BAB II

LANDASAN TEORI

A. Deskripsi Teori

1. Peramalan (*Forecasting*)

Peramalan atau forecasting adalah proses memprediksi masa depan seakurat mungkin dengan memanfaatkan semua informasi yang tersedia, termasuk data historis dan pengetahuan tentang kejadian-kejadian masa depan yang mungkin mempengaruhi prakiraan tersebut. Pemilihan metode peramalan yang tepat sangat tergantung pada data yang tersedia. Langkah-langkah dalam melakukan peramalan (Athanasopoulos & J Hyndman, 2018) antara lain:

1) Mendefinisikan masalah
2) Mengumpulkan informasi
3) Analisis pendahuluan (eksplorasi)
4) Memilih dan memasang model
5) Menggunakan dan mengevaluasi model peramalan

Ahli lain juga menyatakan bahwa peramalan atau *forecasting* merupakan kegiatan memprediksi atau memperkirakan apa yang terjadi di masa yang akan datang berdasarkan data historis pada masa lalu menggunakan model matematika. Dalam melakukan peramalan, hasil yang didapatkan tidak selalu tepat sesuai dengan kenyataan. Salah satu cara yang dapat digunakan dalam menentukan ukuran kesalahan secara statistik yaitu dengan *Mean Absolute Percentage Error* (MAPE). Untuk mengevaluasi hasil prediksi, digunakan *Mean Absolute Percentage Error* (MAPE), yang merupakan rata-rata dari persentase kesalahan antara data aktual dan hasil prediksi. MAPE digunakan untuk menilai proses prediksi, mengukur akurasi dengan mencocokkannya dengan data runtun waktu, dan hasilnya dinyatakan dalam persentase. Semakin kecil nilai MAPE dari hasil prediksi, semakin mendekati data aslinya. Secara matematis, MAPE ditunjukkan dalam persamaan (Datumaya Wahyudi Sumari et al., 2020):

$$MAPE = \left(\frac{1}{n}\right)\sum_{t=1}^{n}\left|\frac{X_t - F_t}{X_t}\right| \quad (6)$$

Dimana:

X_t = Data aktual pada periode (t) tertentu

F_t = Nilai prediksi pada periode (t) tertentu

n = jumlah data uji

2. *Time series*

Data berkala atau *time series* adalah data yang terkumpul dari waktu ke waktu untuk menggambarkan kecenderungan atau tren suatu kondisi. Umumnya, interval waktu antara pengumpulan data secara konsisten sama. Contoh data *time series* (Wayan, 2002) antara lain:

1) Pertumbuhan ekonomi suatu negara pertahun
2) Jumlah produksi minyak perbulan
3) Indeks harga saham perhari

Analisis *time series* bertujuan utama untuk meramalkan masa depan, yang melibatkan pembuatan model untuk merepresentasikan data deret waktu. Model ini, dalam bentuk persamaan matematika atau proses, menghasilkan data deret waktu buatan. Langkah-langkah dasarnya dalam peramalan mencakup (Athanasopoulos & J Hyndman, 2018):

1) Memilih model deret waktu yang sesuai
2) Mengestimasi model tertentu yang cocok dengan pola siklus data
3) Membuat perkiraan nilai yang diharapkan di masa depan berdasarkan perilaku model
4) Menetapkan batas kepercayaan prediksi untuk memperkirakan kisaran di mana nilai observasi di masa depan kemungkinan besar berada, asumsinya jika modelnya akurat

Dalam *time series* terdapat beberapa komponen, antara lain (Athanasopoulos & J Hyndman, 2018):

1) Trend adalah pola jangka panjang dalam data yang menunjukkan peningkatan atau penurunan secara terus-menerus.
2) Pola musiman terjadi ketika pola data dipengaruhi oleh faktor-faktor musiman, seperti perubahan dalam setahun atau dalam satu periode waktu tertentu, seperti hari dalam seminggu.
3) Siklik terjadi ketika data menunjukkan fluktuasi naik turun yang tidak teratur. Jika fluktuasi tersebut tidak memiliki frekuensi yang tetap, maka pola tersebut dianggap siklik. Namun, jika fluktuasi tersebut memiliki frekuensi tetap dan terkait dengan aspek kalender tertentu, maka pola tersebut dianggap musiman. Secara umum,

siklus memiliki rentang waktu yang lebih panjang daripada pola musiman, dan fluktuasinya cenderung lebih variatif.

3. Metode *Autoregressive Integrated Moving Average*

3.1 Model Autoregressive (AR)

Model *Autoregressive* adalah sebuah model regresi yang memanfaatkan deret waktu asli untuk menghasilkan deret lag. Dalam regresi linear berganda, outputnya merupakan hasil kombinasi linear dari beberapa variabel masukan. Namun, dalam model autoregressive, outputnya adalah titik masa depan yang dapat dijelaskan sebagai kombinasi linear dari titik data pada masa lalu. Jumlah masa lalu yang digunakan untuk memprediksi masa depan ditentukan oleh jendela lag yang disebut p. Model *autoregressive* dapat dilambangkan dengan persamaan:

$$Y_t = l + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + e \qquad (7)$$

Dimana l merupakan level dalam dataset, dan e adalah koefisien yang harus diestimasi dari data tersebut. Ini dikenal sebagai model *autoregressive* dengan lag p atau model AR(p). Dalam model AR(p), deret lag digunakan sebagai prediktor tambahan untuk memperhitungkan nilai variabel dependen yang masih berhubungan dengan nilai deret waktu aslinya, y_t (Kotu & Deshpande, 2019).

Ahli lain juga menjelaskan bahwa *autoregressive* merupakan bentuk regresi yang digunakan untuk mengukur tingkat keeratan atau *association* antara X_t dengan X_{t-k}, jika pengaruh dari time lag 1,2,3…, dan seterusnya sampai $k-1$ dianggap terpisah. Pada model ini menunjukkan nilai X_t sebagai fungsi linier dari sejumlah X_t aktual sebelumnya. Bentuk umum peramalan model *autoregressive* yaitu:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \cdots + \phi_p X_{t-p} + e_t \qquad (8)$$

Dimana:

$X_{t-1}, X_{t-2}, X_{t-3} \ldots X_{t-p}$ = variabel yang menentukan

X_t = variabel yang diramalkan

ϕ_p = parameter AR orde ke-p

e_t = galat

Dari persamaan (4) masih terdapat beberapa bentuk persamaan yang lain, tergantung pada derajat susunan (orde) dari p, sehingga dapat ditulis bahwa AR (p). Jika p = 1, maka AR (1), dan seterusnya (Makridakis & Wheelwright, 1998).

3.2 Model Moving Average (MA)

Model *moving average* menggambarkan nilai saat ini sebagai jumlah berbobot dari gangguan kecil atau *white noise* dari waktu sebelumnya atau tergantung pada nilai-nilai kesalahan masa lalu. Bentuk umum model *moving average* adalah:

$$X_t = e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q} \qquad (9)$$

Dimana:

e_t = kesalahan pada saat t

$e - e_{t-1} - \cdots - e_{t-q}$ = nilai-nilai kesalahan terdahulu

θ_q= parameter MA dari orde ke-q

Deret waktu dikatakan stasioner jika dapat direpresentasikan sebagai model *moving average* yang konvergen, yaitu memiliki orde yang terbatas. Oleh karena itu, diperlukan pembatasan terhadap parameter $\theta_2(1{,}2{,}3 \ldots, q)$ seperti pada model *autoregressive*. Jika $\theta(B)$ dinyatakan sebagai fungsi dalam B, maka persamaan karakteristiknya yaitu:

$$\theta(B) = 0 \text{ atau } 1 - \theta_1 B = \theta_2 B^2 - \cdots - \theta_q B^q = 0 \qquad (10)$$

Model MA (q) dikatakan konvergen apabila nilai mutlak akar-akar persamaan karakteristik dalam persamaan di atas lebih dari satu, artinya $\theta(B) > 1$ (Makridakis & Wheelwright, 1998).

3.3 Model Autoregressive Moving Average

Model *Autoregressive Moving Average* (ARMA) menggabungkan model *autoregressive* (AR) dan *moving average* (MA). Dengan demikian, autokorelasinya mempertimbangkan nilai-nilai berturut-turut dari variabel yang diprediksi maupun kesalahan pada masing-masing periode sebelumnya. Model AR dan MA digabungkan dalam satu persamaan yang dikenal sebagai ARMA (p, q). Persamaan umum model ARMA yaitu:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + e_1 + \theta_1 e_{t-1} - \theta_2 e_{t-2} \ldots - \theta_q e_{t-q} \quad (11)$$

Dimana:

$X_{t-1}, X_{t-2}, X_{t-3} \ldots X_{t-p}$ = variabel yang menentukan

X_t = variabel yang diprediksi

$e - e_{t-1} - \cdots - e_{t-q}$ = nilai-nilai kesalahan sebelumnya

e_t = kesalahan pada saat t

θ_q= parameter MA orde ke-q

Contoh:

1) Model ARMA (1,1)

$$X_t = \phi_1 X_{t-1} + e_t - \theta_1 e_{t-1} \quad (12)$$

2) Model ARMA (2,1)

$$X_t = \phi_1 X_{t-1} + \phi_1 X_{t-2} + e_t - \theta_1 e_{t-1} \quad (13)$$

3) Model ARMA (2,2)

$$X_t = \phi_1 X_{t-1} + \phi_1 X_{t-2} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \quad (14)$$

Pada model ARMA, syarat kestasionerannya mengikuti model AR (p), sedangkan model konvergensinya mengikuti model MA (q) (Makridakis & Wheelwright, 1998).

3.4 Model Autoregressive Integrated Moving Average (ARIMA)

Model *Autoregressive Integrated Moving Average* (ARIMA) adalah salah satu model peramalan data deret waktu yang paling populer, pertama kali dikembangkan oleh Box dan Jenkins pada tahun 1970. Model ARIMA menggabungkan model *autoregressive* yang terdiferensiasi dengan model *moving average*.

Persamaan umumnya adalah:

$$Y'_t = l + \alpha_1 Y'_{t-1} + \alpha_2 Y'_{t-2} + \cdots + \alpha_p Y'_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \quad (15)$$

Dimana:

l = level dalam data set

$e - e_{t-1} - \cdots - e_{t-q}$ = nilai-nilai kesalahan sebelumnya

a = koefisien yang harus dipelajari dari data

ϕ_p = parameter AR orde ke-p

Y_t = nilai *time series* pada waktu t

Y'_t = *time series* yang telah terdiferensiasi sebanyak d kali agar menjadi stasioner

Bagian AR dari ARIMA menunjukkan bahwa deret waktu mundur berdasarkan data masa lalunya, sementara bagian MA menunjukkan bahwa kesalahan prediksi merupakan kombinasi linier dari kesalahan masa lalu. Bagian I ARIMA menunjukkan bahwa nilai data telah diganti dengan nilai-nilai yang berbeda sebanyak d kali untuk mencapai stasioneritas, yang merupakan syarat utama model ARIMA.

Persamaan di atas menunjukkan bahwa prediktornya adalah titik data p yang tertinggal untuk bagian *autoregressive* dan kesalahan q yang tertinggal untuk bagian *moving average*, semuanya terdiferensiasi. Prediksinya adalah selisih y_t pada orde ke- d. Ini disebut model ARIMA (p, d, q). Menentukan koefisien α dan θ untuk p, d, q tertentu adalah tugas utama ARIMA saat mempelajari data dari deret waktu. Menentukan nilai p, d, q bisa rumit, tetapi bisa mencoba berbagai kombinasi dan mengevaluasi performa model. Setelah model ARIMA ditentukan dengan nilai p, d, q, koefisien dalam persamaan di atas harus diperkirakan. Metode paling umum untuk estimasi adalah melalui estimasi kuadrat terkecil untuk persamaan regresi, tetapi MLE (*Maximum Likelihood Estimation*) mencari koefisien model yang memaksimalkan peluang menemukan data aktual.

ARIMA adalah model umum, dengan beberapa model khusus yang sering digunakan, seperti:

1) ARIMA (0,1,0) dinyatakan sebagai $y_t = y_{t-1} + e$. Ini adalah model *naive* dengan *error*, juga dikenal sebagai model *random walk*
2) ARIMA (0,2,0) dinyatakan sebagai $y_t = y_{t-1} + e + c$. Ini adalah model *random walk* dengan trend konstan, yang dikenal sebagai *random walk with drift*
3) ARIMA (0,0,0) dinyatakan sebagai $y_t = e$ atau *white noise*
4) ARIMA (p, 0,0) adalah model *autoregressive* (Kotu & Deshpande, 2019)

4. Stasioner dan Non-Stasioner

Pada *time series* dengan tren atau musiman, nilainya dipengaruhi oleh waktu. Deret waktu disebut stasioner ketika nilainya tidak tergantung pada waktu. Contohnya, *white noise* adalah deret waktu yang stasioner. Suhu harian di suatu lokasi tidak stasioner karena adanya kecenderungan musiman dan pengaruh waktu. Namun, komponen *noise* dalam suatu deret waktu bersifat stasioner. Deret waktu yang stasioner tidak memiliki kemampuan untuk meramalkan apa pun karena sifatnya yang benar-benar acak (Athanasopoulos & J Hyndman, 2018).

Deret waktu non-stasioner dapat diubah menjadi deret waktu stasioner melalui teknik yang disebut *differencing*. *Differencing series* adalah perbedaan antara titik data yang berurutan dalam deret tersebut.

$$Y'_t = Y_t - Y_{t-1} \tag{16}$$

Dimana:

Y_t = nilai *time series* pada t

Y'_t = *time series* yang telah diturunkan sebanyak d kali sehingga menjadi stasioner

Ini disebut pembedaan orde pertama. Terkadang, satu kali *differencing* masih menghasilkan deret waktu yang tidak stasioner. Dalam kasus tersebut, diperlukan *differencing* orde kedua, yaitu perbedaan antara dua titik data berurutan dalam deret waktu yang telah berbeda dengan orde pertama. Secara umum, *differencing* orde d digunakan untuk mengubah deret waktu non-stasioner menjadi deret waktu stasioner. (Athanasopoulos & J Hyndman, 2018).

5. Penyusunan Model *Time series*

Penyusunan model *time series* terdiri atas tiga tahap, antara lain:

1) Identifikasi model

Dalam mengidentifikasi model, kita perlu menentukan apakah data berkala yang akan digunakan bersifat stasioner atau non-stasioner. Jika data non-stasioner, langkah pertama adalah melakukan penstasioneran menggunakan metode *differencing*. Stasioneritas tercapai ketika data tidak menunjukkan peningkatan atau penurunan, sehingga tetap berada pada sumbu horizontal. Artinya, frekuensi data berada di sekitar nilai rata-rata yang konstan dan tidak bergantung pada waktu. Setelah data menjadi stasioner, kita harus menentukan bentuk model yang akan digunakan dengan membandingkan koefisien autokorelasi dan autokorelasi parsial dari data tersebut, untuk dicocokkan dengan distribusi yang sesuai dengan model ARIMA.

2) Pendugaan dan pengujian parameter

Setelah berhasil menetapkan identifikasi model sementara, langkah selanjutnya adalah menduga parameter.

Model matematika dan bentuk peramalannya adalah sebagai berikut:

$$X_t = \phi_0 + \phi_1 X_{t-1} + e_t - \theta_1 e_{t-1} \quad (17)$$

dan

$$\hat{X} = \phi_0 + \phi_1 X_{t-1} - \theta_1 e_{t-1} \quad (18)$$

Dimana:

$X_{t-1}, X_{t-2}, X_{t-3} \ldots X_{t-p}$ = variabel yang menentukan

X_t = variabel yang diramalkan

$e - e_{t-1} - \cdots - e_{t-q}$ = nilai-nilai kesalahan sebelumnya

e_t = kesalahan pada saat t

θ_q= parameter MA orde ke-q

Kita harus menentukan ϕ_1 dan θ_1 dengan menggunakan software Eviews, yaitu dengan meminimumkan jumlah kuadrat nilai sisa sebagai kriteria untuk memilih nilai yang optimal.

3) Pengujian parameter serta peramalan model

Pengujian parameter bertujuan untuk memastikan kelayakan model yang digunakan. Ini dilakukan dengan menguji autokorelasi residual $\varepsilon_t = X_t - \hat{X}_t$, untuk memastikan bahwa residual secara signifikan berbeda dari nol. Jika perbedaannya tidak signifikan, model tersebut dianggap tidak layak. Dalam hal ini, perlu kembali ke langkah pertama dan kedua untuk menentukan model alternatif (Makridakis & Wheelwright, 1998).

6. Demam Berdarah

Demam berdarah dengue adalah penyakit infeksi yang bisa berakibat fatal, dengan potensi untuk merenggut nyawa penderitanya dalam waktu singkat jika tidak segera ditangani. Penyakit ini disebabkan oleh virus dengue yang berasal dari famili *Flaviviridae* dan genus *Flavivirus*. Virus ini memiliki empat serotipe, yaitu DEN-1, DEN-2, DEN-3, dan DEN-4, yang masing-masing dapat menyebabkan gejala berbeda pada manusia. Di Indonesia, serotipe yang paling sering menyebabkan infeksi parah adalah DEN-3. Demam berdarah dengue tidak menular melalui kontak langsung antar manusia. Penularannya hanya terjadi melalui nyamuk, sehingga penyakit ini digolongkan ke dalam kelompok penyakit *arthropod borne diseases*. Nyamuk yang paling sering menyebabkan wabah demam berdarah adalah *Aedes aegypti* subgenus *Stegomyia*. Selain itu, jenis nyamuk lain seperti Ae. *albopictus*, Ae. *polynesiensis*, anggota dari Ae. *scutellaris complex*, dan Ae. (*Finlaya*) *niveus* juga dapat menyebarkan virus demam berdarah.

Nyamuk *Aedes aegypti* cenderung berkembang biak di genangan air bersih seperti penampungan air, bak mandi, pot bunga, dan gelas. Tempat-tempat ini sering kali tidak disadari sebagai lokasi favorit nyamuk tersebut. Oleh karena itu, populasi nyamuk ini meningkat selama musim hujan. Penularan penyakit demam berdarah dengue menjadi lebih mudah karena berbagai faktor. Tingginya mobilitas manusia dapat meningkatkan penyebaran demam berdarah dengue. Selain itu, kepadatan penduduk yang tidak merata juga berperan; daerah dengan kepadatan tinggi lebih rentan terhadap penyebaran penyakit ini. Upaya pemberantasan nyamuk sering kali tidak efektif karena hanya nyamuk dewasa yang diberantas, sementara jentik atau telur nyamuk dibiarkan tumbuh dan berkembang. Akibatnya, vektor penyakit demam berdarah dengue dengan cepat kembali berkembang dan menjadi perantara penyakit ini (Satari & Mei, 2004).

B. Kerangka Berpikir

Metode ARIMA adalah teknik statistik yang digunakan untuk memprediksi data *time series*. Langkah-langkah penerapannya (Haritsah, 2015), yaitu:

1) Mengidentifikasi model yang paling cocok dengan data, yang dilakukan dengan menganalisis plot data serta plot ACF (*Autocorrelation Function*) dan PACF (*Partial Autocorrelation Function*) dari data tersebut. Dalam hal ini mencakup langkah-langkah

peramalan mendefinisikan masalah, mengumpulkan informasi, dan analisis pendahuluan (eksplorasi).

2) Melakukan estimasi parameter model. Dalam hal ini mencakup langkah peramalan memilih dan memasang model.

3) Menguji model dengan memeriksa plot residual dan melakukan uji normalitas. Dalam hal ini mencakup langkah peramalan menggunakan dan mengevaluasi model peramalan.