

PEMODELAN DAN PERAMALAN DATA DERET WAKTU DENGAN METODE SEASONAL ARIMA

ANNISA UL UKHRA

*Program Studi Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia,
annisaulkhra25@gmail.com*

Abstrak. Penelitian ini mempelajari suatu metode dalam meramalkan data *time series* yang berupa data musiman. Metode yang dipakai dalam penelitian ini adalah *Seasonal Autoregressive Integrated Moving Average (SARIMA)*. Metode ini merupakan gabungan dari proses *Autoregressive (AR)* Musiman dan *Moving Average (MA)* Musiman. Metode ini diaplikasikan pada data jumlah penumpang suatu maskapai penerbangan dari tahun 1949 – 1959. Sehingga diperoleh ramalan jumlah penumpang untuk tahun berikutnya. Hasil tersebut dibandingkan dengan data aktual dimana data yang diperoleh tidak jauh berbeda dengan data aktual.

Kata Kunci: Metode SARIMA, AR Musiman, MA Musiman

1. Pendahuluan

Analisis deret waktu pada dasarnya digunakan untuk melakukan analisis data yang mempertimbangkan pengaruh waktu. Data dikumpulkan secara periodik berdasarkan urutan waktu, bisa dalam jam, hari, minggu, bulan, kuartal dan tahun. Analisis deret waktu dapat dilakukan untuk membantu dalam menyusun perencanaan ke depan. Sebagai contoh, dalam kasus peramalan banyaknya penumpang suatu maskapai penerbangan dalam rentang waktu tertentu, yang diperlukan hanya data jumlah penumpang pada tahun-tahun sebelumnya. Model deterministik adalah model yang nilai observasi mendatang dapat dihitung atau dapat diramalkan secara pasti melalui suatu fungsi berdasarkan observasi masa lampau, tetapi peramalan hanya berlaku untuk data yang ada saja. Untuk menentukan metode peramalan pada data deret waktu perlu diketahui pola dari data tersebut sehingga peramalan data dapat dilakukan dengan metode yang sesuai. Pola data dapat dibedakan menjadi empat jenis, yaitu pola musiman, siklis, *trend*, dan *irregular* [5]. Pola musiman merupakan fluktuasi dari data yang terjadi secara periodik dalam kurun waktu satu tahun, seperti triwulan, kuartalan, bulanan, mingguan, atau harian. Pola siklis terjadi apabila datanya dipengaruhi oleh fluktuasi ekonomi jangka panjang, seperti yang berhubungan dengan siklus bisnis. Pola ini sulit dideteksi dan tidak dapat dipisahkan dari pola *trend*. Pola *trend* merupakan kecenderungan arah data dalam jangka panjang, dapat berupa kenaikan maupun penurunan. Sedangkan pola *irregular* merupakan kejadian yang tidak terduga dan bersifat acak, tetapi kemunculannya dapat mempengaruhi fluktuasi data *time series*.

Untuk data model stokastik terdapat beberapa model yang dapat digunakan seperti *AR*, *MA*, *ARMA*, *ARMA*, *SARIMA* dan lainnya. Jika data mempunyai pola musiman, maka metode yang lebih tepat adalah *Seasonal Autoregressive Integrated Moving Average (SARIMA)*. Metode inilah yang menjadi pokok bahasan pada paper ini.

2. Metodologi

Langkah-langkah pemodelan *Seasonal Autoregressive Integrated Moving Average (SARIMA)* adalah:

- (1) Proses identifikasi model.
- (2) Pendugaan parameter model.
- (3) Pemeriksaan residual(sisaan).
- (4) Menggunakan model untuk peramalan jika model memenuhi syarat.

3. Model Autoregressive Integrated Moving Average (ARIMA)

3.1. Model Autoregressive (AR)

Model *AR(p)* adalah model dimana X_t merupakan fungsi dari data di masa yang lalu, yakni $t-1, t-2, \dots, t-p$. Persamaan AR diberikan oleh

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + e_t, \quad (3.1)$$

3.2. Model Moving Average (MA)

Model *MA(q)* adalah model untuk memprediksi X_t sebagai fungsi dari kesalahan prediksi di masa lalu (past forecast error) dalam memprediksi X_t . Persamaan *MA* diberikan oleh:

$$X_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}. \quad (3.2)$$

3.3. Model ARIMA

Model *ARIMA* dilakukan pada data stasioner atau data yang didifferencing sehingga data telah stasioner. Secara umum, model *ARIMA* dinotasikan sebagai *ARIMA(p, d, q)*. Model ini merupakan gabungan dari model *ARMA(p, q)* dan proses *differencing*, yaitu

$$\phi_p(B)(1-B)^d X_t = \theta_q(B)e_t. \quad (3.3)$$

4. Metode Seasonal Autoregressive Integrated Moving Average (SARIMA)

4.1. Proses Moving Average(MA) Musiman

Bentuk umum dari proses *Moving Average* musiman periode S tingkat Q atau $MA(Q)^S$ didefinisikan sebagai berikut.

$$X_t = e_t - \Theta_1 e_{t-S} - \Theta_2 e_{t-2S} - \dots - \Theta_Q e_{t-QS}, \quad (4.1)$$

dimana e_t bersifat saling bebas terhadap X_{t-1}, X_{t-2}, \dots yang berdistribusi normal dengan *mean* 0 dan *varian* σ^2 .

Sebagai contoh dari model $MA(Q)^S$ akan dijelaskan dalam model $MA(1)^{12}$. Suatu proses X_t dikatakan mengikuti $MA(1)^{12}$ jika X_t mengikuti model

$$X_t = e_t - \Theta_1 e_{t-12}. \tag{4.2}$$

4.2. Proses Autoregressive(AR) Musiman

Bentuk umum dari proses Autoregressive musiman periode S tingkat P atau $AR(P)^S$ didefinisikan sebagai

$$X_t = \Phi_1 X_{t-S} + \Phi_2 X_{t-2S} + \dots + \Phi_P X_{t-PS} + e_t. \tag{4.3}$$

dimana e_t bersifat saling bebas X_{t-1}, X_{t-2}, \dots yang berdistribusi normal dengan *mean* 0 dan *varian* σ^2 .

Sebagai contoh dari model $AR(P)^S$ akan dijelaskan dalam model $AR(1)^{12}$. Suatu proses X_t dikatakan mengikuti $AR(1)^{12}$ jika X_t mengikuti model

$$X_t = \Phi_1 X_{t-12} + e_t. \tag{4.4}$$

4.3. Model Seasonal ARIMA

Musiman adalah kecenderungan mengulangi pola tingkah gerak dalam periode musim, biasanya satu tahun untuk data bulanan. Model *ARIMA* Musiman merupakan model *ARIMA* yang digunakan untuk menyelesaikan *time series* musiman yang terdiri dari dua bagian, yaitu bagian tidak musiman (non-musiman) dan bagian musiman. Bagian non-musiman dari metode ini adalah model *ARIMA*.

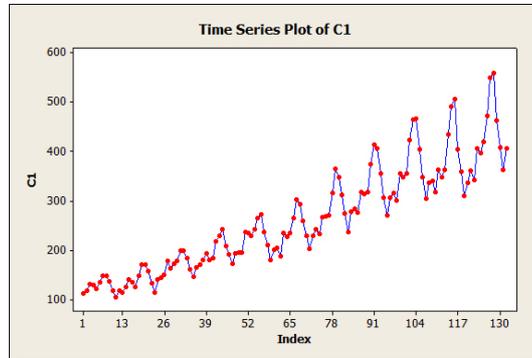
Secara umum bentuk model *ARIMA* musiman atau *ARIMA* $(p, d, q)(P, Q, S)^S$ adalah:

$$\phi_p(B)\Phi_P(B^S)(1 - B)^d(1 - B^S)^D X_t = \theta_q(B)\Theta_Q(B^S)e_t, \tag{4.5}$$

5. Peramalan Data Deret Waktu Musiman dengan Menggunakan Metode *Seasonal Autoregressive Integrated Moving Average (SARIMA)*

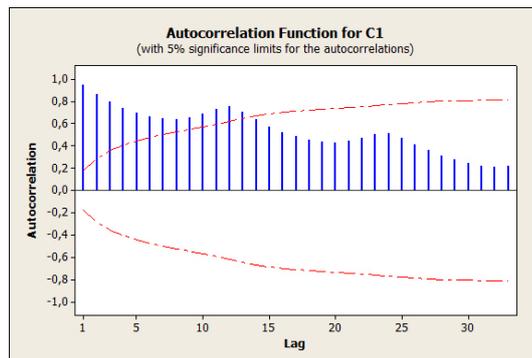
Data yang digunakan dalam penerapan metode ini adalah data *AirPassengers* (dalam ribuan) perbulan dari suatu maskapai penerbangan pada periode 1949 – 1959. Langkah pertama yang dilakukan adalah proses identifikasi model. Proses identifikasi model pertama kali diuji apakah data stasioner atau tidak yaitu dengan melihat plot data asli serta ACF dan PACF dari data asli. *Time series* plot untuk data tersebut adalah:

Gambar 1 memperlihatkan bahwa data dipengaruhi pola *trend* sekaligus pola musiman karena plot menunjukkan fluktuasi meningkat, yaitu gerakan dari kiri bawah ke kanan atas pada grafik dan berulang pada bulan tertentu Gambar 1 memperlihatkan bahwa pola musiman lebih kuat daripada pola *trend* sehingga pola *trend* tampak tidak begitu jelas. Gambar 2 dan Gambar 3 memperlihatkan terjadi autokorelasi pada data, yaitu adanya bar yang melebihi garis putus-putus



Gambar 1. Time Series dan AirPassengers

sehingga data tidak stasioner. Oleh karena itu, data *AirPassengers* perlu dilakukan *differencing* untuk menstasionerkan data tersebut.

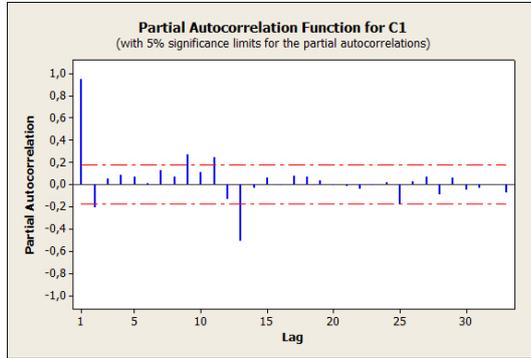


Gambar 2. Grafik ACF AirPassengers

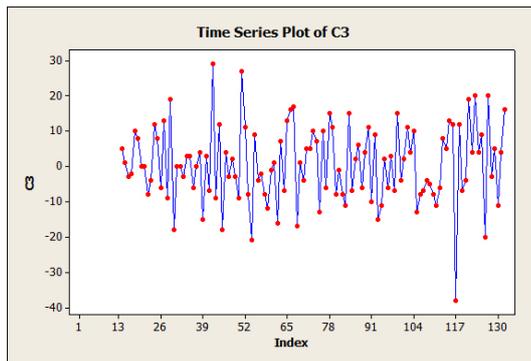
Gambar 4 merupakan plot data *AirPassengers* setelah data dilakukan *differencing* pertama dan musiman lag 12. Dari plot tersebut terlihat bahwa data telah stasioner dalam rata-rata dan varians setelah *differencing* pertama dan musiman lag 12 karena fluktuasi datanya horizontal sepanjang sumbu waktu dan berarti nilai $d = 1, D = 12$.

Setelah dilakukan *differencing* pertama dan musiman lag 12, tahap selanjutnya adalah pendugaan parameter model dengan menggunakan cara mencoba-coba (*trial and error*) yaitu menguji beberapa nilai yang berbeda. Pendugaan parameter dapat dilakukan dengan melihat grafik ACF dan PACF data setelah dilakukan *differencing* pertama dan musiman lag 12. Gambar 5 dan Gambar 6 berikut merupakan grafik dari ACF dan PACF setelah dilakukan *differencing* pertama dan musiman lag 12.

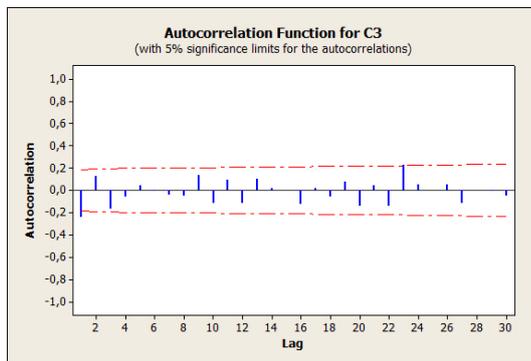
Dari Gambar 5 dan Gambar 6 akan didapat beberapa kandidat model *SARIMA*



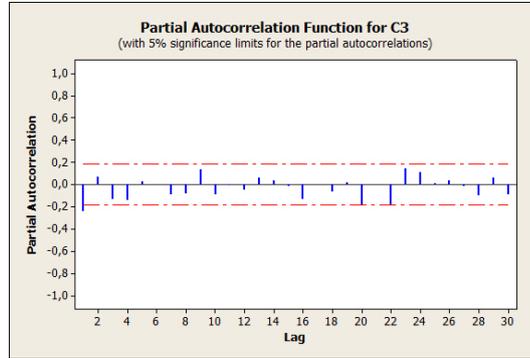
Gambar 3. Grafik *PACF AirPassengers*



Gambar 4. Plot data *AirPassengers* setelah dilakukan *differencing* pertama dan musiman lag 12



Gambar 5. Grafik *ACF AirPassengers* setelah dilakukan *differencing* pertama dan musiman lag 12



Gambar 6. Grafik *PACF AirPassangers* setelah dilakukan *differencing* pertama dan musiman lag 12

dengan melihat bar yang melebihi garis putus-putus. Pada grafik *ACF* terlihat bahwa pada bar pertama garis vertikal melewati garis putus-putus begitu juga dengan grafik *PACF* pada bar pertama garis vertikal juga melewati garis putus-putus. Dengan demikian diperoleh beberapa kandidat model *SARIMA* data *AirPassengers* yang diperlihatkan pada Tabel 1.

Tabel 1. Kandidat Model *SARIMA* Data *AirPassengers* dengan Metode *Seasonal ARIMA*

| Model | MSE | Keterangan |
|--------------------------------|-------|----------------------------|
| $ARIMA(1, 1, 1)(0, 1, 0)^{12}$ | 110.4 | AR(1) dan MA(1) signifikan |
| $ARIMA(1, 1, 0)(0, 1, 0)^{12}$ | 110.3 | AR(1) signifikan |
| $ARIMA(0, 1, 1)(0, 1, 0)^{12}$ | 111.0 | MA(1) signifikan |

Dari tabel di atas dapat disimpulkan bahwa model *SARIMA* yang terbaik untuk data *AirPassengers* adalah $ARIMA(1, 1, 0)(0, 1, 0)^{12}$ karena memiliki nilai MSD terkecil yaitu 110.3. Berdasarkan Gambar 7 AR(1) signifikan karena mempunyai nilai $p < 0.05$. Hal ini menunjukkan bahwa model dapat digunakan.

Jadi model *Seasonal ARIMA* untuk data *AirPassengers* adalah

$$[1 - (-0.2473B)(1 - B)(1 - B^{12})]X_t = e_t,$$

dengan model peramalannya adalah

$$\begin{aligned} e_t &= (1 - (-0.2473B))(1 - B)(1 - B^{12})X_t \\ &= (1 + 0.2473B)(1 - B^{12} - B + B^{13})X_t \\ &= (1 - B^{12} - B + B^{13} + 0.2473B - 0.2473B^{13} - 0.2473B^2 + 0.2473B^{14})X_t \\ &= (1 - 0.7527B - 0.2473B^2 - B^{12} + 0.7527B^{13} + 0.2473B^{14})X_t. \end{aligned}$$

```

Final Estimates of Parameters

Type      Coef  SE Coef   T      P
AR 1     -0,2473  0,0904  -2,74  0,007
Constant  0,6398  0,9628   0,66  0,508

Differencing: 1 regular, 1 seasonal of order 12
Number of observations: Original series 132, after differencing 119
Residuals:   SS = 12906,4 (backforecasts excluded)
              MS = 110,3  DF = 117

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

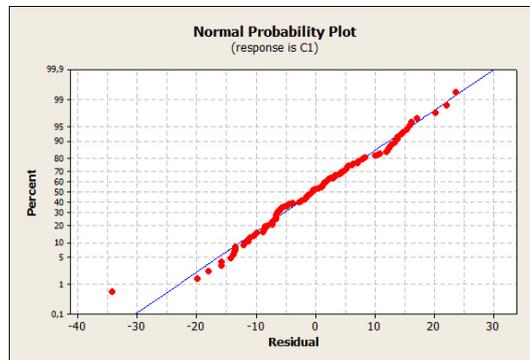
Lag       12    24    36    48
Chi-Square 9,3  27,3  34,9  50,7
DF        10    22    34    46
P-Value   0,502  0,200  0,423  0,295
    
```

Gambar 7. Hasil analisis data *PACF AirPassengers* dengan metode *Seasonal ARIMA*

Jadi

$$X_t = 0.7527X_{t-1} + 0.2473X_{t-2} + X_{t-12} - 0.7527X_{t-13} - 0.2473X_{t-14} + e_t. \quad (5.1)$$

Selanjutnya dilakukan proses diagnostik untuk melihat tingkat kesalahan model, yaitu dengan melihat kenormalan distribusi residunya.



Gambar 8. Grafik *Normal Probability Plot* residu dari data *PACF AirPassengers*

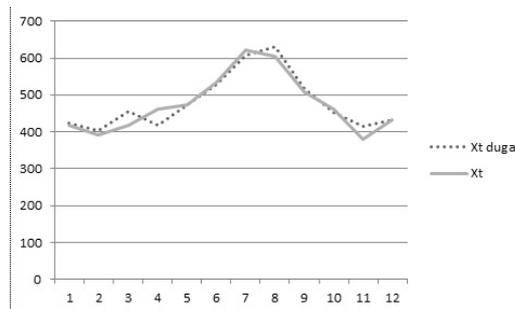
Gambar 8 memperlihatkan residu mengikuti garis diagonal, yang berarti residu berdistribusi normal. Karena residu bersifat random dan berdistribusi normal, maka residu memenuhi asumsi *white noise*. Dari persamaan (4.2.1) didapatkan hasil peramalan data *AirPassengers* untuk tahun 1960 yang diberikan pada Tabel 2.

Gambar 9 dan Gambar 10 berikut memperlihatkan grafik perbandingan data hasil peramalan dengan data asli dan plot dari hasil peramalan untuk tahun 1960.

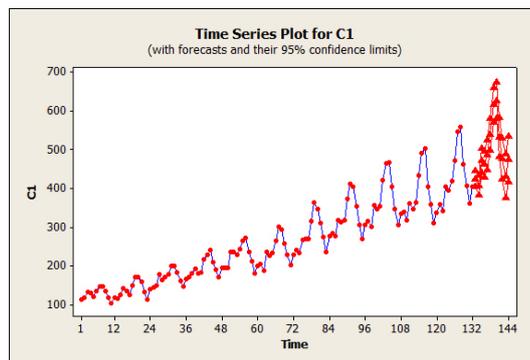
Pada Gambar 9 terlihat bahwa grafik data peramalan dengan data aktual sangat berdekatan. Yang berarti bahwa hasil data peramalan hampir mendekati data aktual. Sedangkan pada Gambar 10 jika data hasil peramalan digabungkan dengan data sebelumnya terlihat bahwa plot data hasil peramalan mengikuti pola dari data sebelumnya.

Tabel 2. Hasil Peramalan Data *AirPassengers* dengan Metode *Seasonal ARIMA* Tahun 1960

| Bulan | Ramalan | Aktual |
|-----------|---------|--------|
| Januari | 424.04 | 413 |
| Februari | 401.72 | 391 |
| Maret | 456.98 | 419 |
| April | 417.9 | 461 |
| Mei | 472.14 | 472 |
| Juni | 527.21 | 535 |
| Juli | 608.28 | 622 |
| Agustus | 630.24 | 606 |
| September | 516.68 | 508 |
| Oktober | 452.49 | 461 |
| November | 413.77 | 380 |
| Desember | 431.9 | 432 |



Gambar 9. Grafik Perbandingan Data Hasil Peramalan dengan Data Asli



Gambar 10. Plot Hasil Peramalan data *PACF AirPassengers* Tahun 1960

6. Kesimpulan

Dari penjelasan diatas dapat diambil kesimpulan bahwa metode *SARIMA* dapat dipakai untuk data yang bersifat musiman yang dapat memberikan hasil peramalan

yang tidak jauh berbeda dengan data aktual.

7. Ucapan Terima kasih

Penulis mengucapkan terima kasih kepada Ibu Dr. Maiyastri, Bapak Dr. Mahdihan Syafwan, Ibu Hazmira Yozza, M.Si, Ibu Izzati Rahmi HG, M.Si, Ibu Dr. Lyra Yulianti, dan ibu Dr. Ferra Yanuar yang telah memberikan masukan dan saran sehingga paper ini dapat diselesaikan dengan baik.

Daftar Pustaka

- [1] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994). *Time Series Analysis: Forecasting and Control Third Edition*. New Jersey: Pearson Prentice Hall
- [2] Brockwell, P.J and Davis, R.A., 1996, *Introduction to Time Series and Forecasting*. Spinger-Verlag New York
- [3] Chatfield, C. (2000). *Time-Series Forecasting*. Boca Raton: Chapman and Hall/CRC
- [4] Cryer, J.D.,. 1986. *Time Series Analysis*. Duxbury Press, Boston.
- [5] Hanke, J.E. and Wichern, D.W. (2005). *Business Forecasting* Eight Edition. New Jersey: Pearson Prentice Hall
- [6] Makridakis, S., Wheelwright, S.C. and McGee, V.E. (1999). *Metode dan Aplikasi Peramalan Jilid 1* (Ir. Untung Sus Ardiyanto, M.Sc. dan Ir. Abdul Basith, M.Sc. Terjemahan). Edisi Kedua. Jakarta: Penerbit Erlangga