2019). Bahaya tersebut yang menjadikan demam tifoid sebagai salah satu penyakit yang memiliki musim tertentu untuk penyebarannya dan haru di tekan angka penyebarannya. Pencegahan awal berupa deteksi prediksi kasus demam tifoid merupakan salah satu langkah maju untuk mencegah penyebarannya.

2.1.2 Suhu Udara

Suhu adalah besaran yang menggambarkan keadaan termal lingkungan yang berkaitan dengan perubahan entropi sistem (Sedrakyan, 2023). Suhu dapat diukur berdasarkan distribusi kecepatan antar molekulnya (Salthammer, 2022). Suhu dapat menjadi variabel penting dalam dunia Kesehatan, dimana tingkat suhu dapat mempengaruhi ekosistem yang terdapat dalam tubuh manusia maupun lingkungan sekitar (Mohammadi, 2020). Suhu lingkungan yang tinggi akan dengan mudah terdistribusikan ke dalam tubuh manusia, hal ini menyebabkan meningkatnya resiko tubuh manusia terserang penyakit (Cheng, 2019). Perpindahan suhu yang sangat cepat juga dapat mengakibatkan mudahnya manusia terserang penyakit seperti demam tifoid

yang dapat berakibat fatal pada manusia.

2.1.3 Pemodelan Matematika

Pemodelan matematika adalah gambaran dari penerjemahan situasi dalam siklus kehidupan nyata ke dalam formulasi matematika dengan tujuan menggambarkan pola tertentu atau prediksi dalam beberapa waktu kedepan (Hwang, 2023). Pemodelan matematika melibatkan pengembangan dari persamaan matematis, baik yang sederhana maupun persamaan yang kompleks sehingga dapat diinterpretasikan system yang sedang terjadi. Pemodelan matematika biasanya juga digunakan untuk visualisasi, simulasi, dan pengujian hipotesis terhadap data yang digunakan. Konsep pemodelan matematika berkembang seiring dengan kemajuan komputasi digital, yang memungkinkan menganalisis data yang lebih kompleks dan presisi yang lebih tinggi (Tedeschi, 2019).

Pemodelan matematika diterapkan di berbagai bidang baik di Teknik, Pendidikan, Kesehatan, dan biomedis. Bidang Pendidikan salah satu contoh dimana konsep pemodelan matematika digunakan untuk menghubungkan situasi abstrak dalam kehidupan nyatanya dan pemodelan juga sering digunakan sebagai nilai evaluasi suatu penelitian (Lyon, 2020). Selain itu, dalam bidang biomedis pemodelan digunakan untuk memberikan prediksi atau sebuah mekanisme kesehatan dan perkembangan penyebaran suatu penyakit (Foy, 2020) (Stefanovski, 2021). Pemodelan matematika juga memberikan manfaat besar dalam menganalisis data melalui model model yang dibangun, hal ini dapat memberikan dampak kepada kemajuan ilmiah dari berbagai disiplin ilmu. Pengembangan model matematika terbaru menjadi pemodelan matematika menjadi alat yang penting untuk mendukung inovasi dan penelitian ilmiah (Abassian, 2020).

2.1.4 Runtun Waktu

Pemodelan runtun waktu adalah data runtut waktu yang dapat digunakan untuk meramalkan nilai masa depan berdasarkan pola dan historis data, dengan mempelajari tren dari berbagai interval waktu (Shen, 2020). Pemodelan runtun waktu juga dapat berarti model prediksi deret waktu yang didasarkan pada model linier sehingga dapat melihat analisis matematika di waktu kemudian (George, 2019). Pemodelan runtun waktu biasanya berbentuk model prediksi seperti autoregressive integrated with moving average (ARIMA), exponential smoothing error, trend, seasonal model, exponential growth model, grey model serta masih banyak model runtun waktu yang digunakan

untuk memodelkan prediksi dari suatu data. Data runtun waktu memiliki keunikan berupa pola data yang dapat berubah seiring dengan banyak nya data yang didapat, trend data yang dapat berubah ubah, serta bentuk data yang fluktuatif sehingga banyak digunakan untuk pemodelan prediksi

2.1.5 *Autoregressive Integrated with Moving Average (ARIMA)*

ARIMA (Autoregressive Integrated Moving Average) adalah model deret waktu yang digunakan untuk memprediksi tren masa depan dengan mengintegrasikan nilai masa lalu dari suatu variabel dan menggabungkan komponen autoregresif, differencing, dan rata-rata bergerak (Dong, 2020). Model ini menjadi populer karena kemampuannya menangkap pola linier dalam data deret waktu, baik untuk prediksi jangka pendek maupun menengah (Lai, 2020). ARIMA digunakan dengan mengidentifikasi pola historis dari data, menyesuaikan parameter model, dan kemudian memproyeksikan tren masa depan berdasarkan pola tersebut. Dengan demikian, ARIMA telah banyak diterapkan dalam berbagai bidang, termasuk keuangan, ekonomi, lingkungan, dan pendidikan untuk membuat prakiraan yang lebih akurat terkait pergerakan harga, tingkat pengembalian, suhu, dan bahkan pendaftaran mahasiswa (Qin, 2019). Persamaan model ARIMA seperti berikut

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (1)$$

y_t : Nilai Variabel target pada waktu t .

$\phi_1 \dots \phi_p$: Koefisien Autoregressive (AR)

$\theta_1 \dots \theta_p$: Koefisien Moving Average (MA)

ϵ_t : Error pada waktu t .

p, d, q : Merupakan orde dari komponen AR, Differencing, dan MA.

Persamaan model ARIMA pada persamaan satu memiliki beragam manfaat salah satunya di sektor keuangan, ARIMA digunakan untuk memprediksi volatilitas indeks saham dan pengembalian aset, serta memberikan hasil yang akurat dalam peramalan jangka pendek di lingkungan yang kompleks, seperti pergerakan pasar saham dan tingkat volatilitas rendah (Hou(2023);Dong(2020)). ARIMA dalam bidang lingkungan membantu memperkirakan curah hujan, suhu, dan pola iklim, memberikan panduan penting bagi perencanaan sumber daya dan mitigasi bencana alam di wilayah rawan banjir (Bora(2023);Lai(2020)). Selain itu, model ini juga digunakan untuk

memproyeksikan indikator ekonomi makro, seperti GDP, dan memberikan pandangan jangka panjang tentang pertumbuhan ekonomi suatu negara, seperti yang dilakukan pada data GDP Mesir (Abonazel, 2019). ARIMA mampu menawarkan keandalan prediksi yang lebih baik dibandingkan metode linier tradisional, terutama dalam menangani data dengan tren dan pola yang berubah-ubah

2.1.6 Autoregressive Integrated with Moving Average with Exogenous Variabel (ARIMAX)

ARIMAX (Autoregressive Integrated Moving Average with Exogenous Variables) adalah pengembangan dari model ARIMA. ARIMAX memungkinkan penambahan variabel eksogen yang dapat mempengaruhi prediksi (Chen, 2022). Penambahan variabel memungkinkan model mempelajari hubungan tambahan dari data regressor terhadap data utama (Maku, 2023). (Rizalde, 2022) Menyebutkan bahwa model ARIMAX lebih baik dalam memprediksi model dibandingkan ARIMA untuk prediksi kasus penyebaran penyakit menular di suatu daerah. Hasil tersebut disebabkan model ARIMAX lebih banyak menangkap informasi yang mungkin mempengaruhi data sehingga lebih baik dalam prediksi kasus terutama penyakit menular (Obi, 2021). Penambahan variabel regressor membuat persamaan ARIMAX berbeda dengan model ARIMA biasanya, seperti berikut :

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \beta x_t + \epsilon_t \quad (2)$$

x_t : Variabel eksogen pada waktu t .

β : Koefisien untuk variabel eksogen.

Persamaan dua menunjukkan penambahan variabel eksogen dalam persamaan model ARIMA menjadi model ARIMAX diharapkan menjadi model yang lebih optimal, sehingga dapat melihat variabel eksogen dalam memprediksi apakah berpengaruh signifikan terhadap hasil prediksi.

2.1.7 Seasonal Autoregressive Integrated with Moving Average (SARIMA)

SARIMA (Seasonal Autoregressive Integrated Moving Average) adalah model yang dikembangkan dari ARIMA dengan menambahkan komponen seasonal, yang digunakan untuk menjelaskan pola yang berulang dari siklus data pada interval waktu tertentu (Prendin, 2022). Seasonal dalam model SARIMA ditambahkan untuk membantu menangkap pola musiman pada data yang tidak bisa dijelaskan secara optimal oleh model ARIMA, sehingga menghasilkan prediksi yang lebih akurat jika data

memiliki seasonalitas (Wang, 2021). Model ARIMA hanya digunakan historis data saja tanpa melihat lebih lanjut pola musiman dari data. SARIMA lebih unggul dikarenakan dapat menangkap pola musiman dengan lebih baik. Persamaan model SARIMA seperti berikut.

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \sum_{k=1}^P \Phi_k y_{t-k-s} + \sum_{l=1}^Q \Theta_l \epsilon_{t-l-s} + \epsilon_t \quad (3)$$

B : Jumlah lag yang signifikan.

S : Panjang periode musiman.

Φ_k : Koefisien AR musiman.

ϕ_i : Koefisien AR non musiman.

Θ_l : Koefisien MA musiman.

θ_j : Koefisien MA non musiman.

Persamaan tiga dapat dilihat pada notasi SARIMA pada umumnya dituliskan sebagai SARIMA(p,d,q)(P,D,Q,s) seperti yang sudah dijelaskan dalam persamaan diatas. Penambahan variabel musiman membuat SARIMA lebih efektif dalam memprediksi model yang memiliki pola musiman. (Hasri, 2023) dalam penelitiannya ditunjukkan bahwa model SARIMA lebih unggul dibandingkan model prediksi lainnya untuk kasus penyakit menular. Penyakit menular memberikan pola yang berulang sehingga memungkinkan model SARIMA menjadi model prediksi paling efektif karena terpengaruh oleh pola dan siklus (Ariyanti, 2023).

2.1.8 Seasonal Autoregressive Integrated with Moving Average with Exogenous Variable (SARIMAX)

SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Factors) adalah model prediksi runtun waktu yang merupakan pengembangan dari ARIMA dengan menambahkan variabel seasonal dan variabel regresor (Farooq, 2022). SARIMAX cocok untuk digunakan untuk data yang memiliki pola musiman serta dipengaruhi oleh variabel eksternal yang terikat. Perbedaan antara SARIMAX dengan ARIMA, ARIMAX, dan SARIMA adalah model ini cocok digunakan untuk data runtun waktu dengan pola musiman yang dipengaruhi oleh variabel eksternal, sehingga model ini lebih efektif dibandingkan dengan model yang

lain untuk kriteria data tersebut (Amelia, 2022). SARIMAX juga menambahkan variabel exogenous dalam persamaannya. Berikut merupakan persamaan dari SARIMAX.

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \sum_{k=1}^P \Phi_k y_{t-k \cdot s} + \sum_{l=1}^Q \Theta_l \epsilon_{t-l \cdot s} + \sum_{m=1}^M \beta_m X_{t,m} + \epsilon_t \quad (4)$$

$\beta_m X_{t,m}$: Pengaruh variabel eksogen pada waktu t dengan koefisien β .

Persamaan empat menunjukkan penambahan variabel eksogen dalam model SARIMA seperti penelitian (Amelia, 2022) yang melakukan peramalan curah hujan dalam suatu wilayah menggunakan SARIMAX. Curah hujan mempertimbangkan faktor musiman dan variabel regresor seperti kecepatan angin serta pembentukan awan konvektif. Penelitian ini menunjukkan bahwa model SARIMAX adalah yang paling efektif dengan memprediksi hujan menggunakan variabel tambahan berupa kecepatan angin yang mempengaruhi curah hujan.

2.1.9 Generalized Space-Time Auto-regressive Integrated Moving Average (GSTARIMA)

GSTARIMA (Generalized Space-Time Autoregressive Integrated Moving Average) adalah model deret waktu pengembangan dari ARIMA, yang digunakan untuk jenis data yang memiliki ketergantungan spasial dan temporal (Saheed., 2024). GSTARIMA cocok untuk data yang dipengaruhi oleh pola musiman dan memiliki korelasi antara lokasi, sehingga variabel dari lokasi sekitar juga dianggap mempengaruhi nilai pada waktu sebelumnya (Rafiqah., 2019). Pengembangan model GSTARIMA dapat dilihat dari persamaan berikut.

$$y_{i,t} = \sum_{j=1}^N w_{ij} \sum_{p=1}^P \phi_p y_{j,t-p} + \sum_{q=1}^P \theta_q \epsilon_{i,t-q} + \epsilon_{i,t} \quad (5)$$

$y_{i,t}$: Nilai di lokasi i pada waktu t .

$w_{i,t}$: Bobot efek spasial antara lokasi i dan j .

N : Jumlah lokasi.

ϕ_p : Koefisien AR *spatio - temporal*.

$\theta_q(B)$: Koefisien MA *spatio - temporal*.

Persamaan

lima menunjukkan dengan menambahkan waktu dan lokasi, GSTARIMA memiliki fleksibilitas yang besar untuk dapat menganalisis data yang memiliki efek spasial. Memungkinkan untuk GSTARIMA lebih akurat untuk memprediksi data spatio-temporal (Ilmi, 2023).

2.1.10 Generalized Space-time Autoregressive Integrated Moving Average with Exogenous Variables (GSTARIMAX)

GSTARIMAX (Generalized Space-time Autoregressive Integrated Moving Average with Exogenous Variables) adalah model pengembangan ARIMA, dimana implementasinya model ini cocok digunakan untuk data yang memiliki ketergantungan spasial, temporal, serta mempertimbangkan variabel eksogen (Sabrina, 2022). Pengembangan model GSTARIMAX dapat dilihat dalam persamaan berikut.

$$y_{i,t} = \sum_{j=1}^N w_{ij} \sum_{p=1}^P \phi_p y_{j,t-p} + \sum_{q=1}^P \theta_q \epsilon_{i,t-q} + \beta x_{i,t} + \epsilon_{i,t} \quad (6)$$

$x_{i,t}$: Variabel eksogen di lokasi i pada waktu t .

β : Koefisien untuk variabel *exogenous* $x_{i,t}$.

Persamaan enam menunjukkan dengan menggabungkan komponen spasial, temporal, dan musiman serta variabel eksogen, membuat GSTARIMAX menjadi model paling fleksibel untuk memprediksi fenomena yang dipengaruhi variabel tersebut (Primageza, 2021). GSTARIMAX biasanya digunakan untuk memprediksi kasus eksternal yang berulang atau temporer seperti penyakit yang dipengaruhi oleh musim. klasifikasi.

2.1.11 Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) memberikan bobot lebih tinggi pada kesalahan yang lebih besar karena setiap suku kesalahan dikuadratkan sebelum di rata-ratakan (Anichebe, 2024). RMSE bersifat serbaguna sehingga dapat diterapkan dalam berbagai domain jenis data, biasanya digunakan untuk evaluasi model runtun waktu (Filippidis, 2024). Persamaan RMSE sebagai berikut:

$$RMSE = \sqrt{\sum \frac{(Y' - Y)^2}{n}} \quad (7)$$

Y' = Nilai Prediksi

Y = Nilai Asli

n = Banyak Data

Fungsi dari persamaan tujuh diatas yang membuat RMSE sangat berguna Ketika kesalahan besar tidak diinginkan (Wardana, 2024). Faktor tersebut

yang membuat RMSE lebih baik digunakan dibandingkan metrik error lainnya (Sharma, 2024).

2.1.12 *Corrected Akaike Information Criterion (AICC)*

Corrected Akaike Information Criterion (AICC) merupakan nilai uji evaluasi yang digunakan untuk pemilihan model dalam analisis statistic. Nilai koreksi AICC membantu mencegah overfitting pada model (Thottoli, 2024). Pemilihan AICC bertujuan untuk mengurangi bias yang dapat terjadi didalam AIC, sehingga AICC dapat menawarkan pendekatan yang lebih seimbang diantara AIC dan BIC (Houdou, 2024). Persamaan nilai uji AICC sebagai berikut:

$$AICC = -2\left(-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(2\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^n \varepsilon_t^2 + 2k + \frac{2k(k+1)}{n-k-1}\right) \quad (8)$$

k = Jumlah Parameter Pada Model

σ^2 = Variansi Residual

ε_t = Residual

n = Jumlah Observasi

k

menunjukkan jumlah parameter dalam model, dimana jumlah parameter ini yang menjadi salah satu alasan penggunaan AICC dalam penelitian ini. AICC cocok untuk menguji model dengan data yang bersifat bulanan. Variansi residual dan residual data serta jumlah observasi digunakan dalam nilai uji AICC. Model dengan nilai AICC terkecil menunjukkan kecocokan antara data yang digunakan dengan model yang dipakai.

2.2 Hasil Penelitian Terdahulu

Beberapa penelitian sebelumnya tentang prediksi kasus menggunakan ARIMA, ARIMAX, SARIMA, SARIMAX, GASTARIMA, dan GSTARIMAX yang menjadi acuan dalam skripsi ini disajikan dalam **Tabel 2.1** berikut :

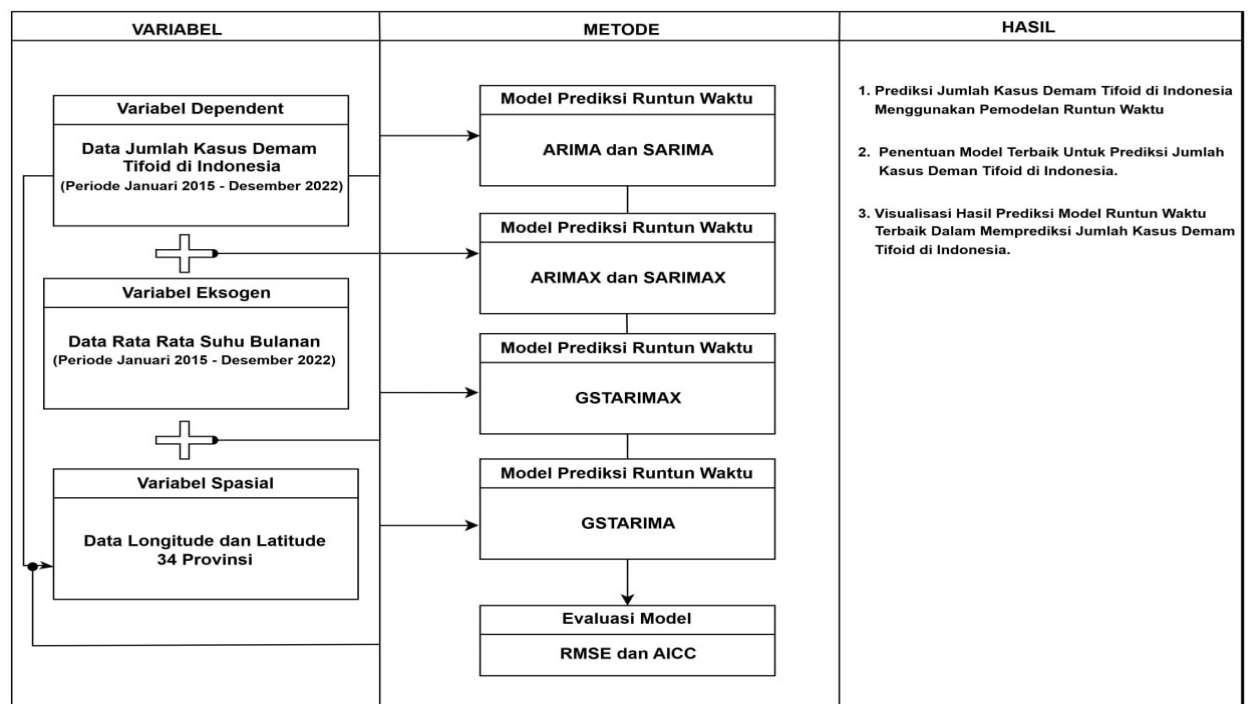
Tabel 2. 1 Hasil Penelitian Terdahulu.

No.	Judul	Hasil	Keterangan/Kesamaan
1	Application of ARIMA Model in Infectious Disease Prediction (Zhangying, 2023)	Model ARIMA menjadi model paling efektif dalam memprediksi kasus penyakit menular di Tiongkok. Lonjakan kasus tertinggi terjadi pada Maret–Juli dan November–Januari 2022. ARIMA cocok untuk memprediksi kasus penyakit menular.	Penggunaan algoritma model prediksi ARIMA untuk prediksi jumlah kasus penyakit menular.
2	Generalized Space-Time Autoregressive Integrated Moving Average (GSTARIMA) for Modelling Rainfall Data in Makassar Region (Ilmi, 2023)	Model GSTARIMA terbaik untuk memprediksi jumlah curah hujan di Makassar (2017–2021). Menggunakan pembobotan kebalikan jarak, nilai RMSE sebesar 132,9. Hasil peramalan sesuai untuk acuan kebijakan menangani curah hujan tinggi.	Penggunaan algoritma model prediksi GSTARIMA dengan titik longitude dan latitude untuk prediksi curah hujan.
3	Early Warning and Prediction of Scarlet Fever in China Using the Baidu Search Index and ARIMAX Model: Time Series Analysis (Tingyan, 2023)	Model ARIMAX mengungguli ARIMA untuk memprediksi kasus demam berdarah di Tiongkok. Nilai error ARIMAX = 1692,1, lebih kecil dibandingkan ARIMA = 2036,92.	Perbandingan antara model ARIMAX dengan ARIMA dalam memprediksi jumlah kasus dengan karakteristik yang sama.
4	Forecasting Natural Gas Production and Consumption in the United States – Evidence from SARIMA and SARIMAX Models (Manigandan, 2021)	Model SARIMAX mengungguli SARIMA dalam prediksi konsumsi gas alam (LNG) di AS. RMSE dan MAPE SARIMAX lebih kecil, konsumsi LNG meningkat hingga 16% pada 2025. Data 1983–2021 digunakan untuk prediksi 2022–2025.	Perbandingan antara model SARIMAX dan SARIMA dalam memprediksi konsumsi LNG.
5	Forecasting the Dynamics of Cumulative COVID-19 Cases (Confirmed, Recovered, and Deaths) for Top-16 Countries Using Statistical Machine Learning Models: ARIMA and SARIMA (Kumar, 2021)	Model SARIMA lebih akurat dibandingkan ARIMA dalam memprediksi kasus dan angka kematian COVID-19 di 16 negara teratas. Hasil membantu perencanaan kebijakan kesehatan menghadapi pandemi.	Perbandingan model SARIMA dengan ARIMA untuk kasus dengan karakteristik penyebaran yang sama.

No.	Judul	Hasil	Keterangan/Kesamaan
6	Pengembangan Model GSTARIMAX dengan Pendekatan SUR untuk Peramalan Data Indeks Harga Konsumen Kelompok Bahan Makanan di Pulau Jawa (Litya, 2019)	Model GSTARIMAX terbaik untuk memprediksi IHK Kelompok Bahan Makanan di Pulau Jawa (Jan 2019–Sep 2019). Menggunakan data IHK dan lokasi kota/pulau, RMSE = 1,00%. Cocok untuk prediksi 1 bulan ke depan.	Penggunaan algoritma model prediksi GSTARIMAX untuk jumlah kasus yang berbeda.

Sumber: diolah oleh peneliti

2.3 Kerangka Pemikiran



Gambar 2. 1 Kerangka Pemikiran

Sumber : diolah oleh peneliti

Gambar 2.1 menjelaskan bahwa penelitian ini menggunakan model runtun waktu Autoregressive Integrated Moving Average (ARIMA) yang dikembangkan menjadi beberapa model turunan dengan memperhatikan seasonal, exogenous variables, dan spatio-temporal effect. Model yang dibangun dalam penelitian ini meliputi ARIMA, ARIMAX, SARIMA, SARIMAX, GSATARIMA, GSTARIMAX dimana dalam membangun keenam model digunakan parameter Autocorrelation dan Partial Autocorrelation untuk menentukan parameter AR dan MA dalam model. Pembangunan model juga dilihat stasioneritas data untuk mengetahui apakah terdapat differencing yang dilakukan dari data awal agar data dapat digunakan untuk dimasukkan kedalam model. Data yang sesuai serta parameter yang telah dianalisis dapat menciptakan model ARIMA, ARIMAX, SARIMA, SARIMAX, GSATARIMA, GSTARIMAX yang optimal untuk melakukan prediksi. Banyaknya pengembangan model ARIMA dengan ditambahkannya berbagai variabel seperti seasonality data, exogenous variables, dan spatio-temporal effect sehingga dapat menghasilkan prediksi data yang akurat serta model terbaik untuk memprediksi kasus pada penelitian ini.

Model prediksi nantinya dievaluasi berdasarkan nilai error terkecil serta memiliki nilai uji kriteria terkecil. Penambahan variabel exogenous berupa data suhu dari masing masing provinsi dalam urutan rata rata suhu per bulan. Kemudian untuk GSTARIMA dan GSTARIMAX akan ditambahkan variabel longitude dan latitude untuk penambahan efek spatio- temporal. Model dengan nilai RMSE terkecil dan nilai AICC terkecil merupakan model terbaik untuk memprediksi kasus demam tifoid.