

APLIKASI PRINCIPAL COMPONENT ANALYSIS (PCA) UNTUK MEREDUKSI FAKTOR-FAKTOR YANG BERPENGARUH DALAM PERAMALAN KONSUMSI LISTRIK

Titik Rahmawati

Program Studi Manajemen Informatika
STMIK Jenderal Achmad Yani Yogyakarta

rahmawati.titik@gmail.com

Abstrak

Penelitian ini bertujuan untuk memberikan alternatif solusi untuk menentukan faktor-faktor yang berpengaruh dalam peramalan konsumsi listrik. Salah satu teknik atau metode yang digunakan adalah PCA (Principal Component Analysis). PCA telah banyak dikenal dalam statistika sebagai metode yang digunakan untuk mengekstrasi struktur dari suatu set data dengan dimensi yang cukup banyak. Principal Component Analysis (PCA) dapat mengurangi besarnya dimensi dari data yang diobservasi menjadi dimensi yang lebih kecil tanpa kehilangan informasi yang signifikan dalam menggambarkan keseluruhan data.

Teknik peramalan yang digunakan adalah metode ARIMA Box Jenkins yang digunakan untuk menentukan lag-lag pada faktor yang berpengaruh pada peramalan konsumsi listrik. Dalam penelitian ini ada beberapa variabel makro yang digunakan dalam peramalan konsumsi listrik yakni: faktor jumlah penduduk, faktor pertumbuhan PDRB, faktor pertumbuhan industri dan data demografi konsumsi listrik yang meliputi pelanggan rumah tangga, pelanggan industri, pelanggan bisnis, pelanggan sosial dan pelanggan publik.

Hasil penelitian ini menunjukkan bahwa penerapan metode Principal Component Analysis (PCA) dapat mengetahui faktor-faktor dominan yang mempengaruhi peramalan konsumsi listrik total¹. Dari hasil simulasi proses ARIMA Box Jenkins untuk peramalan listrik konsumsi total¹ dihasilkan suatu model ARIMA (0, 1, 0).

Kata Kunci: PCA (*Principal Component Analysis*), Peramalan, ARIMA.

1. Pendahuluan

Konsumsi listrik Indonesia setiap tahunnya terus meningkat sejalan dengan peningkatan pertumbuhan ekonomi nasional. Oleh karena itu prakiraan kebutuhan listrik di Indonesia sangat diperlukan agar dapat menggambarkan kondisi kelistrikan saat ini dan masa mendatang. Penghematan Listrik perlu dilakukan agar kebutuhan listrik disemua sektor dapat terpenuhi. Selain itu perlu adanya sistem distribusi tenaga listrik yang efisien, hal ini dapat dilakukan dengan cara penempatan travo harus disesuaikan dengan kebutuhan *real time* suatu wilayah.

Perkiraan kebutuhan listrik di masing-masing wilayah pada setiap saat dipengaruhi oleh beberapa faktor yang berbeda-beda sehingga menyulitkan pihak PLN dan beberapa pihak terkait untuk dapat mengidentifikasi faktor-faktor

apa saja yang berpengaruh dan menjadi penyebab utama dalam konsumsi listrik pada masing-masing wilayah, sehingga dengan adanya identifikasi yang tepat terhadap faktor-faktor yang berpengaruh dalam konsumsi listrik pada suatu wilayah dapat dijadikan sebagai bahan pertimbangan bagi pihak PLN atau beberapa pihak terkait dalam usaha kampanye hemat listrik.

Berdasarkan ulasan singkat tersebut maka diperlukan suatu cara atau teknik yang dapat digunakan untuk menentukan faktor-faktor utama apa saja yang berpengaruh dalam konsumsi listrik pada suatu wilayah.. Salah satu solusi yang dapat digunakan adalah dengan menggunakan analisis komponen utama (*Principal Component Analysis*). Melalui penggunaan analisis komponen utama ini akan dihasilkan variabel-variabel baru yang merupakan kombinasi linier dari variabel-variabel bebas asal dan antar variabel baru ini bersifat saling bebas. Variabel-variabel baru ini disebut komponen utama dan selanjutnya diregresikan dengan variabel tidak bebas.

2. Tinjauan Pustaka

2.1 Peramalan (*Forecasting*)

Peramalan merupakan suatu unsur penting dalam memprediksi ketidakpastian masa depan sebagai upaya membantu perusahaan untuk mengambil keputusan yang lebih baik. Untuk melakukan peramalan, dibutuhkan data lampau (historis) dan memanipulasi data tersebut untuk mencari polanya yang secara efektif sehingga dapat ditarik ke masa depan. Salah satu teknik peramalan yang digunakan yaitu Metode Box-Jenkins (*ARIMA – Autoregressive Integrated Moving Average*). Metode ini telah dipelajari secara mendalam dan dikembangkan oleh George Box dan Gwilyn Jenkins (1976). Model Autoregresif (AR) pertama kali dikembangkan oleh Yule (1926) dan kemudian dikembangkan oleh Walker (1931), sedangkan model *Moving Average* (MA) pertama kali digunakan oleh Slutsky (1937) (Makridakis, et al., 1999).

Model *ARIMA (Autoregressive Integrated Moving Average)* yang dikembangkan oleh George Box dan Gwilyn Jenkins (1976) merupakan yang tidak mengasumsikan pada pola tertentu pada data historis yang diramalkan dan tidak mengikutkan variabel bebas pada pembentukannya. Model *ARIMA* merupakan model gabungan antara *autoregressive* (AR) dan *moving average* (MA) dimana model ini mampu mewakili deret waktu yang stasioner dan non-stasioner (Hanke, et al., 2000).

Makridakis, et al. (1999) menjelaskan bahwa model *Autoregressive Integrated Moving Average* (ARIMA) merupakan metode yang telah dikembangkan oleh George Box dan Gwilym Jenkins yang diterapkan untuk analisis deret berkala, peramalan dan pengendalian. Metode ini paling berbeda dari metode peramalan lain karena tidak mensyaratkan suatu pola data tertentu supaya model dapat bekerja dengan baik. Apabila metode ini digunakan untuk data deret berkala yang bersifat dependen (terikat) atau berhubungan satu sama lain secara statistik maka metode ini akan bekerja dengan baik.

Model ARIMA ini dikenal dengan metode Box-Jenkins. Metode ini mendasarkan analisis pada data masa lalu (*let the data speak for themselves*) dan tidak memperhatikan variabel-variabel yang lain, sehingga disebut juga dengan metode yang *atheoric* atau metode yang tidak berdasarkan pada teori, seperti pada persamaan regresi biasa.

Metode ARIMA dinotasikan sebagai:

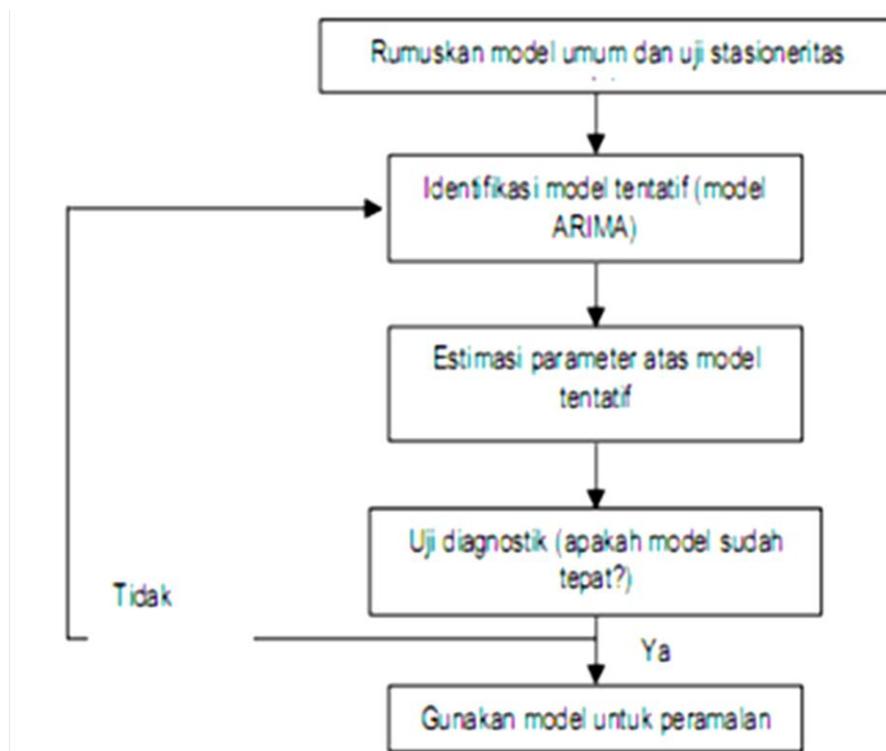
$$ARIMA(p, d, q) \dots\dots\dots(1)$$

dengan,

p = orde atau derajat *autoregressive* (AR),

d = orde atau derajat *differencing* (pembedaan), dan

q = orde atau derajat *moving average* (MA).



Gambar 1 Metodologi Box-Jenkins untuk model ARIMA (Kardoyo dan Kuncoro, 2002)

2.2 *Principal Component Analysis (PCA)*

Dalam statistika Analisis Komponen Utama atau *Principal Component Analysis (PCA)* adalah teknik yang digunakan untuk menyederhanakan suatu data dengan cara mentransformasi linier sehingga terbentuk sistem koordinat baru dengan varians maksimum. PCA dapat digunakan untuk mereduksi dimensi suatu data tersebut secara signifikan (Smith, 2002).

PCA juga dikenal dengan transformasi *Karhunen-Loeve* (untuk menghormati Kari Karhunen dan Michel Loeve) atau transformasi *Hotelling* (untuk menghormati Horold Hotelling). PCA adalah suatu teknik handal untuk mengekstraksi struktur dari suatu set data dengan dimensi yang cukup banyak. PCA dapat mengurangi besarnya dimensi dari data yang diobservasi menjadi dimensi yang lebih kecil tanpa kehilangan informasi yang signifikan dalam menggambarkan keseluruhan data (Kim, 1996).

Permasalahan dalam PCA adalah menemukan *eigenvalue* dan *eigenvector*. Prosedur PCA pada dasarnya adalah bertujuan untuk menyederhanakan variabel yang diamati dengan cara menyusutkan (mereduksi) dimensinya. Hal ini dilakukan dengan cara menghilangkan korelasi diantara variabel bebas melalui transformasi variabel bebas asal ke variabel baru yang tidak berkorelasi sama sekali atau yang biasa disebut dengan *principal component*. *Eigenvector* dengan *eigenvalue* yang besar memiliki peranan yang paling penting dalam proses transformasi. Oleh karena itu mereduksi dimensi dengan cara membuang *eigenvector* dan *eigenvalue* yang sangat kecil tidak akan membuat kehilangan data yang penting (Smith, 2002).

Cara untuk mendapatkan *Principal Component* dari sebuah matriks dilakukan dalam beberapa langkah. Langkah pertama yang dilakukan adalah mengurangi setiap nilai dari matriks dengan rata-rata nilai dari setiap masing-masing dimensi. Langkah berikutnya adalah menghitung matriks kovariannya dan dilanjutkan dengan mencari *eigenvector* dan *eigenvalue* dari matriks kovarian tersebut. Kemudian *eigenvector* yang ada diurutkan berdasarkan *eigenvalue* dari yang paling besar ke yang paling kecil. Dengan memilih *eigenvector* yang memiliki *eigenvalue* terbesar maka kita telah mendapatkan *Principal Component* dari matriks awal dan kita dapat membentuk *feature vector*. *Feature vector* berguna untuk mendapatkan data dari matriks awal yang bersesuaian dengan *eigenvector* yang telah kita pilih dengan cara mengalikan transpose dari *feature vector* dengan transpose dari matriks yang datanya sudah dikurangi dengan nilai rata-rata.

Soemartini (2008) menyatakan bahwa keuntungan penggunaan PCA dibandingkan metode lain adalah sebagai berikut:

1. Dapat menghilangkan korelasi secara bersih (korelasi = 0).
2. Dapat digunakan untuk segala kondisi data/penelitian.
3. Dapat dipergunakan tanpa mengurangi jumlah variabel asal.
4. Walaupun metode regresi dengan PCA ini memiliki tingkat kesulitan yang tinggi akan tetapi kesimpulan yang diberikan lebih akurat dibandingkan dengan penggunaan metode lain.

2.3 Penyelesaian dengan *Principal Component Analysis (PCA)*

Seringkali dengan beberapa *Principal Component (PC)* sudah cukup untuk menjelaskan struktur data asli. Jika data dimensi asli sulit untuk direpresentasikan melalui grafik, maka dengan dua PC atau satu PC bisa digambarkan melalui grafik (Santosa, et al., 2007).

Misalkan terdapat data dengan susunan sebagai berikut:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \dots\dots\dots (2)$$

Dimana n adalah jumlah variabel/atribut dan m adalah banyaknya observasi. Beberapa cara untuk mentransformasi data adalah:

1. *Centering*

Mengurangi setiap data dengan rata-rata dari setiap atribut yang ada dengan rumus sebagai berikut:

$$\hat{X} = X - \bar{X} \dots\dots\dots (3)$$

\hat{X} adalah vektor hasil setelah *centering*, X adalah vektor kolom dan \bar{X} adalah rata-rata dari kolom yang bersangkutan. Proses tersebut dilakukan untuk semua kolom dari $i = 1$ sampai $i = n$, dari data yang sudah di-*centering* dapat dilakukan beberapa operasi untuk mendapatkan beberapa besaran baru.

a. Matriks *Scatter*

Scatter adalah jarak antara variabel X_1 atau X_2 atau X_i dengan X_3 dan seterusnya. *Scatter* antara variabel dengan variabel yang lain bisa dihitung dengan rumus:

$$Scatter = \hat{X} \hat{X} \dots\dots\dots (4)$$

b. Matriks *Kovarian*

Seperti *Scatter* tapi setiap entri dalam matriks dibagi dengan jumlah data m .

$$\text{Kovarian}(C) = \frac{\hat{X}\hat{X}'}{m-1} \dots\dots\dots(5)$$

2. Normalisasi

Setelah *centering* bisa dilanjutkan dengan proses berikutnya yaitu membagi setiap data yang sudah di *centering* dengan standar deviasi dari atribut yang bersangkutan.

$$\hat{X} = \frac{X - \hat{X}}{\sigma_x} \dots\dots\dots(6)$$

3. *Scaling*

Scaling adalah prosedur merubah data sehingga berada dalam skala tertentu. Skala ini bisa antara (0,1), (-1,1) atau skala lain yang dikehendaki dengan batas bawah (BB) adalah 0 dan batas atas (BA) adalah 1. Jika nilai maksimum tiap kolom adalah X_{max} dan nilai minimumnya adalah X_{min} , untuk mengubah ke skala baru, untuk setiap data bisa dilakukan operasi:

$$A\hat{X} = \frac{X - X_{min}}{X_{max} - X_{min}} * (BA - BB) + BB \dots\dots\dots(7)$$

3. Metode Penelitian

3.1 Analisis Sistem Peramalan Konsumsi Listrik

Penelitian ini digunakan untuk membantu menentukan faktor-faktor apa saja yang berpengaruh dalam peramalan konsumsi listrik pada suatu area (wilayah) dengan periode tahun tertentu. Faktor-faktor yang digunakan dalam peramalan konsumsi listrik antara lain faktor jumlah penduduk, faktor pertumbuhan Produk Domestik Regional Bruto (PDRB), faktor pertumbuhan industri dan data demografi konsumsi listrik yang meliputi pelanggan rumah tangga, pelanggan industri, pelanggan bisnis, pelanggan sosial dan pelanggan publik.

Target berisi data periode mendatang berdasarkan rumus Arima Box Jenkins yang ditentukan melalui proses *autocorrelation function* (ACF) dan *partial autocorrelation function* (PCF) yang digunakan untuk menentukan *lag-lag* yang mempengaruhi periode-periode mendatang.

3.2 Proses Reduksi Faktor dengan PCA

Problem dalam PCA adalah menemukan *eigenvalue* dan *eigenvector*. Prosedur PCA pada dasarnya adalah bertujuan untuk menyederhanakan variabel yang diamati dengan cara menyusutkan (mereduksi) dimensinya. Cara untuk mendapatkan *Principal Component* dari sebuah matriks dilakukan dalam beberapa langkah yaitu:

1. Mengurangi setiap nilai dari matriks dengan rata-rata nilai dari setiap masing-masing dimensi. Ukuran matrik ditentukan oleh data masukan yang disesuaikan dengan jenis peramalannya. Pada proses PCA ditentukan berdasarkan seleksi periode tahun data masukannya yang terdiri dari data pelanggan listrik, data konsumsi listrik, jumlah penduduk, pertumbuhan PDRB dan pertumbuhan industri. Kemudian matriks tersebut dicari rata-rata nilai dari masing-masing dimensi dan setiap nilai dari matrik tersebut dikurangi dengan rata-ratanya.
2. Menghitung matriks kovariannya.
Sebelum menghitung matriks kovariannya terlebih dahulu harus dihitung matriks *scatter*-nya. setelah didapatkan nilai matriks *scatter*-nya baru dihitung matriks kovariannya.
3. Mencari *eigenvector* dan *eigenvalue* dari matriks kovarian tersebut.
Proses dari PCA menggunakan vektor-vektor yang disebut dengan *eigenvector* dan nilai-nilai yang disebut dengan *eigenvalue* untuk mendapatkan fitur yang paling signifikan pada *dataset*.

3.3 Proses Arima Box Jenkins

Proses Arima Box Jenkins dimaksudkan untuk menentukan *lag-lag* atau suku-suku mana yang berpengaruh pada peramalan dengan menggunakan empat tahapan utama proses Box Jenkins sebagai berikut:

1. Rumuskan model umum dan uji stasioneritasnya.
Kebutuhan listrik pada suatu periode ditentukan oleh periode-periode yang lalu. Proses PCA digunakan untuk mendapatkan hasil berupa faktor-faktor yang berpengaruh terhadap peramalan, dari faktor yang berpengaruh tersebut kalau datanya belum stasioner harus *differencing* terlebih dahulu sehingga datanya menjadi stasioner.
2. Identifikasi model tentatif (model Arima).
Model Arima (p, d, q) dapat diidentifikasi menggunakan korelogram (*correlogram*) dan kolerogram parsial (*partial correlogram*)

3. Estimasi model

Estimasi parameter autoregresi dan komponen *moving average* yang ada di dalam model dengan menggunakan metode kuadrat terkecil (*least square*) ataupun estimasi nonlinier.

4. Uji diagnostik

Diagnosis terhadap kualitas model apakah sudah sesuai dengan datanya. Caranya adalah dengan menguji apakah residual hasil estimasi sudah bersifat *white noise*. Bila residualnya sudah *white noise*, berarti modelnya sudah tepat dan dapat digunakan untuk *forecast* terhadap data masa yang akan datang, bila residual hasil estimasi belum *white noise* maka kita harus mencari model Arima lainnya.

3.4 Normalisasi Data

Data yang diperoleh kemudian dinormalisasi ke dalam bentuk $0 < x < 1$. Normalisasi data adalah proses pengubahan data menjadi *data range*-nya antara 0 sampai 1 karena fungsi aktivasi yang digunakan adalah fungsi *tansig*.

Rumus yang digunakan pada proses ini adalah:

$$A\hat{X} = \frac{X - X \min}{X \max - X \min} * (BA - BB) + BB \dots\dots\dots(8)$$

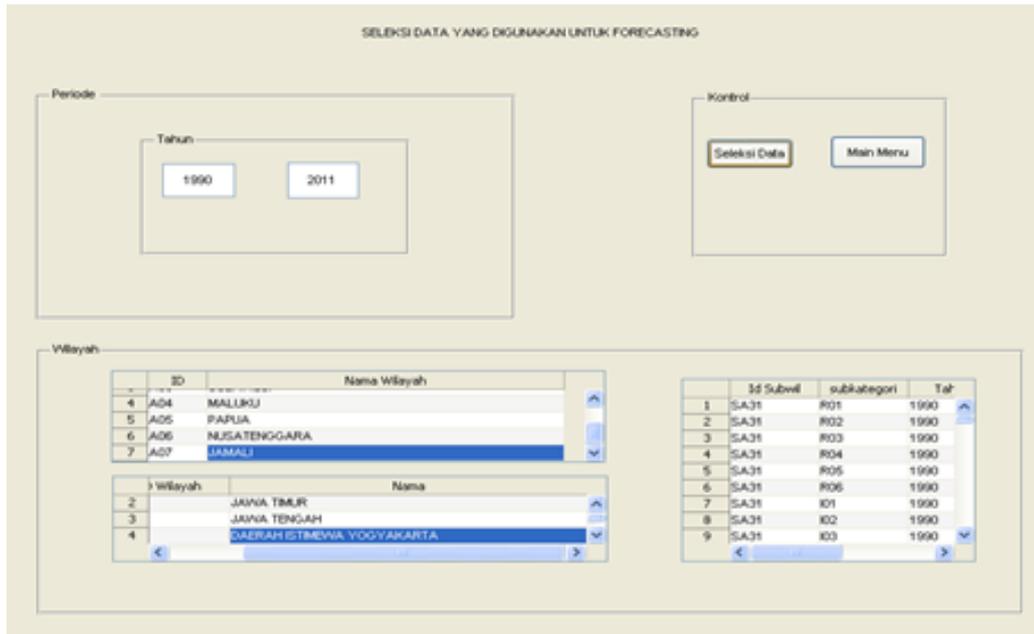
Keterangan:

- $A\hat{X}$ = data yang telah dinormalisasi
- BA = batas atas
- BB = batas bawah
- X_{max} = nilai maksimum
- X_{min} = nilai minimum
- X = data yang belum dinormalisasi

4. Hasil dan Pembahasan

4.1 Seleksi Data

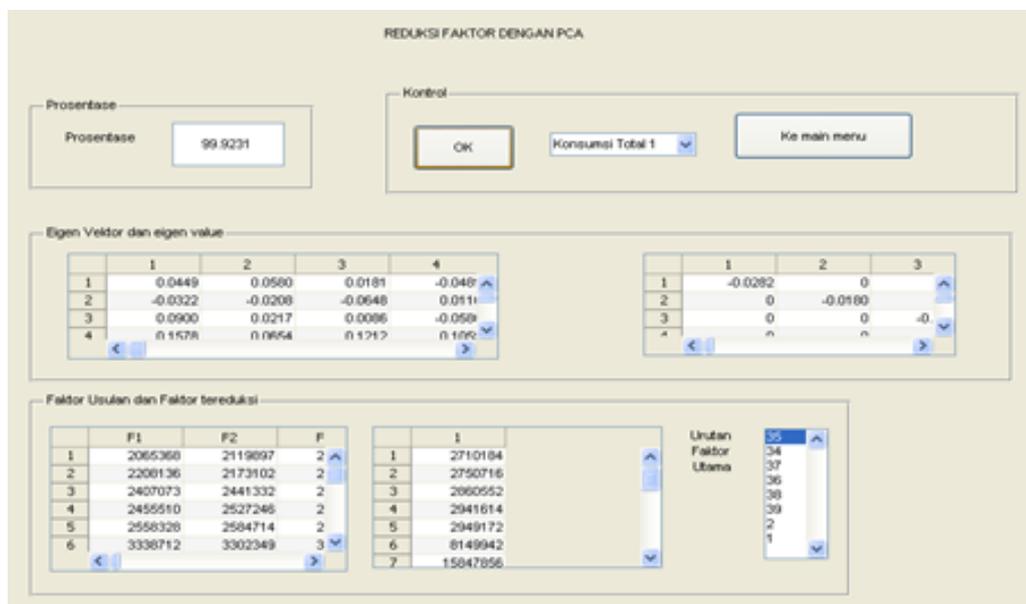
Seleksi terhadap data yang akan digunakan dalam proses peramalan terdiri dari pemilihan wilayah dan subwilayah, pemilihan periode tahun dari data histori yang akan digunakan untuk proses peramalan. Tampilan menu seleksi data untuk *forecasting* dapat dilihat pada Gambar 2.



Gambar 2 Tampilan menu seleksi data

4.2 Hasil Proses Reduksi Faktor dengan PCA

Reduksi faktor (PCA), bagian ini berisi tentang prosentase dari data yang akan digunakan untuk peramalan, kategori peramalan, faktor usulan dan faktor tereduksi serta *eigen vector* dan *eigen value*. Tampilan reduksi faktor (PCA) seperti terlihat pada Gambar 3.



Gambar 3 Tampilan menu reduksi faktor (PCA)

Reduksi faktor dengan menggunakan metode PCA seperti yang terlihat pada Gambar 3 sudah dapat menampilkan faktor-faktor dominan yang mempengaruhi konsumsi listrik pada suatu area (wilayah) tertentu dan periode

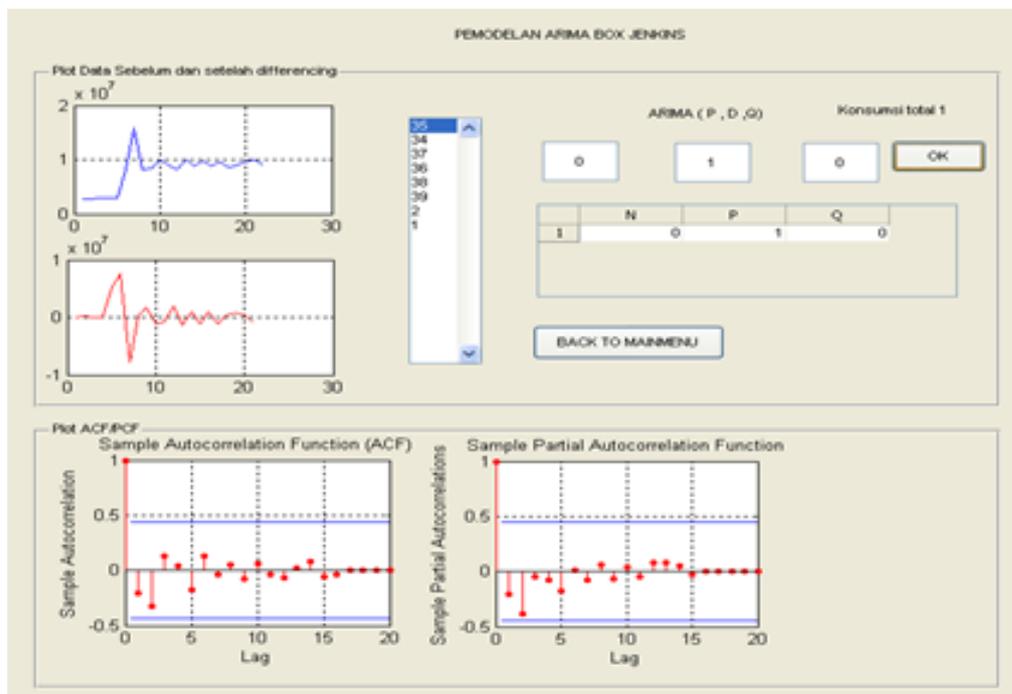
tahun tertentu berdasarkan pilihan kategori yang ingin diramalkan. Berdasarkan urutan faktor utama inilah yang dapat dijadikan masukan ke proses Arima Box Jenkins untuk menentukan *lag-lag* yang berpengaruh terhadap peramalan.

Pada proses PCA sudah dapat diketahui faktor-faktor dominan yang mempengaruhi konsumsi listrik pada suatu area (wilayah) tertentu sesuai dengan seleksi periode tahun yang telah dipilih pada saat seleksi data. Data *input* yang digunakan untuk proses PCA terdiri dari jumlah penduduk, pertumbuhan PDRB, pertumbuhan industri dan data demografi konsumsi listrik.

Proses PCA dilakukan untuk mereduksi faktor yang berpengaruh terhadap peramalan dan setelah dilakukan proses PCA urutan faktor utama yang berpengaruh terhadap peramalan listrik konsumsi *total1* adalah faktor F35 (jumlah pelanggan R-1/TR 900 VA), F34 (jumlah pelanggan R-1/TR 450 VA), F37 (jumlah pelanggan R-1/TR 2200 VA), F36 (jumlah pelanggan R-1/TR 1300 VA), F38 (jumlah pelanggan R-2/TR >2200 – 6600), F39 (jumlah pelanggan R-3/TR > 6600), F2 (konsumsi pelanggan R-1/TR 900 VA), F1(konsumsi pelanggan R-1/TR 450 VA).

4.3 Hasil Pemodelan Arima Box Jenkins

Pemodelan Arima (Box Jenkins), bagian ini berisi plot data sebelum dan setelah *differencing* dalam bentuk grafik, model Arima (p, d, q), dan tampilan plot ACF/PCF seperti pada Gambar 4.



Gambar 4 Tampilan menu pemodelan Arima Box Jenkins

Data *input* yang digunakan pada proses Arima Box Jenkins diperoleh dari hasil proses PCA yaitu nilai *eigenvector* yang diurutkan berdasarkan nilai *eigenvalue* dari yang paling besar ke yang paling kecil. *Eigenvector* yang memiliki nilai *eigenvalue* yang paling besar yang nantinya akan dicari *lag-lag* yang berpengaruh terhadap peramalan. Syarat pada proses Arima yaitu data harus stasioner untuk menstasionerkan data harus dideferencing terlebih dahulu.

Hasil dari proses Arima Box Jenkins untuk peramalan listrik konsumsi *total1* adalah (0, 1, 0) artinya bahwa model sudah didiferen sebanyak 1 kali, memiliki 0 komponen *autoregresif* dan 0 komponen *moving average*.

5. Kesimpulan

Hasil simulasi menunjukkan bahwa pada proses reduksi faktor dengan menggunakan *Principal Component Analysis* (PCA) sudah dapat diketahui faktor-faktor dominan yang mempengaruhi peramalan konsumsi listrik *total1* dan pemodelan Arima Box Jenkins sudah dapat digunakan untuk menentukan *lag-lag* data *input*. Dari hasil simulasi proses ARIMA Box Jenkins untuk peramalan listrik konsumsi *total1* dihasilkan suatu model ARIMA (*Autoregressive Integrated Moving Average*) 0, 1, 0.

Daftar Pustaka

- Hanke, J. E., Reitsch, A. G. & Wichern, D. W., 2001. *Business Forecasting*. Upper Saddle River, NJ: Prentice Hall.
- Kardoyo, H. & Kuncoro, M., 2002, Analisis Kurs Valas dengan Pendekatan Box-Jenkins: Studi Empiris Rp/US\$ dan Rp/Yen, 1983.2-2000.3, *Jurnal Ekonomi Pembangunan*, 7(1), pp. 7-20.
- Kim, K., 1996. Face Recognition using Principle Component Analysis. *International Conference on Computer Vision and Pattern Recognition*, pp. 586-591.
- Makridakis, S., Wheelwright, S. C. & McGee, V. E., 1999. *Metode dan Aplikasi Peramalan*. Jakarta : Binarupa Aksara.
- Santoso, I., Effendi, U. & Fauziya, C., 2007. Penerapan Jaringan Syaraf Tiruan Untuk Peramalan Permintaan Komoditas Karet di PT.Perkebunan Nusantara XII Surabaya. *Jurnal Teknologi Pertanian*, 8(1), pp. 46-54.
- Smith, L. I., 2002. *A Tutorial on Principal Component Analysis*. [Online] Available at: http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf [Accessed 14/02/2012].
- Soemartini, 2008. Principal Component Analysis (PCA) Sebagai Salah Satu Metode untuk Mengatasi Masalah Multikolinearitas. *Tesis*. Bandung: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Padjadjaran.