BAB II

TINJAUAN PUSTAKA

2.1 Analisis *Time Series*

Time series atau deret waktu merupakan pengamatan satu atau beberapa variabel yang diambil secara beruntun terhadap interval waktu yang tetap (Wei, 2006). Pada tahun 1976 George Box dan Gwilyn Jenkins memperkenalkan analisis *time series* untuk pertama kalinya. Analisis *time series* adalah salah satu prosedur statistika yang digunakan pada peramalan kejadian di masa depan. Analisis *time series* menggunakan data yang terpaut oleh waktu, sehingga korelasi antara kejadian saat ini dengan periode waktu sebelumnya akan terjadi. Selain berhubungan antara waktu *time seris* juga terdapat kemungkinan adanya hubungan antara dimensi lain seperti wilayah ataupun dimensi lain yang saling berkaitan.

2.2 *Multivariat Time Series*

Multivariat time series merupakan salah satu analisis *time series* yang melibatkan banyak variabel di dalam modelnya. Data *time series* terkadang memiliki keterkaitan antara satu variabel dengan variabel lainnya, seperti pada data jumlah penumpang kereta api yang memuat wilayah sebagai variabel *multivariat* . Sehingga, dibutuhkan analisis yang lebih dari satu data *time series* yaitu dengan menggunakan *multivariat time series*. Proses dalam *multivariat time* *series* sama dengan prose dalam *univariat time series* yaitu dengan memperhatiakan stasioneritas data (Hapsari, 2017).

2.3 Variasi Kalender

Model variasi kalender merupakan model *time series* yang digunakan untuk meramalkan data berdasarkan pola musiman dengan periode bervariasi. Di sebagian besar negara-negara Islam, data *time series* bulanan di bidang ekonomi dan bisnis dapat diketahui dengan dua jenis efek kalender, yaitu efek hari kerja atau efek hari perdagangan di setiap bulan, yang biasa disebut sebagai efek perdagangan hari dan efek hari libur seperti tahun baru cina, natal, dan idul fitri dimana penentuan hari raya tersebut berbeda dengan kelender Masehi. Indonesia merupakan salah satu negara yang merasakan variasi kalender terutama saat memasuki tahun baru dan hari raya. Saat tahun baru dan hari raya kebanyakan masyarakat Indonesia yang bekerja atau bertempat tinggal di luar kota kelahiraan melakukan tradisi pulang kampung (Mudik) sehingga kebutuhan transportasi akan meningkat dari biasanya, sehingga pemerintah atau perusahaan transportasi perlu melakukan evaluasi dan persiapan untuk kenyamanan penumpang.

2.4 *Forcasting* (Peramalan)

Forcasting adalah suatu ilmu yang digunakan untuk meramalkan atau memprediksi krjadian di waktu mendatang. Peramalan dapat dilakukan dengan cara institusi subjektif atau dengan model matematis (Heizer & Render, 2011).

Peramalan biasa digunakan dalam hal perencanaan penjualan ataupun perencanaan pembangunan sbagai tambahan informasi.

2.5 *Autoregressive Integreted Moving Average* (ARIMA)

ARIMA (*Autoregressive Integrated Moving Average*) atau disebut dengan metode deret berkala *Box-Jenkins*. Nilai data pada waktu mendatang bersifat linier dari data historis dan random error. ARIMA dibagi dalam tiga klasifikasi yaitu:

2.5.1 Model *Autoregressive* (AR)

Autoregressive merupakan peramalan regresi yang veriabel *time lag* (selang waktu) dan nilai-nilai sebelumnya saling.model AR dapat dikatakan mengikuti proses AR apabila lag-lag pada plot ACF menurun secara eksponensial dan lag yang signifikan tidak sama dengan nol pada plot PACF kemudian digunakan indikasi parameter p. Bentuk umum model *autoregressive* dengan orde ke-p atau model ARIMA (p,0,0) dinyatakan sebagai berikut:

$$Zt = \mu + \emptyset 1 Zt{-}1 + \emptyset 2 Zt{-}2 + \cdots + \emptyset p Zt{-}p + et \qquad \ldots(2.1)$$

Dimana:

μ = nilai konstan

Zt = nilai pengamatan pada waktu t

$\emptyset$p = koefisien orde p

et = nilai galat pada saat ke-t

2.5.2 Model *Moving Average* (MA)

Moving Average biasa disebut rata-rata bergerak adalah deret berkala pada waktu t yang dipengaruhi dengan galat saat ini dan galat yang terbobot pada historis waktu sebelumnya. Suatu deret berkala dikatakan mengikuti proses apabila lag-lag pada plot ACF tidak sama dengan nol dan lag yang signifikan menurun secara eksponensial pada plot PACF kemudian digunakan indikasi parameter q. Bentuk umum model *moving average* orde ke-q atau ARIMA (0,0,q) dapat ditulis sebagai berikut:

$$Zt = \mu + et - \theta 1et-1 - \theta 2et-2 - \cdots - \theta qet-q \qquad \ldots(2.2)$$

Dimana:

θq = nilai koefisien orde q

et = nilai galat pada saat t

2.5.3 Model *Autoregresive Moving Average* (ARMA)

Model peramalan deret berkala jenis ini dapat berbentuk *autogregresive* (AR), rata-rata bergerak (MA) atau kombinasi antara keduanya (ARMA). Bentuk umum model ARMA (p,q) dapat ditulis sebagai berikut:

$$\emptyset p\ (B)Z_t = \mu + \theta q\ (B)e_t \qquad \ldots(2.3)$$

Dimana:

$\emptyset p\ (B) = (1 - \emptyset 1B - \emptyset 2B\ 2 - \cdots - \emptyset\ \ )$

$\theta q\ (B) = (1 - \theta 1B - \theta 2B\ 2 - \cdots -\ \ )$

2.5.4 Model *Autoregressive Integrated Moving Average* (ARIMA)

ARIMA (*Autoregressive Integrated Moving Average*). Model ARIMA adalah model ARMA (p,q) yang *nonstasioner*. Model ini membutuhkan suatu proses pembedaan agar data stasioner. Model ARMA dilakukan pembedaan dengan ordo ke-d yaitu $Zt\ d = (1 - B)\ dZt$ sehingga $Z1$, $Z2$, … menjadi deret berkala yang *stasioner*, maka model ARMA (p,q) menjadi model ARIMA (p,d,q). Bentuk model umum ARIMA (p,d,q) sebagai berikut:

$$\emptyset p\ (B)(1 - B)\ dZt = \mu + \theta q\ (B)et \qquad \ldots(2.4)$$

Dimana:

$\emptyset p\ (B) = (1 - \emptyset 1B - \emptyset 2B\ 2 - \cdots - \emptyset\ )$

$\theta q\ (B) = (1 - \theta 1B - \theta 2B\ 2 - \cdots - \ )\ (1 - B)$

dZt = pembeda dengan ordo ke-d

et = nilai galat pada saat t

Metode ARIMA *Box-Jenkins* memiliki 3 langkah analisis yakni identifikasi model, estimasi parameter dan *diagnostic checking*.

2.6 ARIMAX (*Autoregresive Integreted Moving Average with Exogenous Variables*)

Model ARIMAX adalah model ARIMA yang diberi tambahan variabel *dammy* atau disebut variabel *exogenous*, variabel *dummy* yang biasa digunakan adalah model ARIMAX dengan variasi kalender. Pemodelan runtun waktu dengan penambahan beberapa variabel yang dianggap memiliki pengaruh secara signifikan terhadap data digunakan untuk memperkuat akurasi dari peramalan yang dilakukan. Menurut Cyer dan Chan (2008) dalam Lee et al. (2010) bahwa model ARIMAX merupakan model ARIMA dengan tambahan variabel. Model ARIMAX dengan efek variasi kalender dituliskan sebagai berikut :

$$Zt = \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_p X_{p,t} + \frac{\theta_q(B)\, \boldsymbol{\Theta}_{\boldsymbol{Q}(\boldsymbol{B}^{s})}}{\emptyset_p(B)\, \Phi_{\boldsymbol{p}(\boldsymbol{1}-\boldsymbol{B})^{\boldsymbol{d}}(\boldsymbol{1}-\boldsymbol{B}^{\boldsymbol{s}})^{\boldsymbol{D}}}} \alpha_t \qquad \ldots(2.5)$$

Dengan:

Zt	: kombinasi linier dari gabungan pengamatan dan sisaan pada waktu-waktu sebelumnya
$X_{1,t}, X_{2,t}, \ldots, X_{n,t}$	: variabel dummy hari-hari khusus
$\beta 1, \beta 2, \ldots, \beta n$	: parameter variabel dummy hari-hari khusus
ϕp	: parameter regresi diri ordo ke-p
θq	: parameter rataan bergerak ordo ke-q
$\phi p\ (B)$	: $(1 - \phi 1B - \phi 2B\ 2 - \cdots - \quad)$
$\theta q\ (B)$	: $(1 - \theta 1B - \theta 2B\ 2 - \cdots - \quad)$

Pemodelan diatas terdiri dari variabel respon time series dan variabel *dummy* yang di bentuk dari variasi kalender. Menurut Lee & Suhartono , langkah penyelesaaian analisis pada model ARIMAX adalah sebagai berikut :

1. Menentukan variabel *dummy* berdasarkan pada variasi kalender
2. Melakukan pemodelan regresi seperti berikut

$$Yt = \beta 0 + \beta\ 1V1, + \beta\ \ 2V2,t + \ldots + \beta\ pVp,t + wt$$

3. Memodelkan residual dari hasil model regresi dengan model ARIMA
4. Melakukan pemodelan ARIMAX
5. Mengecek signifikansi parameter

2.7 VARIMA (*Vector Autorgresive Integreted Moving Average*)

Data *time series* atau runtunn waktu sering kali terdiri dari beberapa variabel atau biasa di sebut data *time series multivariate*. Variabel *multivariate* pada data *time series* misalnya berupa lokasi yang berbeda, atau jenis produk yang berbeda yang dapat dihitung dan terbatas. Menurut (Wei, 2006) model VARIMA dituliskan sebagai berikut :

$$\Phi p\ (B)D(B)\ \dot{Z}_t = \Theta q\ (B)\ at \qquad \ldots(2.6)$$

Dimana

$\Phi p\ (B) = \Phi_0 - \Phi_1 B - \Phi_2 B^2 - \ldots - \Phi_P B^p$

Dan

$\Theta p\ (B) = \Theta_0 - \Theta_1 B - \Theta_2 B^2 \ldots - \Theta_P B^q$

$BZ_t = Z_{t-1}$

Dengan :

$\dot{Z}_t$ adalah vector pengamatan dengan $Z_t = [\ Z_{1,t},\ Z_{2,t}, \ldots,\ Z_{n,t}]$ Φp (B) dan Θq(B) berturut-turut adalah suatu matriks *autoregressive* dan *moving average* polinomial orde *p* dan *q* dengan ukuran matrik *nonsingular* yaitu *mxm*. Untuk kasus *nonsingular*, matriks varians kovarians Σ dari $_t$ a adalah definit positif.

Jika q = 0 maka proses menjadi vektor AR (p).

$$\Phi_P(B)Z_t = a_t \qquad \ldots(2.7)$$

jika p = 0 maka proses menjadi vektor MA (q)

$$\Phi_P(B)Z_t = \Theta_q(B)\ a_{t-q} \qquad \ldots(2.8)$$

2.7.1 Uji Stasioneritas Terhadap Mean dan Varian

Suatu data dikatakan stasioner pada data *time series* jika nilai *mean* dan *varian* konstan atau tidak mengalami perubahan yang sistematik. Markidakis *et al* (1992) menyatakan bentuk visual dari suatu plot data *time series* seringkali cukup untuk menyakinkan bahwa data tersebut adalah stasioner atau tidak stasioner. Akan tetapi, secara formal untuk mengidentifikasi kestasioneran data dilakukan dengan uji *Augmented Dickey Fuller* (ADF) atau dengan melihat skema matriks korelasi silang MACF dan MPACF. Jika plot MACF dan MPACF turun secara sambat maka data belum stasioner terhadap *mean* sehingga perlu dilakukan *differencing* atau pembedaan. Sebaliknya, data belum stasioner terhadap *varian* jika nilai batas atas dan batas bawah pada lamda kurang dari nol, sehingga perlu dilakukan transformasi *Box Cox* agar data stasioner.

2.7.1.1 Uji *Augmented Dickey Fuller* (ADF)

Pada pengujian data runtun waktu sayarat penting yang harus terpenuhi adalah stasioneritas data. Data stasioner adalah data yang menunjukkan mean, varians dan autovarians (pada variasi lag) tetap sama pada waktu kapan saja data itu dibentuk atau dipakai, artinya dengan data yang stasioner model time series dapat dikatakan lebih stabil. Untuk menguji apakah data bersifat stasioner atau tidak, umumnya digunakan uji akar unit. Terdapat banyak uji akar unit, tetapi yang paling umum dan banyak dipakai adalah *Augmented Dickey Fuller Test* (ADF). Konsep pengujian *Augmented Dickey Fuller Test* adalah jika suatu data *time series* tidak stasioner pada orde nol, I(0), maka stasioneritas data tersebut bisa dicari melalui order berikutnya sehingga diperoleh tingkat stasioneritas pada order ke-n (*first difference)* atau I(1), atau *second difference* atau I(2), dan seterusnya (Purnomo, 2010: 39).

Persamaan uji *Augmented Dickey Fuller* (ADF) adalah sebagai berikut :

$$\Delta Yt = \beta_1 + \beta_2 t + \delta Y_{t-1} + \gamma_i \sum_{i=1}^{m} \Delta Y_{t-i} + \varepsilon_t \qquad \ldots(2.9)$$

Dimana:

ΔYt : *first difference* dari Y

β_1 : nilai konstan atau *intercept*

β_2 : koefisien regresi untuk trend

δ : koefisien regresi untuk *lag* Y

γ : koefisien regresi untuk *difference lag Y*

ε : *error*

m : lag

t : waktu

Dengan hipotesis :

H_0 : $\delta = 0$ (Terdapat akar unit, variable Y tidak stasioner)

H_1 : $\delta \neq 0$ Tidak terdapat akar unit, variable Y stasioner)

Statistik Uji:

$$t_{\delta} = \frac{\hat{\delta} - \delta_0}{se\left(\hat{\delta}\right)} \qquad \ldots(2.10)$$

Jika t_δ lebih besar dari nilai kritis ADF maka gagal tolak hipotesis nol, yang berarti terdapat akar unit (data tidak stasioner). Dan jika t_δ lebih kecil dari nilai kritis ADF maka tolak hipotesis nol, tidak terdapat akar unit (data stasioner).

2.7.1.2 *Matrix Autocorrelation Function* (MACF)

Suatu vektor deret waktu sebanyak n pengamatan Z_1, Z_2,..., Z_n matriks korelasi sampel dinyatakan sebeagai berikut :

$$\hat{\rho}(k) = [\hat{\rho}ij(k)] \qquad (2.11)$$

dimana $\hat{\rho}$ij(k) adalah korelasi silang sampel dari komponen deretke-i dan ke-j yaitu :

$$\hat{\rho}ij(k) = \frac{\sum_{t=1}^{n-k}(Z_{i,t}-\bar{Z}_i)(Z_{j,t+k}-\bar{Z}_j)}{[\sum_{t=1}^{n}(Z_{i,t}-\bar{Z}_i)^2\sum_{t=1}^{n}(Z_{j,t+k}-\bar{Z}_j)^2]^{1/2}} \qquad (2.12)$$

dimana $\bar{Z}_i$ dan $\bar{Z}_j$ adalah rata-rata sampel dari komponen deret yang bersesuaian. Fungsi matriks korelasi sangat diperlukan untuk mengidentifikasi batas orde MA, bila matriks korelasi bernilai nol setelah lag ke-q maka model yang bersesuaian yaitu MA(q). Apabila dimensi dan vektornya semakin besar maka bentuk matriks dan grafik akan semakin kompleks sehingga menyulitkan dalam pengidentifikasian.

Untuk memudahkan maka digunakan simbol yang dinotasikan dengan (+), (-), dan (.) pada matriks korelasi (i,j). Simbol (+) diartikan sebagai $\hat{\rho}$ij(k) lebih besar dari 2 kali standar eror dan menunjukkan memiliki hubungan korelasi positif. Simbol (-) diartikan sebagai $\hat{\rho}$ij(k) kurang dari -2 kali standar eror dan menunjukan memiliki hubungan korelasi negatif. Sedangkan simbol (.) diartikan sebagai $\hat{\rho}$ij(k) ± 2 kali standar eror dan menunjukkan tidak adanya korelasi (Wei, 2006).

2.7.1.3 *Matrix Partial Autocorrelation Function* (MPACF)

Fungsi matriks parsial korelasi sampel sangat diperlukan dalam mengidentifikasi model AR. Koreasi antara Z(t) dan Z(t+K) dapat diketahui setelah ketergantungan linier pada variabel Z(t+1), Z(t+2), ... , Z(t+k+1) dihilangkan. Sebagaimana dirumuskan:

$$\phi_{kk} = \frac{Cov\,[(Z_t - \hat{Z}_t),(Z_{t+k} - \hat{Z}_{t+k})]}{\sqrt{Var(Z_t - \hat{Z}_t)}\sqrt{Var(Z_{t+k} - \hat{Z}_{t+k})}} \qquad (2.13)$$

dimana $\hat{Z}(t)$dan $\hat{Z}(t+k)$ adalah rata-rata kesalahan kuadrat minimum pada estimasi regresi linoer dari Z(t) dan Z(t+k) yang berdasarkan pada Z(t+1), Z(t+2), ... , Z(t+k-1). Korelasi spasial antara Z(t) dan Z(t+k) sama dengan koefisien regresi terakhir Z(t) dan Z(t+k) pada lag ke-k (Wei, 2006).

2.7.2 Identifikasi Model

Identifikasi model VARIMA mirip dengan identifikasi pada model time series univariat dengan melihat nilai *Akaike's Information Criterion* (AIC). Nilai AIC digunakan untuk memilih model terbaik dengan kriteria nilai AIC yang paling minimum, yang secara matematis dapat ditulis sebagai berikut

$$AIC = -2\log(L)+2k \qquad \ldots(2.14)$$

Dengan k adalah julmah parameter pada model, dan L adalah parameter yang diestimasi menggunakan nilai *Maximum Likelihood Estimation* (MLE) (Mossad, 2015).

2.7.3 Metode *Maximum Likelihood Estimation* (MLE)

Fungsi MLE untuk model VARIMA adalah

$$\ln L(\Phi \sum \frac{T}{2} ln 2\pi - \frac{T}{2} ln|\Sigma| - \frac{1}{2} \sum_{t=2}^{n} a_t a_t \Sigma^{-1} \qquad ...(2.15)$$

dengan

$a_t = (Z_t - \Phi_1 Z_{t-1}$ **- ….** $- \Phi p Z_{t-p} + \Phi p a_{t-1} + \ldots + \Phi q a_{t-p})$

setelah parameter-parameter taksiran di peroleh maka parameter tersebut kemudian diuji dengan uji signifikansi parsial dan uji signifikansi simultan meggunakan t-test dan F-test.

2.7.4 *Diagnostic Checking*

Pada pengujian *diagnosting checking* ini mrnggunakan uji *L-jung Box* yang bertujuan untuk menguji signifikansi residual secara keseluruhan. Hipotesis yang digunakan adalah sebagai berikut :

H_0 : Vector residual memenuhi asumsi *white noise*

H_1 : Vector residual tidak memenuhi asumsi *white noise*

H_0 ditolak jika p-value < α

Setelah dilakukan uji *L-jung Box* maka kemudian dilakukan uji asumsi normalitas residual menggunakan uji *Kolmogorov Smirnov* dengan hipotesis sebagai berikut :

H0 : $S(x) = F(x)$ (*error* mengikuti distribusi normal)

H1 : $S(x) \neq F(x)$ (*error* tidak mengikuti distribusi normal)

H0 ditolak jika p-value $< \alpha$

2.8 VARIMAX (*Vector Autoregresive Integreted Moving Average With Exogenous Variabel*)

Model VARIMAX (p,q,r) merupakan model yang terbentuk dari model VARIMA yang memiliki kasus khusus yaitu jika asumsi *white noise* tidak terpenuhi atau adanya *outlier*. Model VARIMAX (p,q,r) dituliskan sebagai berikut :

$$\boldsymbol{\nabla^d}\, \mathrm{Y}\, (\boldsymbol{t}) = \boldsymbol{c} + \sum_{i=0}^{r} \Theta_i^*(\boldsymbol{t}-\boldsymbol{i}) + \sum_{j=i}^{p} \Phi_j \nabla^{\mathrm{d}}(\boldsymbol{t}-\boldsymbol{j})\boldsymbol{y}_{\mathrm{t-j}} - \sum_{k=1}^{q} \Theta_{k^e}\, (\boldsymbol{t}-\boldsymbol{k}) + \boldsymbol{e}(\boldsymbol{t}) \quad \text{...(2.16)}$$

Dimana :

$\boldsymbol{\nabla^d}$	: operator pembedaan pada orde-d
c	: vektor dari *intercept* model VARIMAX
Θ_i^*	: matriks dari parameter eksogen dengan lag-i (i=1,2,…,r)
Φ_j	: matriks dari parameter *autoregressive* dengan lag-j (j=1,2,…,p)
Θ_k	: matrik dari parameter rataan bergerak dengan lag-k (k=1,2,…,q)

2.8.1 Asumsi pada model VARIMAX

Pada model VARIMAX terdapat bebeerapa asumsi yang harus dipenuhi yaitu:

a. Stasioneritas

Stasionerita adalah keadaan dimana fluktuasi data berada pada ssekitar nilai rata-rata yang konstan, tidak tergantung pada waktu dan varians. Apabila data tidak stasioner dalam rataan maka dilakukan proses *differencing*, dan apabila tidak stasione dalam varin maka dilakukan proses *Box-Cox* sehingga didapt varians yang konstan.

b. Residual *white noise*

White noise memiliki deret yang terdiri dari peubah acak yang berurutan dan tidak saling berkorelasi (Zt) = 0 untuk setiap waktu, (Zt) = $\sigma 2$ untuk setiap waktu dan γh = $(Zt{+}h, Zt)$ = 0untuk $h \neq 0$.

c. Multinormal Residual

Pada model VARIMAX diharapkan memiliki residual yang berdistribusi normal, untuk itu dilakukan pengecekan multinomial residual dengan uji *Chi-Square* atau dengan melihat plot multinomial residual. Apabila plot membentu garis lurus maka residual berdistribusi normal.

2.9 Kereta Api

Terdapat beberapa jenis transportasi di dunia ini, mulai dari transportasi udara, darat hingga laut. Jenis transportasi yang mudah dan sering digunakan ialah transportasi darat seperti, bus,kendaraan bermotor, dan kereta api. Dari berbagai jenis transportasi darat tersebut, kereta api merupakan jenis transportasi yang mempunyai lintasan khusus yang bernama rel, sehingga kereta api merupakan transportasi yang sering digunakan karena nyaman dan terhindar dari kemacetan. Kereta api biasa digunakan oleh masyarakat untuk menunjang mobilitas perekonomian, seperti pada bidang perdagangan. Selain lebih nyaman dan cepat, cara membeli tiket kereta api saat ini juga lebih mudah. Untuk mendapat tiket kereta api saat ini kita tidak perlu repot-repot lagi mengantri di loket, tiket kereta api dapat dibeli malalui online pada situs web misalnya KAI Acsess, ataupun pada minimarket-minimarket terdekat. Karena mudahnya acsess tiket kereta api sehingga masyarakat saat ini lebih memilih transportasi tersebut. Sehingga jumlah penumpang kereta api semakin meningkat setiap tahunnya.

Di Indonesia perusahaan kereta api telah di kuasai oleh pemerintah yaitu PT. KAI. PPID PT Kereta Api Indonesia (Persero) atau PT KAI terbentuk pada tahun 2010 yang diatur melalui Surat Keputusan (SK) Direksi PT Kereta Api Indonesia (Persero) Nomor: KEP. U/OT.003/VI/3/KA-2010 tentang Perubahan dan Tambahan (P&T) atas Keputusan Direksi Nomor: KEP/U/OT.003/VI/1/KA-2009 tanggal 5 Juni 2009 tentang Organisasi dan Tata laksana Sekretariat Perusahaan di Lingkungan Kantor Pusat PT Kereta Api Indonesia (Persero). Unit organisasi di bidang Keterbukaan Informasi Publik di lingkungan PT KAI ini ditangani oleh *Manager Public Information Care* yang secara struktur berada di bawah *Vice President Public Relations* yang berada di bawah garis komando *Executive Vice President Corporate Secretary* (EVP Sekretaris Perusahaan). *Manager Public Information Care* atau PPID Pusat PT KAI ini mempunyai tugas pokok dan tanggung jawab atas pengelolaan informasi publik yang mencakup penyimpanan, pendokumentasian, penyediaan dan/atau pelayanan informasi terhadap pengguna informasi publik. Sedangkan PPID di daerah melekat pada Manager Humas daerah sesuai wilayah kewenangannya , ada 13 Manager Humas daerah (9 Daop dan 4 Divre) di seluruh PT KAI. Seluruh aturan perusahaan yang mengatur tentang pembentukan PPID di PT KAI ini mengacu pada Undang-undang No.14 tahun 2008 tentang Keterbukaan Informasi Publik.