Data Mining Peramalan Konsumsi Listrik dengan Pendekatan *Cluster Time Series* sebagai *Preprocessing*

M. Alfan Alfian Riyadi, Kartika Fithriasari, dan Dwiatmono
Jurusan Statistika, Fakultas MIPA, Institut Teknologi Sepuluh Nopember
Jalan Arief Rahman Hakim, Surabaya 60111
e-mail: kartika_f@statistics.its.ac.id, dwiatmono@statistika.its.ac.id

Abstrak-- ***Kondisi big data dan data time series memiliki permasalahan tersendiri didalam mengolah suatu data. Terlebih lagi data tersebut juga multivariabel. Salah satu permasalahan yang terjadi adalah ketika proses identifikasi model yang sesuai untuk tiap series. Beberapa metode time series seperti ARIMA dan ANN membutuhkan proses identifikasi untuk menentukan orde ARIMA dan input ANN yang akan digunakan. Melakukan identifikasi satu per satu tiap series tidak mungkin dilakukan. Untuk itu perlu dilakukan preprocessing data salah satunya dengan menggunakan cluster. Metode ukuran kesamaan dalam cluster time series salah satunya adalah autocorrelation based distance. Dari masing-masing cluster yang dihasilkan dipilih salah satu anggota untuk dilakukan permodelan. Diharapkan model yang dihasilkan mewakili anggota cluster secara keseluruhan. Metode peramalan yang digunakan pada penelitian kali ini adalah ARIMA dan ANN dengan studi kasus data benchmark konsumsi listrik di Portugal. Hasil yang diperoleh adalah dihasilkan sebanyak tujuh cluster dengan anggota cluster terbanyak pada cluster ke empat yakni sebanyak 120 client. Selanjutnya model peramalan dengan menggunakan ANN lebih baik dibandingkan ARIMA. Diperoleh sebanyak 259 dari 348 client yang menyatakan bahwa permodelan dengan menggunakan ANN lebih baik dibandingkan ARIMA***

Kata Kunci—Breast ANN, ARIMA, Autocorrelation based distance, Cluster time series

I. PENDAHULUAN

Kondisi *big data* dan data *time series* memiliki permasalahan tersendiri didalam mengolah suatu data. Terlebih lagi data tersebut juga multivariabel. Biasanya untuk kasus data *time series* dilakukan suatu permodelan peramalan guna menduga kejadian yang berada dimasa yang akan datang. Beberapa metode peramalan yang dapat digunakan adalah ARIMA (*Autoregressive Integrated Moving Average)* dan ANN (*Artificial Neural Network)*. Namun karena data yang besar dan multivariabel memiliki permasalahan tersendiri dengan metode tersebut. Karena pada metode tersebut perlu dilakukan identifikasi guna menentukan orde ARIMA dan input ANN yang sesuai, sehingga sangat sulit bila dilakukan satu per satu [1],[2]. Untuk itu perlu dilakukan *preprocessing data* guna mempermudah menentukan orde ARIMA dan input ANN yang sesuai.

Salah satu *preprocessing data* pada *data mining* adalah melakukan *cluster.* Tujuan dari *cluster* adalah untuk mencari kesamaan karakteristik dari *set* data dengan memaksimumkan ketidaksamaan antar *cluster* dan meminimumkan kesamaan di dalam *cluster*. Ukuran kesamaan sangat diperlukan untuk *cluster* khususnya untuk data *time series* [3].

Metode ukuran kesamaan dalam *cluster time series* salah satunya adalah *autocorrelation based distance*. Penelitian dengan menggunakan metode ini pernah dilakukan untuk kasus pengelompokan *interest rate* beberapa negara. Pada penelitian tersebut terbentuk lima *cluster* [4].

Pada penelitian ini dilakukan *preprocessing data* dengan menggunakan *cluster time series* guna mempermudah proses identifikasi model ARIMA dan ANN yang sesuai. Ukuran kesamaan yang digunakan pada penelitian ini adalah *autocorrelation based distance* dengan studi kasus konsumsi listrik untuk setiap *client* di Portugal. Selanjutnya dilakukan perbandingan hasil ramalan dari ARIMA dan ANN untuk masing-masing *cluster*.

II. TINJAUAN PUSTAKA

A. Analisis Cluster Time Series

Analisis *cluster* atau analisis kelompok merupakan sebuah metode analisis untuk mengelompokkan objek-objek pengamatan menjadi beberapa kelompok sehingga akan diperoleh kelompok dimana objek-objek dalam satu kelompok mempunyai banyak persamaan sedangkan dengan anggota kelompok yang lain memiliki banyak perbedaan [5]. Ukuran kesamaan merupakan suatu hal yang paling utama dalam melakukan analisis *cluster*. Untuk kasus *cluster* pada data time series, salah satu ukuran kesamaan yang dapat digunakan adalah *autocorrelation-based distance*. Proses stokastik $\{Xt\}_{t=1}^{T}$, dimana *T* adalah banyaknya data *time series*. Jarak *autocorrelation* terdiri dari vektor $\rho_{X_T} = \left(\rho_{1X_T}, \rho_{2X_T} \dots, \rho_{LX_T}\right)'$ yang menyatakan vektor *autocorrelation X* dari lag ke-1 sampai lag ke-*L* dan $\rho_{Y_T} = \left(\rho_{1Y_T}, \rho_{2Y_T} \dots, \rho_{LY_T}\right)'$juga merupakan vektor *autocorrelation Y* dari lag ke-1 sampai lag ke-*L*. Selanjutnya vektor *autocorrelation* dari X_T dan Y_T di estimasi oleh $\hat{\rho}_{X_T}$ dan $\hat{\rho}_{Y_T}$, untuk beberapa *L* seperti $\rho_{iX_T} \approx 0$, dengan *i*=1,2,..,*L* dan $\rho_{iY_T} \approx 0$ untuk setiap $i > L$.

Kemudian formula dari ukuran tersebut ditunjukkan pada persamaan (1) sebagai berikut [4]:

$$d_{ACF}(X_T, Y_T) = \sqrt{\left(\hat{\rho}_{X_T} - \hat{\rho}_{Y_T}\right)' \Omega^{-1} (\hat{\rho}_{X_T} - \hat{\rho}_{Y_T})} \quad (1)$$

dimana:

$d_{ACF}(X_T, Y_T)$ = jarak *autocorrelation* vektor X_T dan Y_T

$\hat{\rho}_{X_T}$ = estimasi vektor *autocorrelation* X_T

$\hat{\rho}_{Y_T}$ = estimasi vektor *autocorrelation* Y_T

Ω = matriks bobot.

Apabila jarak ACF tidak menggunakan bobot maka matriks bobot berupa matriks identitas.

B. Algoritma Complete Lingkage

Algoritma *complete linkage* merupakan algoritma hirarki untuk membentuk *cluster* berdasarkan jarak terjauh antar objek [5]. Persamaan guna menentukan jarak antara kelompok (i,j) dengan k yaitu pada persamaan (2) berikut

$$d_{(i,j)k} = \max (d_{ik}, d_{jk}) \quad (2)$$

dimana:

d_{ik} = jarak antara kelompok i dan k

d_{jk} = jarak antara kelompok j dan k

$d_{(ij)k}$ = jarak antara kelompok ij dan kelompok

C. Model ARIMA

Model ARIMA (p, d, q) yang dikenalkan oleh Box dan Jenkins dengan orde p sebagai operator dari AR, orde d merupakan *differencing*, dan orde q sebagai operator dari MA. Dimana bentuk persamaan *differencing* $Z_t = Y_t - Y_{t-1}$. Model ini digunakan untuk data *time series* yang telah di *differencing* atau sudah stasioner dalam *mean*, dimana d adalah banyaknya hasil *differencing*. bentuk persamaan untuk model ARIMA adalah:

$$\varphi_p(B)(1-B)^d Y_t = \theta_0 + \theta_q a_q(B) a_t \quad (3)$$

Parameter yang dihasilkan pada model ARIMA kemudian diuji tingkat signifikansi. Berikut adalah rumusan hipotesis pengujian signifikansi parameter model ARIMA:

Hipotesis:

H_o: $\delta = 0$ (parameter model tidak signifikan)

H_1: $\delta \neq 0$ (parameter model signifikan)

Statistik uji:

$$t = \frac{\hat{\delta}}{\mathrm{S.E}(\hat{\delta})} \quad (4)$$

dengan keputusan untuk menolak H_o apabila $|t_{hitung}| > t_{\frac{\alpha}{2}, n-n_p}$ atau apabila *p-value* $< \alpha$, yang berarti bahwa parameter signifikan[1].

D. Pengujian Diagnostik Residual Model ARIMA

Terdapat dua hal yang dilakukan guna menguji diagnostik residual model ARIMA yakni residual *white noise* dan berdistribusi normal. Pengujian asumsi residual *white noise* merupakan pengujian yang digunakan untuk melihat apakah residual yang dihasilkan sudah independen atau tidak. Perumusan hipotesis asumsi residual *white noise* sebagai berikut.

H_0: $\rho_1 = \rho_2 = \cdots = \rho_K = 0$ (residual bersifat *white noise*)

H_1: paling tidak terdapat satu k di mana $\rho_k \neq 0$ (residual tidak bersifat *white noise*)

Statistik uji:

$$Q = n(n+2) \sum_{k=1}^{K} \frac{\rho_k^2}{n-k} \quad (5)$$

dengan,

n = $p + q$

ρ_k = koefisien autokorelasi sisaan pada lag ke-k

K = *lag* maksimum.

Keputusan untuk menerima hipotesis nol didasarkan pada apabila Q bernilai lebih kecil daripada X^2_{k-p-q} pada taraf nyata α di mana p dan q adalah ordo dari ARIMA atau apabila *p-value* dari statistik uji Q bernilai lebih besar daripada taraf nyata α [6]. Kemudian berikut merupakan rumusan pengujian hipotesis untuk residual berdistribusi normal.

Hipotesis:

H_0: $F(x) = F_0(x)$ (sisaan berdistribusi normal)

H_1: $F(x) \neq F_0(x)$ (sisaan tidak berdistribusi normal)

Statistik uji:

$$D_{hit} = \sup_x \left| F_n(x) - F_0(x) \right| \quad (6)$$

dengan:

$F_n(x)$: fungsi distribusi frekuensi kumulatif yang dihitung dari data sampel

$F_0(x)$: fungsi distribusi frekuensi kumulatif distribusi normal

$\sup_x$: nilai maksimum semua *x* dari $|F_n(x) - F_0(x)|$

Apabila nilai $D_{hit} > D_{(a,n)}$, maka dapat diputuskan bahwa H_0 ditolak dan dapat dikatakan bahwa sisaan tidak berdistribusi normal [7].

E. Artificial Neural Network (ANN)

*Artificial Neural Networ*k (ANN) adalah sistem pemrosesan informasi yang memiliki karakteristik mirip dengan jaringan syaraf biologi khususnya otak manusia. ANN merupakan jaringan dari unit perhitungan sederhana yang disebut *neuron* dimana sangat terkoneksi dan terorganisasi didalam *layer.* Setiap *neuron* mengolah informasi dari *input* yang kemudian diteruskan menjadi *output* [2]. ANN ditentukan oleh 3 hal:

1. Pola hubungan antar neuron (disebut arsitektur jaringan)
2. Metode untuk menentukan bobot penghubung (metode *training/learning*/algoritma)
3. Fungsi aktivasi

Pada ANN terdapat algoritma pembelajaran salah satunya *backpropagation*. *Backpropagation* merupakan algoritma pembelajaran *neural network* dan biasanya digunakan oleh *perceptron* dengan banyak lapisan untuk mengubah bobot-bobot yang terhubung dengan neuron-neuron pada lapisan tersembunyinya [8]. Pembelajaran dari jaringan *backpropagation* terdiri dari tiga tahapan yaitu menghitung arah maju dari pola input pembelajaran (*feedforward)*, menghitung *error backpropagation* dan menentukan peubah bobot [9].

F. Kriteria Kebaikan Model

Kriteria kebaikan model digunakan untuk menentukan model mana yang terbaik. Salah satu kriteria kebaikan model yang umum digunakan adalah MAPE (*Mean Absolute Percentage Error*). Keunggulan dari kriteria kebaikan ini yakni hasil jumlahan *error* tidak saling mengeliminasi karena nilai tersebut diabsolutkan

[10]. Rumus untuk mendapatkan nilai MAPE adalah sebagai berikut:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| 100\% \qquad (7)$$

dengan:
Y_t = nilai sesungguhnya
$\hat{Y}_t$ = nilai ramalan
n = banyak ramalan.

III. METODOLOGI PENELITIAN

A. *Sumber Data Penelitian*

Sumber data pada penelitian ini adalah data yang didonor oleh Artur Trindade dengan judul *Electricity Load Diagrams 2011-2014* yang diunggah pada *website* donor data http://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014 dengan unit eksperimen *client* listrik dan satuan yang digunakan adalah Kwh (*Kilowatt per Hour*).

B. *Langkah Analisis*

Dalam melakukan penelitian harus dilakukan analisis yang tepat. Berikut ini merupakan langkah-langkah penelitian:

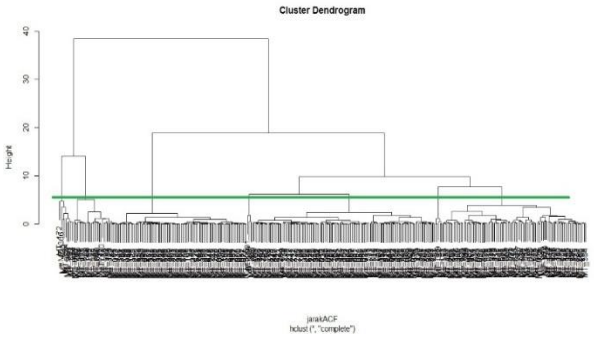
1. Melakukan agregasi data konsumsi listrik yang mulanya tiap 15 menit menjadi tiap hari.
2. Melakukan analisis *cluster time series.*
3. Membuat *time series plot* untuk masing-masing *cluster* dengan mengambil salah satu contoh anggota *cluster*. Kemudian melakukan analisis berdasarkan *time series plot* yang telah dibuat.
4. Membagi data menjadi *in sample* dan *out sample*. Data *in sample* yang digunakan adalah konsumsi listrik dari tanggal 1 Januari hingga 30 November 2014 sedangkan *out sample* dari tanggal 1 Desember hingga 21 Desember.
5. Mengidentifikasi model ARIMA untuk tiap *cluster* dengan mengambil salah satu *series* pada tiap *cluster.* Pengambilan salah satu series diambil sembarang tanpa ada kriteria tertentu.
6. Memodelkan dengan ANN dengan *input* model ARIMA terbaik.
7. Membandingkan hasil permodelan dengan ARIMA dan ANN.
8. Menghitung kesesuaian model untuk setiap anggota *cluster* dengan menyesuaikan nilai MAPE yang dihasilkan dari ARIMA dan ANN terbaik.
9. Mendapatkan kesimpulan.

IV. ANALISIS DAN PEMBAHASAN

Bab berikut ini akan dilakukan penentuan model peramalan konsumsi listrik tiap *client* yang terbaik dengan analisis *cluster time* series sebagai *preprocesssing data.*

A. *Analisis Cluster Time Series*

Jarak yang digunakan pada penelitian ini adalah *autocorrelation based distance* dengan algoritma penentuan anggota *cluster* yakni algoritma *complete linkage*. Maktriks bobot yang digunakan pada penelitian kali ini adalah matriks identitas hal tersebut dikarenakan tidak adanya dasar untuk menentukan bobot. Hasil dari penentuan banyak *cluster* yang dihasilkan berdasarkan dendogram yang disajikan pada Gambar 1.



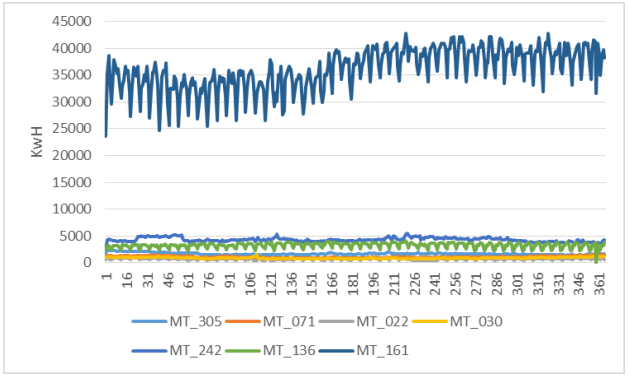
Gambar 1 Dendogram *Complete Linkage Client*

Dari Gambar 1 diperoleh sebanyak tujuh *cluster* yang digunakan pada penelitian ini. Banyak anggota tiap *cluster* ditampilkan pada Tabel 1 berikut.

Tabel 1. Banyak Anggota Tiap *Cluster*

Cluster	Banyak Anggota	*Mean*	Koef. Varians
1	4	2131,83	0,818
2	85	27511,61	4,840
3	9	4719,64	2,342
4	120	16384,23	3,591
5	95	8227,38	3,562
6	30	2407,22	0.910
7	5	29875,13	0,680

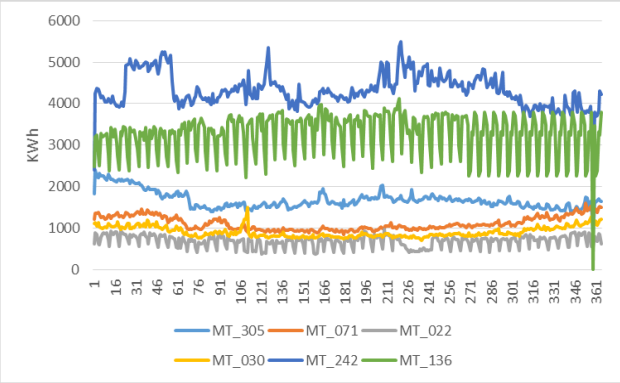
Cluster ke empat memiliki anggota paling banyak yakni 120 *client.* Sedangkan *cluster* satu memiliki anggota paling sedikit yakni 4 *client*. Kemudian rata-rata konsumsi tertinggi terjadi pada *cluster* tujuh yaitu sebesar 29875,13 kwh dan *cluster* satu memiliki rata-rata konsumsi terendah yaitu 2131,83 kwh. Koefisien variansi menunjukkan variabilitas dari data. Dapat dilihat bahwa *cluster* dua memiliki variansi yang paling tinggi sedangkan yang terendah terjadi pada *cluster* 7. Selanjutnya dari masing-masing *cluster* dilihat *time series plot* untuk mengetahui karakteristik pola data yang dihasilkan tiap *cluster.* Pada penelitian ini *cluster* pertama diwakili oleh *client* 305. Kemudian berturut-turut sampai *cluster* ketujuh yakni *client* 71, 22, 30, 242,136 dan 161. Pemilihan *client* yang digunakan, diambil secara sembarang tanpa mempertimbangkan kriteria tertentu.



Gambar 2 *Time Series Plot* Tiap *Cluster*

Cluster tujuh memeliki tingkat konsumsi listrik yang paling tinggi dibandingkan dengan *cluster* yang lain. Juga terlihat dengan jelas pola musiman yang dihasilkan yakni

tiap tujuh hari. Namun karena pada Gambar 2 tidak terlihat dengan jelas hasil *time series plot* untuk *cluster* yang lainnya, maka pada Gambar 3 disajikan *time series plot* untuk tiap *cluster* tanpa melibatkan *cluster* tujuh.



Gambar 2 *Time Series Plot* Tiap *Cluster* Tanpa Melibatkan *Cluster Tujuh*

Cluster tiga dan enam memiliki pola data yang cenderung sama namun berbeda tingkat konsumsi listrik tiap pada harinya. Pada *cluster* enam terjadi penurunan konsumsi listrik yang tinggi. Penurunan konsumsi tersebut, terjadi pada tanggal 25 Desember, dimana pada tanggal tersebut bertepatan dengan hari Natal.

B. Identifikasi Model ARIMA

Identifikasi model ARIMA tiap *cluster* dilakukan dengan cara mengambil salah satu anggota *cluster* kemudian dari anggota tersebut diidentifikasi guna menentukan orde ARIMA mana yang sesuai untuk *cluster* tersebut. Anggota *cluster* yang digunakan untuk identifikasi sama dengan yang digunakan untuk membuat *time series plot* yakni *cluster* pertama diwakili oleh *client* 305. Kemudian berturut-turut sampai *cluster* ketujuh yakni *client* 71, 22, 30, 242,136 dan 161. Berikut merupakan hasil dari identifikasi model ARIMA terbaik untuk setiap *cluster*.

Tabel 2. Model ARIMA Terbaik Tiap *Cluster*

Cluster ke-	Model ARIMA
1	$(1,1,1)(2,1,1)^7$
2	$(0,1,1)(3,1,1)^7$
3	$(2,0,0)(2,1,0)^7$
4	$(0,1,1)(3,1,1)^7$
5	$(0,1,2)$
6	$(2,0,0)(2,1,1)^7$
7	$(1,1,2)(1,1,1)^7$

Namun jika dilihat dari hasil pengujian asumsi residual pada Tabel 3 dan 4, dari semua model ARIMA yang dihasilkan hanya *cluster* tujuh yang memenuhi asumsi white noise dan distribusi normal. Tetapi penelitian ini tidak mempermasalahkan asumsi residual yang tidak terpenuhi. Sebab, pada kasus peramalan yang lebih diutamakan adalah kehandalan model untuk memperoleh hasil ramalan yang tepat [11].

Tabel 3. Uji Residual *White Noise* Model ARIMA Terbaik

Cluster ke-	*White Noise Lag*		
	6	12	18
1	0,9775	0,9952	0,9001
2	0,8302	0,7055	0,5255
3	0,1215	0,2460	0,3185
4	0,0609	0,2216	0,4681
5	0,9833	0,9919	0,9577
6	0,9096	0,4584	0,6385
7	0,4003	0,0907	0,2583

Tabel 3. Uji Residual *White Noise* Model ARIMA Terbaik (Lanjutan)

Cluster ke-	*White Noise Lag*		
	24	30	36
1	0,8233	0,3991	0,2272
2	0,7401	0,8041	0,9076
3	0,0112	0,0023	0,0030
4	0,5207	0,6370	0,7529
5	0,7477	0,8934	0,9426
6	0,7251	0,9203	0,4328
7	0,3893	0,9418	0,2827

Tabel 3. Uji Residual Distribusi Normal Model ARIMA Terbaik

Cluster ke-	*P-Value*	Keterangan
1	0,0000	Tolak H_0
2	0,0000	Tolak H_0
3	0,0000	Tolak H_0
4	0,0000	Tolak H_0
5	0,0000	Tolak H_0
6	0,0000	Tolak H_0
7	0,2259	Gagal tolak H_0

C. Identifikasi Model ANN

Permodelan peramalan menggunakan ANN dilakukan dengan dua cara yakni dengan menggunakan *input* model orde AR ARIMA terbaik dan menggunakan *input autoregressive* dari satu hingga tujuh. Banyak *neuron* untuk permodelan ANN dengan input orde AR ARIMA terbaik adalah lima sedangkan permodelan ANN dengan *input autoregressive*, maksumun orde AR adalah tujuh dengan kombinasi *hidden neuron* sebanyak sepuluh. Berikut merupakan hasil permodelan ANN dengan *input* ARIMA terbaik.

Tabel 4. Kebaikan Model ANN *Input* ARIMA

Cluster ke-	*Input*	MAPE (%)	
		In Sample	*Out Sample*
1	Y_{t-1}, Y_{t-7}, Y_{t-14}	2,5448	3,2739
2	Y_{t-1}, Y_{t-7}, Y_{t-14}, Y_{t-21}	2,7184	2,8503
3	Y_{t-1}, Y_{t-2}, Y_{t-7}, Y_{t-14}	10,2091	7,9868
4	Y_{t-1}, Y_{t-7}, Y_{t-14}, Y_{t-21}	3,5342	2,5821
5	Y_{t-1}, Y_{t-2}	3,0515	3,8540
6	Y_{t-1}, Y_{t-2}, Y_{t-7}, Y_{t-14}	3,0048	0,8512
7	Y_{t-1}, Y_{t-2}, Y_{t-7}	3,1102	1,9052

Berdasarkan Tabel 4. Diperoleh bahwa model ANN untuk *cluster* enam menghasilkan nilai yang lebih kecil dibandingkan pada *cluster* lainnya. Sedangkan untuk *cluster* tiga baik dari MAPE *in sample* maupun *out sample* menghasilkan nilai yang paling besar dibandingkan dengan *cluster* lainnya. Pada *cluster* lima, model ANN yang dihasilkan menggunakan input yang paling sedikit yakni sebanyak dua *input*. Sedangkan *input* terbanyak

terjadi pada *cluster* dua, tiga dan enam dengan *input* sebanyak empat. Selanjutnya dilakukan permodelan ANN dengan menggunakan *input autoregressive*. Hasil dari ANN dengan menggunakan input *autoregressive* disajikan pada Tabel 4 berikut.

Tabel 5. Kebaikan Model ANN *Input Autoregressive*

Cluster ke-	*Input*	Banyak *Hidden Neuron*	MAPE (%) *In Sample*	MAPE (%) *Out Sample*
1	AR(7)	7	2,4375	4,4006
2	AR(2)	5	2,5776	7,6898
3	AR(7)	5	7,5720	15,4738
4	AR(6)	10	3,4858	4,8014
5	AR(6)	8	2,7803	3,8653
6	AR(2)	4	4,5430	12,9927
7	AR(6)	1	3,8289	5,6574

Apabila dibandingkan dengan *input* ARIMA, permodelan ANN dengan menggunakan *input* ANN dengan menggunakan *input autoregressive* menghasilkan nilai MAPE yang lebih besar. Sehingga permodelan peramalan dengan ANN untuk tiap *cluster* menggunakan *input* ARIMA.

D. Perbandingan Model ARIMA dan ANN

Dari model ARIMA dan ANN terbaik, kemudian dibandingkan guna mengetahui model mana yang cenderung lebih cocok digunakan untuk setiap anggota *cluster*. Dari Tabel 6 diperoleh bahwa jika dilihat dari MAPE *out sample*, model ANN baik pada *cluster* 1,2,3 dan 4 sisanya pada *cluster* 5,6,7 model ARIMA lebih baik. Namun model ARIMA terbaik untuk setiap *cluster* tidak memenuhi asumsi distribusi normal dan *white noise*.

Tabel 6. MAPE Model ARIMA dan ANN Terbaik

Cluster ke-	MAPE *In Sample (%)*		MAPE *Out Sample (%)*	
	ARIMA	ANN	ARIMA	ANN
1	2,1907	2,5448	4,2837	3,2739
2	2,2974	2,7184	5,0097	2,8503
3	7,9265	10,2091	11,9393	7,9868
4	3,0707	3,5342	2,6332	2,5821
5	2.8206	3,0515	3,5380	3,8540
6	2,6224	3,0048	0,0011	0,8512
7	2.8047	3,1102	1,6194	1,9052

Selanjutnya dilakukan konfirmasi terkait berapa banyak anggota *cluster* yang lebih baik dimodelkan dengan ARIMA dan ANN. Pengecekan ini didasari dari nilai MAPE untuk tiap anggota *cluster*. Hasil dari analisis ini disajikan pada Tabel 7.

Tabel 7 Banyak *Client* yang Sesuai dengan Model Peramalan

Cluster ke-	Banyak Anggota	ARIMA	ANN
1	4	1	3
2	85	11	74
3	9	4	5
4	120	29	91
5	95	23	72
6	30	19	11
7	5	2	3
Total	348	89	259

Ditinjau dari Tabel 7 model ANN secara keseluruhan menghasilkan hasil yang lebih baik dibandingkan dengan ARIMA. Dari 348 *client*, 259 diantaranya model ANN menghasilkan MAPE yang lebih kecil dibandingkan dengan model ARIMA. Namun, khusus untuk *cluster* enam, konsumsi listrik *client* pada *cluster* tersebut lebih banyak yang sesuai dimodelkan dengan model ARIMA karena dengan model tersebut lebih menghasilkan MAPE yang lebih kecil. Tetapi pada *cluster* lima, Tabel 6 menyatakan bahwa model ARIMA merupakan model terbaik untuk *cluster* lima namun setelah di konfirmasi ternyata sebagian besar *client* pada *cluster* lima lebih cocok dimodelkan dengan ANN. Hal tersebut disebabkan karena model ARIMA tiap anggota *cluster* dipaksa mengikuti hasil dari identifikasi salah satu series dari *cluster* tersebut. Sehingga MAPE yang dihasilkan pada model ARIMA untuk tiap anggota *cluster* menjadi lebih besar.

V. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis dan pembahasan didapatkan beberapa kesimpulan yakni, analisis *cluster* yang telah dilakukan terdapat tujuh *cluster*. Selanjutnya *Cluster* empat memiliki anggota paling banyak yakni 120 *client.* Sedangkan *cluster* satu memiliki anggota paling sedikit yakni 4 *client*. Kemudian rata-rata konsumsi tertinggi terjadi pada *cluster* tujuh yaitu sebesar 29875,13 kwh dan *cluster* satu memiliki rata-rata konsumsi terendah yaitu 2131,83 kwh. Berikutnya *cluster* dua memiliki variansi yang paling tinggi sedangkan yang terendah terjadi pada *cluster* 7. Bila ditinjau dari *time series plot* diperoleh bahwa *cluster* tiga, enam, dan tujuh memiliki pola musiman yang jelas. Namun berbeda tingkat konsumsi listrik yang dihasilkan. Pada *cluster* enam terjadi penurunan tingkat konsumsi listrik yang sangat signifikan yakni pada tanggal 25 Desember dimana pada hari tersebut bertepatan dengan hari natal. Kemudian Permodelan menggunakan ANN dan ARIMA menunjukkan bahwa model ANN unggul pada *cluster* 1,2,3 dan 4. Sedangkan model ARIMA unggul pada *cluster* 5,6,7. Namun model ARIMA terbaik untuk setiap *cluster* tidak memenuhi asumsi distribusi normal dan *white noise*. Setelah dikonfirmasi ulang yakni dengan memodelkan masing-masing *client* di tiap *cluster* dengan menggunakan model ARIMA dan ANN terbaik diperoleh bahwa dari 348 *client*, 259 diantaranya lebih cocok dimodelkan dengan ANN dibandingkan dengan ARIMA. Hal tersebut disebabkan karena model ARIMA tiap anggota *cluster* dipaksa mengikuti hasil dari identifikasi salah satu series dari *cluster* tersebut. Sehingga MAPE yang dihasilkan pada model ARIMA untuk tiap anggota *cluster* menjadi lebih besar.

Saran untuk penelitian selanjutnya dapat menggunakan metode ukuran kesamaan dan algoritma pembentuk *cluster* yang lain. Sebab berbeda ukuran kesamaan dan berbeda algoritma sangat memungkinkan terjadi perbedaan hasil anggota *cluster* yang terbentuk. Pada model peramalan yang digunakan pada penelitian ini juga dapat dikembangkan dengan membandingkan model peramalan yang lain. Sehingga hasil model peramalan yang diperoleh dapat menghasilkan ramalan yang lebih akurat lagi.

DAFTAR PUSTAKA

[1] Wei, W.W.S., (2006). *Time Analysis Univariate and Multivariate Methods*. Addison Wesley Publishing Company, Inc.
[2] Zhang, G.P., (2004). *Neural Network in Business Forcasting.* Idea Group Publishing.
[3] Kleist, Caroline, (2015). *Time Series Data Mining Method: A Review*, Humboldt-Universitat zu Berlin.
[4] Montero, Pablo & Jose A. Vilar. (2014). *TSclust: An R Package for Time Series Clustering*. Journal of Statistical Software Vol 62, Issue 1.
[5] Johnson, R.A and Winchern, D.W. 2007. *Applied Multivariate Analysis*. (Sixth Edition), New Jersey: Prentice Hall Inc.
[6] Cryer, J. D. (2008). Time Series Analysis with Application in R (Second Edition). New York: Springer Science Bussines Media.
[7] Daniel, W.W., (1989). *Statistika Nonparametrik Terapan.* Georgia State University. Jakarta: PT Gramedia.
[8] Kusumadewi, S. (2004). *Membangun Jaringan Syaraf Tiruan (menggunakan MATLAB & Excel Link)*. Yogyakarta: Graha ilmu.
[9] Fausett, Lauren. (1994). *Fundamental of Neural Network: Architectures, Algorithm and Applicalions*. Prantice Hall.
[10] Makridakis S., Wheelwright, Mc Gee, (1999). *Metode dan Aplikasi Peramalan*. Jakarta: Bina Rupa Aksara.
[11] Kostenko, A. V. and Rob J. Hyndman (2008). *Forecasting Without Significance Test?*.