

BAB 3 DATA DAN METODOLOGI PENELITIAN

3.1 Pengantar

Penelitian ini bertujuan untuk melihat apakah metode *Artificial Neural Network* (ANN) dapat lebih akurat memprediksi harga saham di Indonesia dibandingkan dengan metode *Multivariate Auto Regressive Integrated Moving Average* (ARIMA). Penelitian ini merupakan kelanjutan dari penelitian yang dilakukan oleh Portugal (1995), Lawrence (1997), dan Iskandar (2005) dalam hal memprediksi harga saham menggunakan model ANN dibandingkan dengan model *time series forecasting* (ARIMA). Prediksi dilakukan dengan menggunakan harga saham waktu sebelumnya, dikombinasikan dengan variabel *intermarket*, seperti harga emas, minyak, kurs, dan indeks bursa saham.

Jenis penelitian yang dilakukan yaitu pengujian hipotesis observasi model dengan menggunakan data sekunder. Pengujian hipotesis ini akan membuktikan apakah ANN lebih akurat daripada *multivariate* ARIMA dalam memprediksi harga saham-saham di Indonesia. Penelitian ini menggunakan data runtun waktu (*time series*), yang mana perubahan data pada saat ini ataupun di masa mendatang bisa merupakan akibat dari perubahan harga pada masa sebelumnya.

3.2 Data Penelitian

Populasi dalam penelitian ini adalah harga saham-saham di Indonesia. Dalam melakukan *forecast* harga saham, penulis menggunakan data harga saham yang tergolong dalam indeks LQ45 sebagai sampel yang dianggap merepresentasikan pergerakan harga saham-saham di Indonesia. Hal ini dikarenakan indeks LQ45 terdiri dari 45 saham-saham berkapitalisasi terbesar, dan teraktif diperdagangkan di Bursa Efek Indonesia. Penulis mendapatkan harga saham-saham indeks LQ45 dari situs Yahoo Finance (finance.yahoo.com, 2010).

Dalam melakukan *forecast*, informasi historis yang mencukupi dapat meningkatkan hasil *forecast* seperti yang diutarakan oleh (Jenkins, 1994). Untuk mendapatkan model *forecast* yang baik, diperlukan minimal 50-100 data historis (Jenkins, 1994). Oleh karena itu, penelitian harus menggunakan data-data harian.

Data saham indeks LQ45 yang dipilih adalah data harian 10 tahun terakhir, mulai dari tahun 5 April 2001 sampai dengan 28 Februari 2010. Data harga saham harian ini mencakup 2000 lebih data historis. Jika terdapat saham-saham yang harga historisnya tidak mencapai 10 tahun, maka semua harga historisnya diikutkan ke dalam sampel.

Harga saham dipengaruhi oleh variabel-variabel makroekonomi / variabel *intermarket* (Bodie, 2008, p. 538). Oleh karena itu, penelitian ini memakai variabel-variabel makroekonomi / *intermarket*. Data yang digunakan adalah data harian. Data-data tersebut adalah harga emas, harga minyak, kurs Rupiah terhadap Dolar Amerika, indeks KLSE (Kuala Lumpur *Stock Exchange*), indeks STI (*Strait Times Index*), indeks DJI (Dow Jones Industrial), dan IHSG (Indeks Harga Saham Gabungan). Data harga harian emas dalam satuan Dollar per troy ounce didapatkan dari situs (www.usagold.com, 2010). Data harga minyak dalam satuan *Dollar per barrel* didapatkan dari (<http://tonto.eia.doe.gov>, 2010). Kurs Rupiah terhadap Dolar Amerika didapatkan dari situs Bank of Canada (www.bankofcanada.ca, 2010). Indeks KLSE, STI, DJI, dan IHSG diambil dari situs Yahoo Finance (finance.yahoo.com, 2010).

Tabel 3.1 menunjukkan statistik deskriptif data *intermarket* yang dipakai membentuk model *forecast*. Statistik deskriptif tersebut menggambarkan rata-rata, *standard error*, median, modus, standard deviasi, varians sampel, kurtosis, *skewness*, jangkauan, data minimum, data maksimum, dan jumlah data.

Tabel 3.1 Statistik deskriptif data *intermarket*

Statistik Deskriptif	Oil Price	Gold Price	USD IDR	IHSG	KLSE	STI	DJI
Mean	54,64	566,22	9382,84	1237,03	931,35	2210,90	10440,55
Standard Error	0,5	5,32	15,18	15,86	4,94	13,76	32,26
Median	53,32	454,50	9218,59	1096,67	898,14	2118,75	10402,04
Mode	61,81	913,00	9250,00	2657,98	1072,69	2758,90	10099,14
Standard Deviation	26,22	251,67	718,57	751,10	233,74	651,64	1527,67
Sample Variance	687,31	63336,95	516335,75	564148,23	54634,70	424629,13	2333761,84
Kurtosis	0,67	-0,87	2,61	-1,08	-0,74	-0,64	-0,33
Skewness	0,93	0,62	1,41	0,51	0,52	0,58	0,27
Range	127,81	956,55	4126,05	2492,78	962,88	2661,95	7617,48
Minimum	17,50	255,95	8123,95	337,48	553,34	1213,82	6547,05
Maximum	145,31	1212,50	12250,00	2830,26	1516,22	3875,77	14164,53
Count	2242	2242	2242	2242	2242	2242	2242

Sumber: data sekunder yang telah diolah kembali

Terlihat bahwa jumlah observasi berjumlah 2242 buah. DJI mempunyai volatilitas tertinggi yang direpresentasikan dengan deviasi standar 1527.67, sedangkan harga minyak mempunyai volatilitas terendah yaitu 26.22. Ini berarti DJI mempunyai risiko tertinggi, dan harga minyak mempunyai risiko terendah. Untuk rata-rata, *standard error*, median, modus, varians, jangkauan, dan data maksimum tertinggi, DJI juga mempunyai data paling tinggi diantara data *intermarket* lainnya.

3.3 Penyusunan Data Sampel

Semua data yang dipakai digabungkan ke dalam satu *spreadsheet* per saham LQ45, dimana baris menunjukkan data dalam harian, disusun dari data terdahulu ke data terakhir. Kolom menunjukkan jenis data tersebut, seperti harga emas, harga minyak, kurs, IHSG, STI, KLSE, DJI, dan harga masing-masing saham LQ45. Satu *spreadsheet* terdiri dari 7 jenis data *intermarket* dan harga masing-masing saham indeks LQ45, sehingga total terdapat 45 buah *spreadsheet*. Data disusun dalam *spreadsheet* seperti Tabel 3.2 untuk masing-masing saham.

Tabel 3.2 Penyusunan Data

Hari _t (0)	Variabel Independen								Variabel Dependen
	oil _(t) (1)	gold _(t) (2)	kurs _(t) (3)	IHSG _(t) (4)	KLSE _(t) (5)	STI _(t) (6)	DJI _(t) (7)	saham _{i(t)} (8)	saham _{i(t+1)} (9)
0	oil ₀	gold ₀	kurs ₀	IHSG ₀	KLSE ₀	STI ₀	DJI ₀	saham _{it}	saham _{i(t+1)}
1	oil ₁	gold ₁	kurs ₁	IHSG ₁	KLSE ₁	STI ₁	DJI ₁	saham _{i(t+1)}	saham _{i(t+2)}
2	oil ₂	gold ₂	kurs ₂	IHSG ₂	KLSE ₂	STI ₂	DJI ₂	saham _{i(t+2)}	saham _{i(t+3)}
...
n	oil _n	gold _n	kurs _n	IHSG _n	KLSE _n	STI _n	DJI _n	saham _{in}	saham _{i(n+1)}

Keterangan:

$t = 0, 1, 2, \dots, n$ (jumlah observasi data);

$i = 1, 2, \dots, 45$ (jumlah saham LQ45)

3.4 Metodologi Penelitian

Penelitian ini disusun berdasarkan metodologi yang secara garis besar meliputi kegiatan berikut:

- a) Penentuan sampel penelitian
- b) Identifikasi variabel
- c) Proses pembentukan model ARIMA dan ANN
- d) *Forecast* menggunakan kedua model
- e) Analisa hasil *forecast* dengan menggunakan rasio *error*
- f) Uji Hipotesis

3.4.1 Penentuan Sampel Penelitian

Penentuan sampel data penelitian dari populasi data telah dijelaskan pada sub bab 3.2 di atas. Populasi data adalah saham-saham di Indonesia, sedangkan sampel data adalah saham-saham anggota indeks LQ45. Penelitian akan menggunakan data sampel ini untuk membentuk model.

3.4.2 Identifikasi Variabel

Identifikasi dan pengukuran variabel dalam penelitian ini adalah sebagai berikut:

- Variabel Dependen
Variabel dependen merupakan variabel yang menjadi perhatian utama dalam penelitian (Sekaran, 2006, p. 117). Variabel dependen pada penelitian ini adalah harga saham ($t+1$), yakni harga saham keesokan hari dari saham indeks LQ45. Pada Tabel 3.2 di atas, variabel dependen adalah variabel pada kolom nomor 9.
- Variabel Independen
Variabel independen adalah variabel yang mempengaruhi variabel dependen, baik secara positif maupun negatif. Jika terdapat variabel independen, variabel dependen juga hadir, dan dengan setiap unit kenaikan dalam variabel independen, terdapat pula kenaikan atau penurunan dalam variabel dependen (Sekaran, 2006, p. 117). Variabel independen dalam penelitian ini adalah:
 - a) harga minyak dunia
 - b) harga emas dunia

- c) kurs Rupiah terhadap Dolar Amerika
- d) IHSG (Indeks Harga Saham Gabungan)
- e) KLSE (Indeks Kuala Lumpur *Stock Exchange*)
- f) STI (*Strait Times Index*)
- g) DJI (*Dow Jones Industrial*)

Penyusunan variabel independen dapat dilihat pada Tabel 3.2 kolom 1 sampai dengan 8.

3.5 Metode Analisis Data Sampel

Metode analisis data yang digunakan dalam penelitian ini adalah metode *forecasting* yang terbagi menjadi 2 bagian, yaitu model *Multivariate* ARIMA dan ANN.

3.5.1 Metode Analisis Data Sampel Model *Multivariate* ARIMA

Penulis menggunakan perangkat lunak SPSS versi 17 untuk membentuk model *multivariate* ARIMA. Model *multivariate* ARIMA yang mendeskripsikan hubungan antara variabel independen dengan variabel dependen mempunyai bentuk rumus (2.11) sampai dengan (2.14) seperti telah dibahas pada Bab 2. Metode analisis *multivariate* ARIMA adalah seperti berikut (Tsay, 2005, p. 16):

- a) Menentukan variabel dependen Y_t dan variabel independen $X_{it} = X_{1t}, X_{2t}, \dots, X_{8t}$. Variabel dependen dan independen yang digunakan telah dijelaskan pada sub bab 3.5 di atas.
- b) Deteksi *outlier* dan keluarkan *outlier* dari data, jika ada.
- c) Jika Y_t tidak stasioner, tentukan *univariate* ARIMA untuk Y_t dengan melakukan transformasi dan *differencing* Y_t sampai didapat model Y_t yang signifikan,
- d) Jika X_{it} tidak stasioner, transformasi dan *differencing* X_{it} sampai didapat model X_{it} yang signifikan.
- e) Mendapatkan model awal untuk Y_t dan X_{it} .
- f) *Test* model *fitness* dengan mengkombinasikan Y_t dan X_{it} yang didapat pada tahap e) menggunakan CLS (*Conditional Least Squares*) sebagai berikut (Tsay, 2005, p. 16):

$$\hat{N}_t(k) = E(N_{t+k} | N_t, N_{t-1}), \text{ anggap } N_t = 0 \text{ untuk } t < 0 \quad (3.1)$$

$$\sigma^2(k) = \sigma^2 \sum_{j=0}^{k-1} \psi_j^2 \quad (3.2)$$

dimana ψ_j^2 adalah koefisien ekspansi deret pangkat $\frac{MA}{(\Delta xAR)}$,

minimalkan $S = (\sum (N_t - \hat{N}_t(1)))^2$.

- g) Keluarkan X_{it} satu persatu dari model jika X_{it} tidak signifikan setelah *test* CLS pada tahap f) yang akan menghasilkan model baru. Ulangi lagi proses f) yakni *test* model *fitness*. Apabila semua X_{it} sudah signifikan, lanjutkan ke proses h).
- h) Keluarkan parameter ARMA yang tidak signifikan dari model.
- i) *Test* model *fitness* menggunakan CLS.
- j) Untuk setiap X_{it} , cari *lag* yang menyebabkan X_{it} signifikan, dan keluarkan parameter ARMA yang tidak signifikan dari model.
- k) *Test* model *fitness* menggunakan CLS. Keluarkan semua parameter yang tidak signifikan dari model.
- l) *Test* model *fitness* dengan ML (*Maximum Likelihood*). Keluarkan semua parameter yang tidak signifikan dari model. Metode ML adalah seperti berikut (Tsay, 2005, p. 16):

$$\hat{N}_t(K) = E(N_{t+k} | N_t - \hat{N}_{t-1}, \dots, N_1) \quad (3.3)$$

$$\text{maksimalkan } \textit{likelihood} \text{ dari } \left\langle N_t - \hat{N}_t(1) \right\rangle_{t=1}^n \quad (3.4)$$

$$L = \ln(S/n) - (1/n) \sum_{j=1}^n \ln(\eta_j) \quad (3.5)$$

$$S = \frac{\sum (N_t - \hat{N}_t(1))^2}{\mu_t}, \quad (3.6)$$

- m) Uji model *multivariate* ARIMA yang didapat menggunakan uji Ljung-Box Q, yang detailnya adalah sebagai berikut:

Hipotesis uji:

- ♦ H_0 : model *multivariate* ARIMA layak untuk digunakan

- ♦ H_1 : model *multivariate* ARIMA tidak layak untuk digunakan

Kriteria uji: tolak hipotesis nol bila nilai signifikansi statistik Ljung-Box $Q < 0.05$. Rumus statistik Ljung-Box Q adalah sebagai berikut (Tsay, 2005, p. 9):

$$LB = n(n+2) \sum_{k=1}^K r^2 / (n-k) \quad (3.7)$$

- n) Jika uji dengan Ljung-Box menyatakan bahwa model signifikan (layak digunakan), maka model *multivariate* ARIMA tersebut dapat dipakai untuk memprediksi / mem-forecast harga saham.
- o) *Forecast* dilakukan dengan menggunakan *out-of-sample* variabel independen sebanyak 5 hari ke depan. Hasil akhir dari tahap ini adalah didapatkan data *forecast* 5 hari untuk masing-masing saham LQ45 menggunakan metode *multivariate* ARIMA. Hasil *forecast* 5 hari ke depan ini nantinya akan dibandingkan dengan hasil *forecast* menggunakan model ANN.

3.5.2 Metode Analisis Data Sampel *Artificial Neural Network*

Metode analisis data yang digunakan untuk menganalisa data dalam model ANN mengikuti langkah-langkah berikut:

- a) Pemilihan variabel input dan output
- b) Pembagian data sampel
- c) Pembentukan model ANN

Penjelasan lebih detail dari langkah-langkah di atas dapat dilihat pada penjelasan di bawah ini.

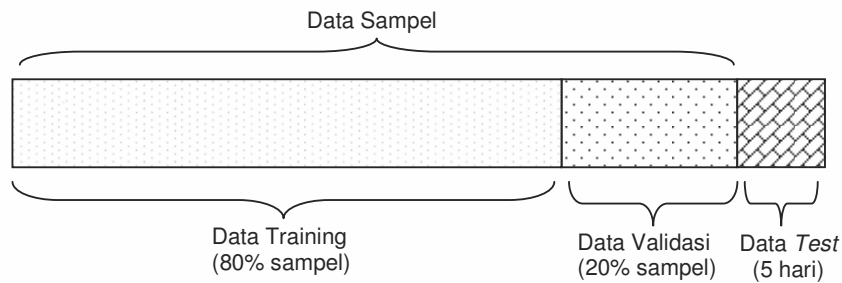
3.5.2.1 Pemilihan Variabel *Input* dan *Output*

Seperti halnya model *multivariate* ARIMA, model ANN juga memerlukan variabel – variabel dependen dan independen. Perbedaannya adalah, dalam terminologi ANN, variabel-variabel independen disebut variabel *input*, sedangkan variabel-variabel dependen disebut variabel *output*. Variabel *output* pada penelitian ini adalah harga saham ($t+1$), yakni harga saham keesokan hari dari saham indeks LQ45. Variabel *input* adalah data *intermarket* dan harga masing-masing saham LQ45, seperti telah dijelaskan pada sub bab 3.5.

3.5.2.2 Pembagian Data Sampel

Dalam analisa data untuk model ANN, data sampel harus dibagi menjadi beberapa sub bagian yang lebih kecil (Demuth, 2009, p. 156). Sub – sub bagian tersebut adalah data untuk *training* model, data untuk validasi model, dan terakhir data untuk *testing* keakuratan model. Porsi terbesar adalah data untuk *training*, dimana data ini digunakan untuk mendefinisikan parameter model (melatih kemampuan model). Data untuk validasi digunakan dengan tujuan untuk menguji kemampuan model selama proses pembentukannya. Sedangkan data untuk *testing* digunakan untuk memeriksa keakuratan *forecast* model jika diuji pada data di luar sampel.

Sebanyak 80% dari sampel digunakan untuk data *training*, sedangkan 20% dari sampel digunakan untuk data validasi. 5 hari data di luar sampel (*out of sample*) digunakan untuk memeriksa kemampuan model dalam mem*forecast* harga saham. Model ANN diharapkan dapat melakukan *forecast* meskipun data *input* yang diberikan (*out of sample*) belum pernah digunakan dalam *training*. Gambar 3.4 menunjukkan pembagian data sampel.



Gambar 3.1 Pembagian data sampel

Sumber: Demuth (2009, p. 156), telah diolah kembali.

3.5.2.3 Pembentukan Model ANN

Pada tahap ini akan dibentuk model ANN. Tujuan membuat model ANN ini adalah untuk membentuk model pergerakan saham-saham LQ45. Data historis harga saham digunakan dengan tujuan untuk memprediksi harga saham di masa depan, sehingga hasil akhir yang diharapkan dari model ANN sudah diketahui, yakni harga saham di masa depan tersebut. Menurut Iskandar(2006), jika hasil akhir sudah diketahui, jenis pelatihan ANN yang cocok adalah jenis *supervised training backpropagation*. Metode *training ANN backpropagation* merupakan algoritma yang sesuai untuk menggeneralisasi suatu pola sebagai *universal approximator* (Iskandar, 2006, p. 136).

Model ANN dibangun dari arsitektur ANN yang terdiri dari *input layer*, *hidden layer*, dan *output layer*. Masing-masing *layer* terdiri dari satu atau lebih *neuron*. Selain itu, parameter lainnya yang harus ditentukan dalam membangun model ANN adalah jenis *backpropagation training*, dan fungsi aktivasi pada masing-masing *layer*.

Kemampuan suatu model ANN bergantung pada jumlah *hidden layer*. Inti dari algoritma *backpropagation* adalah *hidden layer*. *Hidden layer* berpengaruh terhadap kemampuan model ANN untuk menggeneralisasi suatu fungsi / pola. Penelitian yang dilakukan oleh Iskandar (2006, p. 136) menunjukkan bahwa cukup diperlukan satu atau dua *hidden layer* saja untuk fungsi aproksimasi ini.

Jumlah *neuron* yang optimal pada *hidden layer* berpengaruh terhadap keakuratan model ANN dalam mem-*forecast* suatu deret data. Jika jumlah *neuron*

terlalu banyak, maka model cenderung tidak akan berfungsi jika diaplikasikan pada data di luar sampel (*overfitting*). Ini berarti model hanya berfungsi baik hanya di ruang sampelnya saja, dan apabila di-*test* menggunakan data yang di luar sampel, maka model ANN akan tidak dapat memprediksi dengan baik. Jika jumlah *neuron* terlalu sedikit, maka yang terjadi adalah model akan tidak dapat memprediksi dengan baik meskipun di-*test* menggunakan data yang berada di dalam sampel.

Parameter lain yang perlu ditentukan dalam membangun model ANN adalah fungsi transfer / fungsi aktivasi di masing-masing *neuron*. Fungsi transfer berfungsi untuk meneruskan data dari satu *neuron* ke *neuron* yang lain. Fungsi transfer yang umum dipakai dalam membentuk model ANN adalah *hyperbolic tangent sigmoid transfer function*, *log-sigmoid transfer function*, dan *linear transfer function*.

ANN bekerja dengan meminimalkan *error* yang terjadi antara harga saham yang sebenarnya dengan harga saham hasil pemrosesan model. Proses meminimalkan *error* ini yang disebut *training* ANN. Telah disebutkan di awal sub bab bahwa metode *training* ANN yang sesuai dengan penelitian ini adalah *backpropagation training*. Dalam membuat model ANN yang terbaik untuk saham-saham LQ45, diperlukan penentuan jenis *training* yang cocok. Ada banyak jenis *backpropagation training*, seperti *gradient descent backpropagation*, *gradient descent backpropagation with momentum*, *conjugate gradient backpropagation with powell-beale restarts*, *scaled conjugate gradient backpropagation*, *bfgs quasi-newton backpropagation*, *levenberg-marquardt backpropagation*, *batch training with weight and bias learning rules*, *bayesian regulation backpropagation*, *gradient descent with adaptive learning rate backpropagation*, *gradient descent with momentum and adaptive learning rate backpropagation*, *one-step secant backpropagation*, *resilient backpropagation*, dan *sequential order incremental training with learning functions* (Demuth, 2009, p. 806).

Untuk menentukan model ANN yang terbaik dalam memprediksi pergerakan harga saham-saham LQ45, penulis melakukan kombinasi percobaan antara jumlah *input layer*, *hidden layer*, *output layer*, jumlah *hidden layer*, fungsi

aktivasi antar *neuron* dan jenis *backpropagation training*. Data yang dikombinasikan tersebut disebut juga kombinasi parameter ANN atau arsitektur ANN. Penulis membuat pemrograman percobaan model ANN dengan menggunakan perangkat lunak Matlab versi 7.7.

Hasil akhir dari percobaan ini adalah ditemukannya suatu kombinasi parameter ANN yang mempunyai RMSE (*Root Mean Squared Error*) terkecil antara harga saham aktual dengan harga saham hasil validasi. Kombinasi parameter ANN dengan RMSE terkecil inilah yang disebut dengan model ANN terbaik untuk memprediksi saham-saham LQ45. Pada akhir tahapan ini, maka pertanyaan penelitian mengenai seperti apakah arsitektur model ANN terbaik yang dapat dengan tepat memprediksi harga saham LQ45 dapat terjawab.

3.5.2.4 Forecast Data

Setelah diperoleh model ANN, tahapan berikutnya adalah melakukan *forecast* dengan *test data*. *Test data* yang dipakai adalah data *intermarket* sebanyak 5 hari berturut-turut setelah tanggal terakhir data sampel. Data ini bukanlah data yang digunakan untuk *training* ataupun validasi model, akan tetapi data di luar sampel. Tujuan kegiatan *forecast* ini adalah untuk menguji keakuratan model ANN terhadap data baru di luar sampel. Kegiatan *forecast* ini dilakukan untuk semua saham LQ45. Hasil akhir dari tahap ini adalah didapatkan data *forecast* 5 hari ke depan untuk masing-masing saham LQ45 menggunakan metode ANN. Hasil *forecast* 5 hari ke depan ini nantinya akan dibandingkan dengan hasil *forecast* menggunakan model *multivariate ARIMA*.

3.5.3 Perbandingan Keakuratan Forecast Dengan Rasio Error

Pada tahapan sebelumnya, telah diperoleh hasil *forecast* menggunakan kedua model, yaitu model ANN dan *multivariate ARIMA*. Tahap ini akan membandingkan tingkat keakuratan kedua model menggunakan rasio-rasio *error*. *Error* dihitung dengan membandingkan selisih antara data harga saham hasil *forecast* dengan data harga saham sebenarnya. Semakin kecil rasio *error*, maka semakin baik model tersebut dalam melakukan *forecast*.

Rasio-rasio *error* yang dipakai adalah RMSE (*Root Mean Squared Error*), MAE (*Mean Absolute Error*), dan MAPE (*Mean Absolute Percentage Error*). RMSE merupakan akar kuadrat rata-rata dari selisih antara *output* model dengan data yang sebenarnya. Rumus MSE adalah sebagai berikut (Buwana, 2006, p. 49):

$$RMSE = \sqrt{\frac{\sum (n_f - n_a)^2}{m}} \quad (3.8)$$

MAE merupakan hasil nilai absolut dari selisih antara nilai *output* model dengan data sebenarnya. Rumus MAE adalah sebagai berikut (Buwana, 2006, p. 49):

$$MAE = \frac{\sum |n_f - n_a|}{m} \quad (3.9)$$

Sedangkan MAPE adalah perhitungan MAE yang hasilnya dalam bentuk persentase, seperti rumus berikut (Buwana, 2006, p. 49):

$$MAPE = \frac{1}{m} * \frac{\sum |n_f - n_a|}{n_a} * 100\% \quad (3.10)$$

Penjelasan rumus tersebut di atas adalah sebagai berikut:

n_f = harga saham *forecast*

n_a = harga saham *aktual*

m = jumlah data (dalam hal ini adalah 5, karena *forecast* 5 hari ke depan).

Tahap ini akan menghasilkan rasio RMSE, MAE, dan MAPE untuk kedua metode (ANN dan *multivariate ARIMA*) yang meliputi 45 buah saham LQ45, untuk selanjutnya dapat dianalisa apakah model ANN lebih akurat dari model *multivariate ARIMA*, ataukah sebaliknya. Pada akhir tahapan ini, maka pertanyaan penelitian mengenai seberapa akurat metode ANN dapat memprediksi harga saham LQ45 dapat terjawab.

3.5.4 Pengujian Hipotesis Keakuratan *Forecast*

Unuk mengetahui apakah keakuratan *forecast* masing-masing model signifikan secara statistik atau tidak, maka pada tahap ini akan dilakukan uji keakuratan *forecast*. Uji yang dilakukan adalah uji Diebold-Mariano. Jika ada suatu deret data dan dua hasil prediksi, *loss criterion* dapat dihitung berdasarkan (MSE / *Mean Squared Error*) untuk menghitung ukuran keakuratan prediksi yang dihipotesiskan sebagai berikut (Diebold, 1995, p. 253):

- Ho: Keakuratan *forecast* tidak signifikan lebih akurat
- H₁: Keakuratan *forecast* signifikan lebih akurat

Kriteria uji: tolak Ho jika nilai signifikansi statistik *p-value* < 0.05.

Detail metode Diebold-Mariano adalah sebagai berikut (Diebold, 1995, p. 253):

Misalkan *forecast error* dari kedua model adalah

$$\mathcal{E}_{t+h|t}^1 = y_{t+h} - y_{t+h|t}^1 \quad (3.11)$$

$$\mathcal{E}_{t+h|t}^2 = y_{t+h} - y_{t+h|t}^2 \quad (3.12)$$

Dimana *h* adalah *forecast h* langkah ke depan, uji Diebold-Mariano dibuat berdasarkan *loss differential* sebagai berikut:

$$d_t = L(\mathcal{E}_{t+h|t}^1) - L(\mathcal{E}_{t+h|t}^2) \quad (3.13)$$

Diebold-Mariano statistik:

$$S = \frac{\bar{d}}{(a \text{var}(\bar{d}))^{0.5}} = \frac{\bar{d}}{(LRV_{\bar{d}}/T)^{0.5}} \quad (3.14)$$

$$\bar{d} = \frac{1}{T_0} \sum_{t=t_0}^T d_t \quad (3.15)$$

$$LRV_s = \gamma_0 + 2 \sum_{j=1} \gamma_j \gamma_j = \text{cov}(d_t, d_{t-j}) \quad (3.16)$$

Pada uji Diebold-Mariano, juga digunakan kernel tipe Bartlett. Bobot dari nilai Bartlett menjamin bahwa estimasi varians bernilai definit positif (Diebold, 1995, p. 253).

3.5.5 Pengujian Hipotesis Penelitian

Tujuan utama dari penelitian ini adalah untuk meneliti apakah metode ANN dapat melakukan *forecast* lebih akurat daripada model *multivariate* ARIMA. Untuk menguji hal itu, maka dilakukan uji hipotesis dengan kriteria hipotesis sebagai berikut:

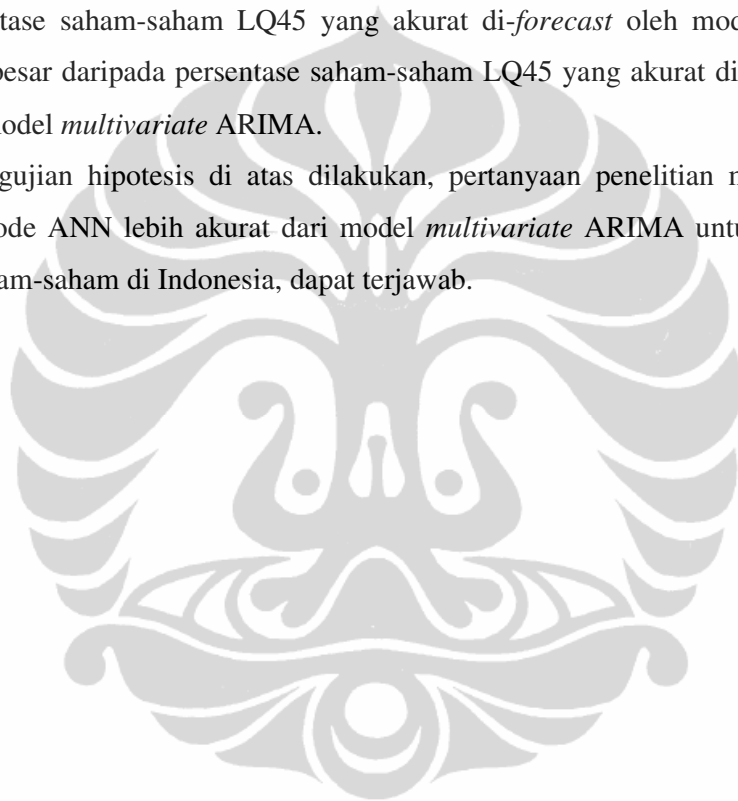
- Hipotesis nol (Ho): ketepatan *forecast* model ANN adalah sama atau tidak akurat (tidak signifikan secara statistik) dalam mem*forecast* perubahan harga saham LQ45 dibandingkan dengan metode *time series forecasting* (*multivariate* ARIMA).

- Hipotesis alternatif (H_a): model ANN dapat lebih akurat dan signifikan secara statistik dalam memforecast perubahan harga saham LQ45 dibandingkan dengan metode *time series forecasting* (*multivariate* ARIMA).

Kriteria uji: tolak H_0 jika hasil *forecast* memenuhi kedua kriteria sebagai berikut:

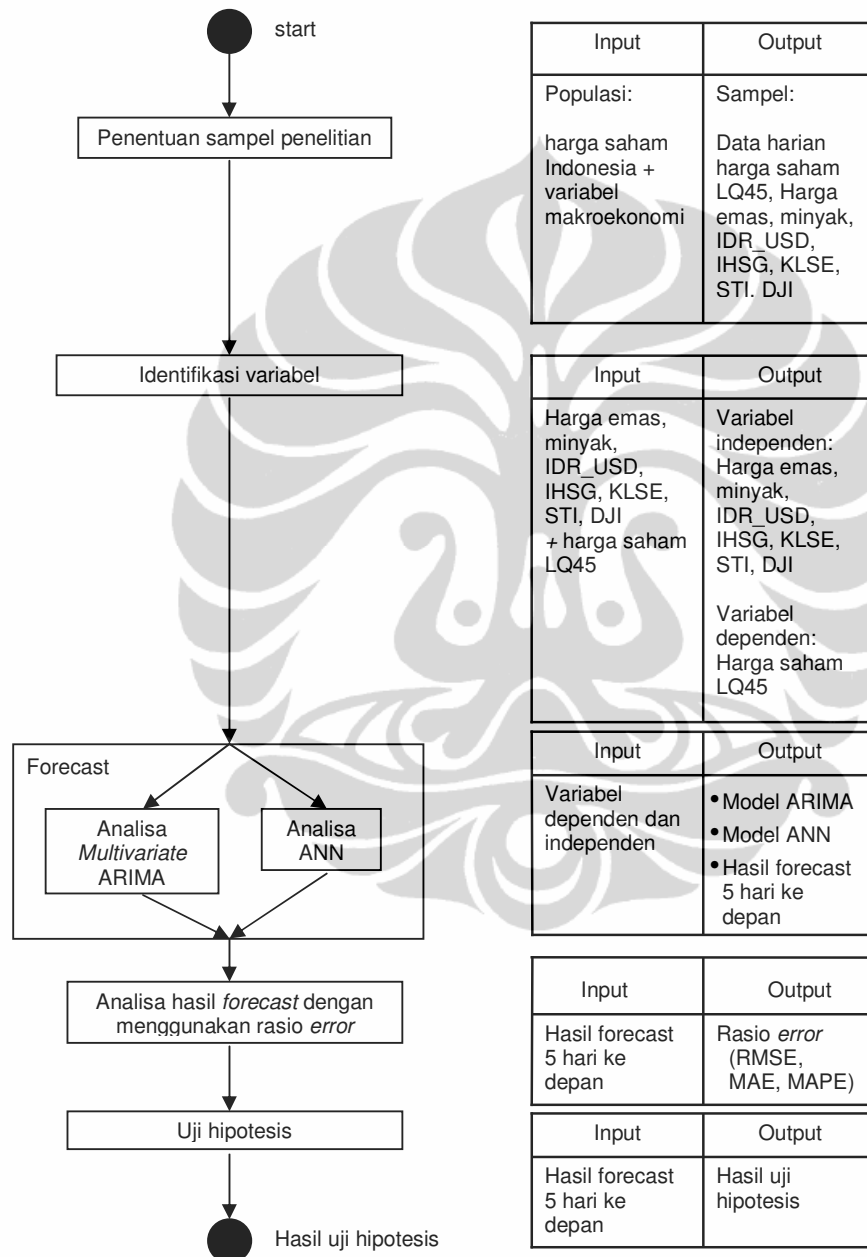
- Hasil uji Diebold-Mariano menunjukkan bahwa model ANN signifikan secara statistik lebih akurat daripada model *multivariate* ARIMA untuk saham-saham LQ45, dan
- Persentase saham-saham LQ45 yang akurat di-*forecast* oleh model ANN lebih besar daripada persentase saham-saham LQ45 yang akurat di-*forecast* oleh model *multivariate* ARIMA.

Setelah pengujian hipotesis di atas dilakukan, pertanyaan penelitian mengenai apakah metode ANN lebih akurat dari model *multivariate* ARIMA untuk memforecast saham-saham di Indonesia, dapat terjawab.



3.5.6 Flowchart Metodologi Penelitian

Penelitian ini disusun berdasarkan metodologi yang secara garis besar digambarkan dalam *flowchart* seperti Gambar 3.2 berikut:



Gambar 3.2 Flow chart metodologi penelitian