

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Penelitian-Penelitian Sebelumnya

Terdapat beberapa penelitian sebelumnya yang dijadikan acuan, yaitu pada tabel 2.1 sebagai berikut:

Tabel 2. 1 *State of The Art* Penelitian

No	Nama Penulis	Judul Penelitian	Kekurangan	Kelebihan
1	Siti Nurwahdani a, dkk (2020) [5]	Prediksi Produksi Beras Di Provinsi Jawa Timur Menggunakan Musiman <i>Autoregressive Terintegrasi Moving Average</i> Dengan Metode <i>Exogenous Input</i> (SARIMAX)	<ul style="list-style-type: none"> <li>• Membutuhkan data <i>time series</i> yang cukup panjang untuk identifikasi model.</li> <li>• Memerlukan asumsi data stasioner dan parameter yang tepat untuk model optimal.</li> </ul>	<ul style="list-style-type: none"> <li>• Mampu memprediksi data <i>time series</i> yang stasioner.</li> <li>• Dapat mengatasi pola trend, musiman, dan fluktuasi.</li> <li>• Memperhitungkan variabel eksogen.</li> <li>• Hasil prediksi akurat dengan MAPE 5,83%.</li> </ul>
2	Diyanti, dkk (2023)[6]	Prediksi Hasil Panen Padi Tahun 2023 menggunakan	<ul style="list-style-type: none"> <li>• Membutuhkan data yang lengkap dan akurat.</li> </ul>	<ul style="list-style-type: none"> <li>• Dapat memprediksi data <i>time series</i> yang linear.</li> </ul>

		Metode Regresi Linier di Kabupaten Indramayu	<ul style="list-style-type: none"> <li>• Tidak akurat jika data tidak memenuhi asumsi-asumsi regresi linier.</li> <li>• Sensitif terhadap <i>outlier</i> dan <i>multicollinearit y</i>.</li> </ul>	<ul style="list-style-type: none"> <li>• Memodelkan hubungan antara variabel dependen dan independen.</li> <li>• Hasil prediksi cukup akurat dengan <i>R2-Score</i> 0,93%.</li> </ul>
3	Willmen TB Panjaitan, dkk (2022)[7]	Prediksi Panen Padi Menggunakan Algoritma <i>k-nearest Neighbour</i>	<ul style="list-style-type: none"> <li>• Memerlukan data yang cukup banyak.</li> <li>• Dapat dipengaruhi oleh data <i>outlier</i> dan <i>noise</i>.</li> <li>• Pemilihan nilai <i>k</i> yang kritis untuk model optimal.</li> </ul>	<ul style="list-style-type: none"> <li>• Mampu memodelkan data <i>time series</i> dengan pola kompleks.</li> <li>• Hasil prediksi cukup akurat dengan RMSE 90532.279 +/- 0.000 dan <i>Absolute Error</i> 80876.833 +/- 40682.076.</li> </ul>
4	Muhammad Ridwan, dkk (2021)[8]	Peramalan Produksi Beras di Provinsi Jawa Tengah menggunakan metode <i>double exponential smoothing (Holts)</i>	<ul style="list-style-type: none"> <li>• Membutuhkan data <i>time series</i> yang cukup panjang untuk identifikasi model.</li> <li>• Pemilihan nilai <i>alpha</i> dan <i>beta</i> yang kritis untuk model optimal.</li> <li>• Tidak mampu memodelkan data <i>time series</i> dengan pola musiman.</li> </ul>	<ul style="list-style-type: none"> <li>• Dapat memprediksi data <i>time series</i> yang stasioner maupun non-stasioner.</li> <li>• Mampu memprediksi data <i>time series</i> yang memiliki pola <i>trend</i>.</li> <li>• Hasil prediksi cukup akurat dengan MAPE 4,83%.</li> </ul>

Berdasarkan penelitian-penelitian sebelumnya yang terlampir pada Tabel 2.1 yang berfokus pada analisis deret waktu, memiliki beberapa kesamaan dengan metode yang digunakan dalam penelitian terdahulu seperti SARIMAX dan *double exponential smoothing*, khususnya dalam kemampuannya memodelkan data yang memiliki pola trend dan non-stasioner. maka peneliti memutuskan melakukan penelitian dengan mengimplementasikan model ARIMA untuk Prediksi Jumlah Beras di Jawa Barat. Pada penelitian yang dilakukan memiliki pembaharuan dari penelitian sebelumnya yaitu pertama, fokus penelitian belum ada yang secara khusus menerapkan ARIMA untuk wilayah Jawa Barat, dan yang kedua terletak pada sumber data sehingga permasalahan yang terdapat pada sumber data tersebut belum pernah diteliti sebelumnya. Dengan demikian, penelitian ini menawarkan kontribusi baru dalam hal metode, fokus geografis, dan pendekatan data, yang memperkaya literatur prediksi produksi beras di Indonesia, khususnya untuk wilayah Jawa Barat.

## **2.2 Definisi Produksi Beras**

Produksi adalah perubahan input menjadi output melalui mekanisme sistem produksi yang meliputi produksi barang dan jasa. Produksi beras berarti melibatkan perubahan input yang digunakan untuk memproduksinya dari biji-bijian menjadi beras yang dapat dikonsumsi oleh konsumen.

Beras adalah produk hasil pengolahan dari tanaman padi (*Oryza sativa*) yang berasal dari sektor pertanian. Di sebagian besar masyarakat Asia, beras adalah makanan pokok utama. Tanaman padi adalah sumber makanan utama bagi hampir setengah penduduk dunia, dan memiliki nilai strategis yang tinggi. Ini terjadi di Indonesia, di mana hampir seluruh penduduknya mendapatkan makanan mereka dari tanaman padi. Tanaman padi memiliki nilai spiritual, budaya, dan politik yang signifikan. Ketahanan pangan dan pemenuhan kebutuhan hidup warga negara sangat bergantung pada tanaman padi. Hal ini ditunjukkan oleh peningkatan produksi beras yang terus meningkat seiring dengan pertumbuhan populasi. Produksi beras di Indonesia sangat dipengaruhi oleh luas area panen dan tingkat

produktivitas. Dengan luas area panen dan produktivitas yang lebih tinggi, produksi beras juga akan meningkat [1].

### 2.3 Definisi Prediksi

Prediksi adalah upaya untuk mengurangi kesalahan dengan menggunakan informasi masa lalu dan saat ini untuk memperkirakan apa yang paling mungkin terjadi di masa depan. Prediksi tidak selalu memberikan jawaban yang pasti untuk apa yang akan terjadi, tetapi mereka berusaha mencari jawaban sedekat mungkin. Salah satu cara untuk mengatasi masalah menentukan jumlah produksi dalam jangka waktu tertentu adalah dengan menggunakan teknik prediksi. Mengetahui prediksi jumlah yang akan diproduksi akan membantu perusahaan dalam pemenuhan bahan baku. Selain itu, dengan mengetahui prediksi jumlah produksi, masalah pemenuhan kebutuhan konsumen terhadap produk tertentu akan diselesaikan. [9].

Prediksi adalah perkiraan tentang apa yang akan terjadi selanjutnya dengan menggunakan informasi dan metode sebelumnya. Hal ini didasarkan pada pola data. Oleh karena itu, prediksi merupakan suatu penilaian yang memuat lebih dari satu perkiraan dan didasarkan pada metode tertentu. Dengan mempelajari pola data saat ini dan pola sebelumnya.

Data runtun waktu dibagi menjadi dua bagian, hal tersebut berguna untuk kepentingan evaluasi peramalan sebagai berikut.

1. Data *training* atau data *in-sample*, digunakan untuk mendapatkan model yang sesuai.
2. Data *testing* atau data *out-sample*, digunakan untuk mengevaluasi kinerja Model

### 2.3.1 Jenis-Jenis Metode Prediksi

Terdapat 2 jenis pendekatan umum untuk metode prediksi yaitu kuantitatif dan kualitatif sebagai berikut [10]:

1. Metode prediksi kualitatif: Ini digunakan ketika data tidak cukup untuk dianalisis dan berpusat pada pendapat dan analisis deskriptif. Contoh metode ini meliputi analisis data ekonomi, riset pasar, wawancara, dan diskusi.
2. Metode Prediksi Kuantitatif: Metode ini berhubungan dengan hitungan matematis dan sering digunakan ketika ada data yang cukup untuk dianalisis. Beberapa metode kuantitatif yang digunakan untuk algoritmanya meliputi metode deret berkala (seri waktu), seperti rata-rata gerak sederhana, rata-rata gerak sederhana, Arima, Sarima, model naif, perhalusan ekspektasi tunggal, perhalusan ekspektasi dua, dan *Holt-Winters* ekspektasi perhalusan.

### 2.3.2 Jangka Waktu Prediksi

Jangka waktu Prediksi dapat dikelompokkan menjadi tiga kategori, yaitu sebagai berikut [10]:

1. Prediksi jangka panjang, yang menggunakan analisis dengan waktu yang lebih lama, biasanya berlangsung lebih dari dua tahun.
2. Prediksi jangka menengah, yang memiliki waktu tiga hingga dua tahun.
3. Prediksi jangka pendek, yang memiliki waktu nol hingga tiga bulan.

### 2.3.3 Proses Prediksi

Dalam proses prediksi, terlepas dari bentuk atau jenis prediksi yang akan dilakukan, terdapat lima langkah utama yang dapat diimplementasikan, yaitu [11] :

1. Formulasi masalah dan pengumpulan data.  
Jika mengadopsi metode prediksi kuantitatif, penting untuk memastikan ketersediaan dan keakuratan data. Jika data tidak benar, maka perlu dilakukan evaluasi ulang terhadap masalah prediksi atau bahkan perlu memeriksa kembali metode prediksi kuantitatif yang

digunakan. Ketersediaan dan keakuratan data adalah elemen kunci dalam memastikan keberhasilan dan keandalan prediksi.

2. Manipulasi dan pembersihan data

mungkin kita dihadapkan dengan kemungkinan memiliki data yang terlalu banyak atau terlalu sedikit, dan sebagian informasi mungkin tidak relevan terhadap permasalahan yang dihadapi. Ada potensi keberadaan nilai yang hilang dalam beberapa data yang perlu diestimasi. Sebagian data mungkin memerlukan pemrosesan tambahan, seperti penggabungan dari berbagai sumber, sementara beberapa data mungkin perlu dikonversi ke unit lainnya. Selain itu, beberapa data mungkin hanya relevan untuk periode waktu tertentu. Pada umumnya, diperlukan usaha untuk memastikan bahwa data tersedia dalam format yang sesuai dengan kebutuhan prosedur prediksi yang spesifik.

3. Pembentukan dan evaluasi model

fokus terletak pada memastikan bahwa data yang telah dikumpulkan sesuai dengan model prediksi yang diinginkan, dan upaya dilakukan untuk meminimalkan jumlah kesalahan prediksi. Proses ini melibatkan pembentukan model prediktif yang memadai, yang kemudian dievaluasi untuk memastikan konsistensi dan keakuratannya. Langkah-langkah evaluasi ini penting untuk menjamin bahwa model dapat memberikan prediksi yang dapat diandalkan sesuai dengan masalah yang dihadapi.

4. Implementasi model (prediksi sebenarnya)

Implementasi model mencakup pembuatan model prediksi aktual ketika data yang diperlukan telah dikumpulkan dan dipilih. Prediksi ini sering kali dibandingkan dengan nilai historis aktual yang diketahui untuk periode waktu sekarang. Proses ini bertujuan untuk mengevaluasi keakuratan model, dan penggunaan nilai historis yang telah terjadi dapat memberikan gambaran seberapa baik model tersebut dapat mereplikasi dan meramalkan kondisi aktual. Perbandingan ini

membantu dalam menilai kinerja model dan mengukur sejauh mana model tersebut dapat diandalkan dalam situasi praktis.

5. Evaluasi prediksi

Setelah model dipilih dengan hati-hati dan parameternya diestimasi dengan tepat, model tersebut akan digunakan untuk membuat prediksi, dan pengguna prediksi akan mengevaluasi manfaat dan kekurangan dari model tersebut seiring berjalannya waktu. Ketika model dipasang pada data yang diketahui, prakiraan tugas peramalan belum selesai. Hanya setelah data untuk periode peramalan tersedia, kinerja model dapat dievaluasi dengan benar.

#### 2.3.4 Karakteristik Prediksi

Karakteristik prediksi yang baik seharusnya memiliki kriteria yaitu sebagai berikut [12]:

1. Ketepatan dan Akurasi

Tujuan utama dari proses prediksi adalah untuk membuat prediksi yang tepat dan akurat. Prediksi yang terlalu rendah dapat menyebabkan kekurangan persediaan, sementara prediksi yang terlalu tinggi dapat menyebabkan biaya operasional yang lebih tinggi. Oleh karena itu, tingkat akurasi dan ketepatan menjadi ukuran penting untuk menilai keberhasilan prediksi.

2. Biaya

Biaya pengembangan model prediksi dan pelaksanaan prediksi dapat berubah jika jumlah produk dan data meningkat. Dalam pengembangan model prediksi, terdapat *trade-off* antara tingkat keakuratan dan biaya yang harus diperhitungkan. Ini karena pengembangan model yang lebih kompleks, meskipun dapat meningkatkan keakuratan prediksi, juga dapat menghasilkan biaya yang lebih tinggi.

### 3. Responsif

Prediksi yang responsif seharusnya konsisten dan tidak dipengaruhi oleh perubahan kebutuhan. Kemampuan untuk beradaptasi dengan cepat terhadap perubahan dalam lingkungan atau permintaan adalah ciri penting dari prediksi yang responsif. Dengan demikian, model prediksi seharusnya dapat mempertahankan tingkat konsistensi yang tinggi saat menanggapi perubahan.

### 4. Sederhana

Salah satu keuntungan utama menggunakan prediksi yang sederhana adalah kemudahan pelaksanaannya. Jika ada masalah dengan metode tersebut, diagnosis dapat dilakukan dengan lebih mudah. Secara umum, lebih baik menggunakan metode yang paling sederhana, yang tetap memenuhi kebutuhan prediksi.

## 2.4 Deret Waktu Model

Deret Waktu adalah suatu rangkaian atau seri nilai dari suatu variabel atau hasil observasi. Konsep *time series* melibatkan metode prediksi yang menggunakan analisis pola hubungan antara variabel yang akan diperkirakan dan variabel waktu. Ini adalah metode statistik yang digunakan untuk menganalisis data yang dihasilkan secara rinci sepanjang waktu.

Deret Waktu dapat digunakan untuk beberapa tujuan, termasuk identifikasi tren, siklus, dan variabilitas dalam data. Selain itu, deret waktu juga dapat digunakan untuk melakukan prediksi masa depan berdasarkan data historis. Beberapa kegunaan *time series* antara lain:

1. Metode Box–Jenkins (ARIMA)
2. Metode Proyeksi trend dengan Regresi.
3. Metode Pemulusan (*Smoothing*)

Tingkat kesalahan adalah komponen penting dari metode prediksi, dan penting untuk diperhatikan dalam setiap proses prediksi. Para peramal berusaha mengurangi tingkat kesalahan agar hasilnya lebih mirip dengan data asli. Analisis

rangkaian waktu adalah jenis analisis yang mengukur dan menjelaskan perubahan atau perkembangan dalam data selama periode waktu tertentu. Metode ini memanfaatkan data masa lalu untuk memahami pola dalam rangkaian waktu, yang memungkinkan untuk meramalkan nilai di masa mendatang [13].

#### 2.4.1 Pola Data

Mempertimbangkan jenis pola data adalah langkah penting dalam memilih deret waktu yang tepat. Oleh karena itu, metode yang paling sesuai dengan pola ini dapat diuji. Plot data terdiri dari empat kategori berbeda yaitu sebagai berikut [14]:

1. Pola Data Tren (*Trend*)

Plot data menunjukkan kecenderungan naik atau turun secara berkesinambungan selama periode waktu tertentu.

2. Pola Data Musiman (*Seasonal*)

Pola musiman terjadi ketika data menunjukkan fluktuasi atau pola berulang dalam periode waktu tertentu, seperti harian, bulanan, atau tahunan.

3. Pola Data Siklus (*Cyclic*)

Siklus menciptakan pola yang lebih luas dan tidak teratur yang melibatkan periode waktu yang lebih panjang, tidak seperti musiman yang bersifat lebih teratur.

4. Pola Data Horizontal (*Horizontal*)

Terjadi ketika nilai data berfluktuasi di sekitar nilai rata-rata yang konstan. Deret seperti itu adalah stasioner terhadap nilai rata-ratanya.

#### 2.5 Identifikasi Model Deret Waktu

Identifikasi model Deret Waktu adalah tahapan penting dalam analisis data, Pemahaman terhadap model *time series* atau seri waktu merupakan kunci penting untuk meramalkan dan memahami tren data sepanjang waktu. Model *time series* memungkinkan untuk mengidentifikasi pola dalam data yang berubah-ubah seiring waktu dan menggunakan pola tersebut untuk membuat prediksi masa depan yang akurat, Untuk tahapan identifikasi Model yaitu sebagai berikut:

### 2.5.1 Proses *Integrated* (Stasioneritas dan Non stasioneritas)

Stasioneritas berarti tidak terdapat perubahan yang drastis pada data. Fluktuasi data berada di sekitar suatu nilai rata-rata yang konstan, tidak tergantung pada waktu dan variansi dari fluktuasi tersebut [15] Data Deret waktu dikatakan stasioner jika rata-rata dan variansinya konstan, tidak ada unsur *trend* dalam data, dan tidak ada unsur musiman.

Ingatlah bahwa sebagian besar deret berkala tidak stasioner, dan aspek AR dan MA dari model ARIMA hanya berkaitan dengan deret berkala yang stasioner. Suatu deret waktu Apabila data tidak stasioner, maka perlu dilakukan transformasi untuk menghasilkan data yang stasioner. Salah satu cara yang sering dipakai adalah metode pembedaan (*differencing*). Untuk menentukan apakah Deret Waktu stasioner atau non stasioner dapat dibantu dengan melihat plot dari *series* atau bentuk *difference*-nya. Proses *differencing* dapat dilakukan untuk beberapa periode sampai data stasioner, yaitu dengan cara mengurangi suatu data dengan data sebelumnya. Yang dimaksud dengan *differencing* adalah untuk menemukan nilai perbedaan atau perubahan dari nilai yang diamati. Untuk memastikan apakah varians stasioner, lakukan *differencing* lagi. Jika tidak, lakukan transformasi logaritma. Untuk melakukan Proses *differencing* menggunakan persamaan sebagai berikut :

$$y_t = z_t - z_{t-1} \text{ atau, } y_t = (1 - B)^d Z_t \quad (2-1)$$

Dimana :

$Y_t$  : nilai variabel pada waktu t setelah *differencing*

$d$  : 1,2, ...(biasanya 1 dan 2)

$Z_t$  : nilai variabel Z pada waktu t

$Z_{t-1}$  : nilai variabel Z pada saat t-1

$B$  : *Backshift* operator (operator mundur)

Kestasioneran data terhadap rata-rata dapat diketahui dengan menggunakan uji *Augmented Dicky Fuller* (ADF). Uji ADF digunakan untuk menguji kestasioneran dalam rata-rata dan untuk memastikan apakah data perlu dilakukan *differencing* atau tidak. Konsep pengujian ADF adalah jika suatu data *time series* tidak stasioner pada orde nol, maka stasioneritas data tersebut dapat dicari melalui order 9 berikutnya. Sehingga diperoleh tingkat stasioneritas pada order ke-n, (*first differencing*) atau *second differencing* dan seterusnya. Uji ADF mempunyai persamaan sebagai berikut[16]:

$$\Delta Y_t = \beta_1 + \beta_2 + \delta Y_{t-1} + \phi_i \sum_{i=1}^k \Delta Y_{t-1} \varepsilon_t \quad (2-2)$$

Dimana :

- $\Delta Y_t$  : *first different* dari Y
- $\beta_1$  : nilai konstan atau *intercept*
- $\beta_2$  : koefisien untuk trend
- $\delta$  : koefisien untuk lag Y
- $\phi$  : koefisien untuk *difference lag* Y
- $\varepsilon$  : error
- k : lag
- t : waktu

Sehingga uji hipotesis yang digunakan untuk melihat apakah data runtun waktu memiliki kondisi stasioner pengujian berdasarkan *p-value* dari uji ADF.

i. Hipotesis

$H_0 : \delta = 0$  (Data tidak stasioner)

$H_1 : \delta \neq 0$  (Data stasioner)

ii. Taraf signifikansi

$\alpha = 5\% = 0,05$

## iii. Daerah kritis

Tolak  $H_0$  jika  $ADF (thitung) > t_{\alpha/2;n}$  atau  $p\text{-value} < \alpha$

## iv. Statistik uji

$$ADF = \frac{\hat{\delta}}{SE\hat{\delta}} \quad (2-3)$$

Keterangan:

$\hat{\delta}$  = penduga kuadrat terkecil dari  $\delta$

$SE\hat{\delta}$  = standar error untuk  $\hat{\delta}$

## v. Keputusan dan kesimpulan

Jika diperoleh nilai  $ADF (thitung) > ttabel$  atau  $p\text{-value} < \alpha$  maka keputusan tolak  $H_0$  dan artinya data stasioner.

Dalam Deret Waktu, stasioneritas dibagi menjadi 2 [17], yaitu:

1. Stasioner dalam *mean* (rata-rata)

Stasioner dalam mean adalah fluktuasi data berada di sekitar suatu nilai rata-rata yang konstan, tidak tergantung pada waktu dan variansi dari fluktuasitersebut. Dari bentuk plot data seringkali dapat diketahui bahwa data tersebut stasioner atau tidak stasioner. Apabila dilihat dari plot *ACF*, maka nilai-nilai autokorelasi dari data stasioner akan turun menuju nol sesudah *time lag* (selisih waktu) kedua atau ketiga.

## 2. Stasioneritas dalam Variansi

Suatu data *time series* dikatakan stasioner dalam variansi apabila struktur data dari waktu ke waktu mempunyai fluktuasi data yang tetap atau konstan dan tidak berubah-ubah. Secara visual untuk melihat hal tersebut dapat dibantu dengan menggunakan plot *time series*, yaitu dengan melihat fluktuasi data dari waktu ke waktu.

### 2.5.2 Autocorrelation Function (ACF)

Fungsi Autokorelasi (ACF) merupakan suatu fungsi yang menunjukkan ukuran korelasi antara observasi pada waktu tertentu dengan observasi pada waktu sebelumnya. Definisi dari fungsi autokorelasi untuk data seri waktu yang stasioner dinyatakan sebagai  $\rho_k$ , di mana untuk nilai lag yang berurutan seperti 1, 2, 3, ..., k. Maka fungsi autokorelasi sebagai berikut [17],

$$\hat{\rho}_k = \frac{Cov(Y_t, Y_{t+k})}{Var(Y_t)} = \frac{\sum_{t=1}^{n-k} (Y_t - \mu_t)(Y_{t+k} - \mu_t)}{\sum_{t=1}^n (Y_t - \mu_t)^2} \quad (2-4)$$

Dimana :

$\hat{\rho}_k$  : koefisien autokorelasi lag ke-k, untuk  $k = 1, 2, 3, \dots$

$n$  : jumlah data pengamatan

$Y_t$  : nilai pengamatan periode ke-t

$\mu_t$  : nilai rata-rata dari data pengamatan atau  $E(Y_t)$

### 2.5.3 Partial Autocorrelation Function (PACF)

Fungsi Autokorelasi Parsial (PACF) merupakan suatu teknik perhitungan yang digunakan untuk mengukur tingkat hubungan langsung antara  $Y_t$  dan  $Y_{t+k}$  dengan mengesampingkan pengaruh dari data yang berada di antara keduanya, yaitu  $Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1}$ . Fungsi PACF didefinisikan sebagai berikut.

$$\hat{\phi}_{kk} = \frac{\hat{\rho}_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \hat{\rho}_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \hat{\rho}_j} \quad (2-5)$$

Dan,

$$\hat{\phi}_{k,j} = \hat{\phi}_{k-1,j} - \hat{\phi}_{kk} \hat{\phi}_{k-1,l-j}, \quad j = 1, 2, 3, \dots, k-1 \quad (2-6)$$

Dimana :

$J$  : 1, 2, 3, ..., k dengan nilai  $\hat{\phi}_{11} = \rho_1$  dan  $\rho_0 = 1$

#### 2.5.4 Estimasi Parameter

Setelah berhasil menetapkan identifikasi model sementara, langkah selanjutnya adalah pendugaan parameter. Parameter dari model tersebut ditaksir, sehingga didapatkan besaran koefisien model. Secara umum, penaksiran parameter model ARIMA *Box-Jenkins* dapat dilakukan dengan menggunakan beberapa metode seperti metode moment, metode *least squared*, metode *maximum likelihood*, dan sebagainya. sebagian besar model ARIMA adalah model nonlinier dan memerlukan penggunaan prosedur fitting model nonlinier. Namun, hal ini biasanya secara otomatis dilakukan oleh paket perangkat lunak yang canggih seperti seperti Mini tab, JMP, dan SAS. Pada beberapa paket perangkat lunak, pengguna mungkin memiliki pilihan metode estimasi dan dapat memilih metode yang paling tepat berdasarkan yang paling tepat berdasarkan spesifikasi masalah[18].

Setelah diperoleh nilai estimasi dari masing-masing parameter, kemudian dilakukan pengujian signifikansi parameter untuk mengetahui apakah model sudah layak atau belum untuk digunakan.

i. Hipotesis

$H_0 : \phi_i = 0$  (Parameter AR tidak signifikan)

$H_1 : \phi_i \neq 0$  dimana  $i = 1, 2, 3, \dots, p$  (Parameter AR signifikan)

ii. Taraf signifikansi

$\alpha = 5\% = 0,05$

iii. Daerah kritis

Tolak  $H_0$  jika nilai  $|t_{hitung}| > t_{\alpha/2, (n-np)}$  dimana  $n$  adalah jumlah data pengamatan dan  $np$  adalah jumlah parameter yang diestimasi.

Tolak  $H_0$  jika  $p\text{-value} < \alpha (0,05)$ .

iv. Statistik uji

$$t_{hitung} = \frac{\hat{\phi}_i}{SE(\hat{\phi}_i)} \quad (2-7)$$

Keterangan:

$\hat{\phi}_i$  = nilai estimasi dari parameter model AR dimana  $i = 1, 2, 3, \dots$

$SE(\hat{\phi}_i)$  = standard error dari parameter model AR

v. Keputusan dan kesimpulan

Berdasarkan hasil  $|t_{hitung}| > t_{tabel}$  atau p-value  $< \alpha (0,05)$ , maka tolak  $H_0$  kesimpulannya adalah parameter model signifikan.

### 2.5.5 Pengujian Diagnostik

Setelah tahap estimasi parameter, langkah selanjutnya adalah melakukan pemeriksaan diagnosa. Pada tahap ini, model yang dianggap cocok untuk digunakan dalam peramalan adalah model yang residunya memenuhi asumsi *white noise*. Residual yang memenuhi asumsi *white noise* berarti bahwa residu dalam pemodelan seri waktu merupakan variabel acak yang tidak saling berkorelasi satu sama lain. Proses *white noise* dapat dilakukan menggunakan uji statistik Ljung-Box untuk melihat adanya autokorelasi residual pada analisis *error*-nya [17]. Hipotesis yang digunakan sebagai berikut:

i. Hipotesis

$H_0 : \rho_1 = \rho_2 = \rho_3 = \dots = \rho_k = 0$  (Residual *white noise*)

$H_1 : \text{minimal ada satu } \rho_j \neq 0, j = 1, 2, 3, \dots, K$  (Residual tidak *white noise*)

ii. Taraf signifikansi

$\alpha = 5\% = 0,05$

iii. Daerah kritis Tolak  $H_0$  jika nilai  $Q > \chi^2_{\alpha; df=K-p-q}$  dengan K adalah lag dan p adalah ordo AR dan q adalah ordo MA.

iv. Statistik uji

$$Q = n(n+2) \sum_{k=1}^K \frac{\rho_k^2}{n-k} \quad (2-8)$$

Keterangan:

Q : parameter berdistribusi chi-square

K : banyak lag yang diuji

$\rho_k$  : *autocorrelation* untuk lag ke-k,  $k = 1, 2, 3 \dots k$

n : jumlah data pengamatan variabel Y

v. Keputusan dan kesimpulan

Berdasarkan hasil statistik uji  $Q < X_{tabel 2}$  atau  $p\text{-value} > \alpha (0,05)$  yang diperoleh artinya gagal tolak  $H_0$ , maka kesimpulannya adalah model memenuhi asumsi *white noise* atau tidak terdapat autokorelasi antar residual.

## 2.6 Metode *Autoregressive Integrated Moving Average* (ARIMA)

ARIMA merupakan suatu metode prediksi *time series* yang berbasis pada analisis pola sintesis data masa lalu digunakan dengan asumsi bahwa data seri waktu bersifat stasioner. Salah satu terobosan terbaru dalam metode prediksi ekonomi adalah *Autoregressive Integrated Moving Average* (ARIMA), yang dikembangkan oleh George Box dan Gwilym Jenkins dan dikenal sebagai "Metode Box-Jenkins" sejak tahun 1976. ARIMA digunakan untuk membangun model struktural dengan merangkum teori ekonomi dan logika, serta menganalisis probabilitas atau sifat stokastik dari data seri waktu [14].

### 2.6.1 *Autoregressive Model* (AR)

*Autoregressive* (AR) adalah model yang didasarkan pada asumsi bahwa data pada periode sekarang dipengaruhi oleh data periode sebelumnya. Model AR biasanya diwakili oleh ordo p (AR(p)) atau model ARIMA (p,0,0) dalam persamaan matematika (2-1).

$$X_t = \emptyset_1 X_{t-1} + \dots + \emptyset_p X_{t-p} + S_t \quad (2-9)$$

Dimana:

$X_t$	: nilai variabel pada waktu ke- $t$
$X_t, X_{t-1}, X_{t-2}, \dots, X_{t-p}$	: nilai masa lalu dari <i>time series</i> yang bersangkutan pada waktu $t, t-1, t-2, \dots, t-p$
$\theta_i$	: koefisien regresi, $i: 1, 2, 3, \dots, p$
$s_t$	: nilai <i>error</i> pada waktu ke- $t$

### 2.6.2 Moving Average Model (MA)

*Moving Average* (MA) yang merupakan bentuk umum model *moving average* ordo  $q$  (MA( $q$ )) atau ARIMA (0,0, $q$ ) dinyatakan dalam persamaan matematika (2-2).

$$X_t = s_t + \theta_1 s_{t-1} + \theta_2 s_{t-2} + \dots + \theta_q s_{t-q}; s_t \sim N(0, \sigma^2) \quad (2-10)$$

Dimana:

$X_t$	: nilai variabel pada waktu ke- $t$
$s_t, s_{t-1}, s_{t-2}, \dots, s_{t-q}$	: nilai-nilai dari <i>error</i> pada waktu $t, t-1, t-2, \dots, t-q$
$\theta_i$	: koefisien regresi, $i: 1, 2, 3, \dots, q$
$q$	: orde MA

### 2.6.3 Autoregressive Moving Average (ARMA)

*Autoregressive Moving Average* (ARMA) terdiri dari dua proses yang berbeda yaitu proses ARMA dan proses ARIMA. Model umum untuk proses ARMA adalah campuran dari ordo  $p$  (AR( $p$ )) dan *moving average* ordo  $q$  (MA( $q$ )) murni yang dinyatakan dalam persamaan matematika (2-3).

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + s_t + \theta_1 s_{t-1} + \quad (2-11)$$

$$\theta_2 s_{t-2} + \dots + \theta_q s_{t-q}$$

Dimana :

$X_t$	: nilai variabel pada waktu ke- $t$
$\phi_i$	: koefisien regresi, $i: 1, 2, 3, \dots, p$
$p$	: orde <i>AR</i>
$q$	: orde <i>MA</i>
$\theta_i$	: parameter model <i>MA</i> ke- $i$ , $i = 1, 2, 3, \dots, q$
$s_t$	: nilai <i>error</i> pada waktu ke- $t$
$s_t, s_{t-1}, s_{t-2}, \dots, s_{t-q}$	: <i>error</i> pada waktu $t-1, t-2, \dots, t-q$ dan $s_t$

diasumsikan *White Noise* dan normal.

#### 2.6.4 Autoregressive Integrated Moving Average (ARIMA)

Apabila pada proses *ARMA* tidak terpenuhi stasioneritasnya, maka model umum *ARIMA* ( $p, d, q$ ) terpenuhi. Secara umum model *ARIMA* ( $p, d, q$ ) untuk suatu data *time series*  $X_t$  adalah sebagai berikut:

$$\phi_p(B)(1 - B)^d y_t = \theta_q(B)e_t \quad (2-12)$$

Persamaan (2-12) dapat ditulis menggunakan operator  $B$  (*backshift*), menjadi:

$$(1 - B)^d (X_t - \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p}) = (s_t + \theta_1 s_{t-1} + \theta_2 s_{t-2} + \dots + \theta_q s_{t-q}) \quad (2-13)$$

Dimana:

$X_t$	: nilai variabel pada waktu ke- $t$
$B$	: operator <i>backshift</i>
$(1 - B)^d X_t$	: <i>time series</i> yang stasioner pada perbedaan ke- $d$
$s_t$	: nilai <i>error</i> pada waktu ke- $t$

$p$	: orde <i>AR</i>
$d$	: orde perbedaan
$q$	: orde <i>MA</i>
$S_t, S_{t-1}, S_{t-2}, \dots, S_{t-q}$	: <i>error</i> pada waktu $t-1, t-2, \dots, t-q$ dan $s_t$ diasumsikan <i>White Noise</i> dan normal.

### 2.6.5 Klasifikasi Model ARIMA

Metode ini terbagi menjadi tiga kelompok dalam model Box-Jenkins (ARIMA), yaitu model *Autoregressive* (AR), model *Moving Average* (MA), dan model campuran ARIMA (*Autoregressive Integrated Moving Average*). Setiap kelompok model memiliki karakteristik yang serupa dengan model pertamanya[19].

Model ARIMA, yang merupakan singkatan dari *Autoregressive Integrated Moving Average*, dengan notasi (p, d, q). Komponen p menandakan proses *autoregressive* dalam model, d menunjukkan tingkat integrasi yang diperlukan untuk data sebelum analisis, dan q menunjukkan proses *moving average*. Jika  $d = 0$  dan  $q = 0$ , model tersebut disebut model *autoregressive* dan dinotasikan AR(p). Sebaliknya, jika  $d = 0$  dan  $p = 0$ , model tersebut adalah model *moving average* dan dinotasikan MA(q). Model yang menggabungkan ketiga proses ini dikenal sebagai model *autoregressive integrated moving average*, atau ARIMA (p, d, q). Kunci untuk membangun model ARIMA yang efektif adalah memastikan bahwa model tersebut memiliki tingkat kesalahan yang rendah. Oleh karena itu, sangat penting untuk secara hati-hati mengidentifikasi model seri waktu yang digunakan. Dalam pengaplikasian metode ARIMA, ada empat langkah kritis yang harus diikuti: identifikasi korelasi, penentuan parameter model, pemeriksaan diagnostik model, dan akhirnya, peramalan, sebagaimana disarankan oleh Montgomery dan Johnson pada tahun 1998 (hal. 190).

*Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA) atau Autoregressive Integrated Moving Average (ARIMA),* dapat dilihat dari bentuk plot ACF dan PACF pada tabel berikut :

**Tabel 2. 2** Identifikasi Tipe Model ARIMA

Type Model	ACF	PACF
MA(q)	<i>Cut-off</i> setelah lag ke- $q$ (nantinya jadi patokan orde ke-berapa)	<i>Dies Down</i> (Menurun) mengikuti bentuk eksponensial atau gelombang sinus menuju nol
AR(p)	<i>Dies Down</i> (Menurun) mengikuti bentuk eksponensial atau gelombang sinus menuju nol Selang-seling (kanan kiri)	<i>Cut-off</i> setelah lag ke- $p$ (nantinya jadi patokan orde ke-berapa)
ARMA(p,q)	<i>Dies Down</i> setelah lag. ( $q-p$ )	<i>Dies Down</i> setelah lag. ( $q-p$ )

Model *Autoregressive Integrated Moving Average (ARIMA)* merupakan salah satu model yang populer dalam peramalan data runtun waktu. Proses ARIMA (p,d,q) merupakan model runtun waktu ARMA(p,q) yang memperoleh *differencing* sebanyak d. Proses ARMA (p,q) adalah suatu model campuran antara *autoregressive* orde p dan *moving average* orde q

- Jika ACF menunjukkan pola *dying down*, dan PACF menunjukkan *cut off*, maka dapat dikatakan model ARIMA berupa AR murni.
- Jika ACF menunjukkan pola *cut off*, dan PACF menunjukkan *dying down*, maka dapat dikatakan model ARIMA berupa MA murni.
- Jika ACF dan PACF menunjukkan *dying down* maka dapat dikatakan model ARIMA berupa gabungan AR dan MA.

## 2.7 Nilai Ketepatan Prediksi

Untuk memberikan nilai ketepatan prediksi, perlu memahami konteks atau model apa yang sedang digunakan. Ketepatan prediksi atau akurasi biasanya diukur dengan membandingkan hasil prediksi model dengan nilai sebenarnya pada data yang telah diuji. Untuk menjelaskan kesalahan (*error*) yang dihasilkan oleh fakta dalam teknik prediksi, ada beberapa pendekatan yang lebih dikenal. Rata-rata dari perbedaan antara nilai aktual dan nilai prediksi diperlukan untuk sebagian besar pengukuran ini. Beberapa metrik umum yang digunakan untuk mengukur ketepatan prediksi antara lain [20]:

### 2.7.1 Mean square error (MSE)

MSE (*Mean square error*) adalah salah satu metode pengujian kesalahan yang paling umum digunakan Untuk mengukur keakuratan hasil prediksi model. MSE juga dapat menunjukkan ukuran kesalahan yang dihasilkan oleh model Prediksi. Cara menghitungnya dengan mengurangi nilai aktual dan hasil prediksi, kemudian mengkuadratkan dan menjumlahkan hasil akhir. Dengan rumus :

$$MSE = \frac{\sum_{t=1}^n (A_t - F_t)^2}{n} \quad (2-14)$$

Dimana :

$A_t$  : Nilai Aktual permintaan

$F_t$  : Nilai hasil peramalan

$n$  : banyaknya data

### 2.7.2 Mean Absolute Percentage Error (MAPE)

MAPE (*Mean Absolute Percentage Error*) adalah alat statistik yang digunakan untuk mengukur keakuratan suatu model statistik dalam melakukan prediksi atau peramalan. Selisih data aktual dengan peramalan dibagi dengan data aktual, kemudian nilainya dimutlakkan (*absolut*). Artinya MAPE akan selalu bernilai positif. Dengan rumus:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right| \times 100\% \quad (2-15)$$

Dimana:

- $n$  : ukuran sampel
- $A_i$  : nilai data aktual
- $F_i$  : nilai data peramalan

Interpretasi nilai MAPE dapat dilihat dari interval nilainya sebagai berikut.

Tabel 2. 3 Interpretasi MAPE

Nilai MAPE (%)	Interpretasi
$\leq 10$	Hasil peramalan sangat akurat
10 - 20	Hasil peramalan baik
20 - 50	Hasil peramalan layak (cukup baik)
$> 50$	Hasil peramalan tidak akurat