

DATA MINING DENGAN REGRESI LINIER BERGANDA UNTUK PERAMALAN TINGKAT INFLASI

Amrin

Program Studi Manajemen Informatika

AMIK Bina Sarana Informatika Jakarta

amrin.ain@bsi.ac.id

Abstract—*In this study will be used multiple linear regression method to predict the monthly inflation rate in Indonesia. In the results of the data analysis is concluded that the model of multiple linear regression obtained in this study is $Y = 0,241X_1 + 0,164X_2 + 0,271X_3 + 0,07X_4 + 0,040X_5 + 0,060X_6 + 0,169X_7 - 0,010$. Then coefficient of regression value is 0,999 and coefficient of determination value is 0,997. the performance of multiple linear regression that formed by the training data and validated by testing data generates prediction accuracy rate is very good with a Mean Absolute Deviation (MAD) is 0.0380, a Mean Square Error (MSE) is 0.0023, and a Root Mean Square Error (RMSE) is 0.0481.*

Keywords : *Inflation, multiple linear regression, Mean Square Error*

Intisari—Pada penelitian ini akan digunakan metode prediksi regresi linier berganda, untuk memprediksi tingkat inflasi bulanan di Indonesia. Dari hasil analisis data yang dilakukan disimpulkan bahwa model regresi linier berganda yang dihasilkan pada penelitian ini adalah $Y = 0,241X_1 + 0,164X_2 + 0,271X_3 + 0,07X_4 + 0,040X_5 + 0,060X_6 + 0,169X_7 - 0,010$. Adapun nilai koefisien regresi sebesar 0,999 dan koefisien determinasi sebesar 0,997 (99,7%). Performa model regresi linier berganda yang dibentuk dari data training dan divalidasi pada data testing memberikan tingkat akurasi prediksi yang cukup baik dengan nilai dengan nilai *Mean Absolute Deviation (MAD)* sebesar 0.0380, *Mean Square Error (MSE)* 0.0023, dan nilai *Root Mean Square Error (RMSE)* sebesar 0.0481.

Kata kunci: Inflasi, regresi linier berganda, rata-rata error kuadrat.

PENDAHULUAN

Inflasi adalah kenaikan harga barang dan jasa secara umum dimana barang dan jasa tersebut merupakan kebutuhan pokok masyarakat atau turunnya daya jual mata uang suatu negara (Badan Pusat Statistik, 2015).

Secara sederhana inflasi diartikan sebagai meningkatnya harga-harga secara umum dan terus menerus. Kenaikan harga dari satu atau dua barang saja tidak dapat disebut inflasi kecuali bila kenaikan itu meluas (atau mengakibatkan kenaikan harga) pada barang lainnya. Kebalikan dari inflasi disebut deflasi (Bank Indonesia, 2015).

Kestabilan inflasi merupakan prasyarat bagi pertumbuhan ekonomi yang berkesinambungan yang pada akhirnya memberikan manfaat bagi peningkatan kesejahteraan masyarakat. Pentingnya pengendalian inflasi didasarkan pada pertimbangan bahwa inflasi yang tinggi dan tidak stabil memberikan dampak negatif kepada kondisi sosial ekonomi masyarakat. Bank Indonesia menjelaskan bahwa ada tiga dampak negatif dari tingginya nilai inflasi dan ketidakstabilan inflasi. Pertama, inflasi yang tinggi akan menyebabkan pendapatan riil masyarakat akan terus turun sehingga standar hidup dari masyarakat turun dan akhirnya menjadikan semua orang, terutama orang miskin, bertambah miskin. Kedua, inflasi yang tidak stabil akan menciptakan ketidakpastian (*uncertainty*) bagi pelaku ekonomi dalam mengambil keputusan. Pengalaman empiris menunjukkan bahwa inflasi yang tidak stabil akan menyulitkan keputusan masyarakat dalam melakukan konsumsi, investasi, dan produksi, yang pada akhirnya akan menurunkan pertumbuhan ekonomi. Ketiga, tingkat inflasi domestik yang lebih tinggi dibanding dengan tingkat inflasi di negara tetangga menjadikan tingkat bunga domestik riil menjadi tidak kompetitif sehingga dapat memberikan tekanan pada nilai rupiah (Bank Indonesia, 2015).

Penelitian mengenai peramalan inflasi di suatu negara mendapatkan perhatian yang positif bagi peneliti makroekonomi. Sebagian besar bank sentral menggunakan inflasi sebagai salah satu pertimbangan untuk mengambil kebijakan moneter. Kebijakan moneter diambil dengan pertimbangan nilai inflasi yang akan datang. Nilai inflasi sekarang, merupakan hasil dari kebijakan yang lalu, mungkin hanya memberikan informasi yang samar-samar. Bagi pemerintah, peramalan inflasi merupakan

jembatan penghubung untuk mengetahui nilai inflasi yang akan datang. Penelitian ini merupakan pengembangan peramalan inflasi di Indonesia yang dapat memberikan input bagi Bank Indonesia sebagai pertimbangan pengambilan kebijakan.

Pada penelitian ini akan digunakan metode prediksi regresi linier berganda, untuk memprediksi tingkat inflasi bulanan di masa yang akan datang. Data yang penulis gunakan adalah data tingkat inflasi bulanan di Indonesia yang dikeluarkan oleh Badan Pusat Statistik dan Bank Indonesia.

Kelebihan metode regresi linier berganda diantaranya melakukan generalisasi dan ekstraksi dari pola data tertentu, mampu mengakuisisi pengetahuan walau tidak ada kepastian, dan mampu melakukan perhitungan secara paralel sehingga proses lebih singkat.

BAHAN DAN METODE

1. Data Mining

Data mining adalah rangkaian proses untuk menggali nilai tambah berupa informasi yang belum terekplorasi dari sebuah basis data, melakukan eksplorasi dengan cara-cara tertentu untuk memanipulasi data menjadi informasi yang lebih berharga dengan cara mengekstraksi dan mengenali pola penting dari basis data (Han & Kamber, 2006). Menurut Daryl Pregibons dalam (Gorunescu, 2011) "*Data mining* adalah perpaduan dari ilmu statistik, kecerdasan buatan, dan penelitian bidang *database*". Nama *data mining* berasal dari kemiripan antara pencarian informasi yang bernilai dari *database* yang besar dengan menambang sebuah gunung untuk sesuatu yang bernilai (Sumathi, 2006). Keduanya memerlukan penyaringan melalui sejumlah besar material, atau menyelidiki dengan cerdas untuk mencari keberadaan sesuatu yang disebut bernilai tadi.

Data Mining merupakan teknologi baru yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data mereka. Beberapa aplikasi data mining fokus pada prediksi, mereka meramalkan apa yang akan terjadi dalam situasi baru dari data yang menggambarkan apa yang terjadi di masa lalu (Witten, Frank, & Hall, 2011).

2. Regresi Linier Berganda

Analisis yang memiliki variabel bebas lebih dari satu disebut analisis regresi linier berganda. Teknik regresi linier berganda digunakan untuk mengetahui ada tidaknya

pengaruh signifikan dua atau lebih variabel bebas ($X_1, X_2, X_3, \dots, X_k$) terhadap variabel terikat (Y). Model regresi linier berganda untuk populasi dapat ditunjukkan sebagai berikut (Supranto, 2004):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + e$$

Model regresi linier berganda untuk populasi diatas dapat ditaksir dengan model regresi linier berganda untuk sampel, yaitu :

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n$$

dengan :

\hat{Y} = nilai penduga bagi variabel Y

b_0 = dugaan bagi parameter konstanta

b_1, b_2, \dots, b_k = dugaan bagi parameter konstanta $\beta_1, \beta_2, \dots, \beta_n$

X = Variabel bebas

a. Uji Asumsi Klasik

1) Uji Multikolinearitas

Uji multikolinearitas digunakan untuk mengetahui ada atau tidaknya penyimpangan asumsi klasik multikolinearitas, yaitu adanya hubungan linear antar variabel independen dalam model regresi (Priyatno, 2008).

2) Uji Autokorelasi

Uji autokorelasi digunakan untuk mengetahui ada atau tidaknya penyimpangan asumsi klasik autokorelasi, yaitu korelasi yang terjadi antara residual pada satu pengamatan dengan pengamatan lain pada model regresi (Priyatno, 2008).

b. Metode Kuadrat Terkecil

Salah satu prosedur pendugaan model untuk regresi linier berganda adalah dengan prosedur *Least Square* (kuadrat terkecil). Konsep dari metode *least square* adalah menduga koefisien regresi (β) dengan meminimumkan kesalahan atau error (Supranto, 2004).

c. Koefisien Determinasi

koefisien determinasi merupakan Alat untuk mengukur tingkat kecocokan/kesempurnaan model regresi atau untuk menyatakan proporsi keragaman total nilai-nilai peubah Y yang dapat dijelaskan oleh nilai-nilai peubah X melalui hubungan linier tersebut (Riana, 2012).

3. Analisis Data

Analisis data yang digunakan dalam penelitian ini adalah Regresi linier berganda. Model ini dipilih untuk mengetahui seberapa besar pengaruh variabel bebas terhadap variabel terikat baik secara parsial maupun bersama-sama. Sebelum model regresi digunakan untuk menguji hipotesis, maka terlebih dahulu dilakukan:

a. Pengujian asumsi klasik

Dalam penelitian ini uji asumsi klasik yang dilakukan adalah:

1. Uji multikolinieritas
 2. Uji autokorelasi
- b. Pengujian Hipotesis

Apabila syarat untuk ditelitinya suatu model regresi telah terpenuhi semua, maka langkah selanjutnya untuk mengetahui diterima atau tidaknya hipotesis yang diajukan dalam penelitian ini, dilakukan analisis data dengan :

1. Uji F
2. Uji T

4. Metode Pengumpulan Data

a. Data Primer

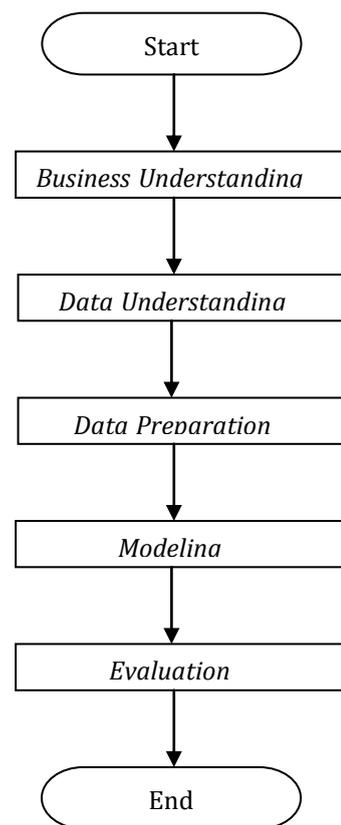
Data primer adalah data yang dikumpulkan langsung oleh peneliti dari responden, dan bukan berasal dari pengumpulan data yang pernah dilakukan sebelumnya. Data primer adalah data yang diperoleh dari sumber-sumber asli. Sumber asli disini diartikan sebagai sumber pertama darimana data tersebut diperoleh.

b. Data Sekunder

Data sekunder adalah data yang telah diolah dari data primer. Contoh: data kependudukan yang dikeluarkan oleh Badan Pusat Statistik (BPS). Pada penelitian ini penulis menggunakan data sekunder, yaitu data inflasi bulanan indonesia menurut kelompok pengeluaran tahun 2006 - 2015 yang dikeluarkan oleh BPS lewat website resminya (www.bps.go.id, 2015). Jenis data sekunder yang lain pada penelitian ini adalah semua pustaka dan laporan penelitian yang telah dilakukan baik dalam bentuk jurnal maupun kumpulan seminar nasional yang berhubungan dengan peramalan tingkat inflasi.

5. Tahap Penelitian

Terdapat beberapa tahap dalam pengolahan data eksperimen, pada penelitian ini menggunakan model *Cross-Standard Industry for Data Mining* (CRISP-DM), yaitu seperti terlihat pada diagram di bawah ini: (Sumathi, 2006)



Sumber: Sumathi (2006)

Gambar 1. Diagram Tahap Penelitian

HASIL DAN PEMBAHASAN

1. Variabel Penelitian

Data yang diperoleh untuk penelitian ini sebanyak 118 *record*, dimana 80% dari data tersebut sebagai data training, dan 20% dari data sebagai data testing. Variabel independen pada penelitian ini terdiri dari tujuh variabel, yaitu: 1. Inflasi Makanan, 2. Inflasi Makanan Jadi, Minuman, Rokok, dan Tembakau, 3. Inflasi Perumahan, Air, Listrik, Gas, dan Bahan Bakar, 4. Inflasi Sandang, 5. Inflasi Kesehatan, 6. Inflasi Pendidikan, Rekreasi, dan Olahraga, 7. Inflasi Transpor, Komunikasi, dan Jasa Keuangan, Sedangkan variabel dependen adalah variabel Inflasi Umum. Perangkat lunak yang digunakan untuk menganalisa adalah *spss versi 17.0*.

2. Uji Asumsi Klasik

a. Uji Multikolinearitas

Hasil uji multikolinearitas dengan menggunakan *spss versi 17.0* diperlihatkan oleh tabel 1 berikut ini:

Tabel 1. Hasil Uji Multikolinearitas

Model		Collinearity Statistics	
		Tolerance	VIF
1	x1	.751	1.332
	x2	.560	1.786
	x3	.644	1.553
	x4	.904	1.106
	x5	.739	1.353
	x6	.934	1.071
	x7	.803	1.245

Sumber: Hasil Pengolahan Menggunakan Spss 17.0

Dari hasil di atas dapat diketahui nilai *variance inflation factor* (VIF) ketujuh variabel independen lebih kecil dari 5, sehingga bisa diduga bahwa antarvariabel independen tidak terjadi persoalan multikolinearitas.

b. Uji Autokorelasi

Hasil uji autokorelasi dengan menggunakan *spss versi 17.0* diperlihatkan oleh tabel 2 berikut ini:

Tabel 2. Hasil Uji Autokorelasi

Model	R	R Square	Durbin-Watson
1	.999 ^a	.997	1.625

Sumber: Hasil Pengolahan Menggunakan Spss Versi 17.0

Dari hasil di atas diketahui nilai D (Durbin-Watson) sebesar 1.625. Selanjutnya nilai tersebut dibandingkan dengan nilai penerimaan H_0 dari tabel D (Durbin-Watson) diantara 1.5151 dan 1.8264. Karena nilai D hitung terletak pada daerah penerimaan H_0 sehingga asumsi tidak ada autokorelasi dapat diterima. Dari tabel di atas juga diketahui nilai koefisien regresi sebesar 0,999 dan koefisien determinasi sebesar 0,997 (99,7%).

3. Uji Koefisien Regresi Linier Berganda

a. Uji Simultan (Uji F)

Adapun langkah-langkah pengujiannya sebagai berikut:

1) Menentukan Formulasi Hipotesis

- $H_0 : b_i = 0$, artinya variabel independen tidak mempunyai pengaruh yang

signifikan secara bersama-sama terhadap variabel dependen.

- $H_0 : b_i \neq 0$, artinya variabel independen mempunyai pengaruh yang signifikan secara bersama-sama terhadap variabel dependen.

2) Menentukan derajat kepercayaan 95% ($\alpha = 0,05$)

3) Menentukan signifikansi

- Nilai signifikansi (*P Value*) < 0,05 maka H_0 ditolak dan H_1 diterima.
- Nilai signifikansi (*P Value*) > 0,05 maka H_0 diterima dan H_1 ditolak.

4) Membuat kesimpulan

- Bila (*P Value*) < 0,05 maka H_0 ditolak dan H_1 diterima. Artinya variabel independen secara simultan (bersama-sama) mempengaruhi variabel dependen.
- Bila (*P Value*) > 0,05 maka H_0 diterima dan H_1 ditolak. Artinya variabel independen secara simultan (bersama-sama) tidak mempengaruhi variabel dependen.

Hasil uji F pada penelitian ini dengan menggunakan *spss versi 17.0* dapat dilihat pada tabel 3 berikut ini:

Tabel 3. Hasil Uji F

Model	df	F	Sig.
1 Regression	7	4357.58	.000 ^a
	6		
Residual	88		
Total	95		

Sumber: Hasil Pengolahan Menggunakan Spss 17.0

Dari hasil uji F pada penelitian ini, didapatkan nilai F sebesar 4357,586 dengan angka signifikansi sebesar 0,000. Dengan tingkat signifikansi 95% ($\alpha = 0,05$), angka signifikansi sebesar 0,000 < 0,05. Atas dasar perbandingan tersebut, maka H_0 ditolak atau berarti variabel independen mempunyai pengaruh yang signifikansi secara bersama-sama terhadap variabel dependen.

b. Uji t

Hasil uji t pada penelitian ini dengan menggunakan *spss versi 17.0* dapat dilihat pada tabel 4 berikut ini:

Tabel 4. Hasil Uji t

Model		t	
		t	Sig.
1	(Constant)	-1.346	.182
	x1	91.191	.000
	x2	12.835	.000
	x3	25.235	.000
	x4	17.434	.000
	x5	2.932	.004
	x6	13.516	.000
	x7	69.771	.000

Sumber: Hasil Pengolahan Menggunakan *Spss 17.0*.

Dari hasil uji t pada penelitian ini, diketahui nilai signifikansi dari ketujuh variabel independen < 0,05, yang berarti H_0 ditolak dan H_1 diterima. Artinya masing-masing dari ketujuh variabel independen mempunyai pengaruh yang signifikan secara parsial terhadap variabel dependen.

4. Analisis Regresi Linier Berganda

Dengan menggunakan *spss versi 17.0*, maka didapatkan hasil uji regresi pada data training seperti pada tabel 5 berikut ini:

Tabel 5. Uji Regresi

Model		Unstandardized Coefficients	
		B	Std. Error
1	(Constant)	-.010	.007
	x1	.241	.003
	x2	.164	.013
	x3	.271	.011
	x4	.070	.004
	x5	.040	.013
	x6	.060	.004
	x7	.169	.002

Sumber: Hasil Pengolahan Menggunakan *Spss 17.0*.

Berdasarkan tabel di atas maka didapatkan persamaan regresi linier berganda sebagai berikut:

$$Y = 0,241X_1 + 0,164X_2 + 0,271X_3 + 0,07X_4 + 0,040X_5 + 0,060X_6 + 0,169X_7 - 0,010$$

Model regresi linier ganda tersebut kemudian diterapkan pada data testing. Hasilnya seperti diperlihatkan oleh tabel berikut ini:

Tabel 2. Prediksi Inflasi dengan Regresi Linier Berganda

Tahun/Bulan	Real	Prediksi	error
2014 Januari	1.0700	1.17	-0.0973
Pebruari	0.2600	0.28	-0.0200
Maret	0.0800	0.07	0.0112
April	-0.0200	-0.08	0.0560
Mei	0.1600	0.14	0.0219
Juni	0.4300	0.46	-0.0297
Juli	0.9300	0.99	-0.0643
Agustus	0.4700	0.46	0.0063
September	0.2700	0.24	0.0287
Oktober	0.4700	0.48	-0.0122
Nopember	1.5000	1.50	0.0012
Desember	2.4600	2.51	-0.0544
2015 Januari	-0.2400	-0.12	-0.1167
Pebruari	-0.3600	-0.38	0.0175
Maret	0.1700	0.15	0.0212
April	0.3600	0.28	0.0796
Mei	0.5000	0.53	-0.0283
Juni	0.5400	0.58	-0.0433
Juli	0.9300	0.95	-0.0218
Agustus	0.3900	0.40	-0.0130
September	-0.0500	-0.09	0.0382
Oktober	-0.0800	-0.13	0.0534

Sumber: Hasil Pengolahan (2016)

5. Performa Model

untuk mengukur ketepatan dan keakuratan model prediksi inflasi bulanan di Indonesia ditentukan dengan menghitung nilai *Mean Square Error* (MSE), *Root Mean Square Error* (RMSE), dan *Mean Absolute Deviation* (MAD). Perbandingan dari kedua model diperlihatkan oleh tabel di bawah ini:

Tabel 5. Performa Model

Model	Multiple Linear Regression
MAD	0.0380
MSE	0.0023
RMSE	0.0481

Sumber: Hasil Pengolahan (2016)

KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah dilakukan dapat disimpulkan bahwa pada hasil penelitian dengan menggunakan uji F diketahui bahwa variabel dependen secara simultan atau bersama-sama berpengaruh signifikan terhadap variabel dependen dari model regresi. Dari hasil uji t pada penelitian ini, diketahui masing-masing dari ketujuh variabel independen mempunyai pengaruh yang signifikan secara parsial terhadap variabel dependen. Model regresi linier berganda yang dihasilkan pada penelitian ini adalah $Y = 0,241X_1 + 0,164X_2 + 0,271X_3 + 0,07X_4 + 0,040X_5 + 0,060X_6 + 0,169X_7 - 0,010$. Adapun nilai koefisien regresi sebesar 0,999 dan koefisien determinasi sebesar 0,997 (99,7%). Performa model regresi linier berganda dalam meramal tingkat inflasi bulanan Indonesia menghasilkan tingkat akurasi dengan nilai *Mean Absolute Deviation* (MAD) sebesar 0.0380, *Mean Square Error* (MSE) 0.0023, dan nilai *Root Mean Square Error* (RMSE) sebesar 0.0481.

DAFTAR PUSTAKA

- Badan Pusat Statistik (BPS). 2015. Inflasi, <http://www.bps.go.id/aboutus.php?id_subyek=03&tabel=1&fl=2> Diunduh pada tanggal 03 Desember 2015.
- Gorunescu, Florin (2011). *Data Mining: Concepts, Models, and Techniques*. Verlag Berlin Heidelberg: Springer.
- Han, J., & Kamber, M. (2006). *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman.
- Priyatno, Dwi. (2008). *Mandiri Belajar SPSS*. Yogyakarta: MediaKom.
- Riana, Dwiza. (2012). *Statistika Deskriptif itu mudah*. Jakarta: Jelajah Nusa.
- Sumathi, & S., Sivanandam, S.N. (2006). *Introduction to Data Mining and its Applications*. Berlin Heidelberg New York: Springer

Supranto, J. 2004. Analisis Multivariat Arti dan Interpretasi. Rineka Cipta. Jakarta.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning and Tools*. Burlington: Morgan Kaufmann Publisher.



Amrin, S.Si, M.Kom. Dompu 10 Agustus 1980. Tahun 2003 lulus dari Program Strata Satu (S1) Jurusan Matematika Universitas Diponegoro Semarang. Tahun 2014 lulus dari Program Strata Dua (S2)

Jurusan Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta. Pekerjaan saat ini sebagai Dosen AMIK BSI Jakarta sejak tahun 2007. Telah menulis beberapa paper di beberapa jurnal diantaranya Jurnal TECHNO STMIK Nusa Mandiri, Jurnal Teknik Komputer AMIK BSI Jakarta dan Jurnal PARADIGMA AMIK BSI Jakarta.