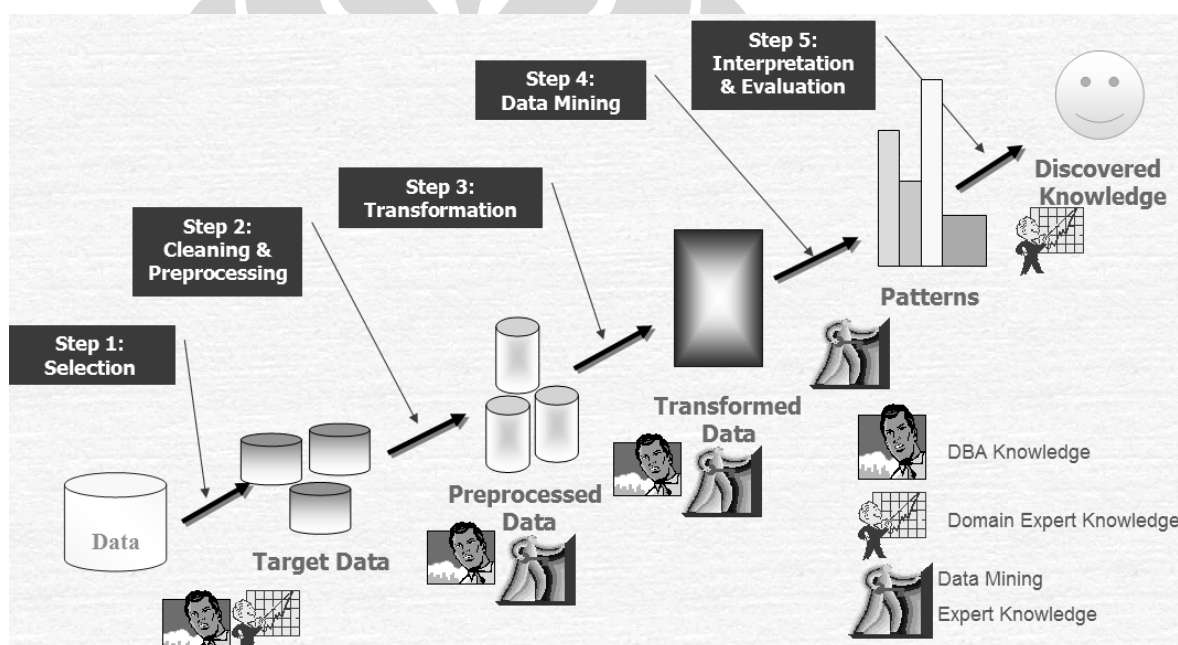


BAB III

METODOLOGI PENELITIAN

Metodologi yang digunakan dalam penelitian ini merujuk pada proses *Knowledge Discovery in Database* (KDD). KDD merupakan nama lain dari *Data Mining* yang biasa digunakan dalam jurnal ilmiah. Langkah-langkah yang terdapat dalam proses KDD dijelaskan dalam gambar 3.1.



Gambar 3.1 Proses *Knowledge Discovery in Database*
[Sumber: Sheng 2002]

Proses KDD pada Gambar 3.1 terdiri dari lima tahap, yaitu:

1. *Selection*; merupakan tahap awal dimana *database administrator* (DBA) bekerjasama dengan *domain expert* untuk menyeleksi data yang dibutuhkan menjadi *target data*.

2. *Cleaning & Preprocessing*; pada tahap ini, DBA bekerjasama dengan *data mining expert* untuk membersihkan *target data* menjadi *preprocessed data*.
3. *Transformation*; sekali lagi, DBA bekerjasama dengan *data mining expert* untuk men-*transform preprocessed data* menjadi *transformed data* yang siap untuk dianalisis.
4. *Data Mining*; pada tahap ini, *data mining expert* menganalisis data untuk menemukan pola yang terdapat di dalamnya.
5. *Interpretation & Evaluation*; pada tahap terakhir, *domain expert* menerjemahkan dan mengevaluasi pola yang terdapat di dalam data sehingga diperoleh *knowledge*.

Bab ini menjelaskan implementasi proses KDD pada penelitian. Dimulai dengan menjelaskan model penelitian yang digunakan. Kemudian, penjelasan tentang tahap ke-1 sampai ke-3 proses KDD yang mencakup data penelitian, mulai dari sumber data sampai menjadi data yang siap untuk dianalisis. Implementasi tahap ke-4 dan ke-5 proses KDD dijelaskan secara lengkap dalam bab selanjutnya.

1.10 Selection

Selection merupakan tahap awal dimana *database administrator* (DBA) bekerjasama dengan *domain expert* untuk menyeleksi data yang dibutuhkan menjadi *target data*. Data yang dipilih terkait dengan model penelitian yang digunakan. Penelitian ini menggunakan Model Broda sebagai model penelitiannya. Model Broda terdiri dari 4 variabel, yaitu: *terms-of-trade*, nilai tukar riil, indeks harga konsumen, dan produk domestik bruto riil.

Model dalam penelitian ini menggunakan model penelitian Broda sebagai dasarnya. Broda meneliti peranan sistem nilai tukar dalam merespon dampak penurunan harga komoditi ekspor sebesar 10% terhadap PDB riil di 75 negara sedang berkembang. Broda mengklasifikasikan negara-negara sedang berkembang berdasarkan sistem nilai tukar yang diterapkan di masing-masing negara. Penelitian Broda menyimpulkan bahwa PDB riil negara-negara sedang berkembang yang menggunakan sistem nilai tukar mengambang tidak terpengaruh secara signifikan (hanya sebesar 0,2%) terhadap penurunan harga komoditi ekspor. Namun, hal sebaliknya terjadi pada negara-negara sedang berkembang yang menggunakan sistem nilai tukar tetap di mana PDB riil negara-negara tersebut mengalami penurunan sebesar 2% setelah terjadinya penurunan harga komoditi ekspor [Broda 2003].

Model penelitian Broda yang terdiri dari empat variabel (*Terms-of-Trade*, nilai tukar riil, indeks harga konsumen, dan produk domestik bruto riil) diadaptasi sedemikian rupa sehingga menjadi sistem persamaan sebagai berikut:

$$TOT_t = a_{10} - a_{12}RER_t - a_{13}CPI_t - a_{14}RGDP_t + a_{11}TOT_{t-1} + a_{12}RER_{t-1} + a_{13}CPI_{t-1} + a_{14}RGDP_{t-1} + \varepsilon_{TOTt}$$

$$RER_t = a_{20} - a_{21}TOT_t - a_{23}CPI_t - a_{24}RGDP_t + a_{21}TOT_{t-1} + a_{22}RER_{t-1} + a_{23}CPI_{t-1} + a_{24}RGDP_{t-1} + \varepsilon_{RERt}$$

$$CPI_t = a_{30} - a_{31}TOT_t - a_{32}RER_t - a_{34}RGDP_t + a_{31}TOT_{t-1} + a_{32}RER_{t-1} + a_{33}CPI_{t-1} + a_{34}RGDP_{t-1} + \varepsilon_{CPIt}$$

$$RGDP_t = a_{40} - a_{41}TOT_t - a_{42}RER_t - a_{43}CPI_t + a_{41}TOT_{t-1} + a_{42}RER_{t-1} + a_{43}CPI_{t-1} + a_{44}RGDP_{t-1} + \varepsilon_{RGDPt}$$

dimana:

TOT : *terms of trade*

RER : nilai tukar riil (*real exchange rate*)

CPI : indeks harga konsumen (*consumer price index*)

RGDP : produk domestik bruto riil (*real gross domestic product*)

ε : *error terms*

sistem persamaan VAR di atas, dengan menggunakan aljabar matriks, dapat ditulis sebagai berikut:

$$\underbrace{\begin{pmatrix} 1 & a_{12} & a_{13} & a_{14} \\ a_{21} & 1 & a_{23} & a_{24} \\ a_{31} & a_{32} & 1 & a_{34} \\ a_{41} & a_{42} & a_{43} & 1 \end{pmatrix}}_A \underbrace{\begin{pmatrix} TOT \\ RER \\ CPI \\ RGDP \end{pmatrix}}_{x_t} = \underbrace{\begin{pmatrix} a_{10} \\ a_{20} \\ a_{30} \\ a_{40} \end{pmatrix}}_{\Gamma_0} + \underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}}_{\Gamma_1} \underbrace{\begin{pmatrix} TOT_{t-1} \\ RER_{t-1} \\ CPI_{t-1} \\ RGDP_{t-1} \end{pmatrix}}_{x_{t-1}} + \underbrace{\begin{pmatrix} \varepsilon_{TOT_t} \\ \varepsilon_{RER_t} \\ \varepsilon_{CPI_t} \\ \varepsilon_{RGDP_t} \end{pmatrix}}_{\varepsilon_t}$$

atau

$$Ax_t = \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t$$

dengan asumsi bahwa A selalu mempunyai invers, maka dengan mengalikan A^{-1} pada persamaan di atas sehingga diperoleh:

$$\underbrace{A^{-1}A}_I \underbrace{x_t}_{x_t} = \underbrace{A^{-1}\Gamma_0}_{B_0} + \underbrace{A^{-1}\Gamma_1}_{B_1} \underbrace{x_{t-1}}_{x_{t-1}} + \underbrace{A^{-1}\varepsilon_t}_{e_t}$$

atau

$$x_t = B_0 + B_1 x_{t-1} + e_t$$

dengan mendefinisikan notasi b_{i0} merupakan elemen i dari vektor B_0 , b_{ij} merupakan elemen yang mewakili baris i dan kolom j matriks B_1 , dan e_{it} merupakan elemen i vektor e_t , maka persamaan tersebut dapat ditulis kembali sebagai berikut:

$$TOT_t = b_{10} + b_{11}TOT_{t-1} + b_{12}RER_{t-1} + b_{13}CPI_{t-1} + \alpha_{14}RGDP_{t-1} + \varepsilon_{TOTt}$$

$$RER_t = b_{20} + b_{21}TOT_{t-1} + b_{22}RER_{t-1} + b_{23}CPI_{t-1} + b_{24}RGDP_{t-1} + \varepsilon_{RERt}$$

$$CPI_t = b_{30} + b_{31}TOT_{t-1} + b_{32}RER_{t-1} + b_{33}CPI_{t-1} + b_{34}RGDP_{t-1} + \varepsilon_{CPIt}$$

$$RGDP_t = b_{40} + b_{41}TOT_{t-1} + b_{42}RER_{t-1} + b_{43}CPI_{t-1} + b_{44}RGDP_{t-1} + \varepsilon_{RGDPt}$$

Sistem persamaan VAR inilah yang akan digunakan sebagai model penelitian. Fokus pembentukan sistem persamaan VAR selanjutnya terletak pada koefisien dan parameter masing-masing variabel serta penentuan *lag* optimal (*lag* 1 dalam sistem persamaan di atas hanya berfungsi untuk menunjukkan adanya hubungan *auto-regressive* dan tidak menunjukkan bahwa *lag* 1 merupakan *lag* optimal sistem persamaan VAR) [Pyndick & Rubinfeld 1998].

1.11 *Cleaning & Preprocessing*

Pada tahap *cleaning* dan *preprocessing* ini, DBA bekerjasama dengan *data mining expert* untuk membersihkan *target data* menjadi *preprocessed data*.

Data yang digunakan dalam penelitian adalah data kwartalan dari kwartal pertama tahun 1999 sampai dengan kwartal ke-4 tahun 2007. Periode ini dipilih karena sudah tergolong stabil setelah krisis ekonomi tahun 1997 – 1998. Data tersebut diperoleh, baik mentah maupun olahan, dari buletin-buletin statistik terbitan Biro Pusat Statistik (BPS) dan *International Financial Statistic* (IFS) terbitan *International Monetary Fund* (IMF).

Data variabel TOT menggunakan data olahan dari Buletin Perdagangan Luar Negeri (Ekspor dan Impor) berbagai terbitan BPS, yang dapat diakses melalui <http://www.bps.go.id>, dengan menggunakan rumus [Salvatore 1998]:

$$N = \frac{P_X}{P_M} \times 100$$

dimana:

- N : *terms-of-trade*
 P_X : indeks harga komoditi ekspor
 P_M : indeks harga komoditi impor

Buletin Perdagangan Luar Negeri (Ekspor dan Impor) terbitan BPS tidak menyediakan data indeks harga komoditi ekspor (P_X) dan indeks harga komoditi impor (P_M). Buletin statistik tersebut menyediakan total nilai ekspor (V_X), total nilai impor (V_M), total volume ekspor (Q_X), dan total volume impor (Q_M). Penghitungan indeks harga komoditi ekspor dihitung dengan menggunakan rumus:

$$P_X = \frac{V_X}{Q_X}$$

dimana:

- P_X : indeks harga komoditi ekspor
 V_X : total nilai komoditi ekspor
 Q_X : total volume komoditi ekspor

Sementara penghitungan indeks harga komoditi impor dihitung dengan menggunakan rumus:

$$P_M = \frac{V_M}{Q_M}$$

dimana:

P_M : indeks harga komoditi impor

V_M : total nilai komoditi impor

Q_M : total volume komoditi impor

Data variabel RER berasal dari *International Financial Statistic* (IFS) terbitan *International Monetary Fund* (IMF) yang dapat di-download melalui <http://imfstatistics.org/imf> dan diolah dengan menggunakan rumus:

$$RER = \frac{WPI_{USA}}{WPI_{INA}} \times NER$$

dimana:

RER : Nilai Tukar Riil (*Real Exchange Rate*)

WPI_{USA} : Indeks Harga Perdagangan Besar (*Wholesales Price Index*) Amerika

WPI_{INA} : Indeks Harga Perdagangan Besar (*Wholesales Price Index*) Indonesia

NER : Nilai Tukar Nominal (*Nominal Exchange Rate*)

Selain RER, data CPI juga dapat di-download melalui website yang sama, perbedaannya adalah data CPI sudah tersedia sehingga tidak perlu diolah lagi.

Seperti data variabel RER, data variabel RGDP juga diolah dari IFS terbitan IMF dengan menggunakan rumus:

$$RGDP = \frac{NGDP}{GDPDeflator} \times 100$$

dimana:

RGDP : Produk Domestik Bruto Riil (*Real Gross Domestic Product*)

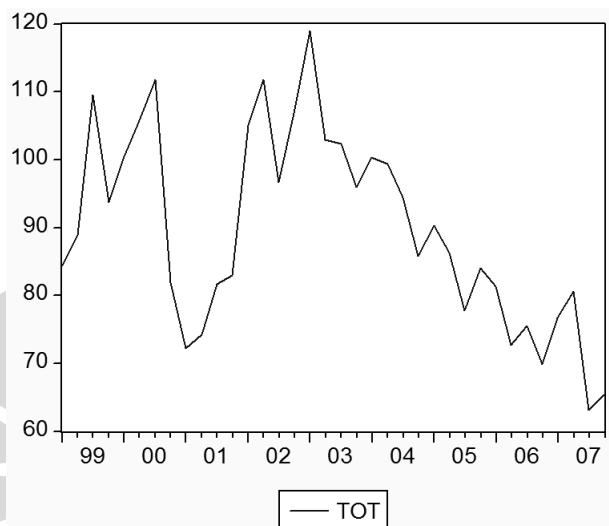
NGDP : Produk Domestik Bruto Nominal (*Nominal Gross Domestic Product*)

GDP Deflator : Deflator Produk Domestik Bruto

1.12 Transformation

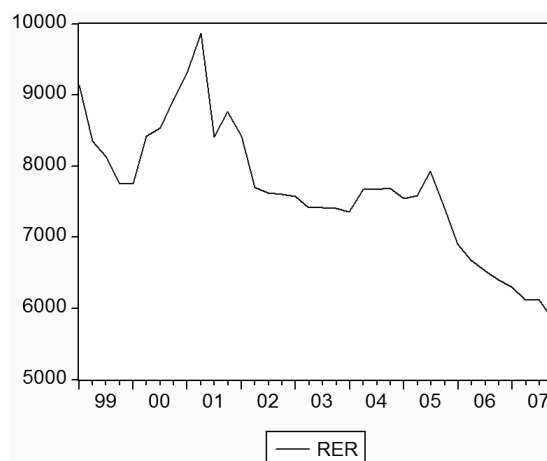
Pada tahap ketiga, *Transformation*, DBA bekerjasama dengan *data mining expert* untuk men-*transform preprocessed data* menjadi *transformed data* yang siap untuk dianalisis. Tahapan penelitian selanjutnya adalah memastikan data penelitian berada dalam kondisi stasioner. Data *time-series* berada dalam kondisi stasioner jika rata-rata dan auto-kovarians data tersebut tidak dipengaruhi oleh waktu. Uji stasioneritas data berguna untuk mencegah terjadinya regresi palsu (*spurious regression*). Cara yang paling mudah untuk melihat stasioneritas data *time-series* adalah dengan melihat pergerakan data tersebut. Jika terlihat adanya trend dalam pergerakan data seiring berjalannya waktu, maka dapat diduga bahwa data tersebut tidak stasioner (*nonstationary*). Jika pada tingkat *level* data masih tidak stasioner, maka data harus di-*transform* ke tingkat *difference*.

Gambar 3.2 sampai 3.5 memperlihatkan pergerakan data keempat variabel (TOT, RER, CPI, dan RGDP) selama periode kuartal pertama tahun 1999 sampai kuartal ke-4 tahun 2007.



Gambar 3.2 Grafik Pergerakan Data TOT Periode 1999.1 – 2007.4
[Sumber: diolah dari Buletin Statistik Perdagangan Luar Negeri (Ekspor dan Impor) BPS]

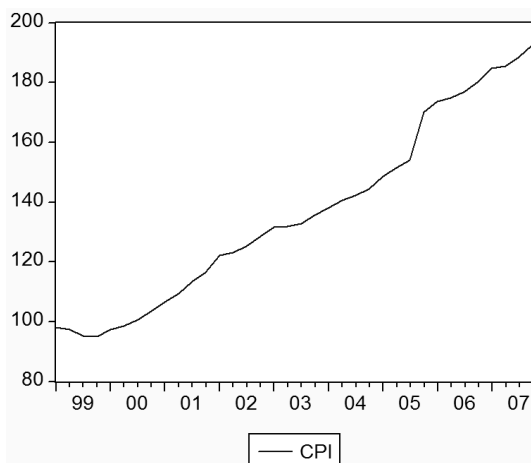
Pergerakan data TOT dalam Gambar 3.2 menunjukkan adanya fluktuasi sepanjang periode tahun 1999 sampai dengan tahun 2007. Walaupun sempat turun drastis pada pertengahan tahun 2000, TOT kembali naik dan cenderung menurun sejak pertengahan tahun 2002 sampai tahun 2007.



Gambar 3.3 Grafik Pergerakan Data RER Periode 1999.1 – 2007.4

[Sumber: diolah dari *International Financial Statistics* IMF]

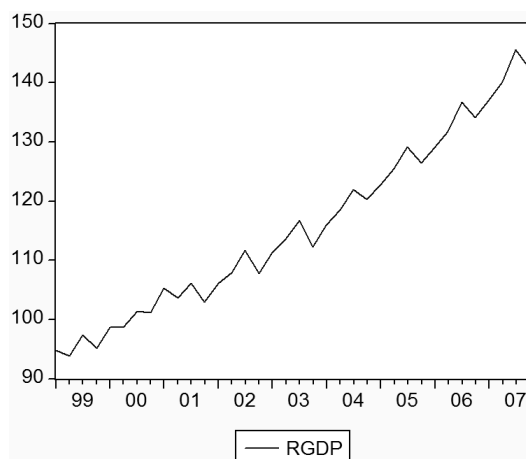
Pergerakan data RER, seperti dapat dilihat dalam gambar 3.3, menunjukkan adanya depresiasi pada tahun 2001. Depresiasi adalah melemahnya nilai tukar mata uang suatu negara terhadap mata uang negara lain. Sebaliknya, jika nilai tukar mata uang suatu negara menguat terhadap mata uang negara lain, maka dikenal dengan istilah apresiasi. Pergerakan data RER setelah tahun 2001 cenderung stabil sampai tahun 2005 yang diikuti dengan apresiasi sampai tahun 2007. Apresiasi ini diduga disebabkan oleh menurunnya pertumbuhan di Amerika Serikat.



Gambar 3.4 Grafik Pergerakan Data CPI Periode 1999.1 – 2007.4

[Sumber: diolah dari *International Financial Statistics* IMF]

Trend meningkat juga terjadi pada variabel CPI. Gambar 3.4 menunjukkan terjadinya kecenderungan ini selama periode kuartal pertama tahun 1999 sampai kuartal ke-4 tahun 2007. Peningkatan CPI, atau Indeks Harga Konsumen (IHK), biasa dikenal dengan istilah inflasi. Sebaliknya, penurunan CPI dikenal dengan istilah deflasi. Peningkatan CPI paling signifikan terjadi pada tahun 2005 yang terjadi karena adanya dua kali peningkatan harga BBM dalam kurun waktu setahun.



Gambar 3.5 Grafik Pergerakan Data RGDP Periode 1999.1 – 2007.4

[Sumber: diolah dari *International Financial Statistics* IMF]

Tidak berbeda dengan data variabel RER dan CPI, kecenderungan peningkatan juga ditemui pada data variabel RGDP. Gambar 3.5 yang terlihat berundak-undak menunjukkan adanya pola musiman yang mencapai puncak pada kuartal ke-3 setiap tahunnya.

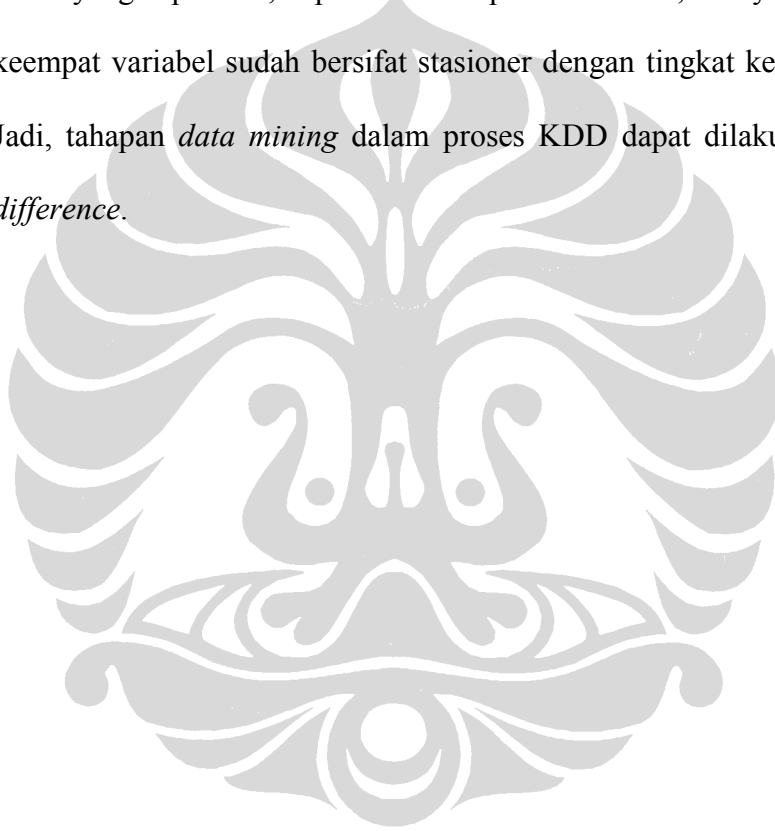
Kesimpulan yang dapat diambil dari grafik 3.2 – 3.5 adalah ditemukan adanya trend meningkat pada data variabel CPI dan RGDP. Data variabel TOT dan RER juga menunjukkan masih adanya trend dalam pergerakan datanya yang fluktuatif. Oleh karena itu, dapat diduga bahwa pergerakan data beberapa variabel masih bersifat non-stasioner.

Uji stasioneritas data dengan cara melihat ada tidaknya trend seperti di atas tidak dapat dijadikan sebagai patokan karena bersifat subyektif. Cara lain yang dapat dilakukan dan menghasilkan kesimpulan yang dapat dipercaya adalah dengan melakukan *unit root test*. Uji ini akan melihat ada tidaknya *unit root* yang merupakan indikasi kestasioneran data *time-series*.

Uji kestasioneran data dalam penelitian ini menggunakan uji Phillips-Perron (PP). Tabel 3.1 memperlihatkan hasil uji PP terhadap keempat variabel

pada tingkat *level* dengan mengikutsertakan trend dan *intercept* dalam pengujian. Hasil uji ini menyimpulkan bahwa data keempat variabel bersifat non-stasioner.

Oleh karena data masih bersifat non-stasioner pada tingkat *level*, maka tahap berikutnya adalah melakukan uji PP pada tingkat *1st difference* dengan hanya mengikut sertakan *intercept* dalam pengujian. Uji ini hanya menyertakan *intercept* dalam pengujian karena trend sudah hilang akibat proses diferensiasi. Hasil yang diperoleh, seperti terlihat pada tabel 3.2, menyimpulkan bahwa data keempat variabel sudah bersifat stasioner dengan tingkat keyakinan di atas 99%. Jadi, tahapan *data mining* dalam proses KDD dapat dilakukan pada tingkat *1st difference*.



Tabel 3.1 Phillips-Perron Test pada tingkat level

Variabel	ADF Test Statistic <i>Level</i>	Test Critical Values			Keterangan H_0 : ada <i>unit root</i> H_1 : tidak ada <i>unit root</i>
		1%	5%	10%	
TOT	-2.852681	-4.243644	-3.544284	-3.204699	Tidak Signifikan pada semua tingkat α tidak dapat menolak H_0 , ada <i>unit root</i> , data non-stasioner
RER	-2.488787	-4.243644	-3.544284	-3.204699	Tidak Signifikan pada semua tingkat α tidak dapat menolak H_0 , ada <i>unit root</i> , data non-stasioner
IHK	-2.629277	-4.243644	-3.544284	-3.204699	Tidak Signifikan pada semua tingkat α tidak dapat menolak H_0 , ada <i>unit root</i> , data non-stasioner
RGDP	-3.933885	-4.243644	-3.544284	-3.204699	Tidak Signifikan pada $\alpha=1\%$ tidak dapat menolak H_0 , ada <i>unit root</i> , data non-stasioner

[sumber: Lampiran 2]

Tabel 3.2 Phillips-Perron Test pada tingkat 1st difference

Variabel	PP Test Statistic <i>1st Difference</i>	Test Critical Values			Keterangan H_0 : ada <i>unit root</i> H_1 : tidak ada <i>unit root</i>
		1%	5%	10%	
TOT	-6.689926	-3.639407	-2.951125	-2.614300	Signifikan pada semua tingkat α H_0 ditolak, tidak ada <i>unit root</i> , data stasioner
RER	-6.304663	-3.639407	-2.951125	-2.614300	Signifikan pada semua tingkat α H_0 ditolak, tidak ada <i>unit root</i> , data stasioner
IHK	-4.879865	-3.639407	-2.951125	-2.614300	Signifikan pada semua tingkat α H_0 ditolak, tidak ada <i>unit root</i> , data stasioner
RGDP	-11.58426	-3.639407	-2.951125	-2.614300	Signifikan pada semua tingkat α H_0 ditolak, tidak ada <i>unit root</i> , data stasioner

[sumber: Lampiran 2]

1.13 Data Mining

Tahapan proses *Knowledge Discovery in Database* (KDD) selanjutnya adalah *data mining*. Penelitian ini menggunakan dua metode berbasis ilmu statistika dan ekonometrika. Metode tersebut adalah *Vector Auto Regression* (VAR) yang ditemukan oleh Christopher Sims dan *General-to-Specific* (GetS) *Modelling* yang ditemukan oleh David F. Hendry.

Tahapan *data mining* dalam proses KDD menggunakan VAR akan berfokus pada penentuan lag optimal yang masih stabil. Stabilitas sistem persamaan VAR dapat dilihat dari nilai *inverse root* karakteristik AR polinomialnya. Suatu sistem persamaan VAR tergolong stabil jika seluruh *roots*-nya memiliki *modulus* lebih kecil dari satu dan semuanya terletak dalam *unit circle*. Penentuan lag optimal akan menggunakan kriteria informasi yang tersedia, seperti: *Likelihood Ratio* (LR), *Final Prediction Error* (FPE), *Akaike Information Criterion* (AIC), *Schwarz Information Criterion* (SIC), dan *Hannan-Quinn Information Criterion* (HQ). Jika kriteria-kriteria tersebut hanya mengacu pada sebuah lag sebagai pilihan, maka lag tersebut adalah lag optimal. Namun, jika menghasilkan beberapa lag sebagai pilihan, maka lag optimal akan dipilih dari sistem persamaan VAR yang memiliki nilai *adjusted R²* paling tinggi pada variabel-variabel utama dalam sistem persamaan.

Penelitian ini juga menggunakan GetS *Modelling* sebagai metode *data mining*. GetS *Modelling* dipilih karena diduga lebih baik daripada VAR dalam memprediksi. VAR memiliki kecenderungan menggeneralisir lag pada sistem persamaan yang terbentuk. Pada GetS, suatu persamaan akan dianalisis secara khusus. Sehingga, persamaan tersebut diperkenankan memiliki lag yang tidak

sama. Namun demikian, persamaan yang diperoleh melalui *GetS Modelling* tetap memenuhi aturan-aturan yang terdapat dalam ilmu statistika dan ekonometrika.

Data mining menggunakan *Vector Auto Regression* (VAR) dan *General-to-Specific* (GetS) *Modelling* akan dibahas lebih dalam pada bab selanjutnya.

