

PENERAPAN ANALISIS DERET WAKTU DAN METODE PERAMALAN PADA DATA KUNJUNGAN PASIEN DI KLINIK PRATAMA ATMA JAYA CISAUK TAHUN 2012 – 2016

Ignatius Danny Pattirajawane, Siti Khodijah, Erfen G. Suwangto
Lembaga Pengembangan Jaringan Klinik Pratama Atma Jaya

email korespondensi: dannyradja@yahoo.co.id

ABSTRAK

Daerah urban merupakan pusat pertumbuhan ekonomi, penduduk, dan juga perubahan gaya hidup, termasuk gaya hidup yang berkaitan dengan masalah-masalah kesehatan. Masalah kesehatan masyarakat dapat muncul karena perubahan lingkungan, sanitasi, gaya hidup urban, tingginya *stressor* dan sebagainya. Berdasarkan data kunjungan masyarakat pada fasilitas layanan kesehatan primer, penelitian ini bermaksud mengidentifikasi pola penyakit yang dikeluhkan masyarakat di daerah urban. Sampel data kunjungan pasien bulanan diambil pada Klinik Pratama Atma Jaya di Cisauk, Tangerang, mulai bulan Juni 2012 hingga Mei 2016. Data dalam plot waktu dengan jumlah total kunjungan pasien dan jumlah infeksi saluran nafas atas (ISPA), dispepsia, dan hipertensi. Pemodelan data menggunakan ARIMA atau penghalusan eksponensial. Setelah proses estimasi parameter dan pemeriksaan kelayakan model diperoleh hasil bahwa data deret waktu jumlah kunjungan pasien dan jumlah penyakit terbanyak (ISPA, dispepsia dan hipertensi) dapat dimodelkan oleh kedua metode ini. Data lebih baik dimodelkan oleh metode ARIMA daripada metode penghalusan eksponensial meskipun metode ini memiliki ramalan yang lebih dekat dengan data aktual untuk taksiran satu periode ke depan.

Kata Kunci: Kesehatan Urban, ARIMA, Penghalusan Eksponensial

PENDAHULUAN

Salah satu fenomena di Asia Tenggara tahun 1950 – 2000 adalah terbentuknya kota-kota primat (*primate cities*) yang ditandai dengan konsentrasi *massif* pembangunan dan penduduk di hanya beberapa kota (Silver, 2008). Kota-kota tersebut umumnya merupakan ibukota negara dan dengan populasi sekitar 8 juta penduduk atau lebih, disebut juga *megacities*. Pada tahun 1990-an Jakarta merupakan megacity terbesar di Asia Tenggara melampaui Manila dan Bangkok.

Hal yang menarik dari perkembangan urbanisasi di Jakarta adalah pertumbuhan wilayah-wilayah di pinggiran sekitar Jakarta. Terdapat pergeseran penduduk yang cukup besar dalam pertumbuhan penduduk antara DKI dan Botabek (Bogor, Tangerang, Bekasi). Pertumbuhan penduduk yang tinggal di wilayah inti Jakarta mengalami perlambatan dari 3,8% menjadi 2,4% sedangkan daerah Botabek yang berbatasan langsung dengan DKI tumbuh (dari 4%) hingga 9% (Dorleans, 2007). Di samping itu bermunculan kota-kota satelit di sekitar Bintaro, Bumi Serpong Damai, Sentul, Cikarang, dan Karawaci.

Fenomena pergeseran dan pertumbuhan penduduk, serta pertumbuhan ekonomi, disertai juga dengan perubahan pola hidup masyarakat urban di sekitar Jakarta terutama yang terkait dengan masalah kesehatan. Masalah-masalah kesehatan dapat muncul karena adanya perubahan lingkungan, masalah sanitasi, gaya

hidup urban, tingginya *stressor*, dan sebagainya. Selain oleh indikator-indikator kesehatan umum seperti angka kematian bayi, angka kematian ibu, cakupan imunisasi, masalah kesehatan masyarakat pada suatu daerah dapat dilihat dari jenis keluhan atau diagnosis penyakit yang ditemukan di fasilitas kesehatan tingkat primer. Green (2004) menyebutkan secara umum memperkirakan dari 1000 penduduk, 800 di antaranya (80%) dikatakan memiliki keluhan kesehatan, 327 (32,7%) mencari pertolongan medis, dan 217 (21,7%) mengunjungi fasilitas pelayanan primer.

Menurut perkiraan Green di atas bahwa sekitar dua puluh persen populasi masyarakat mengunjungi fasilitas layanan primer, artinya gaya hidup berkaitan dengan kesehatan masyarakat diantaranya adalah fenomena kunjungan mereka ke fasilitas kesehatan. Penelitian ini bermaksud untuk mendeteksi pola masalah kesehatan (keluhan atau penyakit) yang ditemukan dari kunjungan pasien di fasilitas layanan primer. Fasilitas pelayanan primer yang diteliti adalah Klinik Pratama Atma Jaya Cisauk yang beroperasi sejak bulan Mei tahun 2012 dan letaknya berbatasan dengan kota satelit Bumi Serpong Damai. Pola penyakit yang muncul diidentifikasi berdasarkan perubahan atau *trend* jumlah penyakit setiap bulan pada fasilitas kesehatan layanan primer.

Salah satu metode statistik untuk mempelajari pola data dari waktu ke waktu adalah analisis deret waktu (*time series analysis*). Analisis deret waktu (*time series analysis*) merupakan metode statistik yang dipergunakan untuk menganalisis data-data yang tersaji berdasarkan urutan waktu. Pada awalnya analisis deret waktu berkembang pesat dalam bidang ekonomi, khususnya ekonometrika (Juanda & Junaidi, 2012; Green, 1993). Penelitian dalam bidang kesehatan dan kedokteran Choi dan Thacker pada tahun 1981 telah menggunakan analisis deret waktu untuk mengevaluasi angka mortalitas influenza (Trottier *et al*, 2006). Sejak itu telah banyak penggunaan analisis deret waktu pada bidang ini baik dalam menganalisis pola penyakit infeksi (Allard, 1998; Trottier *et al*, 2006) maupun non-infeksi (Fazekas, 2004).

Ada beberapa tujuan orang melakukan analisis deret waktu. Pertama, untuk mempelajari pola dari data yang disajikan berdasarkan urutan waktu dengan cara membuat model matematis; kedua, untuk melakukan peramalan (*forecasting*) guna mengantisipasi kejadian yang akan datang (Alt, Hung, Lap-Ming, 1998). Dalam bidang kesehatan dan kedokteran peramalan umumnya dapat digunakan untuk memprediksi jumlah kasus suatu penyakit dalam beberapa periode ke depan entah itu mingguan, bulanan atau tahunan. Sebelum peramalan dibuat, maka langkah penting yang perlu dilakukan adalah melakukan permodelan matematis terhadap data-data yang sudah dimiliki.

Metode yang sederhana dan cukup lama dipergunakan para praktisi peramalan hingga sekarang adalah metode penghalusan eksponensial. Pada peramalan dengan menggunakan metode ini, kita sudah memiliki sejumlah data historis terlebih dahulu yang digunakan dalam peramalan. Dalam perhitungan peramalan tersebut data tiap waktu memiliki bobot, dimana semakin tua datanya semakin menyusut bobot datanya secara multiplikatif, karena itu dinamakan penghalusan eksponensial (Makridakis, Wheelright, McGee, 1999; Box, Jenkins, Reinsel, 1994).

Metode analisis deret waktu lain yang cukup luas dipergunakan adalah metode Box-Jenkins. Pada tahun 70-an George E. P. Box dan Gwilym M. Jenkins mensistematisasi analisis deret waktu melalui pendekatan sangat terkenal yang disebut dengan nama mereka yakni metode Box-Jenkins atau ARIMA (*autoregressive integrated moving average*). Buku mereka *Time Series Analysis – Forecasting and Control* terbit pada tahun 1970 dan direvisi pada tahun 1976 dan tahun 1994 bersama penulis ketiga Gregory C. Reinsel, telah menjadi acuan otoritatif dalam analisis deret waktu ARIMA. Dua proses penting dalam deret waktu Box-Jenkins adalah rata-rata bergerak (*moving average* = MA) dan otoregresi (*autoregression* = AR). (Djauhari, 2015; Montgomery, Jennings, Kulahci, 2008).

Masalah penelitian ini adalah:

1. Bagaimana pola jumlah kunjungan dan jumlah penyakit pasien setiap bulan di Klinik Atma Jaya Penjaringan dari bulan Juni 2012 hingga bulan Mei 2016?
2. Bagaimana model matematis yang cocok terhadap data historis jumlah kunjungan pasien dan jumlah penyakit tersering dengan menggunakan metode ARIMA dan penghalusan eksponensial?
3. Berapa ramalan jumlah kunjungan pasien dan jumlah penyakit terbanyak untuk 1 bulan ke depan yakni bulan Juni 2016? Dan ramalan dengan metode mana lebih akurat?

Tujuan penelitian ini adalah:

1. Membuat dan menganalisis plot deret waktu dari data jumlah kunjungan dan jumlah penyakit pasien setiap bulan di Klinik Atma Jaya Penjaringan dari bulan Juni 2012 hingga bulan Mei 2016,
2. Memperoleh model matematis terbaik terhadap data historis jumlah kunjungan pasien dan jumlah penyakit tersering dengan menggunakan metode ARIMA dan penghalusan eksponensial dan mengevaluasinya secara statistik,
3. Menggunakan model matematis yang diperoleh untuk meramal jumlah kunjungan pasien dan jumlah penyakit terbanyak untuk 1 bulan ke depan yakni bulan Juni

2016 serta mengevaluasi keakuratan ramalan dari masing-masing metode melalui galat-galat yang dihasilkan.

METODE PENELITIAN

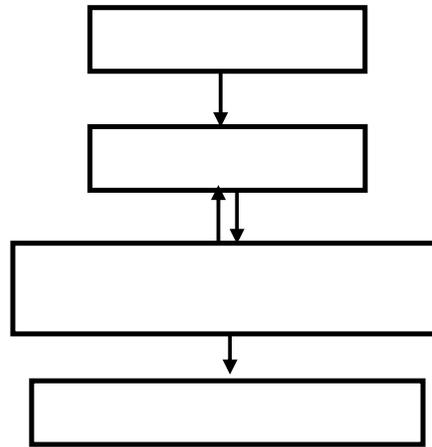
Penelitian ini merupakan penelitian berdasarkan data primer dimana sampel diambil dari kunjungan pasien Klinik Pratama Atma Jaya Cisauk Kabupaten Tangerang, Cisauk adalah wilayah urban yang berbatasan langsung dengan Kota Tangerang Selatan dan memiliki luas wilayah 26,91 Km² atau 2.80% dari luas keseluruhan Kabupaten Tangerang. Ia juga berbatasan dengan kota satelit Jakarta modern Bumi Serpong Damai (BSD).Kecamatan Cisauk memiliki jumlah penduduk sebanyak 76.622 Jiwa dengan komposisi jumlah penduduk laki-laki sebanyak 39.066 jiwa dan tercatat ada 37.556 jiwa penduduk perempuan (BPS Kabupaten Tangerang, 2015).

Penelitian ini bermaksud mempelajari pola jumlah kunjungan pasien yang berbobat serta penyakit-penyakit terbanyak per bulan dengan menggunakan analisis deret waktu. Data yang diambil adalah data kunjungan pasien Klinik Pratama Atma Jaya Cisauk mulai bulan Juni 2012 hingga bulan Mei 2016. Data yang diterima berupa data yang disajikan dalam aplikasi Microsoft Excel yang kemudian diolah dengan aplikasi SPSS 21 dan *Eview* 9. Pada tahap awal dilakukan *Cleaning* atau pengecekan kembali data yang dimasukkan untuk melihat kemungkinan-kemungkinan adanya kesalahan pengentrian, memastikan keseragaman istilah diagnosis penyakit yang akan diolah, kemudian memisahkan data pasien umum dan mengeluarkan data pasien gigi, selanjutnya menghitung total kunjungan pasien.

Dalam penelitian ini dikelompokkan tiga kelompok diagnosis penyakit terbanyak yang akan diteliti yakni dispepsia mencakup semua diagnosis kerja oleh dokter yang menyebutkan dispepsia dan gastritis; infeksi saluran nafas atas (ISPA)mencakup diagnosis kerja ISPA, pharingitis, tonsilopharingitis, dan tonsillitis; serta Hipertensi. Pada awalnya penelitian ini bermaksud untuk meneliti pola kasus demam yang mencakup keluhan demam atau infeksi yang diagnosis sebagai dengue atau tifoid, namun karena banyak nilai observasi yang nol kasus-kasus ini dikeluarkan dari penelitian ini.

Data yang telah diolah akan disajikan dalam plot waktu berdasarkan jumlah total kunjungan pasien dan jumlah kasus-kasus tersering seperti ISPA, Dispepsia dan Hipertensi. Setelah itu akan diperiksa korelogram dan uji akar satuan untuk menilai stasionaritas data deret waktu. Kemudian akan diestimasi parameter-parameter model dengan metode analisis deret waktu seperti ARIMA atau penghalusan eksponensial serta dilakukan pemeriksaan residu atas model.

Langkah-langkah melakukan permodelan dan peramalan deret waktu menurut Box-Jenkins (1994), Juanda dan Junaidi (2012) dan Rosadi (2012) secara garis besar dapat dibagi dalam empat langkah seperti pada bagan di bawah ini.



Gambar 1. Diagram langkah-langkah permodelan dan peramalan deret waktu

Pada langkah pertama, setelah data bersih (*clean*), dilakukan identifikasi model, dengan cara menginspeksi plot data, menguji stasionaritas data dan memeriksa korelogram yakni menilai pola plot nilai koefisien fungsi otokorelasi (ACF) dan koefisien fungsi otokorelasi parsial (PACF). Langkah ini penting untuk mengarahkan perkiraan kita akan model matematis yang cocok. Dalam menginspeksi penting untuk mengidentifikasi adanya unsur trend atau unsur musiman dalam plot deret waktu. Kedua unsur tersebut membuat deret waktu menjadi non-stasioner.

Unsur trend dapat diidentifikasi ketika plot deret waktu memperlihatkan pola peningkatan atau penurunan (umumnya linear). Sedangkan pola musiman dapat diidentifikasi ketika plot deret waktu memperlihatkan ada pengulangan pola yang periodik. Terdapat uji statistik yang dapat membantu menyimpulkan apakah deret waktu yang dimaksud stasioner atau tidak yakni uji akar unit. Metode uji akar unit yang paling sering digunakan dan dipakai dalam penelitian ini adalah Augmented Dickey-Fuller (ADF).

Selanjutnya dilakukan pemeriksaan pola nilai koefisien fungsi otokorelasi (ACF) dan koefisien fungsi otokorelasi parsial (PACF) deret waktu yang akan dimodelkan. Hal ini bermanfaat, terutama saat menggunakan metode ARIMA, untuk model matematisnya apakah berjenis proses otoregresi (*autoregressive / AR*), rata-rata bergerak (*moving average / MA*) atau kombinasi keduanya (ARMA).

Deret waktu yang stasioner memiliki pola nilai-nilai ACF dan PACF yang berada dalam nilai batas atas dan batas bawah semuanya. Sedangkan deret waktu yang non-stasioner pada korelogram ACF memperlihatkan pola peluruhan yang lambat dan pada

korelogram PACF memperlihatkan nilai yang sangat menonjol pada lag 1 jauh melampaui nilai-nilai batasnya dan langsung teredam cepat di dalam nilai-nilai batasnya mulai dari lag 2. Box-Jenkins (1994), Makridakis *et al* (1999) dan Montgomery *et al*(2008) dalam buku mereka memuat pedoman secara visual pola ACF dan PACF untuk membantu para peneliti menentukan model matematis ARIMA.

Dalam analisis deret waktu, secara khusus dengan metode ARIMA, sebelum deret waktu dimodelkan secara matematis pertama-tama penting untuk menstasionerkan data deret waktu terlebih dahulu. Dalam proses ini mungkin juga dapat dipertimbangkan transformasi data yang dapat menggunakan fungsi transformasi yang sering digunakan seperti akar kuadrat atau logaritma natural. Penstasioneran data biasanya dilakukan dengan membuat selisih (*difference*) baik dalam jeda (*lag*) pertama atau hingga maksimum kedua.

Pada permodelan matematis dengan metode Box-Jenkins, ARMA (p,q) dengan p menunjukkan derajat proses otoregresif AR(p) dan q menunjukkan derajat rata-rata bergerak MA(q), memiliki bentuk umum sebagai berikut:

$$X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \dots\dots(1)$$

Pada persamaan (1) di atas δ adalah parameter konstanta, $\theta_1, \theta_2, \dots, \theta_q$ adalah parameter-parameter koefisien otoregresi, $\phi_1, \phi_2, \dots, \phi_p$ merupakan parameter-parameter koefisien galat dari rata-rata bergerak. Semua parameter-parameter tersebut di atas harus diestimasi.

Bila deret waktu tidak stasioner, maka dimungkinkan dilakukan operasi selisih d yang umumnya maksimum dilakukan hingga tingkat kedua. Hal ini mengumumkan metode Box-Jenkins ARIMA (p,d,q) dengan d menunjukkan derajat selisih yang dilakukan yang dapat mengambil nilai $d = 0, 1, \text{ dan } 2$. Untuk $d = 1$, maka kita akan memiliki bentuk umum ARIMA (p,1,q) sebagai berikut

$$Z_t = \delta + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \dots(2)$$

dengan $Z_t = X_t - X_{t-1}, Z_{t-1} = X_{t-1} - X_{t-2}, \dots, Z_{t-p} = X_{t-p} - X_{t-(p+1)}$ merupakan variabel-variabel baru yang diperoleh dari operasi selisih satu langkah dengan nilai sebelumnya.

Untuk permodelan matematis dengan metode penghalusan eksponensial setelah menginspeksi plot deret waktu, maka kita memperkirakan persamaan matematis dengan mengidentifikasi adanya unsur trend atau musiman dan apakah hubungannya aditif atau multiplikatif. Hubungan aditif diperkirakan ketika plot deret

waktu berfluktuasi dengan amplitudo yang tetap, sedangkan hubungan multiplikatif ketika amplitudo fluktuasi plot deret waktunya berubah teratur.

Untuk kemudahan permodelan matematis deret waktu dengan metode penghalusan eksponensial maka kita dapat memanfaatkan taksonomi yang dibuat Pegels dengan memperhatikan unsur trend, unsur musiman dan hubungan multiplikatif atau aditif. Secara sistematis permodelan tersebut dapat disajikan dalam bentuk matriks berikut:

Tabel 1. Permodelan Matematis dengan Metode Penghalusan Eksponensial Klasifikasi Pegels

	Unsur Musiman			
	1 (tidak ada)	2 (aditif)	3 (multiplikatif)	
Unsur Trend	A (tidak ada)	$P = X_t$	$P = X_t - C_{t-L}$	$P = X_t/D_{t-L}$
		$Q = S_{t-1}$	$Q = S_{t-1}$	$Q = S_{t-1}$
	B (aditif)	$P = X_t$	$P = X_t - C_{t-L}$	$P = X_t/D_{t-L}$
		$Q = S_{t-1} + A_{t-1}$	$Q = S_{t-1} + A_{t-1}$	$Q = S_{t-1} + A_{t-1}$
	C (multiplikatif)	$P = X_t$	$P = X_t - C_{t-L}$	$P = X_t/D_{t-L}$
		$Q = S_{t-1}B_{t-1}$	$Q = S_{t-1}B_{t-1}$	$Q = S_{t-1}B_{t-1}$

Sumber: Makridakis *et al* (1999)

Pada Tabel 1 di atas, X_t adalah nilai-nilai data aktual, S_t adalah nilai-nilai data yang dihaluskan dan L adalah periode musiman. Beberapa hubungan yang perlu dirinci di antaranya: $S_t = \alpha P + (1 - \alpha)Q$ yang merupakan data penghalusan, $A_t = \beta(S_t - S_{t-1}) + (1 - \beta)A_{t-1}$ yang merupakan unsur trend aditif, $B_t = \gamma(S_t/S_{t-1}) + (1 - \gamma)B_{t-1}$ yang merupakan unsur trend multiplikatif, $C_t = \delta(X_t - S_t) + (1 - \delta)C_{t-L}$ yang merupakan unsur musiman aditif, dan $D_t = \theta(X_t/S_t) + (1 - \theta)D_{t-L}$ yang merupakan unsur musiman aditif. Parameter-parameter $\alpha, \beta, \gamma, \delta, \theta$ adalah parameter-parameter yang perlu diestimasi dan bernilai di antara 0 dan 1.

Langkah kedua dalam permodelan adalah mengestimasi parameter. Dua metode tersering yang dilakukan dalam melakukan estimasi parameter adalah metode *maximum likelihood* dan untuk model-model linear metode kuadrat terkecil [Box-Jenkins (1994), Greene (1993), Montgomery *et al* (2008), Soejoeti (2006), Djauhari (2007)]. Pada metode ARIMA parameter-parameter yang diestimasi perlu dilakukan uji signifikansinya dengan statistik t pada tingkat signifikansi umumnya 0,05. Bila ditemukan taksir koefisien yang tidak signifikan secara statistik maka parameter tersebut harus dikeluarkan dari persamaan model.

Namun untuk model matematis menggunakan metode penghalusan eksponensial parameter-parameternya harus dipilih, sering kali dengan cara coba-coba. Meski demikian beberapa ahli menganjurkan pemilihan untuk nilai parameter α berada di antara 0,01 dan 0,3 (Djauhari, 2007). Sedangkan Makridakis *et al* (1999)

mengajukan α berada dalam interval tertutup 0,1 dan 0,2, namun Montgomery *et al* (2008) menganjurkan di antara interval tertutup 0,1 dan 0,4.

Pada penelitian ini seluruh perhitungan estimasi-estimasi parameter dilakukan dengan menggunakan aplikasi komputer SPSS 21. Pada metode ARIMA yang paling mengikuti langkah-langkah penelitian seperti pada diagram di Gambar 1 estimasi parameter dipilih dengan mempertimbangkan signifikansi statistiknya. Oleh sebab itu permodelannya tidak menggunakan fasilitas *expert modeler* dari SPSS 21 yang melakukan pemilihan otomatis model yang dianggap oleh aplikasi terbaik. Sedangkan metode penghalusan eksponensial diperoleh dari menggunakan fasilitas *expert modeler* SPSS 21 dan nilai-nilai parameter yang diestimasi tidak diperhatikan signifikansi statistiknya.

Langkah ketiga setelah dilakukan estimasi parameter adalah melakukan pemeriksaan / evaluasi diagnostik. Pada langkah ini yang dilakukan adalah menilai pola korelogram ACF dan PACF residu permodelan [Montgomery *et al* (2008), Juanda-Junaidi (2012), Rosadi (2012)]. Residu adalah selisih antara nilai observasi aktual dengan nilai prediksi dari modelnya. Model yang baik memiliki pola korelogram ACF dan PACF residu yang berada di antara nilai batas atas dan batas bawahnya. Pola korelogram residu yang demikian mengarahkan kita kepada kesimpulan bahwa tidak terdapat korelasi residu antar lag sehingga residunya terdistribusi normal (random).

Langkah terakhir adalah menggunakan model dalam peramalan. Untuk peramalan kita menggunakan model persamaan (2) dengan parameter-paramater yang sebelumnya telah diestimasi, namun variabel independen waktunya mulai dihitung setelah waktu data observasi terakhir. Jika n adalah jumlah observasi, maka ramalan diperuntukan untuk memprediksi nilai yang keluar mulai waktu $n + 1, n + 2, \dots$ dan seterusnya. Persamaan ramalan pada model $ARIMA(p, 1, q)$ tanpa musiman untuk waktu ke- $n + \tau$ adalah

$$Z_{n+\tau} = \delta + \phi_1 Z_{n+\tau-1} + \phi_2 Z_{n+\tau-2} + \dots + \phi_p Z_{n+\tau-p} + \varepsilon_{n+\tau} - \theta_1 \varepsilon_{n+\tau-1} - \theta_2 \varepsilon_{n+\tau-2} - \dots - \theta_q \varepsilon_{n+\tau-q} \dots \dots \dots (2a)$$

Sedangkan untuk peramalan menggunakan metode penghalusan eksponensial untuk m langkah ke depan dapat dilihat pada Tabel 2.

Tabel 2. Nilai ramalan dengan metode penghalusan eksponensial m langkah ke depan

	Unsur Musiman			
		1 (tidak ada)	2 (aditif)	3 (multiplikatif)
Unsur Trend	A (tidak ada)	S_t	$S_t - C_{t-L+m}$	$P = S_t D_{t-L+m}$
	B (aditif)	$S_t + mA_t$	$X_t + mA_{t-1} + C_{t-L+m}$	$(S_t + mA_t)D_{t-L+m}$
	C (multiplikatif)	$P = S_t B_t^m$	$P = S_t B_t^m + C_{t-L+m}$	$S_t D_{t-L+m} B_t^m$

Ramalan dengan menggunakan model matematis perlu dinilai keakuratannya. Dalam penggunaan dua atau lebih model kita memerlukan kriteria untuk membandingkan keakuratan ramalan model yang satu dengan yang lain. Kriteria penilaian tersebut ialah galat-galat (*errors*) yang dihasilkan. Tentu semakin kecil galat-galatnya semakin akurat ramalannya.

Galat ramalan diperoleh dari selisih nilai observasi dengan nilai ramalannya. Galat ramalan yang sering digunakan adalah galat ramalan 1 langkah ke depan yakni memeriksa ramalan yang dibuat pada waktu ke-($t - 1$) dengan nilai observasinya pada waktu ke- t

$$e_t(1) = X_t - \hat{X}_t(t - 1) \dots\dots\dots(3a)$$

$e_t(1)$ adalah simbol untuk galat ramalan 1 langkah ke depan pada waktu ke- t , X_t adalah nilai observasi pada waktu ke- t , $\hat{X}_t(t - 1)$ adalah nilai ramalan waktu ke- t yang dibuat pada waktu ke-($t - 1$). Bila ramalan dibuat untuk beberapa waktu ke depan kita dapat menulis galat ramalan sebagai e_t .

Ada beberapa ukuran untuk menilai kumpulan galat ramalan apabila ramalannya dibuat untuk beberapa waktu ke depan. Yang pertama adalah galat rata-rata (*mean error* = ME)

$$ME = \frac{1}{n} \sum_{t=1}^n e_t, t = 1, 2, \dots, n \dots\dots\dots(4a)$$

Pada ramalan yang baik kita berekspektasi nilai ME mendekati nol. Bila ME merupakan rata-rata galat, untuk melihat variabilitas galatnya dapat digunakan rata-rata deviasi atau galat mutlak (*mean absolute deviation* = MAD atau *mean absolute error* = MAE) dan galat rata-rata kuadrat (*mean squared error* = MSE)

$$MAD = \frac{1}{n} \sum_{t=1}^n |e_t|, t = 1, 2, \dots, n \dots\dots\dots(4b)$$

$$MSE = \frac{1}{n} \sum_{t=1}^n [e_t]^2, t = 1, 2, \dots, n \dots\dots\dots(4c)$$

Baik pada ME, MAD dan MSE nilainya dipengaruhi skala dari variabel yang diteliti. Sebagai contoh jumlah kasus akan berbeda skala ukurannya dengan suhu. Untuk mengatasi hal ini dikembangkan ukuran-ukuran relatif seperti galat persentase rata-rata (*mean percent error* = MPE) dan galat persentase rata-rata mutlak (*mean absolute percent error* = MAPE)

$$MPE = \frac{1}{n} \sum_{t=1}^n r e_t, t = 1, 2, \dots, n \dots\dots\dots(5a)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n |r e_t|, t = 1, 2, \dots, n \dots\dots\dots(5b)$$

Dalam hal ini $re_t(1)$ merupakan persentase dari ME sehingga persamaan (4a) dapat tulis ulang menjadi galat ramalan relatif

$$re_t = \left(\frac{X_t - \hat{X}_t}{X_t} \right) 100 = \left(\frac{e_t}{X_t} \right) 100 \dots\dots\dots(3b)$$

Metode peramalan yang dipilih adalah metode peramalan yang memberikan galat-galat ramalan yang terkecil. Dalam hal ini dua ukuran penting yang dapat digunakan apabila kita ingin memilih metode peramalan yang terbaik dari dua atau beberapa metode yang tersedia adalah kriteria informasi Akaike (AIC) atau kriteria informasi Schwarz/Bayesian (SIC atau BIC) (Makridakis *et al*, 1999). Kedua ukuran tersebut dapat dinyatakan secara matematis dengan persamaan sebagai berikut

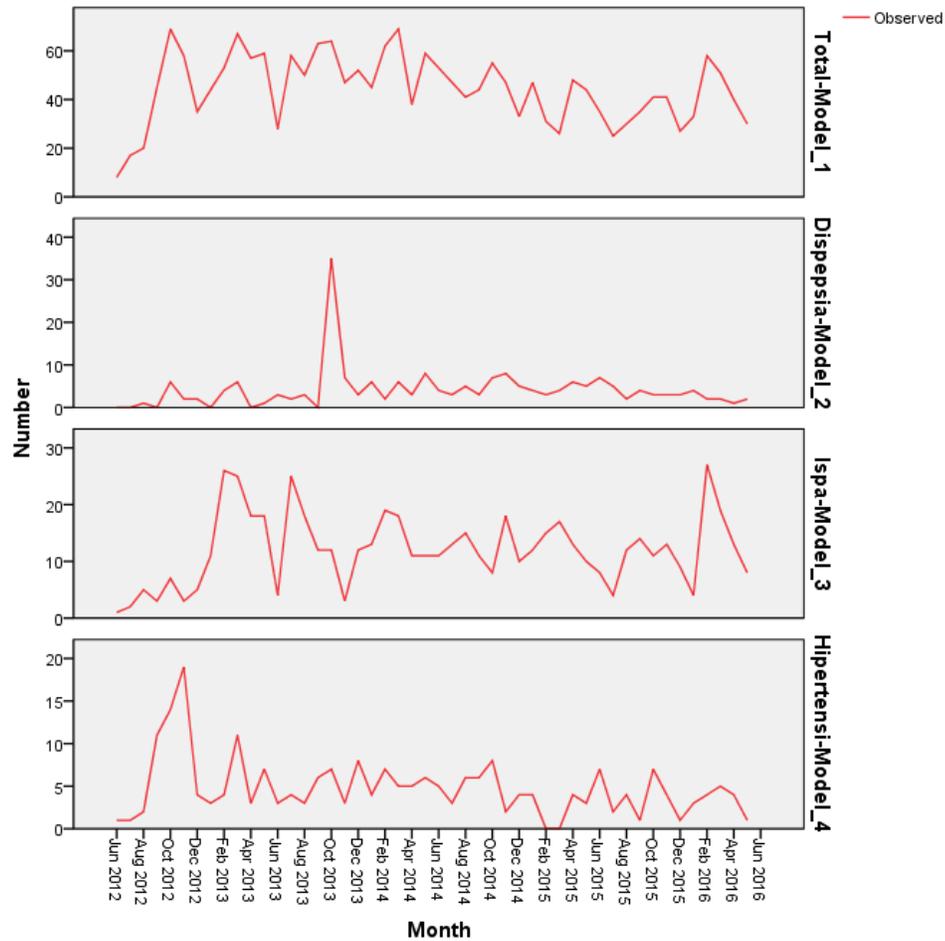
$$AIC = \ln \left(\frac{\sum_{t=1}^T e_t}{T} \right) + \frac{2p}{T} \dots\dots\dots(6a)$$

$$SIC (BIC) = \ln \left(\frac{\sum_{t=1}^T e_t}{T} \right) + \frac{p \ln T}{T} \dots\dots\dots(6b)$$

Dalam kedua persamaan di atas T adalah banyaknya nilai peramalan yang digunakan dan p adalah banyaknya parameter yang digunakan. Metode peramalan yang dipilih adalah metode peramalan yang memberikan nilai AIC atau BIC yang terkecil. Baik AIC atau BIC mempertimbangkan jumlah parameter (p). Metode peramalan yang menghasilkan banyak parameter akan menerima “hukuman” (*penalized*) berupa peningkatan nilai kriteria informasinya sehingga memperburuk nilainya terhadap metode peramalan lain yang menggunakan parameter yang lebih sedikit.

HASIL DAN PEMBAHASAN

Langkah pertama melakukan permodelan dan peramalan adalah menampilkan plot data untuk menilai secara visual adanya kecenderungan peningkatan atau penurunan (trend) dan pola musiman pada data. Visualisasi plot data ini sangat membantu dalam menerka struktur persamaan matematis baik dalam model penghalusan eksponensial, maupun ARIMA.



Gambar 2. Plot-plot data jumlah kunjungan pasien dan jumlah kasus penyakit di Klinik Pratama Atma Jaya Cisauk per bulan bulan Juni 2012 – Mei 2016. Mulai atas ke bawah jumlah total kunjungan, jumlah kasus dispepsia, kasus infeksi saluran nafas atas (Ispa), dan hipertensi.

Pada plot data deret waktu jumlah kunjungan pasien total yang ditampilkan pada Gambar 2 teratas terlihat adanya trend penurunan yang cukup panjang mulai dari pertengahan tahun 2013. Ini mengindikasikan adanya unsur trend dan konstanta sebagai elemen-elemen yang harus diperhitungkan dalam pemeriksaan kestasioneran. Hal yang sama terlihat pada plot data deret waktu jumlah kasus hipertensi (gambar paling bawah). Kedua plot deret waktu memperlihatkan amplitudo fluktuasi yang kurang lebih sama.

Pada plot deret waktu data jumlah kasus dispepsia hanya mengandung unsur konstanta sedikit di atas nol dan tidak terdapat unsur trend. Kecuali pada nilai observasi pada bulan Oktober 2013 yang memperlihatkan lonjakan, semua data lainnya membentuk pola fluktuasi dengan amplitudo yang relatif stabil. Sedangkan plot deret waktu data kasus ISPA berpola ireguler dalam amplitudo fluktuasi maupun trend naik-turun mengindikasikan proses yang non-stasioner.

Setelah membuat plot akan dilakukan pemeriksaan kestasioneran data dengan statistik uji akar unit dan memeriksa korelogram. Pilihan jenis uji akar unit yang dilakukan adalah uji statistik Augmented Dickey-Fuller (ADF) dengan hipotesis nol H_0 deret waktu tidak stasioner mengikuti distribusi t-student. Daerah penerimaan H_0 adalah bila nilai mutlak uji statistik lebih kecil daripada nilai kritisnya.

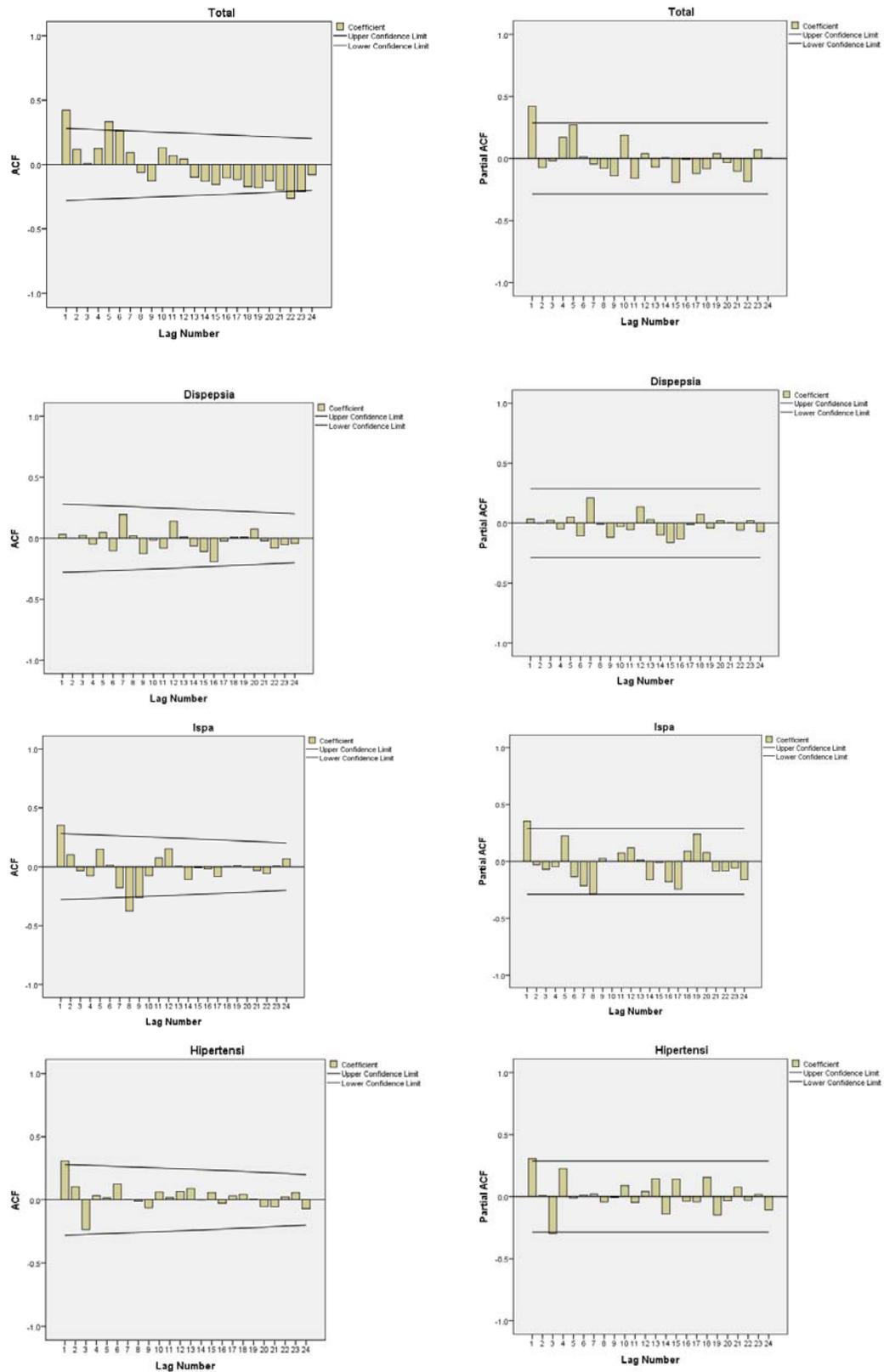
Dalam melakukan uji akar unit ADF perlu diperhatikan pola plot data yang dinilai dari visualisasi data pada Gambar 1 di atas. Untuk deret waktu jumlah kunjungan total turut dimasukkan dalam persamaan yang diuji adalah unsur konstanta dan trend (kemiringan). Hal ini berlaku sama dengan deret waktu jumlah kasus hipertensi. Untuk deret waktu jumlah kasus dispepsia karena tidak terlihat trend hanya diikuti sertakan unsur konstanta. Sedangkan pada deret waktu jumlah kasus ISPA yang irregular tidak dimasukkan unsur trend maupun konstanta.

Tabel 3. Nilai Uji Akar Unit Augmented Dickey-Fuller terhadap Deret Waktu Jumlah Kunjungan Pasien / Kasus di Klinik Atma Jaya Cisauk Tahun 2012 – 2016

Deret Waktu	ADF	Nilai Kritis	Probabilitas
Total	-4.895495	- Tingkat 1% (-4.165756) - Tingkat 5% (-3.508508) - Tingkat 10% (-3.184230)	0.0013
Dispepsia	-6.527046	- Tingkat 1% (-3.577723) - Tingkat 5% (-2.925169) - Tingkat 10% (-2.600658)	0.0000
Hipertensi	-5.497267	- Tingkat 1% (-4.165756) - Tingkat 5% (-3.508508) - Tingkat 10% (-3.184230)	0.0002
Ispa	-0.150781	- Tingkat 1% (-2.619851) - Tingkat 5% (-1.948686) - Tingkat 10% (-1.612036)	0.6258

Keterangan : Data diolah dengan Eview 9.

Tabel 3 menampilkan nilai ADF data deret-deret waktu jumlah kunjungan pasien dan kasus penyakit di Klinik Atma Jaya Cisauk. Bila nilai statistik ADF lebih negatif daripada nilai kritisnya atau nilai mutlak statistiknya lebih besar daripada nilai mutlak nilai kritisnya, maka deret waktu tersebut stasioner. Sebaliknya bila nilai mutlak statistik ADF lebih kecil daripada nilai mutlak nilai kritisnya, maka deret waktu yang diuji tidak stasioner. Terlihat bahwa hanya data deret waktu jumlah kasus ISPA yang non-stasioner sedangkan ketiga yang lain stasioner.



Gambar 3. Korelogram koefisien fungsi otokorelasi (kolum kiri) dan korelogram koefisien fungsi otokorelasi parsial (kolom kanan) deret waktu dari atas ke bawah: jumlah kunjungan total, jumlah kasus dispepsia, jumlah kasus ISPA dan jumlah kasus hipertensi.

Pemeriksaan stationaritas deret waktu juga dapat dinilai dari menilai korelogramnya. Dari korelogram terlihat data deret waktu jumlah kasus dispepsia sudah stasioner dengan seluruh nilai koefisien fungsi otokorelasi (ACF) dan otokorelasi parsial (PACF) tidak ada yang melewati nilai batas atas maupun batas bawah. Untuk deret waktu jumlah kunjungan total dan jumlah kasus hipertensi terdapat beberapa lonjakan di beberapa lag (jeda) pada korelogram ACF, namun pada korelogram PACF hanya ada lonjakan pada lag 1 (dan sedikit di lag 2 pada hipertensi). Ini mengarahkan persamaan model ke proses otoregresi AR (1) pada analisis dengan metode ARIMA.

Untuk data deret waktu ISPA sebenarnya tampilan korelogram ACF dan PACF mirip seperti deret waktu jumlah kunjungan total dan jumlah kasus hipertensi, namun nanti akan kita lihat bahwa deret waktu ini tidak cocok dimodelkan dengan AR (1). Perlu dicatat bahwa uji stasionaritas dengan statistik ADF memperlihatkan bahwa dengan tanpa memasukan unsur konstanta dan trend dalam persamaan, deret waktu tidak stasioner. Hal ini mengarahkan pilihan model pada ARIMA (1,1,0), (0,1,1) atau (1,1,1).

Langkah berikutnya dalam permodelan deret waktu adalah mengestimasi parameter-parameter. Pertama dengan metode ARIMA. Untuk deret waktu jumlah kunjungan total dan hipertensi berdasarkan identifikasi model dan pemeriksaan stasionaritas kita mengajukan model AR(1) atau ARIMA (1,0,0) dengan konstanta. Sedangkan dispepsia yang sudah stasioner ARIMA (0,0,0) dengan konstanta. Uji signifikansi statistik menunjukkan hasil yang memuaskan untuk model-model tersebut dengan nilai $p < 0,05$ (lihat Tabel 4).

Sedangkan untuk memodelkan data deret waktu ISPA perlu usaha tambahan yang dikeluarkan sebab fasilitas *expert modeller* SPSS 21 tidak otomatis memberikan estimasi parameter yang signifikan. Pertama-tama data harus ditransformasi akar kuadrat dan setelah itu dilakukan operasi selisih pertama dan baru kemudian data diestimasi dengan baik dengan MA(1) tanpa konstanta. Ini mengarahkan kita kepada model non-stasioner ARIMA (0,1,1) tanpa konstanta. Semua model-model ARIMA di atas tidak mengandung unsur musiman.

Tabel 4. Estimasi Paramater Model Matematis ARIMA

Deret Waktu	Parameter	Estimasi	SE	t	Sig.
Total	Konstanta	43.376	3.556	12.198	.000
	AR Lag 1	.492	.129	3.812	.000
Dispepsia	Konstanta	4.063	.732	5.550	.000
ISPA (transformasi akar kuadrat)	Beda (d)	1			

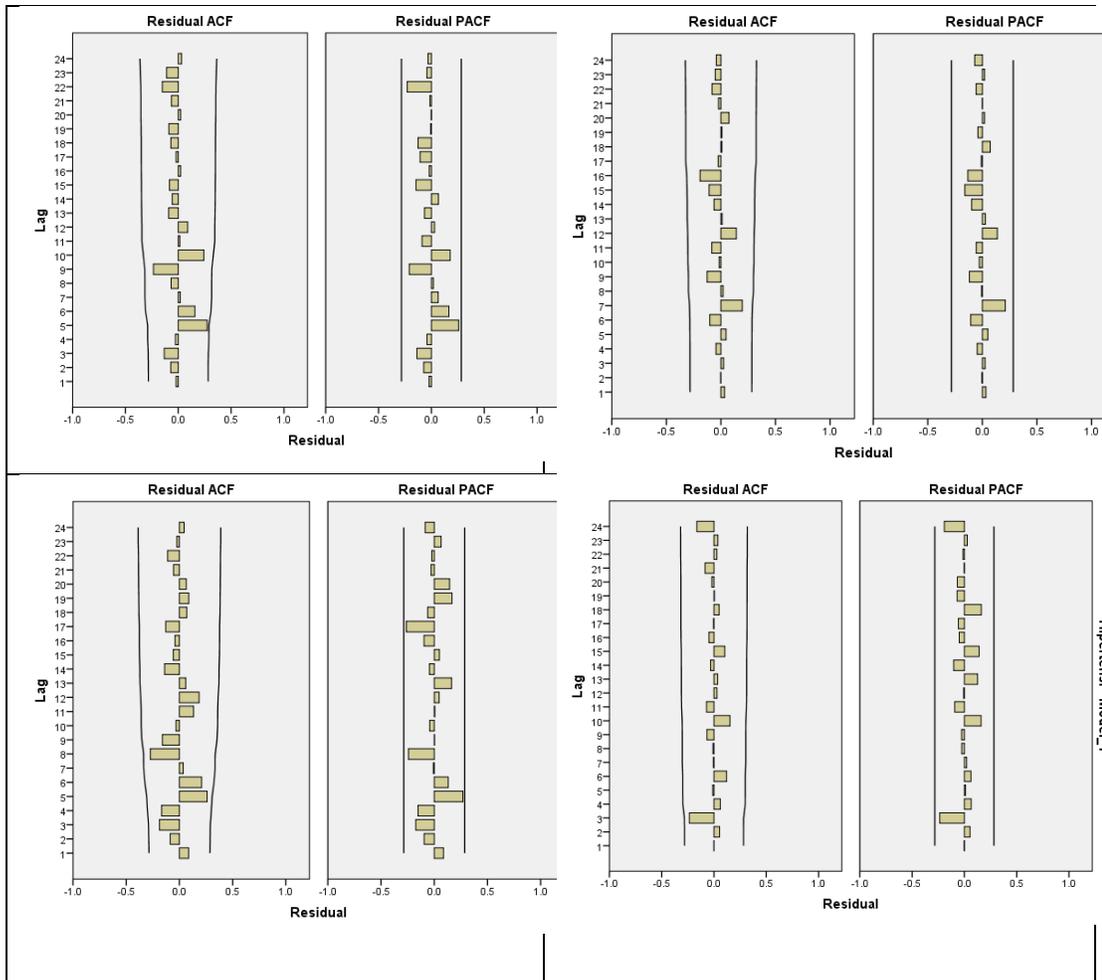
Deret Waktu	Parameter	Estimasi	SE	t	Sig.
	MA Lag 1	.611	.120	5.108	.000
Hipertensi	Konstanta	2.010	.165	12.189	.000
	AR Lag 1	.324	.142	2.285	.027

Peramalan dengan metode penghalusan eksponensial umumnya parameter-parameternya dipilih sejak awal. Fasilitas *expert modeller* SPSS 21 dapat memilih parameter-parameter otomatis yang dianggap paling optimal (dari segi meminimuman galat-galat). Model yang dipilih aplikasi SPSS 21 adalah musiman sederhana (tanpa trend). Namun apabila diperiksa signifikansi statistiknya hasilnya buruk (lihat Tabel 5). Hanya estimasi parameter alfa untuk deret waktu kunjungan total dan ISPA yang signifikan dengan $p < 0,05$. Namun demikian nilai-nilai estimasi parameter untuk alfa berada dalam selang interval yang diajukan Montgomery *et al* yakni $0.1 \leq \alpha \leq 0.4$.

Tabel 5. Estimasi Parameter Model Matematis dengan Metode Penghalusan Eksponensial

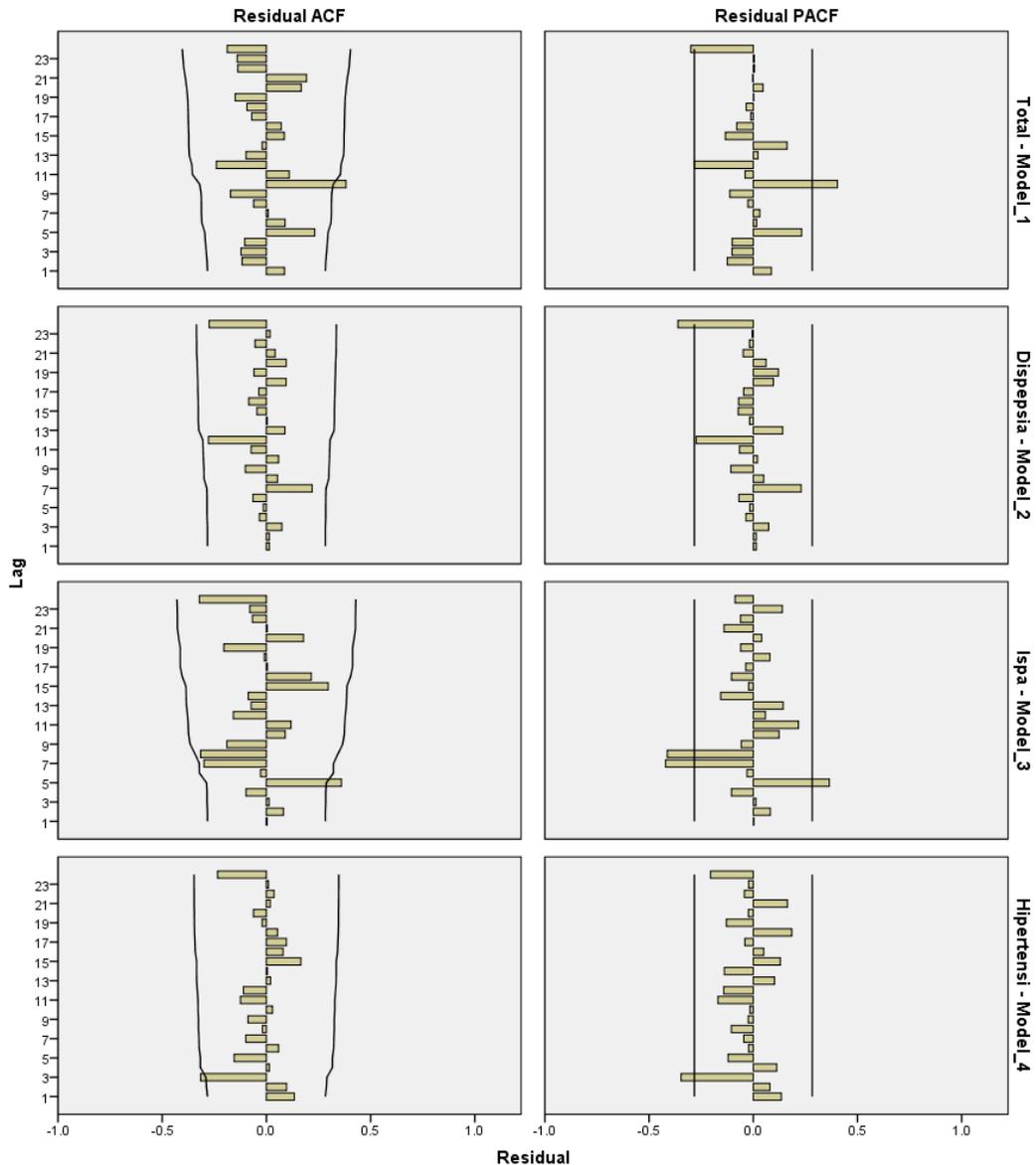
Model Penghalusan Eksponensial Musiman Sederhana			Estimasi	SE	t	Sig.
Total	Tanpa Transformasi	Alpha (Tingkat)	.400	.121	3.315	.002
		Delta (Musim)	.000	.273	.001	1.000
Dispepsia	Tanpa Transformasi	Alpha (Tingkat)	.100	.061	1.636	.109
		Delta (Musim)	3.426E-006	.204	1.678E-005	1.000
Ispe	Tanpa Transformasi	Alpha (Tingkat)	.300	.100	3.015	.004
		Delta (Musim)	.000	.229	.000	1.000
Hipertensi	Tanpa Transformasi	Alpha (Tingkat)	.100	.065	1.540	.131
		Delta (Musim)	5.173E-005	.131	.000	1.000

Setelah dilakukan estimasi parameter dan permodelan matematis, maka langkah berikutnya adalah memeriksa kualitas model dengan memeriksa korelogram ACF dan PACF residu model. Semua model yang diperoleh dari metode ARIMA memiliki nilai-nilai ACF dan PACF di dalam nilai batas atas dan bawah (lihat Gambar 4) sehingga permodelan-permodelan tersebut dianggap baik karena tidak mengandung korelasi nilai residu pada setiap lag. Beda dengan korelogram yang dihasilkan dengan metode penghalusan eksponensial di mana semua modelnya, kecuali ACF untuk model dispepsia, memiliki nilai ACF dan PACF yang melampaui nilai-nilai batas atas atau batas bawahnya (Gambar 5).



Gambar 4. Koefisien fungsi ACF dan PACF residual untuk model ARIMA terhadap deret waktu jumlah kunjungan total, jumlah kasus dispepsia, ISPA dan Hipertensi. Seluruh nilainya di dalam batas atas dan batas bawah yang memperlihatkan tidak ada korelasi pada tiap lag atau distribusinya bersifat random.

Transformasi akar kuadrat pada data memberikan sebaran residu yang lebih baik pada metode penghalusan eksponensial, namun tidak semuanya. Demikian pula transformasi tersebut tidak dapat memperbaiki signifikansi parameter-parameter metode penghalusan eksponensial. Dengan demikian peneliti tidak melakukan upaya lebih lanjut untuk memperbaiki model yang dihasilkan oleh metode penghalusan eksponensial.



Gambar 5. Koefisien fungsi ACF dan PACF residual untuk model penghalusan eksponensial terhadap deret waktu jumlah kunjungan total, jumlah kasus dispepsia, ISPA dan Hipertensi. Terlihat terdapat nilai yang melewati batas atas dan batas bawah yang memperlihatkan ada korelasi pada jeda-jeda (*lags*) tertentu atau distribusinya tidak bersifat random.

Namun demikian dalam langkah berikutnya yakni dalam penggunaan model dalam peramalan kedua metode memberikan hasil yang mengejutkan. Nilai peramalan yang dihasilkan oleh metode penghalusan eksponensial lebih baik daripada yang dihasilkan oleh metode ARIMA. Hal ini dilihat dari dua aspek. Pertama, nilai ramalannya lebih mendekati data aktual atau nilai observasi; kedua, interval peramalan (selisih antara nilai batas atas UCL dan nilai batas bawah LCL) yang dihasilkan oleh metode penghalusan eksponensial lebih kecil daripada yang dihasilkan oleh metode ARIMA (lihat Tabel 6).

Tabel 6. Nilai peramalan (*forecast*) untuk bulan Juni 2016 serta nilai batas atas dan bawah pada setiap model dari dua metode. Nilai observasi adalah nilai riil pada bulan Juni 2016.

Model	Penghalusan Eksponensial			ARIMA			Nilai Observasi
	Forecast	UCL	LCL	Forecast	UCL	LCL	
Total	21	42	1	37	62	11	20
Dispepsia	3	3	-6	4	14	-6	0
ISPA	5	5	-4	13	29	2	9
Hipertensi	3	3	-3	4	10	-3	2

Keakuratan hasil peramalan juga dapat dibandingkan dengan melihat besarnya galat-galat yang dihasilkan (lihat Tabel 7 dan 8). SPSS 21 dapat pula menampilkan perhitungan nilai maksimum galat persentase mutlak (MaxAPE) dan nilai maksimum galat mutlak (MaxAE). Pada tabel 7 dan tabel 8 terlihat bahwa galat-galat serta nilai kriteria informasi bayesian yang dihasilkan oleh metode penghalusan eksponensial lebih baik dibandingkan oleh metode ARIMA.

Tabel 7. Statistik untuk menilai kebaikan ramalan model penghalusan eksponensial

Model	RMSE	MAPE	MAE	MaxAPE	MaxAE	BIC
Total	10.410	21.082	8.162	96.763	22.919	4.847
Dispepsia	4.336	53.866	2.169	365.913	23.662	3.095
Ispa	4.573	42.488	3.440	335.257	12.381	3.202
Hipertensi	3.072	71.766	2.259	469.654	10.900	2.406

Tabel 8. Statistik untuk menilai kebaikan ramalan model ARIMA

Model	RMSE	MAPE	MAE	MaxAPE	MaxAE	BIC
Total	13.009	32.261	10.124	442.204	35.376	5.293
Dispepsia	5.071	63.249	2.544	306.250	30.938	3.328
Ispa	6.539	53.738	4.744	385.755	18.607	3.838
Hipertensi	3.398	76.709	2.485	369.926	11.349	2.608

Fenomena model yang baik dengan nilai parameter-parameternya yang signifikan secara statistik serta memiliki korelogram residu yang random namun memberikan ramalan yang belum tentu baik sudah disadari oleh para praktisi dan peneliti peramalan. Fenomena ini yang disebut dengan *overfitting* atau terlalu besar energi yang dicurahkan untuk membentuk model matematis yang cocok untuk data-data historis (Montgomery *et al*). Akhirnya persamaan matematis yang dihasilkan terlalu kompleks dan terlalu cocok dengan data historis sehingga memberikan ekstrapolasi ramalan yang baik.

KESIMPULAN

Data deret waktu jumlah kunjungan total pasien di klinik pratama Atma Jaya Cisauk periode Juni 2012 – Mei 2016 dapat dimodelkan dengan baik oleh metode ARIMA (1,0,0) dengan konstanta. Demikian pula data deret waktu jumlah kasus hipertensi. Data deret waktu jumlah kasus dispepsi dimodelkan dengan baik oleh model ARIMA (0,0,0) dengan konstanta. Sedangkan data deret waktu ISPA dimodelkan oleh ARIMA (0,1,1) tanpa konstanta setelah dilakukan transformasi akar kuadrat. Penilaian baiknya model dilihat dari signifikannya secara statistik parameter-parameter yang diestimasi dan korelogram nilai residu yang memperlihatkan distribusi yang bersifat random.

Permodelan matematis dengan metode penghalusan eksponensial memberikan model yang lebih buruk dibandingkan metode ARIMA bila dinilai dari kriteria signifikansi statistik parameter-parameternya dan korelogram residunya.

Permodelan matematis dengan metode penghalusan eksponensial memberikan nilai ramalan yang lebih akurat daripada dengan metode ARIMA dipandang dari kriteria kecilnya galat-galat yang dihasilkan dan lebih sempitnya interval kepercayaan peramalannya. Temuan ini menguatkan pandangan yang menyatakan bahwa model matematis yang baik memodelkan data historis belum tentu memberikan nilai peralaman yang akurat disebabkan karena *overfitting*.

DAFTAR PUSTAKA

- Allard, R. (1998). Use of time series analysis in infectious disease surveillance. *Bulletin of the World Health Organization*, 76 (4), 327-333.
- Alt, F. B., Hung, K., Lap-Ming Wun. (1998). Time Series Analysis. *Handbook of Statistical Method for Enginneers and Scientist* (Harrison M. Wadsworth, ed.), 2nd Edition. McGraw-Hill : New York.
- Badan Pusat Statistik [BPS] Kabupaten Tangerang. (2015). *Statistik Daerah Kecamatan Cisauk 2015*. Badan Pusat Statistik Kabupaten Tangerang : Tangerang.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C. (1994). *Time Series Analysis – Forecasting and Control*, 3rd Ed. Holden-Day : Oakland, California.
- Djauhari, M. (2015). *Metode Peramalan*, Edisi 2. Penerbit Universitas Terbuka: Tangerang Selatan.
- Dorleans. (2007). *Jakarta Batavia: esai Sosio-Kultural*, Kees Grijns & Peter J. M. Nas (Ed.). Banana.
- Fazekas, M. (2004). *Application of Time Series Models on Medical Research*. 6th International Conference on Applied Informatics : Eger, Hungary, January 27th-31th.

- Green, L. A. (2004). The Research Domain of Family Medicine. *Annals of Family Medicine*, Vol. 2, Supplement 2.
- Greene, W. H. (1993). *Econometric Analysis*, 2nd Ed. Macmillan Publishing Co. : New York.
- Juanda, B., dan Junaidi. (2012). *Ekonometrika Deret Waktu: Teori dan Aplikasi*. IPB Press: Bogor.
- Makridakis, S., Wheelright, S. C., McGee, V. E. (1999). *Metode dan Aplikasi Peramalan, Jilid 1* (Untung Sus Andriyanto & Abdul Basith, terj.). Penerbit Erlangga: Jakarta.
- Montgomery, D. C., Jennings, C. L., Kulahci, M. (2008). *Introduction to Time Series Analysis and Forecasting*. John Wiley & Sons: Hoboken, New Jersey.
- Rosadi, D. (2012). *Ekonometrika dan Analisis Runtun Waktu dengan Eview*. Andi Offset.
- Silver, C. (2008). *Planning The Megacity : Jakarta in Twentieth Century*. Routledge.
- Soejoeti, Z. (2006). *Pengantar Statistika Matematis 2*, Edisi 1. Penerbit Universitas Terbuka : Tangerang Selatan.
- Trottier, H., Philippe, P., Roy, R. (2006). Stochastic modeling of empirical time series of childhood infectious diseases data before and after mass vaccination. *Emerging Themes in Epidemiology* 2006, 3 (9). BioMed Central Ltd.