BAB II

TINJAUAN PUSTAKA

2.1 Pendahuluan

Pada Bab II akan dijelaskan pengertian–pengertian dan teori dasar yang digunakan sebagai landasan pembahasan pada bab selanjutnya. Teori yang akan dibahas pada Bab II ini secara garis besar yaitu meliputi analisis data deret waktu, pengantar metode ARIMA Box–Jenkins dan analisis spektral.

Cuaca adalah keadaan fenomena fisik dari atmosfer (yang berhubungan dengan suhu, tekanan udara, angin, awan, kelembaban udara, radiasi, jarak pandang/*visibility*, dsb) di suatu tempat dan pada waktu tertentu. Contoh Pengamatan cuaca dilakukan setiap hari. Dalam hal ini data yang diambil sebagai pengamatan yaitu data cuaca dari kota Surabaya dari periode waktu oktober 2004 sampai dengan desember 2011.

Cuaca juga merupakan salah satu faktor yang menentukan keberhasilan pada sektor pertanian. Dalam skala waktu perubahan cuaca akan membentuk pola atau siklus tertentu, baik harian, musiman, tahunan maupun siklus beberapa tahunan. Selain perubahan yang berpola tertentu, aktivitas manusia menyebabkan pola iklim berubah secara berkelanjutan, baik dalam skala global maupun skala lokal. Sehingga dengan menelaah beberapa faktor penentu keberhasilan sektor pertanian dalam hal ini salah satunya adalah faktor cuaca, maka diharapkan bisa bermanfaat dalam menentukan musim tanam dan mengetahui secara dini mengenai serangan hama pertanian yang dapat muncul pada kondisi cuaca tertentu.

Pada dasarnya setiap nilai dari hasil pengamatan (data) selalu dapat dikaitkan dengan waktu pengamatannya. Hanya pada saat melakukan analisis, variabel waktu dengan pengamatan sering tidak dipersoalkan. Dalam hal kaitan variabel waktu dengan pengamatan diperhatikan, sehingga data dianggap sebagai fungsi atas waktu, maka data seperti ini disebut data deret waktu (*time series*). Banyak persoalan dalam ilmu terapan yang datanya merupakan data deret waktu, salah satunya dalam bidang ilmu fisika mengenai klimatologi.

2.2 Analisis Deret Waktu

Pada dasarnya setiap nilai dari hasil pengamatan (data) selalu dapat dikaitkan dengan waktu pengamatannya. Hanya pada saat melakukan analisis, variabel waktu dengan pengamatan sering tidak dipersoalkan. Dalam hal kaitan variabel waktu dengan pengamatan diperhatikan, sehingga data dianggap sebagai fungsi atas waktu, maka data seperti ini disebut data deret waktu (*time series*). Banyak persoalan dalam ilmu terapan yang datanya merupakan data deret waktu, misalnya dalam bidang ilmu

a. Ekonomi : banyak barang terjual dalam setiap hari, keuntungan perusahaan dalam setiap tahun dan total nilai ekspor dalam setiap bulan.
b. Fisika : curah hujan bulanan, temperatur udara harian dan gerak partikel.
c. Demografi : pertumbuhan penduduk, mortalitas dan natalitas.
d. Biomedis : denyut nadi, proses penyembuhan dan pertumbuhan mikroba.

Karena data deret waktu merupakan regresi data atas waktu, dan salah satu segi (*aspect*) pada data deret waktu adalah terlibatnya sebuah besaran yang dinamakan Autokorelasi (*autocorrelation*), yang secara konsep sama dengan korelasi untuk data bivariat dalam analisis regresi biasa. Signifikansi (keberartian) autokorelasi menentukan analisis regresi yang harus dilakukan pada data deret waktu. Jika autokorelasi tidak signifikan (dalam kata lain data deret waktu tidak berautokorelasi), maka analisis regresi yang harus dilakukan adalah analisis regresi sederhana biasa, yaitu analisis regresi data atas waktu. Sedangkan jika signifikan (berautokorelasi) harus dilakukan analisis regresi data deret waktu, yaitu analisis regresi antar nilai pengamatan. Segi lain dalam data deret waktu adalah kestasioneran data yang diklasifikasikan atas stasioner kuat (*strickly stationer*) dan stasioner lemah (*weakly stationer*) dan kestasioneran ini merupakan kondisi yang diperlukan dalam analisis data deret waktu, karena akan memperkecil kekeliruan baku.

Dalam teori statistika, setiap data deret waktu dibangun atas komponen trend (T), siklis (S), musiman (M, untuk data bulanan) dan residu (R) atau sering disebut komponene acak. Bentuk hubungan antara nilai data dengan komponen–komponennya tersebut bisa bermacam–macam dan bentuk hubungan yang sering digunakan adalah linier dan multiplikatif. Jika Y_t pengamatan pada waktu $-t$ dan hubungan dengan komponennya linier, maka persamaannya adalah

$$Y_t = T_t + S_t + M_t + R_t, \text{ jika } t : \text{bulanan} \quad \textbf{(2.1)}$$

$$Y_t = T_t + S_t + R_t, \text{ jika } t : \text{tahunan} \quad \textbf{(2.2)}$$

dan multiplikatif, maka persamaannya

$$Y_t = T \cdot S \cdot M \cdot R, \text{ jika} t\text{: bulanan} \quad \textbf{(2.3)}$$

$$Y_t = T \cdot S \cdot R, \text{ jika } t\text{: tahunan} \quad \textbf{(2.4)}$$

Sebagai akibat dari terdapatnya komponen–komponen dalam data deret waktu dan terjadinya hubungan antar komponen, adalah berautokorelasinya antar pengamatan sehingga dapat dibangun sebuah hubungan fungsional yang dinamakan regresi deret waktu.

Deret waktu adalah suatu rangkaian atau seri dari nilai–nilai suatu variabel atau hasil observasi yang dicatat dalam jangka waktu yang berurutan (Atmaja, 2009:29). Metode deret waktu (*time series*) adalah metode peramalan dengan menggunakan analisis pola hubungan antara variabel yang akan diperkirakan dengan variabel waktu atau analisis deret waktu, beberapa metode yang sering digunakan antara lain:

1. Metode *Smoothing*
2. Metode Box–Jenkins (ARIMA)
3. Metode Proyeksi *Trend* dan Regresi

Hal yang perlu diperhatikan dalam melakukan peramalan adalah pada galat (*error*), yang tidak dapat dipisahkan dalam metode peramalan. Untuk mendapatkan hasil yang mendekati data asli, maka seorang peramal berusaha membuat galatnya sekecil mungkin.

2.3 Kawasan dalam Analisis Deret Waktu

Sudah dikemukakan pada penjelasan di atas bahwa data deret waktu adalah data yang merupakan fungsi atas waktu dan setiap data deret waktu dibangun oleh komponen trend, siklis, musiman (untuk data bulanan) dan komponen acak atau residu. Sehingga berdasarkan konsep tersebut, analisis data deret waktu dapat dilakukan dalam dua kawasan (*domain*), yaitu kawasan waktu (*time domain*) dan kawasan frekuensi (*frequency domain*). Dalam kawasan waktu adalah telaah signifikansi autokorelasi, kestasioneran data, penaksiran parameter model regresi deret waktu, dan peramalan (*forecasting*). Sedangkan dalam kawasan frekuensi adalah telaahan frekuensi tersembunyi, yaitu frekuensi komponen siklis yang sulit diperoleh dalam kawasan waktu, dengan tujuan untuk mengetahui hal–hal istimewa atau kondisi tertentu pada data.

Salah satu analisis data deret waktu yang sering digunakan dalam kawasan waktu (*time domain*) adalah peramalan dengan metode ARIMA Box-Jenkins. Sedangkan dalam penelitian ini akan digunakan analisis data deret waktu dalam kawasan frekuensi (*frequency domain*) dengan menggunakan metode analisis spektral. Analisis spektal merupakan metode dalam analisis deret waktu yang jarang dibahas, padahal perananya sangat besar dalam melengkapi informasi mengenai cirri (*characters*) data deret waktu. Analisis ini membahas mengenai cara menelaah periodisitas data tersembunyi (*hidden periodecities*) yang sulit diperoleh pada saat kajian dilakukan pada kawasan waktu. Kajian periodisitas data perlu dilakukan untuk memberikan informasi mengenai karakteristik dari data deret waktu tersebut, dan harus dilakukan pada kawasan frekuensi melalui analisis spektral (Mulyana, 2004).

2.4 Metode ARIMA Box-Jenkins

Seperti yang sudah dijelaskan di atas bahwa salah satu analisis deret waktu yang sering digunakan dalam kawasan waktu (*time domain*) adalah peramalan dengan metode ARIMA Box-Jenkins. Pengertian *time series* disini adalah deret atau urutan observasi atau pengamatan dan biasanya urutan ini berdasarkan waktu (Wei, 1994). Analisis diperkenalkan pertama kali pada tahun 1970 oleh George E.P. Box dan Gwilym M. pendekatan *time series* dapat menggunakan metode analisis fungsi autokorelasi dan fungsi autokorelasi parsial untuk mempelajari perubahan data deret waktu. Untuk model parametrik seringkali dikenal dengan analisis domain waktu ARIMA (Von Storch dan Zwier, 1999).

Pola cuaca/iklim di Indonesia cenderung membentuk pola musiman, oleh karena itu perlu diramalkan melalui model musiman ARIMA (p,d,q) $(P,D,Q)^S$ (Nuryadi, 2005), dengan persamaan umum sebagai berikut:

$$\phi_p(B)\phi_P(B^S)(1-B)^d(1-B^S)^D Z_t = \theta_q(B)\theta_Q(B^S)a_t \qquad \textbf{(2.5)}$$

dimana:

p, d, q = orde AR (*Autoregresisive*), MA (*Moving Average*) dan pembedaan (*differencing*) non musiman,

P, D, Q = orde AR (*Autoregresisive*), MA (*Moving Average*) dan pembedaan (*differencing*) musiman,

$\phi_p(B) = (1-\phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)$,

$\phi_P(B^S) = (1-\phi_1 B^S - \phi_2 B^{2S} - \ldots - \phi_P B^{PS})$,

$(1-B)^d$ = orde pembedaan (*differencing*) non musiman,

$\theta_q(B) = (1-\theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)$,

$\Theta_Q(B^S) = (1-\theta_1 B^S - \theta_2 B^{2S} - \ldots - \theta_Q B^{QS})$,

$Z_t = Z_t - \mu$.

Metode ARIMA merupakan metode yang tidak melibatkan variabel prediktor. Bey (2003) menyebutkan bahwa ARIMA adalah salah satu metode stokastik yang sangat bermanfaat untuk membangkitkan proses (data) deret waktu dimana setiap kejadian saling berkorelasi. Namun demikian metode ARIMA sangat ketat terhadap asumsi (data dan residual *white noise*) dan digunakan untuk data yang berpola linier (Sutikno, 2005).

2.5 Analisis Spektral

Pada penjelasan sebelumnya sudah diterangkan bahwa analisis data deret waktu dapat dilakukan dalam dua kawasan (*domain*), yaitu kawasan waktu (*time domain*) dan kawasan frekuensi (*frequency domain*). Salah satu contoh analisis dalam domain waktu (*time domain*) adalah peramalan (*forecasting*) dengan metode ARIMA Box- Jenkins. Sedangkan dalam penelitian ini akan membahas analisis deret waktu yang dilakukan dalam kawasan frekuensi (*frequency domain*), yaitu dengan menganalisis periodisitas data tersembunyi (*hidden periodecities*) yang sulit diperoleh pada saat kajian dilakukan pada kawasan waktu. Metode yang digunakan adalah analisis spektral.

Analisis spektral atau dinamakan juga analisis spektrum, dikenalkan oleh A.Schuster seorang pekerja sosial, pada akhir abad ke–20 dengan tujuan mencari periode tersembunyi dari data. Pada saat ini analisis spektral digunakan pada persoalan penaksiran spektrum untuk seluruh selang frekuensi. M. S. Bartlett dan J.W. Tukey, mengembangkan analisis spektral modern sekitar tahun ketiga abad ke–20 dan teorinya banyak digunakan para pengguna di bidang klimatologi, teknik kelistrikan, meteorologi dan ilmu kelautan.

Tidak seperti metode ARIMA yang sangat ketat terhadap asumsi (data dan residual *white noise*), kelebihan dari analisis spektral yaitu penggunaanya lebih praktis, karena dari data yang diperoleh bisa langsung diolah tanpa adanya asumsi-asumsi yang harus dipenuhi terlebih dahulu.

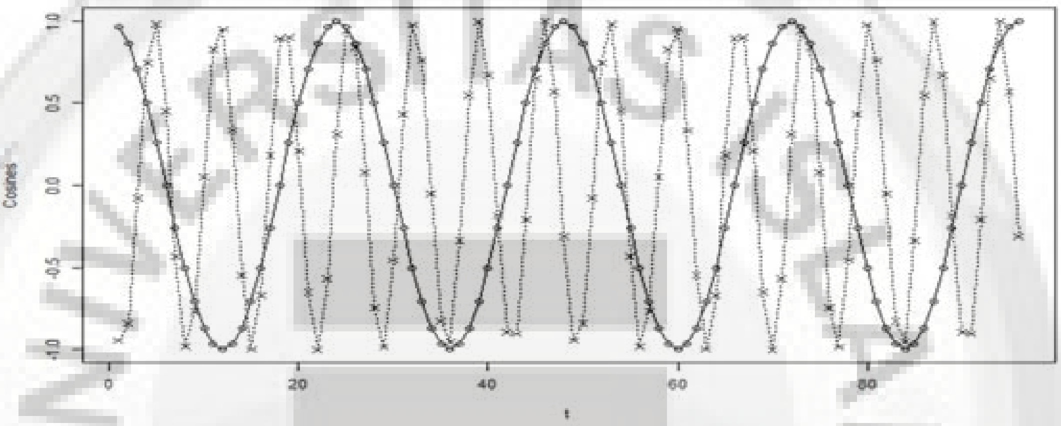
Analisis spektral termasuk dalam kawasan frekuensi untuk menelaah periodesitas tersembunyi, yaitu periodesitas yang sulit ditemukan dalam kawasan waktu. Analisis ini dilakukan jika diperlukan informasi mengenai periodesitas hal–hal yang bersifat khusus, untuk melengkapi hasil analisis dalam kawasan waktu. Analisis Spektral juga merupakan suatu metode untuk melakukan transformasi sinyal data dari domain waktu ke domain frekuensi, sehingga kita bisa melihat pola periodiknya untuk kemudian ditentukan jenis pola yang terlihat. Pada dasarnya metode ini merupakan analisis statistik inferensial (dapat disimpulkan) yang berdasarkan konsep frekuensi. Secara visual digambarkan dengan spektrum. Konsep dasar analisis spektral yakni mengitung dan menggambarkan periodogram dari data.

2.6 Model dalam Analisis Spektral

Secara umum persamaan kurva cosinus yaitu

$$Y = R\cos(2\pi ft + \Phi) \qquad \textbf{(2.8)}$$

dimana $R(>0)$ adalah amplitudo, f (frekuensi) dan Φ adalah fase dari kurva. Jika kurva berulang setiap $1/f$ unit waktu, maka $1/f$ disebut periode dari gelombang cosinus.

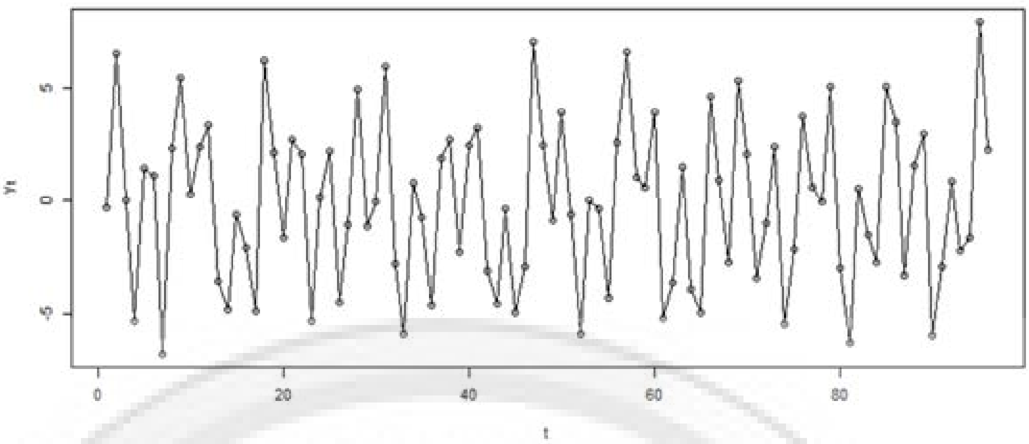


Gambar 2.1 Kurva Kosinus dengan *n* = 96 dan Dua Frekuensi dan Fase

Pada Gambar 2.1 menunjukkan dua kurva cosinus diskrit dengan waktu berjalan dari 1 sampai 96. Dari contoh gambar pola frekuensi diketahui masing–masing pada frekuensi ke 4/96 dan 14/96. Kurva frekuensi rendah memiliki fase nol, sedangkan kurva frekuensi yang lebih tinggi memiliki fase 0.6π.

Gambar 2.2 menunjukkan grafik kombinasi linier dari dua kurva cosinus. Pada kurva tersebut frekuensi rendah memiliki nilai pengganda sebanyak 2, sedangkan kurva frekuensi yang lebih tinggi memiliki nilai pengganda sebanyak 3 dan fase yang diketahui yaitu 0.6π. Secara maatematis persamaan dari Gambar 2.2 yaitu sebagai berikut

$$Y_t = 2\cos\left(2\pi t\frac{4}{96}\right) + 3\cos\left[2\pi\left(t\frac{14}{96} + 0.3\right)\right] \qquad \textbf{(2.9)}$$



Gambar 2.2 Kombinasi Linier dari Dua Kurva Cosinus

Pada Persamaan (2.8) tidak sesuai untuk mengestimasi karena parameter R dan Φ tidak dalam bentuk linier. Jadi Persamaan (2.8) diubah dalam bentuk persamaan sebagai berikut

$$Y = R\cos(2\pi ft + \Phi) = A\cos(2\pi ft) + B\sin(2\pi ft) \qquad \textbf{(2.10)}$$

dimana $$R = \sqrt{A^2 + B^2}, \qquad \Phi = a\tan(-B/A) \qquad \textbf{(2.11)}$$

dan sebaliknya, $$A = R\cos(\Phi), \qquad B = -R\sin(\Phi) \qquad \textbf{(2.12)}$$

Untuk mengestimasi model pada Persamaan (2.10) dapat dilakukan dengan menggunakan metode regresi kuadrat terkecil biasa, dimana $\cos(2\pi ft)$ dan $\sin(2\pi ft)$ digunakkan sebagai variabel prediktor.

Jika Y dipengaruhi oleh m kurva cosinus dengan amplitudo sembarang, maka model matematis dari frekuensi dan fase dapat ditulis sebagai berikut

$$Y_t = A_0 + \sum_{j-1}^{m}\left[A_j\cos(2\pi f_j t) + B_j\sin(2\pi f_j t)\right] \qquad \textbf{(2.13)}$$

Model pada Persamaan (2.13) tidak lain adalah model matematis yang dapat dituliskan dalam bentuk persamaan regresi sebagai berikut

$$Y_t = \alpha_0 + \alpha_1 X_{1t} + \beta_1 Z_{1t} + \cdots + \alpha_m X_{mt} + \beta_m Z_{mt} + \varepsilon_t \qquad \textbf{(2.14)}$$

dimana $X_{jt} = \cos(2\pi f_j t)$ dan $Z_{jt} = \sin(2\pi f_j t)$. Koefisien regresi $\alpha_0, \alpha_1, \ldots, \alpha_m$ dan $\beta_1, \ldots, \beta_m$ pada model dapat diduga dengan menggunakan metode regresi kuadrat terkecil biasa.

2.7 Periodogram

Konsep dasar analisis spektral yakni mengitung dan menggambarkan periodogram dari data. Tujuan dari perhitungan dan penggambarannya sendiri yaitu untuk mengetahui berapa banyak jumlah ujung garis periodogram (*peak*) dari data. Selain itu, pada plot periodogram terdapat besaran nilai frekuensi (f) dari data yang akan digunakan sebagai perhitungan dalam pembentukan model. Grafik periodogram sendiri yaitu plot hubungan antara fungsi spektrum kuasa atas frekuensinya.

Perhitungan periodogram dapat dilakukan berdasarkan ukuran sampel. Untuk sampel ganjil dengan $n = 2k + 1$, periodogram I pada frekuensi $f = j/n$ untuk $j = 1, 2, \ldots, k$, persamaannya didefinisikan sebagai berikut

$$I(f) = I\left(\frac{j}{n}\right) = \frac{n}{2}\left(\hat{A}_j^2 + \hat{B}_j^2\right) \qquad \textbf{(2.15)}$$

Menurut Cryer (2008:321), dimana

$$\hat{A}_j = \frac{2}{n}\sum_{t-1}^{n} Y_t \cos(2\pi t j/n), \; \hat{B}_j = \frac{2}{n}\sum_{t-1}^{n} Y_t \sin(2\pi t j/n) \text{ dan } \hat{A}_0 = \overline{Y}. \qquad \textbf{(2.16)}$$

$$\hat{A}_0 = \overline{Y}. \qquad \textbf{(2.17)}$$

Sedangkan jika ukuran sampelnya adalah genap dengan $n = 2k$, dimana $j = 1, 2, \ldots, k-1$ maka persamaannya adalah sebagai berikut

$$\hat{A}_k = \frac{1}{n}\sum_{t-1}^{n}(-1)^t Y_t \text{ dan } \hat{B}_k = 0 \tag{2.18}$$

Pada Persamaan (2.16) dan (2.17) masih berlaku untuk $j = 1,\ 2,\ \ldots, k-1$, tetapi dengan catatan bahwa $f_k = k/n = 1/2$ (frekuensi ekstrim). Karena pada frekuensi ekstrim, dimana $j = k$, berlaku Persamaan sebagai berikut

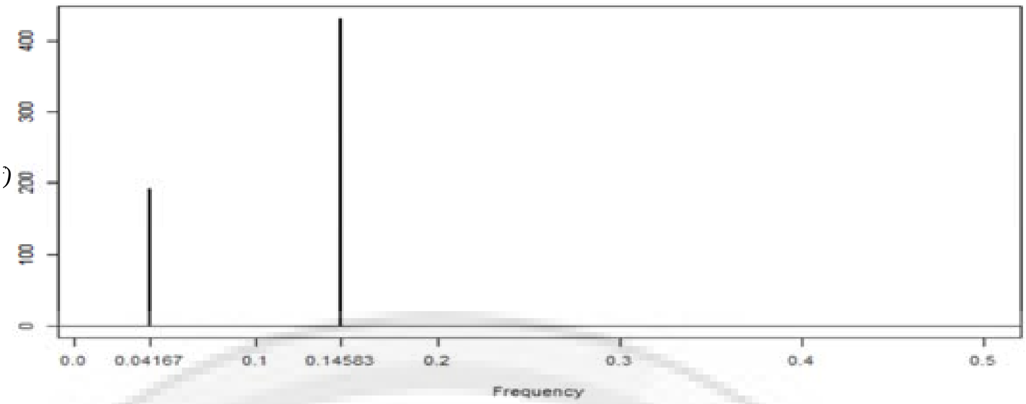
$$I\left(\frac{1}{2}\right) = n\left(\hat{A}_k\right)^2 \tag{2.19}$$

Ketinggian dari periodogram menunjukkan kekuatan relatif pasangan cosinus–sinus pada berbagai frekuensi dalam pola keseluruhan data deret waktu. Sebagai contoh, Gambar 2.3 menunjukkan grafik dari dari periodogram untuk plot deret waktu pada Gambar 2.2. Ketinggian garis pada gambar menunjukan dua buah ujung garis (*peak (m)*) yang menandakan adanya dua komponen cosinus–sinus. Pada Gambar 2.3, grafik menunjukkan nilai dari frekuensi pertama yaitu $4/96 \approx 0.04167$ dan nilai dari frekuensi kedua yaitu $14/96 \approx 0.14583$ yang ditandai pada sumbu frekuensi. Artinya jika kita terapkan ke dalam model Persamaan (2.14) akan menjadi model sebagai berikut

$$Y_t = \alpha_0 + \alpha_1 X_{1t} + \beta_1 Z_{1t} + \cdots + \alpha_m X_{mt} + \beta_m Z_{mt} + \varepsilon_t$$

dimana $X_{jt} = \cos(2\pi f_j t)$ dan $Z_{jt} = \sin(2\pi f_j t)$ maka

$$Y_t = \alpha_0 + \alpha_1 \cos(2\pi(0.0416)t) + \beta_1 \sin(2\pi(0.0146)t) + \alpha_2 \cos(2\pi(0.14583)t)$$
$$+ \beta_2 \sin(2\pi(0.14583)t) + \varepsilon_t$$



Gambar 2.3 Periodogram dari Deret Waktu pada Gambar 2.2

Untuk seri dalam jangka waktu yang lama, perhitungan untuk mencari nilai dari koefisien regresi akan lebih rumit. Dengan demikian untuk mempermudah melakukan perhitungan maka digunakan bantuan software statistika agar lebih efisien.

2.8 Koefisien Determinasi

Koefisien detetrminasi (*goodness of fit*) yang dinotasikan dengan R^2 merupakan suatu ukuran yang menginformasikan baik atau tidaknya suatu model yang ditaksir. Atau dengan kata lain, angka tersebut dapat mengukur seberapa dekat nilai yang ditaksir dengan data sesungguhnya (Nachrowi, 2006).

Nilai koefisien determinasi menunjukkan besar variasi dari respon yang dapat dijelaskan oleh model. Adapun persamaan untuk menghitung nilai koefisien deteminasi adalah sebagai berikut

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \bar{Y}_i\right)^2} \qquad \textbf{(2.20)}$$

dimana :

R^2 = Koefisien Determinasi

n = Jumlah Data

Y_i = Nilai Aktual dari Data

$\overline{Y}_i$ = Nilai Rata-rata dari Data

$\hat{Y}_i$ = Nilai Prediksi dari Data

Nilai R^2 berada di antara nilai 0 sampai 1. Nilai R^2 sama dengan 0, menunjukkan bahwa variasi respon tidak dapat dijelaskan oleh model sama sekali. Nilai R^2 sama dengan satu, menunjukkan bahwa variasi respon secara keseluruhan dapat dijelaskan oleh model. Nilai R^2 x 100% merupakan besarnya pengaruh prediktor terhadap variasi respon dalam model. Pengaruh akan semakin kuat dan signifikan jika nilai R^2 mendekati nilai 1.