

# Peramalan Jumlah Penumpang Kereta Api Kelas Ekonomi Kertajaya Menggunakan ARIMA dan ANFIS

Ilafi Andalita dan Irhamah

Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Institut Teknologi Sepuluh Nopember (ITS)  
Jl. Arief Rahman Hakim, Surabaya 60111  
Email : irhamahn@gmail.com

**Abstrak**—Kebutuhan masyarakat terhadap penggunaan alat transportasi kereta api kelas ekonomi perlu diperhatikan mengingat jenis kereta api tersebut telah dipilih masyarakat berdasarkan kenyamanan dan terjangkau tarif tiket. Peramalan jumlah penumpang kereta api kelas ekonomi Kertajaya merupakan salah satu upaya penting untuk mengetahui kebutuhan pengguna transportasi tersebut. Namun, *time series* jumlah penumpang kereta api Kertajaya yang mempunyai fluktuasi tinggi terbukti bersifat nonlinear berdasarkan uji linieritas. Pemodelan menggunakan pendekatan linier seperti ARIMA Box-Jenkins tidak selalu memberikan peramalan yang baik karena terikat beberapa asumsi dalam membangun model. Oleh karena itu penelitian ini mengusulkan penggunaan metode ANFIS yang diharapkan memberikan kinerja lebih baik dalam pemodelan nonlinear dan dibandingkan dengan hasil dari ARIMA. Penggunaan ANFIS untuk peramalan jumlah penumpang selama 14 periode ke depan memberikan akurasi ramalan yang lebih tinggi daripada ARIMA karena MAPE dan RMSE yang dihasilkan lebih kecil. Model ANFIS terbaik dihasilkan dari *input* jumlah penumpang pada satu, tujuh, dan delapan hari sebelumnya dan fungsi keanggotaan pi.

**Kata Kunci**— ANFIS, ARIMA, nonlinear, penumpang

## I. PENDAHULUAN

Kereta api merupakan salah satu alat transportasi yang biasa digunakan oleh masyarakat di Indonesia. Badan Pusat Statistik [1] menyebutkan Jumlah penumpang kereta api kumulatif pada Januari s.d Agustus 2014 mencapai 178.0 juta orang atau meningkat sebesar 32.44 % bila dibandingkan dengan periode yang sama pada tahun 2013. Kenaikan jumlah penumpang kereta api terjadi pada seluruh wilayah yaitu: Jabodetabek, Jawa non-Jabodetabek, dan Sumatera. Jumlah penumpang kereta api masing-masing wilayah naik sebesar 37.35%, 19.53%, dan 25.62 %.

Badan Usaha Milik Negara (BUMN) yang menyelenggarakan jasa angkutan perkeretaapian adalah PT Kereta Api Indonesia (PT. KAI). PT. KAI berkewajiban menyelenggarakan *Public Service Obligation* (PSO) yaitu salah satunya dengan memberikan subsidi kepada beberapa kereta api. PT. KAI memprioritaskan KRL dan KA ekonomi jarak dekat untuk pemberian subsidi [2]. Hal ini mengakibatkan tarif untuk KA Ekonomi Jarak Jauh yaitu kereta api Kertajaya mengalami beberapa penyesuaian tarif. Perubahan tarif yang berkali-kali dalam satu tahun membuat masyarakat masih belum terbiasa dengan kondisi tersebut. Hal tersebut berakibat pada jumlah penumpang kereta api yang berfluktuatif di sepanjang tahun 2014 dan 2015. Sehubungan dengan kondisi perkeretaapian, maka penting untuk meramalkan jumlah penumpang kereta api Ekonomi Kertajaya untuk mengetahui

kebutuhan masyarakat. Namun, data jumlah penumpang termasuk ke dalam *time series* yang nonlinear. Pendekatan linier seperti ARIMA Box-Jenkins menganggap suatu *time series* berasal dari proses yang linier, maka ARIMA Box-Jenkins tidak selalu memberikan peramalan yang baik terhadap *time series* yang nonlinear karena terikat beberapa asumsi untuk membangun model. Adapun alternatif metode adalah ANFIS yang merupakan salah satu teknik *soft computing*. ANFIS menggabungkan kemampuan belajar dari jaringan syaraf tiruan dengan kemampuan adanya aturan *fuzzy if-then* yang dapat menghasilkan pemodelan yang lebih baik. Gyu-jin, Jin-pyung, Jee-hyong, & Moon-hyun [3] melakukan peramalan jangka pendek terhadap jumlah penumpang kereta pada dua terminal metropolitan Korea dan memberikan hasil bahwa model ANFIS memiliki ketepatan lebih dari 93%. Perbandingan model ARIMA dan ANFIS pernah dilakukan oleh Galavi, Mirzaei, Shui, dan Valizadeh [4] dengan meramalkan ketinggian air di Sungai Klang Malaysia. Analisis menunjukkan bahwa peramalan dengan menggunakan kedua model memiliki nilai MAPE sebesar <0.5% dan pemodelan ANFIS menghasilkan MAPE lebih kecil. Maka, penelitian ini mengusulkan model ANFIS sebagai metode untuk peramalan jumlah penumpang. Adapun metode ANFIS menggunakan *input* dari pendekatan ARIMA. Hasil pemodelan ANFIS dibandingkan dengan pendekatan linier yaitu ARIMA untuk meramalkan jumlah penumpang.

## II. TINJAUAN PUSTAKA

### A. *Time series*

Suatu metodologi statistik yang dapat menganalisis suatu *time series* disebut sebagai analisis *time series*. Setiap pengamatan dinyatakan sebagai variabel random  $Z_t$  yang didapatkan berdasarkan indeks waktu tertentu  $t$  sebagai urutan waktu pengamatan dengan  $t=1,2,3,\dots,n$  sehingga suatu *time series* yang dinyatakan sebagai himpunan variabel random adalah  $\{Z_1, Z_2, Z_3, \dots, Z_n\}$ . Salah satu metode untuk analisis *time series* adalah prosedur Box-Jenkins [5].

### B. *Prosedur ARIMA Box Jenkins*

Prosedur ARIMA Box-Jenkins sebagai analisis *time series* terdiri dari beberapa prosedur yang pertama adalah identifikasi model. Identifikasi model dilakukan untuk mengidentifikasi kebutuhan transformasi pada data yang belum stasioner dalam variansi, *differencing* pada data yang belum stasioner dalam rata-rata, dan menentukan orde  $p$  dan  $q$  yang tepat untuk model. Prosedur kedua adalah estimasi

parameter. Parameter yang telah diperoleh dapat dilakukan pengujian signifikansi parameter. Adapun hipotesis dan statistik uji yang digunakan adalah sebagai berikut.

Tabel 1 Hipotesis dan Statistik Uji	
Hipotesis	
AR(p)	MA(q)
$H_0: \phi_i = 0$	$H_0: \theta_j = 0$
$H_1: \phi_i \neq 0, i = 1, 2, \dots, p$	$H_1: \theta_j \neq 0, j = 1, 2, \dots, q$
Statistik Uji	
AR(p)	MA(q)
$t_{hitung} = \frac{\hat{\phi}_i}{s.e.(\hat{\phi}_i)}$ (1)	$t_{hitung} = \frac{\hat{\theta}_j}{s.e.(\hat{\theta}_j)}$ (2)

$H_0$  ditolak apabila  $|t_{hitung}| \geq t_{\frac{\alpha}{2}, n-r}$ , dengan  $n$  menunjukkan banyaknya data dan  $r$  menunjukkan banyak parameter pada model [6]. Selanjutnya pemeriksaan diagnostik residual *white noise* dan distribusi normal. Uji *Kolmogorov-Smirnov* digunakan dalam pengujian asumsi distribusi normal. Berikut ini merupakan hipotesis yang digunakan.

- $H_0$ : Data mengikuti distribusi normal
- $H_1$ : Data tidak mengikuti distribusi normal

Statistik uji yang digunakan adalah sebagai berikut:

$$D = \max\{D^+, D^-\} \tag{2}$$

Berdasarkan statistik uji yang diperkenalkan oleh Kolmogorov, maka diketahui bahwa  $D^+ = \max_i \{i/n - Z_{(i)}\}$ ,  $D^- = \max_i \{Z_{(i)} - (i-1)/n\}$ , serta  $Z_{(i)} = F(x_{(i)})$ . Adapun  $F(x)$  adalah fungsi distribusi peluang dari distribusi normal dan  $x_{(i)}$  merupakan orde statistik ke- $i$  dari sampel random dengan  $1 \leq i \leq n$  dan  $n$  merupakan ukuran sampel. Apabila  $D > D_{(1-\alpha; n)}$ , maka hipotesis nol ditolak dengan  $D_{(1-\alpha; n)}$  merupakan nilai pada tabel *Kolmogorov-Smirnov* [7].

- $H_0: \rho_k = 0 \quad \forall k$
- $H_1: \rho_k \neq 0 \quad \exists k, k = 1, 2, \dots, K$

Adapun statistik uji yang digunakan adalah :

$$Q = n(n+2) \sum_{k=1}^K (n-k)^{-1} \hat{\rho}_k^2 \tag{3}$$

dengan  $n$  merupakan ukuran sampel dan  $\rho_k$  menunjukkan autokorelasi residual pada lag ke- $k$ . Hipotesis nol ditolak apabila  $Q > \chi^2_{(1-\alpha), df=K-m}$  dengan  $m=p+q$  [5]. Prosedur selanjutnya adalah pemilihan model terbaik. Kriteria yang digunakan adalah *Mean Absolute Percentage Error* (MAPE) dan *Root Mean Square Error* (RMSE). Berikut ini kriteria berdasarkan kesalahan peramalan *out-sample* [5].

$$MAPE = \left( \frac{1}{n} \sum_{t=1}^n \left| \frac{Z_t - \hat{Z}_t}{Z_t} \right| \right) \times 100\% \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Z_t - \hat{Z}_t)^2} \tag{5}$$

Berdasarkan persamaan tersebut,  $n$  merupakan banyaknya *out sample*,  $Z_t$  merupakan data aktual ke- $t$  dan  $\hat{Z}_t$  merupakan peramalan ke- $t$  dari *out-sample*.

Pemodelan ARIMA musiman multiplikatif Box Jenkins dapat ditunjukkan sebagai berikut ini.

$$\Phi_p(B^s) \phi_p(B) (1-B)^d (1-B^s)^D \hat{Z}_t = \theta_q(B) \Theta_Q(B^s) a_t \tag{6}$$

Adapun persamaan menunjukkan  $\Phi_p(B^s)$  merupakan koefisien komponen *Autoregressive* (AR) musiman dengan periode  $s$  dan orde  $P$  sedangkan  $\Theta_Q(B^s)$  koefisien komponen *Moving Average* (MA) musiman dengan periode  $s$  dan orde  $Q$ . Selain itu, koefisien komponen regular untuk AR orde  $p$  adalah  $\phi_p(B)$  dan untuk MA orde  $q$  adalah  $\theta_q(B)$ . Perbedaan musiman orde  $D$  dan periode musiman  $s$  ditunjukkan oleh  $(1-B^s)^D$  sedangkan untuk perbedaan regular orde  $d$  ditunjukkan oleh  $(1-B)^d$ .

C. Adaptive Neuro Fuzzy Inference System (ANFIS)

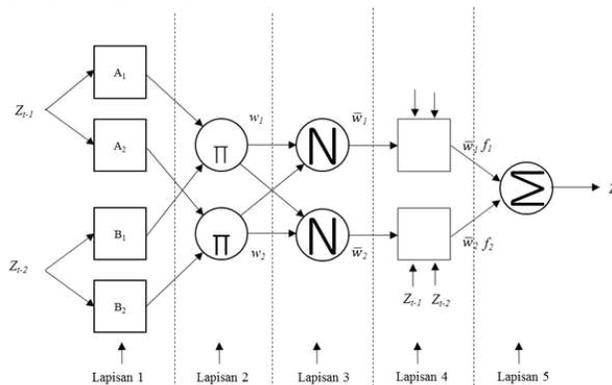
Beberapa kerangka bernalar yang membangun suatu *soft computing* adalah *neural network* dan sistem fuzzy.. Kelebihan dari *neural network* adalah melakukan pembelajaran dan adaptasi. *Neural Network* memproses suatu informasi dengan merepresentasikan otak manusia. Aturan *if-then* dari sistem fuzzy membentuk komponen kunci dari *fuzzy inference system* yang dapat secara efektif memodelkan keahlian manusia dalam aplikasi yang spesifik, Namun sistem fuzzy tidak memiliki suatu kemampuan adaptasi untuk menangani perubahan lingkungan. Maka, penggabungan konsep pembelajaran *neural network* pada *fuzzy inference systems* menghasilkan pemodelan neuro-fuzzy yang merupakan teknik penting dalam *soft computing*. Salah satu pemodelan neuro-fuzzy adalah model *Adaptive Neuro Fuzzy Inference Systems* (ANFIS) [8]. Adapun arsitektur ANFIS yang digunakan dalam penelitian ini berdasarkan aturan *fuzzy if-then* tipe Takagi dan Sugeno, misalnya:

if  $Z_{t-1}$  adalah  $A_1$  and  $Z_{t-2}$  adalah  $B_1$ , then

$$f_1 = p_1 Z_{t-1} + q_1 Z_{t-2} + r_1$$

if  $Z_{t-1}$  adalah  $A_1$  and  $Z_{t-2}$  adalah  $B_2$ , then

$$f_2 = p_2 Z_{t-1} + q_2 Z_{t-2} + r_2$$



Gambar 1. Arsitektur ANFIS yang Setara

Lima lapisan pada arsitektur ANFIS Sugeno dapat dijelaskan sebagai berikut ini.

- Lapisan 1  
Setiap node  $i$  pada lapisan 1 ini merupakan node adaptif dengan fungsi node sebagai berikut:  
 $O_{1,i} = \mu_{A_i}(Z_{t-1})$ , untuk  $i=1,2$  atau  
 $O_{1,i} = \mu_{B_{i-2}}(Z_{t-2})$ , untuk  $i=3,4$  (7)

Berdasarkan fungsi tersebut,  $\mu_A$  merupakan nilai keanggotaan dari himpunan fuzzy A Fungsi keanggotaan A dapat berupa fungsi keanggotaan dengan parameter yang sesuai, misalnya fungsi keanggotaan dengan bentuk kurva Pi:

$$pi(Z_{t-1}; a, b, c, d) = \begin{cases} 0, & Z_{t-1} \leq a \\ 2((Z_{t-1} - a)/(b - a))^2, & a \leq Z_{t-1} \leq (a + b)/2 \\ 1 - 2((Z_{t-1} - b)/(b - a))^2, & (a + b)/2 \leq Z_{t-1} \leq b \\ 1, & b \leq Z_{t-1} \leq c \\ 1 - 2((Z_{t-1} - c)/(d - c))^2, & c \leq Z_{t-1} \leq (c + d)/2 \\ 2((Z_{t-1} - d)/(d - c))^2, & (c + d)/2 \leq Z_{t-1} \leq d \\ 0, & Z_{t-1} \geq d \end{cases} \quad (8)$$

dengan  $\{a_i, b_i, c_i, d_i\}$  adalah himpunan parameter dan parameter yang terdapat pada lapisan ke-1 dikenal dengan nama *premise parameters* atau parameter non-linier.

2. Lapisan 2

*Output* pada lapisan ke-2 merupakan hasil perkalian operator AND berikut ini.

$$O_{2,i} = w_i = \mu_A(Z_{t-1}) \mu_B(Z_{t-2}) \text{ untuk } i=1,2 \quad (9)$$

Setiap node merepresentasikan  $w_i$  sebagai *the firing strength of the rule..*

3. Lapisan 3

Node ke-*i* menghitung rasio ke-*i* dari *rule's firing strength* terhadap jumlah keseluruhan *rule's firing strengths* seperti yang ditunjukkan fungsi berikut ini.

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2} \text{ untuk } i=1,2 \quad (10)$$

*Output* dari lapisan 3 yaitu  $\bar{w}_i$  atau disebut juga sebagai *normalized firing strengths*.

4. Lapisan 4

Lapisan 4 terdiri dari setiap node adaptif dengan suatu fungsi node sebagai berikut:

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i Z_{t-1} + q_i Z_{t-2} + r_i) \quad (11)$$

Berdasarkan fungsi node tersebut, diketahui  $\bar{w}_i$  merupakan *normalized firing strengths* dari lapisan 3 dan  $y_i$  merupakan hasil operasi parameter dalam node ini. Adapun parameter  $\{p_i, q_i, r_i\}$  merupakan parameter pada lapisan 4 yang disebut juga *consequent parameters* atau parameter linier.

5. Lapisan 5

Node tetap menghitung jumlah keseluruhan *output* sebagai hasil terakhir dari seluruh sinyal masuk. Adapun *output* keseluruhan dapat ditunjukkan oleh persamaan berikut ini.

$$O_{5,i} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i f_i} \quad (12)$$

### III. METODOLOGI PENELITIAN

A. *Sumber Data dan Variabel Penelitian*

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari PT. KAI DAOP VIII Pasar Turi, Surabaya. Adapun data yang digunakan adalah data jumlah penumpang Kereta Api Kelas Ekonomi Kerta-

jaya mulai 1 Mei 2014 s.d 6 Mei 2015 sebanyak 371 data. Variabel yang digunakan dalam penelitian ini adalah data jumlah penumpang harian Kereta Api Kertajaya.

B. *Langkah Analisis*

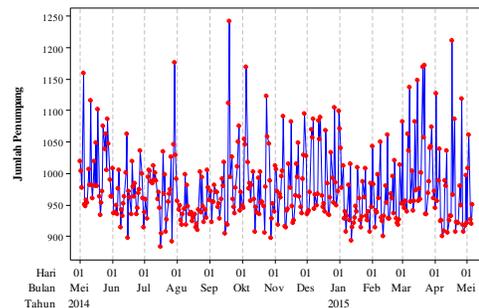
Berikut ini ditunjukkan langkah analisis untuk pemodelan ARIMA.

1. Melakukan identifikasi model.
  2. Melakukan estimasi dan signifikasi parameter.
  3. Melakukan pemeriksaan diagnostik.
  4. Memilih model yang terbaik. Beberapa alternatif model yang didapatkan diseleksi menggunakan kriteria kebaikan model MAPE dan MSE.
- Langkah analisis untuk ANFIS sebagai berikut ini.
1. Menentukan variabel *input* berdasarkan orde *p* dari lag yang signifikan pada plot PACF berdasarkan data yang telah stasioner atau dari penjabaran model ARIMA.
  2. Menentukan jumlah fungsi keanggotaan.
  3. Menentukan jenis fungsi keanggotaan.
  4. Melakukan iterasi sampai nilai parameter konvergen dengan error yang minimum.
  5. Melakukan peramalan dari masing-masing kombinasi jenis dan jenis *input*.
  6. Memilih model yang terbaik berdasarkan kriteria kebaikan MAPE dan RMSE.

### IV. ANALISIS DAN PEMBAHASAN

A. *Pemodelan Jumlah Penumpang Kereta Api Kertajaya Menggunakan ARIMA Box-Jenkins*

Pemodelan jumlah penumpang menggunakan ARIMA diawali dengan identifikasi model melalui plot *time series*. Jumlah penumpang kereta api Kertajaya mencapai 1242 orang perhari dan paling sedikit terdapat 884 orang perhari. Rata-rata jumlah penumpang sebesar 884 orang perhari dengan variansi yang cukup besar yaitu 3430.49. Berdasarkan visualisasi, diindikasikan bahwa data jumlah penumpang tidak stasioner dalam variansi dan rata-rata yang dapat diketahui dari penyebaran jumlah penumpang yang tidak terjadi disekitar variansi dan rata-rata.

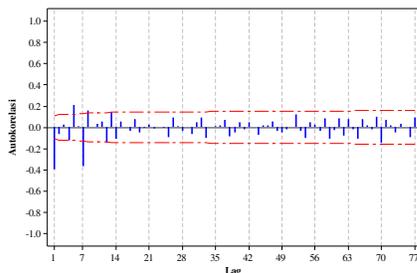


Gambar 2. Plot Time Series Jumlah Penumpang Kereta Api Kertajaya

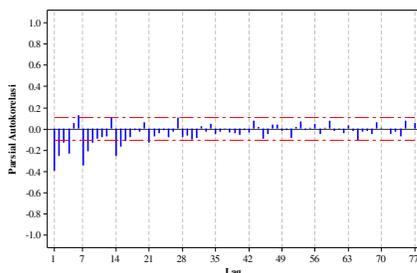
Berdasarkan transformasi *Box-Cox*, diketahui bahwa batas bawah dan atas kendali lambda adalah tidak terhingga sedangkan nilai pembulatan adalah -5, maka data jumlah penumpang memang tidak stasioner dalam variansi. Namun, dalam hal ini lebih baik tidak melakukan transformasi karena tidak diketahui lambda yang tepat untuk transformasi. Pemeriksaan pada Plot PACF dan ACF mengindikasikan un-

tuk dilakukan pembedaan musiman tujuh dengan orde  $D=1$  dan reguler dengan orde  $d=1$ .

Visualisasi PACF diketahui bahwa pada lag ke-1,2,3, dan 6 signifikan sedangkan pada plot ACF terjadi lag yang signifikan pada lag ke-1 dan kelipatan tujuh sehingga model sementara yang dapat dibentuk adalah model musiman ARIMA  $([1,2,3,6],1,1) (0,1,1)^7$ . Selain itu, lag signifikan pada lag ke-4 dan ke-6 pada plot PACF sedangkan ACF signifikan pada lag ke-1 dan 7 atau kelipatan tujuh sehingga model sementara adalah ARIMA  $([4,6],1,1) (0,1,1)^7$ . Selanjutnya, model sementara juga dapat dibentuk dari plot ACF dengan lag yang signifikan pada lag ke-1, ke-4, dan ke-7 atau musiman tujuh sehingga dapat dibentuk model ARI-MA  $(0,1, [1,4]) (0,1,1)^7$ . Prosedur selanjutnya adalah estimasi parameter model sementara.



(a)



(b)

**Gambar 3** Plot (a) ACF dan (b) PACF Data Jumlah Penumpang dengan Pembedaan  $(1-B)(1-B^7)$

Identifikasi model menghasilkan model musiman ARIMA multiplikatif yaitu model ARIMA  $([1,2,3,6],1,1) (0,1,1)^7$ , ARIMA  $([4,6],1,1) (0,1,1)^7$  serta model ARIMA  $(0,1, [1,4])(0,1,1)^7$ . Estimasi parameter pada ketiga model ARIMA menunjukkan bahwa parameter signifikan pada taraf signifikansi 5%. Parameter pada masing-masing model signifikan karena  $|t| > t_{\frac{0.05}{2}, 357-r}$  atau dapat diketahui dari  $p$ -

value yang kurang dari taraf signifikansi  $\alpha = 0.05$ . Setelah dilakukan estimasi, maka tahap selanjutnya adalah pemeriksaan residual model

Pengujian residual kali ini menguji hipotesis nol yaitu residual telah *white noise* sehingga berdasarkan *chi square* hingga lag ke-48 yang menunjukkan kondisi bahwa  $X^2_{hitung} < X^2_{(1-0.05), db}$  dan juga dapat ditandai dengan adanya  $p$ -value yang lebih dari taraf signifikansi  $\alpha = 0.05$ , maka dapat diputuskan untuk gagal menolak hipotesis nol, sehingga residual ketiga model ARIMA telah *white noise*.

Pemeriksaan diagnostik selanjutnya adalah pemeriksaan diagnostik residual berdistribusi normal. Residual yang

didapatkan pada model ini adalah sebanyak data *in-sample* dikurangi dengan delapan ( $357-8=349$ ) sehingga diketahui kuantil dari statistik kolmogorof smirnov adalah sebesar

$$D_{(1-0.05; n=349)} = \frac{1.36}{\sqrt{349}} = 0.0728 \text{ dalam taraf signifikansi } 5\%.$$

Pada pengujian ini, hipotesis yang diuji adalah bahwa distribusi dari residual model adalah distribusi normal. Pengujian menghasilkan kondisi  $D > D_{(1-0.05; 349)}$  sehingga diputuskan untuk menolak hipotesis nol, berarti residual dari model tidak berdistribusi normal sehingga diindikasikan adanya *outlier* pada data jumlah penumpang.

Model ARIMA  $([4,6],1,1) (0,1,1)^7$  memiliki enam *outlier* yang signifikan dengan lima *outlier* tipe *additive* dan satu *outlier* tipe *level shift*, namun deteksi *outlier* dengan taraf signifikansi 0.135 % masih menemukan 20 *outlier* sehingga dilakukan proses memasukkan model *outlier* kedalam model ARIMA satu persatu untuk dilakukan signifikansi parameter. Namun hingga *outlier* terakhir dimasukkan, tidak ada satupun dari dua puluh *outlier* yang signifikan terhadap model sehingga residual tetap tidak berdistribusi normal karena kurva yang terbentuk adalah kurva leptokurtik. Kostenko dan Hyndman [9] dalam peramalan yang dilakukannya dalam bidang bisnis. Peramalan dapat lepas dari uji signifikansi, namun tetap memperhatikan suatu model yang dapat meramalkan dengan baik berdasarkan kriteria kebaikan model.

Pemodelan tidak hanya dilakukan dengan satu model, namun ditemukan enam model masing-masing tiga model dari data yang diterapkan pembedaan reguler dan musiman serta tiga model dengan pembedaan musiman saja. Berikut ini merupakan perbandingan keenam model tersebut.

**Tabel 2.** Pemilihan Model Terbaik

Model ARIMA	Kriteria <i>Out-sample</i>	
	RMSE	MAPE(%)
$([1,2,3,6],1,1)(0,1,1)^7$	65.073	4.728
$([4,6],1,1)(0,1,1)^7$	<b>64.262</b>	<b>4.668</b>
$(0,1,[1,4])(0,1,1)^7$	65.382	4.803
$(1,0,1) (0,1,1)^7$	69.134	5.147
$(1,0,[2]) (0,1,1)^7$	69.748	5.201
$(2,0,[14]) (1,1,0)^7$	68.572	5.130

Berdasarkan kriteria *out-sample*, ketiga model hasil dari pemodelan dengan data yang telah mengalami pembedaan musiman dan reguler cenderung menghasilkan RMSE yang lebih kecil dari pada pemodelan dari data yang mengalami pembedaan musiman saja. Berdasarkan kriteria data *out-sample*, diketahui bahwa model ARIMA  $([4,6],1,1) (0,1,1)^7$  memiliki kemampuan yang lebih baik dibandingkan kelima model ARIMA lain dalam meramalkan jumlah penumpang sampai dengan 14 periode ke depan. Hasil peramalan untuk model ARIMA terbaik ditunjukkan pada Tabel 7. Adapun model ARIMA  $([4,6],1,1)(0,1,1)^7$  adalah:

$$Z_t = Z_{t-1} - 0.1190Z_{t-4} + 0.1190Z_{t-5} - 0.1116Z_{t-6} + 1.1116Z_{t-7} - Z_{t-8} \\ + 0.1190Z_{t-11} - 0.1190Z_{t-12} + 0.1116Z_{t-13} - 0.1116Z_{t-14} - 0.7281a_{t-1} \\ - 0.7281a_{t-7} - 0.7414a_{t-7} + 0.5398a_{t-8} + a_t$$

**B. Pemodelan Jumlah Penumpang Kereta Api Kertajaya Menggunakan ANFIS**

Pemodelan ANFIS dilakukan berdasarkan data jumlah penumpang yang nonlinier. Pemodelan ANFIS dapat di-

lakukan setelah melakukan uji linieritas menggunakan uji Terasvirta dan White.

Pengujian Terasvirta dan White menunjukkan nilai *chi-square* berturut-turut sebesar 10.0616 dan 11.4871 yang lebih besar dari pada  $\chi^2_{(1-0.05)db} = 5.911$  serta hasil pengujian juga dapat ditandai dengan adanya *p-value* yang lebih kecil dari taraf signifikansi  $\alpha = 0.05$ . Keduanya menunjukkan bahwa data jumlah penumpang adalah nonlinier. Pemodelan ANFIS dilakukan dengan menggunakan kombinasi *input* dari pemodelan ARIMA yang terbaik yaitu model ARIMA ([4,6],1,1) (0,1,1)<sup>7</sup>. Kombinasi variabel *input* yang didapatkan dari penjabaran model ARIMA ([4,6],1,1) (0,1,1)<sup>7</sup> adalah  $Z_{t-8}, Z_{t-7}, Z_{t-1}; Z_{t-7}, Z_{t-1}; Z_{t-8}, Z_{t-7}; Z_{t-7}, Z_{t-4}, Z_{t-1}; Z_{t-7}, Z_{t-6}, Z_{t-1}; Z_{t-7}, Z_{t-4}, Z_{t-7}$ , dan  $Z_{t-4}, Z_{t-6}$

Jumlah fungsi keanggotaan yang digunakan adalah sebanyak dua karena apabila lebih dari dua terdapat beberapa *input* yang berada di luar batas spesifikasi. jenis fungsi keanggotaan yang digunakan adalah bentuk kurva *Gaussian*, *generalized bell*, *trapezoidal*, *triangular*, dan *pi* karena kelima jenis fungsi keanggotaan memiliki bentuk kurva yang hampir sama. Pemodelan ANFIS digunakan jumlah *epoch* yang disesuaikan yaitu 500 *epoch*. Penjelasan detail mengenai pemodelan ANFIS dengan lima lapisan hanya diberikan untuk jenis *input*  $Z_{t-8}, Z_{t-7}, Z_{t-1}$  yang diterapkan pada pemodelan ANFIS fungsi keanggotaan *Pi*.

Lapisan ke-1 merupakan lapisan yang memproses suatu himpunan bilangan klasik yang dalam hal ini merupakan data jumlah penumpang. Adapun jenis *input*  $Z_{t-8}, Z_{t-7}, Z_{t-1}$  pada data *in-sample* yang diterapkan pada pemodelan ANFIS antara lain ditunjukkan oleh Tabel 3 berikut ini.

Tabel 3. Input Data In-sample

Data ke-	$Z_{t-8}$	$Z_{t-7}$	$Z_{t-1}$	Target Output ( $Z_t$ )
1	1019	1004	954	1007
2	1004	977	1007	982
3	977	1160	982	1116
⋮	⋮	⋮	⋮	⋮
349	925	933	923	922

Lapisan ke-1 menghasilkan parameter nonlinier atau bisa disebut parameter *premise* yang digunakan untuk mengubah bilangan klasik menjadi bilangan fuzzy. Proses adaptasi terjadi pada lapisan ini sesuai dengan parameter yang dihasilkan. Adapun parameter tersebut dapat digunakan untuk menentukan derajat keanggotaan dari tiga *input* yang digunakan memiliki masing-masing sejumlah dua fungsi keanggotaan sehingga dihasilkan 6 kelompok derajat keanggotaan.

Tabel 4. Parameter Nonlinier

Input	a	b	c	d
Input1 mf1 (A1)	633.4	776.6	1044.0	1045.0
Input1 mf2 (A2)	997.6	1150.0	1349.0	1493.0
Input2 mf1 (B1)	633.4	776.6	1076.0	1088.0
Input2 mf2 (B2)	985.0	1179.0	1349.0	1493.0
Input3 mf1 (C1)	633.4	776.6	938.6	1130.0
Input3 mf2 (C2)	1038.0	1039.0	1349.0	1493.0

*Output* yang dihasilkan pada lapisan ke-1 yaitu derajat keanggotaan yang diperoleh dari operasi parameter non-

linier berdasarkan fungsi keanggotaan *Pi* lalu diproses kembali pada lapisan ke-2. Proses pada lapisan ke-2 adalah perhitungan *the firing strength*. Adapun perhitungan *the firing strength* adalah dengan mengalikan semua derajat keanggotaan 1 dengan memperhatikan *the firing strengt of rule* di mana aturan tersebut berasal dari logika AND. Berikut ini delapan aturan yang berasal dari jumlah *input* dipangkatkan jumlah fungsi keanggotaan ( $2^3=8$ )

- if  $Z_{t-8}$  is A1 and  $Z_{t-7}$  is B1 and  $Z_{t-1}$  is C1, then  $w_{1,t} = \mu_{A1} \mu_{B1} \mu_{C1}$
- if  $Z_{t-8}$  is A1 and  $Z_{t-7}$  is B1 and  $Z_{t-1}$  is C2, then  $w_{2,t} = \mu_{A1} \mu_{B1} \mu_{C2}$
- if  $Z_{t-8}$  is A1 and  $Z_{t-7}$  is B2 and  $Z_{t-1}$  is C1, then  $w_{3,t} = \mu_{A1} \mu_{B2} \mu_{C1}$
- if  $Z_{t-8}$  is A1 and  $Z_{t-7}$  is B2 and  $Z_{t-1}$  is C2, then  $w_{4,t} = \mu_{A1} \mu_{B2} \mu_{C2}$
- if  $Z_{t-8}$  is A2 and  $Z_{t-7}$  is B1 and  $Z_{t-1}$  is C1, then  $w_{5,t} = \mu_{A2} \mu_{B1} \mu_{C1}$
- if  $Z_{t-8}$  is A2 and  $Z_{t-7}$  is B1 and  $Z_{t-1}$  is C2, then  $w_{6,t} = \mu_{A2} \mu_{B1} \mu_{C2}$
- if  $Z_{t-8}$  is A2 and  $Z_{t-7}$  is B2 and  $Z_{t-1}$  is C1, then  $w_{7,t} = \mu_{A2} \mu_{B2} \mu_{C1}$
- if  $Z_{t-8}$  is A2 and  $Z_{t-7}$  is B2 and  $Z_{t-1}$  is C2, then  $w_{8,t} = \mu_{A2} \mu_{B2} \mu_{C2}$

*The firing strength* yang dihasilkan dari proses pada lapisan ke-2 kemudian menjadi *input* untuk lapisan ke-3. Proses pada lapisan ke-3 merupakan proses normalisasi untuk mendapatkan *normalized firing strength*. Perhitungan *normalized firing strength* adalah dengan membagi antara  $w_{i,t}$  dengan jumlah total dari  $w_{i,t}$  dengan  $i=1,2,\dots,8$  dan  $t=349$ .

Tabel 5. Parameter Linier

Input	P	q	r	s
Aturan 1	-1061.0	353.0	714.9	-5335.0
Aturan 2	-764.0	-244.4	716.2	3797.0
Aturan 3	-916.1	350.2	420.4	3794.0
Aturan 4	-248.5	-235.6	415.9	-2703.0
Aturan 5	-1061.0	1036.0	451.2	3795.0
Aturan 6	-917.2	1040.0	-168.2	-2702.0
Aturan 7	-252.4	423.6	-174.0	1923.0

Sama halnya dengan lapisan ke-1, lapisan ke-4 bersifat adaptif karena menghasilkan parameter sesuai dengan lima fungsi keanggotaan yang digunakan. Parameter linier dihasilkan sebanyak empat parameter yaitu  $p_i, q_i, r_i,$  dan  $s_i$  dengan  $i=1,2,\dots,8$  sesuai dengan aturan ditunjukkan oleh Tabel 5 sehingga pada lapisan ke-5 terjadi proses penjumlahan semua *input* sebagai berikut ini.

$$\begin{aligned} \hat{Z}_t &= \bar{w}_{1,t}f_1 + \bar{w}_{2,t}f_2 + \bar{w}_{3,t}f_3 + \bar{w}_{4,t}f_4 + \bar{w}_{5,t}f_5 + \bar{w}_{6,t}f_6 + \bar{w}_{7,t}f_7 + \bar{w}_{8,t}f_8 \\ &= \bar{w}_{1,t} (p_1Z_{t-8} + q_1Z_{t-7} + r_1Z_{t-1} + s_1) + \bar{w}_{2,t} (p_2Z_{t-8} + q_2Z_{t-7} + r_2Z_{t-1} + s_2) \\ &\quad + \bar{w}_{3,t} (p_3Z_{t-8} + q_3Z_{t-7} + r_3Z_{t-1} + s_3) + \bar{w}_{4,t} (p_4Z_{t-8} + q_4Z_{t-7} + r_4Z_{t-1} + s_4) \\ &\quad + \bar{w}_{5,t} (p_5Z_{t-8} + q_5Z_{t-7} + r_5Z_{t-1} + s_5) + \bar{w}_{6,t} (p_6Z_{t-8} + q_6Z_{t-7} + r_6Z_{t-1} + s_6) \\ &\quad + \bar{w}_{7,t} (p_7Z_{t-8} + q_7Z_{t-7} + r_7Z_{t-1} + s_7) + \bar{w}_{8,t} (p_8Z_{t-8} + q_8Z_{t-7} + r_8Z_{t-1} + s_8) \end{aligned}$$

Proses penjumlahan semua *input* yang masuk di lapisan ke-5 menghasilkan peramalan untuk jumlah penumpang sesuai dengan *input* yang digunakan dan target *output* dari *in-sample*. Selain itu pemodelan ANFIS juga diterapkan pada data *out-sample* sebanyak 14 periode. Pemodelan ANFIS dilakukan sebanyak 35 kali yang didasarkan pada lima jenis fungsi keanggotaan serta tujuh jenis *input*. Namun dalam pemodelan ANFIS, *input*  $Z_{t-7}, Z_{t-4}, Z_{t-1}$  dan  $Z_{t-7}, Z_{t-6}, Z_{t-1}$  tidak dapat diterapkan pada pemodelan karena terdapat beberapa *input* berada di luar batas spesifikasi sehingga tidak dapat dilanjutkan dalam pemodelan. Dalam hal ini jenis *input* yang digunakan sebanyak lima jenis *input* sehingga pemodelan ANFIS

dilakukan sebanyak 25 kali. Adapun untuk melakukan pemilihan model ANFIS terbaik digunakan MAPE dan RMSE.

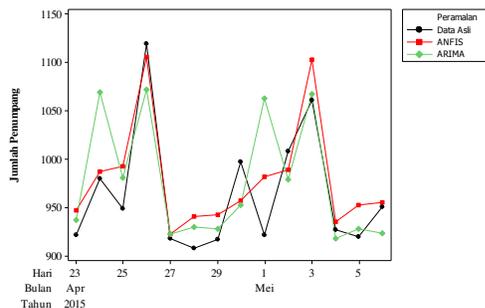
**Tabel 6.** Pemilihan Model ANFIS terbaik

Jenis Input	kriteria	Gauss	Gbell	Trap	Tri	Pi
$Z_{t-8}, Z_{t-7}, Z_{t-1}$	MAPE	4.27	<b>3.94</b>	4.19	<b>3.81</b>	<b>3.81</b>
	RMSE	57.04	52.18	53.64	<b>51.90</b>	<b>51.30</b>
$Z_{t-7}, Z_{t-1}$	MAPE	4.79	4.89	4.91	4.74	4.98
	RMSE	59.06	58.77	59.53	58.46	60.21
$Z_{t-8}, Z_{t-7}$	MAPE	4.81	4.85	4.85	5.74	4.58
	RMSE	58.35	58.04	58.59	66.21	57.80
$Z_{t-7}, Z_{t-4}$	MAPE	6.14	6.28	6.49	6.68	6.03
	RMSE	75.60	78.67	82.22	88.59	76.74
$Z_{t-7}, Z_{t-6}$	MAPE	5.44	4.50	5.60	4.64	4.80
	RMSE	65.98	56.89	67.78	59.62	61.49

Input  $Z_{t-8}, Z_{t-7}, Z_{t-1}$  cenderung memberikan RMSE dan MAPE yang kecil pada peramalan data *out-sample*. Fungsi keanggotaan *Generalized Bell*, *Triangular*, dan *Pi* memberikan nilai MAPE paling kecil berturut-turut adalah 3.94%, 3.81%, dan 3.81%. Secara umum, pemodelan ANFIS dengan jumlah input sebanyak tiga yaitu  $Z_{t-8}, Z_{t-7}, Z_{t-1}$  dan sejumlah dua fungsi keanggotaan serta jenis fungsi keanggotaan *Pi*, *Triangular*, dan *Generalized Bell* memberikan hasil peramalan *out sample* yang lebih baik ditinjau berdasarkan nilai MAPE, namun fungsi keanggotaan *Pi* mampu memberikan MAPE dan RMSE yang terkecil. Sehingga pemodelan ANFIS terbaik adalah menggunakan jenis input  $Z_{t-8}, Z_{t-7}, Z_{t-1}$  dengan dua fungsi keanggotaan dan menggunakan jenis fungsi keanggotaan berbentuk *Pi*.

**C. Perbandingan Pemodelan ARIMA dan ANFIS**

Perbandingan model dilakukan pada model ANFIS dengan input  $Z_{t-8}, Z_{t-7}, Z_{t-1}$ , yang menggunakan jenis fungsi keanggotaan berbentuk *Pi* dan model ARIMA  $([4,6],1,1) (0,1,1)^7$ . Kriteria kebaikan model yaitu MAPE dan RMSE yang diterapkan pada data *out-sample* menunjukkan bahwa MAPE dan RMSE model ANFIS lebih kecil yaitu sebesar 3.81% dan 51.30 daripada model ARIMA sebesar 4.67% dan 64.26, sehingga model ANFIS memiliki kemampuan peramalan jumlah penumpang kereta api kertajaya yang lebih baik dibandingkan dengan model ARIMA.



**Gambar 4.** Perbandingan Peramalan Model ANFIS dan ARIMA

Hasil peramalan dapat divisualisasikan dengan plot *time series* pada Gambar 4 yang diketahui bahwa hasil peramalan menggunakan model ANFIS lebih mendekati data asli jumlah penumpang mulai Kamis, 23 April 2015 sampai dengan Rabu, 6 Mei 2015. Namun, kedua model telah mampu meramalkan jumlah penumpang selama 14 periode ke depan dengan MAPE yang dihasilkan adalah kurang dari 5 %. Adapun peramalan jumlah penumpang untuk 14 periode ke depan dapat ditunjukkan oleh Tabel 7.

**Tabel 7.** Hasil Peramalan Jumlah Penumpang

Periode	Hari	Data Asli	ANFIS	ARIMA
23 April 2015	Kamis	922	947	937
24 April 2015	Jumat	980	987	1069
25 April 2015	Sabtu	949	993	981
26 April 2015	Minggu	1119	1106	1072
27 April 2015	Senin	918	922	923
28 April 2015	Selasa	908	941	930
29 April 2015	Rabu	917	943	928
30 April 2015	Kamis	997	957	953
1 Mei 2015	Jumat	922	982	1063
2 Mei 2015	Sabtu	1008	989	979
3 Mei 2015	Minggu	1061	1102	1067
4 Mei 2015	Senin	927	935	918
5 Mei 2015	Selasa	920	953	928
6 Mei 2015	Rabu	951	956	924

**V. KESIMPULAN**

Hasil penelitian dapat disimpulkan sebagai berikut:

1. Jumlah penumpang kereta api Kertajaya mencapai 1242 orang perhari dan paling sedikit terdapat 884 orang perhari. Adapun jumlah penumpang memiliki pola yang sama dan berulang pada satu minggu atau dapat disebut sebagai musiman tujuh. Model ARIMA  $([4,6],1,1) (0,1,1)^7$  memiliki MAPE dan RMSE yang lebih kecil yaitu 4.67 % dan 64.26 dibandingkan lima model ARIMA yang lain.
2. Model ANFIS dengan input  $Z_{t-8}, Z_{t-7}, Z_{t-1}$  dan fungsi keanggotaan bentuk kurva *Pi* memiliki akurasi yang lebih baik dibandingkan model ANFIS yang lainnya yaitu dengan MAPE 3.81% dan RMSE 51.30.
3. Model ANFIS dengan input  $Z_{t-8}, Z_{t-7}, Z_{t-1}$  dan fungsi keanggotaan *Pi* menghasilkan akurasi yang lebih tinggi daripada model ARIMA  $([4,6],1,1) (0,1,1)^7$  karena MAPE dan RMSE yang dihasilkan lebih kecil. Data jumlah penumpang merupakan *time series* nonlinier yang sulit apabila dianalisis menggunakan pendekatan ARIMA yang merupakan pendekatan linier. Berdasarkan model terbaik yaitu ANFIS, jumlah penumpang dipengaruhi oleh jumlah penumpang pada satu, tujuh, dan delapan hari sebelumnya.

**DAFTAR PUSTAKA**

[1] BPS. (2014, Oktober ), “Perkembangan Pariwisata dan Transportasi Nasional,” Berita Resmi Statistik [online]. pp. 1-10. Available: <http://www.bps.go.id>

[2] KAI. (2014, Maret 06), “Pemerintah Subsidi Penumpang KA Kelas Ekonomi,” PT. Kereta Api Indonesia [online]. Available : <https://www.kereta-api.co.id>

[3] J. Gyu-jin, K. Jin-pyung, L. Jee-hyong, and K. Moon-hyun, “ANFIS based Short-term Prediction of Passenger Volume in Urban Rail Transit “. International Symposium on Advanced Intelligent systems (ISIS). USA: Springer (2011). USA: Springer.

[4] H. Galavi, M. Mirzaei, L. T. Shui, and N. Valizadeh, “Klang River-Level Forecasting Using ARIMA and ANFIS “. *American Water Works Association Journal* (2013). E496-E506.

[5] W. Wei, “Time Series Univariate and Multivariate Methods”.Canada: Addison Wesley (2006).

[6] J. D. Cryer, and K. S. Chan, “Time Series Analysis with Applications in R “. New York: Springer (2008).

[7] M. A. Stephen, “Aspects of Goodness of Fit. California: Technical Report, Department of Statistics, Stanford University. (1993).

[8] R. J-S. Jang, C. Sun, and E. Mizutani, “Neuro-Fuzzy and Soft Computing. London: Prentice Hall (1997).

[9] A. V. Kostenko, dan R. J. Hyndman, “Forecasting Without Significance Test?” [online]. Available: <http://robjhyndman.com/papers/ssst2.pdf>. (2008)