

**MACHINE LEARNING PADA ANALISIS DATA INFLASI INDONESIA
MENGUNAKAN METODE REGRESI LOGISTIK**

***MACHINE LEARNING IN INDOONESIAN INFLATION DATA ANALYSIS
USING LOGISTIC REGRESSION METHOD***

Nadia Armelia Amanda¹, Marhaeni²

Program Studi Teknik Informatika, Fakultas Sains dan Teknologi Informasi
Institut Sains dan Teknologi Nasional
Jl. Moh. Kahfi II, Bhumi Srengseng Indah, Jakarta Selatan, 12640
Telp. (021) 7874647, Fax. (021) 7866955
nadiaarmelia.na@gmail.com¹, enimarhaeni24@gmail.com²

ABSTRAK

Inflasi memiliki arti kecenderungan kenaikan harga barang dan jasa secara umum yang berlangsung terus menerus. Bila harga barang dan jasa lebih tinggi di dalam negeri, maka inflasi meningkat. Kenaikan harga barang dan jasa juga menyebabkan turunnya nilai uang. Penelitian ini bertujuan menganalisis data inflasi memanfaatkan *machine learning* dengan metode regresi logistik sebagai algoritma dalam menentukan naik dan turunnya nilai inflasi pada negara Indonesia. Hasil analisis menunjukkan bahwa nilai inflasi selama tahun 2020 sampai dengan 2022 mengalami dengan nilai akurasi yang cukup tinggi sebesar 75% ini membuktikan bahwa inflasi diprediksi mengalami kenaikan yang datanya diambil dari sampel per bulan selama 3 tahun tersebut.

Kata Kunci : Inflasi, *Machine Learning*, Analisis Regresi Logistik

ABSTRACT

Inflation means a tendency to increase the prices of goods and services in general which continues continuously. If the prices of goods and services are higher in the country, then inflation increases. An increase in the price of goods and services also causes a decrease in the value of money. This research aims to analyze inflation data using machine learning with the logistic regression method as an algorithm in determining the rise and fall of inflation values in Indonesia. The results of the analysis show that the inflation value during 2020 to 2022 experienced a fairly high accuracy value of 75%. This proves that inflation is predicted to increase, the data was taken from monthly samples for these 3 years.

Keywords : *Inflation, Machine Learning, Logistic Regression Analysis*

1. PENDAHULUAN

Tahun lalu tepatnya tahun 2022 di Bulan September Indonesia sempat mengalami kenaikan inflasi dibandingkan dengan beberapa bulan sebelumnya. Sebagai akibatnya menyebabkan kenaikan harga pada berbagai sektor komoditas (Fernando, 2022). Hal tersebut menyebabkan Indonesia perlu adanya untuk memperkirakan nilai inflasi pada masa yang mendatang. Jika tingkat inflasi berada pada level yang terlalu tinggi, maka hal ini mampu mengindikasikan bahwa perekonomian suatu negara sudah tumbuh terlalu cepat. Karena itu dibutuhkan kenaikan suku bunga acuan oleh bank sentral supaya pertumbuhan tersebut bisa direm, sebagai akibatnya kenaikan inflasi tidak terlalu mencekik masyarakat. Inflasi terjadi karena keterbatasan komoditas yang beredar serta mengakibatkan ditahannya barang dan jasa yang ditawarkan oleh para pengusaha, sebagai akibatnya konsumen akan memperoleh barang dan jasa tadi dengan harga yang tinggi. Akibat inflasi yang mengalami

gejolak tinggi sangat berpengaruh bagi pemerintah diantaranya menyebabkan nilai uang turun serta ekspor turun sebagai akibatnya pendapatan negara berkurang dan terjadi kekacauan ekonomi karena terganggunya stabilitas ekonomi.

Inflasi artinya kesamaan harga-harga barang dan jasa termasuk faktor-faktor produksi, diukur menggunakan satuan mata uang, yang semakin menarik secara awam serta terus menerus. Kebalikan dari inflasi adalah deflasi. Namun, keadaan yang umum terjadi di perekonomian sekarang ini adalah inflasi. Kenaikan harga sebab inflasi bukan berasal dari faktor teknologi, sifat-sifat barang, serta dampak *trend*. Melainkan berasal dari harga yang terjadi di pasar antara pihak-pihak yang bebas.

Berdasarkan (Yanuar, 2018) statistika memiliki peranan sangat penting bagi perkembangan ilmu-ilmu lain untuk mengambil kesimpulan, menguji hipotesis atau teori, memahami fenomena, menganalisis eksperimen, menentukan keputusan, dan lain

sebagainya. *Machine learning* yang adalah salah satu cabang berasal *artificial intelligence* (kecerdasan buatan) saat ini terus mengalami perkembangan dan semakin terkenal.

Perkembangan ilmu statistika dan *machine learning* ini tentunya sebab tidak terlepas dari faktor utamanya, yaitu data. *Machine learning* memiliki setidaknya 2 tujuan primer, yaitu pemecahan persoalan dalam memprediksi masa depan (*unobserved event*) dan/atau memperoleh ilmu pengetahuan (*knowledge discovery*). *Statistical machine learning* mengacu pada teknik untuk memprediksi masa depan dan menerima pengetahuan dari data secara rasional. Untuk bisa mendapatkan tujuan tersebut, *statistical machine learning* bisa sebagai alat atau metode yang tepat.

Dengan memanfaatkan statistika *machine learning* diharapkan menjadi salah satu jalan dalam memprediksi inflasi dengan bantuan teknologi yang telah ada dan berkembang saat ini. Metode Regresi Logistik salah satu teknik dari *statistical machine learning* yang akan dipergunakan untuk memprediksi permasalahan inflasi di Indonesia yang akan mendatang pada tahun berikutnya.

2. METODOLOGI PENELITIAN

Dijelaskan (Budiharto, 2018), proses penelitian merupakan langkah-langkah yang mana penentuan minat atau topik riset hingga publikasi hasil penelitian dilakukan secara baik. Proses penelitian melibatkan berbagai langkah yang terstruktur seperti pada tahapan penelitian yang akan dijelaskan sebagai secara jelas sebagai berikut.

Identifikasi Masalah dan Tujuan

Penelitian biasanya diawali dengan adanya masalah yang teridentifikasi di sekitar, lingkungan atau di bidang kerja, diikuti dengan tujuan yang ingin dicapai dari penyelesaian permasalahan tersebut. Pada tahap ini masalah yang dihadapi adalah permasalahan kenaikan dan penurunan data inflasi yang dapat dianalisis dengan bantuan *machine learning* menggunakan metode regresi logistik dengan tujuan agar kedepannya pemanfaatan teknologi dapat terus dipergunakan dan terus berkembang.

Studi Literatur

Pada tahap ini penelitian akan diperkuat dengan informasi dan referensi yang dikumpulkan dari berbagai sumber seperti penelitian terdahulu, buku cetak, *e-book*, artikel jurnal *online*, *website*, artikel berita, dan literatur lainnya berkaitan dengan analisis data

inflasi menggunakan metode regresi logistik. Tahapan ini bertujuan dalam pemahaman teori terkait penelitian yang dilakukan dan hasil tahap ini dapat ditemukan dalam tinjauan pustaka.

Machine Learning

Dari (Pangaribuan et al., 2021, 3) *Machine Learning* merupakan sebuah metode berbasis komputer yang tidak perlu diatur dahulu oleh manusia dan dapat belajar dengan bantuan data dan akan semakin pintar seiring dengan banyaknya data yang telah diolah (belajar melalui pengalaman). Metode ini sering digunakan dalam menyelesaikan kasus klasifikasi dan *clustering* dan biasanya digunakan untuk menangani data dalam skala besar atau *big data*.

Pada dasarnya, *machine learning* memungkinkan komputer untuk memeriksa dan membedakan pola tanpa harus memprogramnya secara terpisah. Ketika digabungkan teknik statistik dan pembelajaran mesin adalah alat yang ampuh untuk menganalisis berbagai jenis data di banyak bidang komputasi, termasuk pengolahan gambar, pemrosesan bahasa, pemrosesan bahasa alami, kontrol robot, dan ilmu dasar seperti biologi, kedokteran, astronomi, fisika serta material.

Metode Regresi Logistik

Menurut (Wulandari et al., 2019, 382), regresi logistik atau kadang disebut menjadi *logit model*, adalah metode analisis yang digunakan untuk mendeteksi korelasi antara variabel dependen dan independen ketika variabel dependen bersifat kategorikal. Ada dua model dalam analisis regresi logistik, yaitu regresi logistik biner dan regresi logistik multinomial. Regresi logistik biner digunakan ketika variabel dependen data bersifat dikotomi. Sementara itu, bila variabel dependen yang digunakan terdiri dari lebih dua kategori, maka model regresi logistik yang sesuai adalah regresi logistik multinomial.

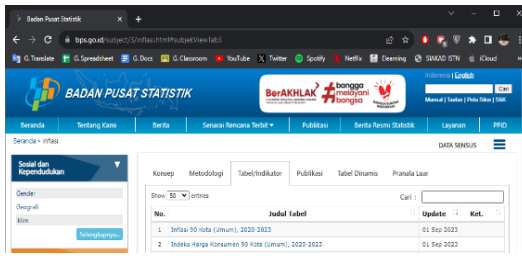
Bentuk umum model probabilitas regresi logistik dengan variabel penjelas p , dirumuskan sebagai berikut:

$$\begin{aligned} \text{logit}(p) &= \beta_0 + \beta_1 X \\ \ln(p/(1-p)) &= \beta_0 + \beta_1 X \\ p/(1-p) &= e^{(\beta_0 + \beta_1 X)} \\ p &= e^{(\beta_0 + \beta_1 X)} / (1 + e^{(\beta_0 + \beta_1 X)}) \\ &= 1 / (1 + e^{-(\beta_0 + \beta_1 X)}) \end{aligned}$$

Pengumpulan Data

Penelitian ini menggunakan data sekunder sehingga proses pengumpulan data yang akan digunakan adalah dengan mencatat dan menganalisis dokumen yang sudah ada pada situs resmi atau jurnal serta buku yang berkaitan

dengan penelitian baik dalam bentuk data statistik, informasi, dan lainnya. Seperti pada gambar 1 di bawah ini dalam situs resmi BPS data yang akan digunakan dapat diambil pada halaman tersebut.



Gambar Indikator Data. Sumber: Badan Pusat Statistik, 2023

Data yang digali adalah data inflasi yang dipublikasikan oleh Badan Pusat Statistik (BPS) Indonesia melalui situs resminya dengan mengambil data inflasi, indeks harga konsumen, nilai tukar uang, dan banyaknya uang yang tersebar dalam perbulannya pada Indonesia di 90 kota dari tahun 2020 hingga 2022 sebagai bahan pertimbangan dalam memperkirakan naik turunnya inflasi pada tahun berikutnya. Pada tahap ini juga data akan ditabulasikan dalam *google spreadsheet* atau *microsoft excel*.

Instrumen Penelitian

Instrumen penelitian yang dipergunakan dalam penelitian ini berupa perangkat keras dan perangkat lunak seperti berikut:

- 1) Perangkat Keras, yang digunakan dalam penelitian adalah laptop ACER Aspire 3 A314 - 21 Series dengan spesifikasi berikut ini:
 - Processor AMD Dual-Core A4-9120E
 - RAM 8 GB DDR4
 - Harddisk 500 GB
 - Monitor 14 inch
- 2) Perangkat Lunak, adapun yang digunakan dalam penelitian ini dalam melakukan analisis sebagai berikut:
 - Windows 10 Pro
 - Python 3 (Bahasa pemrograman)
 - Jupyter Notebook (Aplikasi web *interpreter* python yang digunakan dalam menganalisis)
 - Google Document (Aplikasi berbasis web dan *online* yang berfungsi sama seperti *microsoft word*, digunakan untuk membuat laporan penelitian)
 - Google Spreadsheet (Aplikasi produk *google* berbasis web dan *online* yang berfungsi sama seperti *microsoft excel*, digunakan untuk mengelompokkan data)

Variabel Penelitian dan Definisinya

Penelitian yang dilakukan menggunakan satu variabel dependen dan tiga variabel

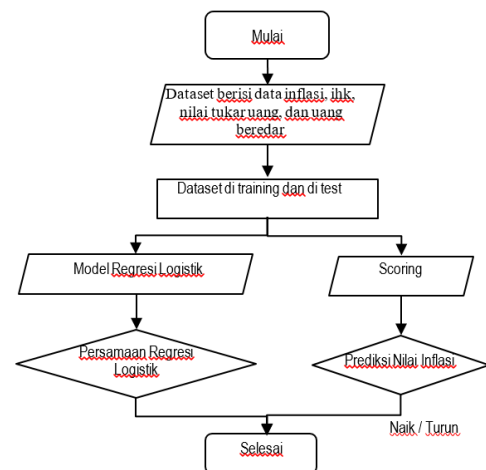
independen. Adapun variabel yang dipergunakan dalam penelitian akan ditampilkan dalam bentuk tabel berikut ini:

Tabel Variabel Penelitian

Variabel		Definisi Variabel
Variabel Dependen (y)	Inflasi	Kecenderungan naiknya harga barang dan jasa pada umumnya yang berlangsung secara terus menerus.
	Indeks Harga Konsumen	Indeks yang menghitung rata-rata perubahan harga dari suatu paket barang dan jasa yang dikonsumsi oleh suatu rumah tangga selama periode tertentu.
Variabel Independen (x)	Nilai Tukar Uang (Kurs)	Jumlah uang domestik atau banyaknya rupiah yang dibutuhkan untuk memperoleh satu unit mata uang asing.
	Jumlah Uang Beredar	Meliputi uang kartal, uang giral, uang kuasi, dan surat berharga yang diterbitkan oleh sistem moneter yang dimiliki sektor swasta domestik dengan sisa jangka waktu sampai dengan satu tahun.

Analisis Regresi Logistik

Analisis pada tahap ini bertujuan untuk mengetahui bagaimana data inflasi diperkirakan naik dan turunnya menggunakan metode regresi logistik sebagai salah satu algoritma dari *machine learning*. Ada dua alasan utama untuk memilih regresi logistik. Pertama, dari sudut pandang matematika, regresi logistik adalah fungsi yang sangat fleksibel dan mudah digunakan. Kedua, mudah dipahami dan ditafsirkan.



Gambar Flowchart Analisis Regresi Logistik. Sumber: Rancangan Penelitian, 2023

Pada gambar *flowchart* di atas adalah alur dalam analisis regresi logistik yang dilakukan penulis menggunakan *machine learning* dengan diawali memanggil perintah untuk memasukkan dataset yang sebelumnya sudah dipersiapkan. Kemudian dataset yang sudah dimasukkan dalam program tersebut dideklarasikan dengan y dan X . Sebagai keterangan y atau data inflasi adalah variabel terikat yang akan di analisis dengan variabel X sebagai variabel bebasnya.

Dalam hal ini y didefinisikan sebagai data inflasi dan X sebagai data indeks harga konsumen (X_1), nilai tukar uang (X_2), dan jumlah uang beredar (X_3). Variabel-variabel tadi akan dibedakan menjadi variabel *training* dan *test* yang nantinya akan dapat menghasilkan model persamaan regresi logistik dan dapat memprediksi naik dan turunnya inflasi dengan *scoring*.

Hal lain yang membedakan regresi logistik dengan regresi linear adalah distribusi variabel respon. Regresi linear mengasumsikan bahwa variabel respon memiliki distribusi normal, sedangkan regresi logistik mengasumsikan bahwa variabel respon tersebut memiliki distribusi bernoulli. Bentuk spesifik model regresi logistik yang akan digunakan:

$$n(x) = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)}$$

Transformasi yang menjadi fokus pada regresi logistik adalah transformasi logit. Transformasi tersebut didefinisikan dalam bentuk $\pi(x)$, sebagai:

$$\begin{aligned} \text{logit}[\pi(x)] &= g(x) \\ &= \ln(\pi(x)/(1 - \pi(x))) \\ &= \beta_0 \\ &+ \beta_1 x_1 + \dots + \beta_p x_p \end{aligned}$$

```
import statsmodels.api as sm
import pandas as pd
dataset = pd.read_csv("20202022.csv")
print(dataset)
X1 = dataset['ihk']
X2 = dataset['nilai-tukar']
X3 = dataset['uang-beredar']
y = dataset['inflasi']
intercept = [1] * 36
penjelas = pd.DataFrame({
    "intercept": intercept, "X1": X1, "X2": X2, "X3": X3})
log_reg = sm.Logit(y, penjelas).fit()
print(log_reg.summary())
```

Gambar Pseudo Code Regresi Logistik.

Sumber: Program 2023

Pada gambar ditampilkan *pseudocode* dalam mencari persamaan regresi logistik, didapati pada program tersebut dataset dideklarasikan dan dipanggil menggunakan *library* yang ada pada *pandas* dengan *function* *read_csv*. Kemudian variabel dalam dataset dideklarasikan dengan X_1 sebagai dataset data *ihk* atau indeks harga konsumen, X_2 sebagai dataset data nilai tukar uang, X_3 sebagai dataset

data uang beredar di Indonesia, dan y sebagai dataset data inflasi.

Intercept, disebut juga konstanta, adalah titik di mana garis regresi memotong sumbu Y . Dalam analisis regresi, *intercept* menunjukkan nilai rata-rata variabel dependen ketika nilai variabel independen sama dengan nol. Artinya, *intercept* adalah nilai ketika variabel independen tidak mempengaruhi variabel dependen. Sehingga pada gambar 3 dideklarasikan $\text{intercept} = [1] * 36$ dimana angka 36 ini menggambarkan jumlah data yang akan dicari regresi logistiknya.

Selanjutnya *pandas* membuat data menjadi seperti tabel dengan perintah `pd.DataFrame()` yang kemudian akan direalisasikan dalam menyajikan data analisis regresi logistik.

Hasil Analisis

Hasil analisis akan didapat bila analisis yang dilakukan berhasil, dalam tahapan ini penulis menganalisis data mengacu dari tahapan sebelumnya sesuai dengan studi literatur, data yang dikumpulkan, dan analisis regresi logistik. Nantinya hasil akan berupa keputusan akhir apakah inflasi akan naik atau turun dengan pengaruh variabel penelitian dan memanfaatkan *machine learning* dalam proses analisisnya serta dapat ditarik kesimpulan dari hasil analisis yang telah diperiksa.

Pembuatan Laporan

Pada tahap ini hasil penelitian meliputi hasil yang telah diuji baik dengan bahasa pemrograman akan disusun dalam bentuk laporan yang menjadi tujuan penelitian ini dilakukan untuk pengembangan penelitian selanjutnya.

3. HASIL DAN PEMBAHASAN

Hasil Analisis Regresi Logistik

Jupyter Notebook adalah salah satu aplikasi yang menyediakan tempat untuk melakukan pemrograman sederhana menggunakan bahasa python. Versi bahasa pemrograman python yang digunakan adalah versi 3.

Mempersiapkan Training Dataset

Sebelum melakukan *test* pada *dataset* yang berupa file CSV (*comma separated values*) perlu dipersiapkannya *training dataset*. Dengan bantuan *pandas* file dataset tadi dimuat dalam bentuk statistik pentingnya pada gambar berikut

```
In [1]: import pandas as pd
dni = pd.read_csv('20202022.csv')
dni.describe()
```

Gambar Memanggil statistik dataset

Out[1]:

	ihk	nilai-tukar	uang-beredar	inflasi
count	36.000000	36.000000	3.600000e+01	36.000000
mean	107.495278	3.923611	7.221971e+06	0.777778
std	2.843283	0.569522	6.622230e+05	0.421637
min	104.330000	3.500000	6.046651e+06	0.000000
25%	105.032500	3.500000	6.762699e+06	1.000000
50%	106.535000	3.625000	7.145311e+06	1.000000
75%	109.207500	4.250000	7.847711e+06	1.000000
max	113.590000	5.500000	8.528022e+06	1.000000

Gambar Tampilan statistik dataset

Gambar Tampilan statistik dataset menampilkan statistik dataset, terlihat ada 36 baris di dalam dataset, artinya ada 36 data inflasi yang diperhitungkan. Ada tiga variabel dengan rata-rata indeks harga konsumen (ihk) adalah 107,49 rupiah, nilai tukar dengan 5,5 sebagai angka tertinggi dan jumlah uang beredar paling sedikit 6 juta rupiah.

```
In [2]: dni.isnull().sum()
Out[2]: ihk          0
        nilai-tukar  0
        uang-beredar 0
        inflasi      0
        dtype: int64
```

Gambar Memeriksa data kosong

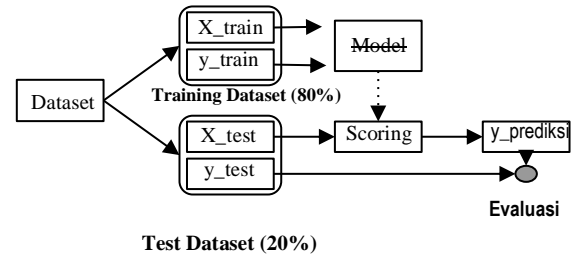
Sebagai salah satu langkah yang perlu dilakukan terhadap sembarang *dataset* yang akan dipergunakan, sangat diperlukan untuk memeriksa data yang kosong (*null*) di dalamnya. Ditampilkan pada gambar Memeriksa data kosong bahwa data sama sekali tidak ada yang kosong dan tidak ada data yang perlu untuk dibuang.

```
In [5]: dni['inflasi'].value_counts()
Out[5]: 1    28
        0     8
        Name: inflasi, dtype: int64
```

Gambar Statistik jumlah data

Dari semua data inflasi, berapa banyak data inflasi yang nilainya naik atau turun tersaji dalam gambar Statistik jumlah data yang menunjukkan bahwa ada 28 nilai inflasi yang masuk ke *class* 1 berarti data inflasi tersebut termasuk dalam data yang nilai inflasinya naik dalam *dataset*.

Penulis menggunakan sebagian besar (80%) dari dataset yang tersedia sebagai *training dataset*, dan sisanya (20%) sebagai *test dataset*. *Training dataset* digunakan untuk membentuk model, kemudian *test dataset* digunakan untuk menguji model tersebut dan mengevaluasi akurasinya ditunjukkan dalam gambar Pembagian dataset ke dalam training dan test.



Gambar Pembagian dataset ke dalam training dan test

Dalam *training dataset*, *feature* atau variabel independen akan tersimpan pada *DataFrame* bernama *X_train* sementara targetnya atau variabel dependen tersimpan di *y_train*. Demikian juga dalam *test dataset*, *feature* akan disimpan dalam *DataFrame* bernama *X_test* dan targetnya disimpan pada *y_test*.

Untuk keperluan *training*, seluruh *feature* yang ada (indeks harga konsumen, nilai tukar uang, dan jumlah uang beredar) akan digunakan. *Training dataset* dan *test dataset* harus dipilih secara acak, hal ini akan dilakukan *pandas* pada program. Fungsi *train_test_split()* dapat dipanggil dengan memasukkan ukuran *test dataset* seperti pada gambar di bawah ini dengan ukuran 0,2 atau 20%.

```
In [6]: import sklearn.model_selection as ms
X = dni[['ihk', 'nilai-tukar', 'uang-beredar']]
y = dni.inflasi
X_train, X_test, y_train, y_test = ms.train_test_split(X, y, test_size=0.2, random_state=0)
```

Gambar menampilkan perintah untuk training dataset

Perlu dipastikan bahwa angka-angka dalam *test dataset* harus berada di dalam jangkauan angka-angka di *training dataset*. Model prediksi tidak akan akurat bekerja apabila harus memprediksi dataset yang jangkauannya belum pernah dilatih dengan *training dataset*.

Pembuatan Model

Scikit-learn digunakan untuk membuat model *Logistic Regression*, yang sudah menyediakan semuanya dengan lengkap. Langkah pertama yang dilakukan adalah impor terlebih dahulu paket yang diperlukan, yaitu *LogisticRegression* di dalam *sklearn.linear_model*. Kemudian untuk melatih model, fungsi *fit()* dipanggil dengan memasukkan *Training dataset* yang sudah disiapkan sebelumnya seperti pada gambar 10 yang menunjukkan rangkaian paket dan fungsi yang dipanggil untuk membuat model regresi logit.

```
In [7]: import sklearn.linear_model as lm
model = lm.LogisticRegression(solver='lbfgs')
model.fit(X_train,y_train)
```

Gambar Model Logistic Regression

Pada titik ini, algoritma *machine learning* sudah berhasil menemukan koefisien untuk masing-masing dari ketiga *feature* di *training dataset* seperti yang ditampilkan gambar Koefisien tiga variabel yang menunjukkan isi dari koefisien *feature* atau variabel tersebut.

```
In [8]: print(model.coef_)
[[2.74053242e-12 1.04822541e-13 1.79867450e-07]]
```

Gambar Koefisien tiga variabel

Selanjutnya, memakai model regresi logit untuk mengeluarkan seluruh hasil prediksi terhadap *test dataset*, data yang sudah siap ada 8 baris data *X_test*. Jadi akan ada 8 baris hasil prediksi (dalam angka nol dan satu) seharusnya dan yang tampil seperti gambar Hasil prediksi test dataset

```
In [9]: y_prediksi = model.predict(X_test)
print(y_prediksi)
[1 1 1 1 1 1 1 1]
```

Gambar Hasil prediksi test dataset

Model regresi logit berhasil membuat prediksi apakah inflasi bernilai naik atau malah turun. Sebagai contoh dengan ditampilkannya lima data inflasi pada gambar Data acak test dataset. Data acak test dataset.

```
In [10]: X_test.head()
Out[10]:
```

	ihk	nilai-tukar	uang-beredar
31	111.57	3.75	7897628.21
20	106.53	3.50	7300920.64
16	106.63	3.50	7004093.08
30	111.80	3.50	7845551.91
22	107.05	3.50	7573319.90

Gambar Data acak test dataset

Di baris pertama ada data no. 31, dengan satuan *ihk* sebesar 111,57, memiliki nilai tukar uang 3,75 dan jumlah uang beredar sebanyak 7.897.628,21 rupiah. Model memprediksi bahwa data inflasi pada no.31 akan mengalami penurunan (*inflasi=0*). Dapat dibenarkan prediksi ini setelah dibandingkan dengan target aslinya di *DataFrame y_test* pada gambar Hasil tes prediksi

```
In [11]: y_test.head(1)
Out[11]: 31    0
          Name: inflasi, dtype: int64
```

Gambar Hasil tes prediksi

Hasil prediksi terbukti bahwa pada data urutan no. 31 nilai inflasi akan turun. Pemeriksaan secara acak seperti ini dapat memberikan kinerja model apakah bagus atau

tidak, namun perlu diukur dengan lebih seksama seberapa baik kinerja model yang sudah dibuat ini.

Mengukur Kinerja Model

Ada beberapa cara dalam mengukur kinerja suatu model yang menghasilkan *binary classification* seperti pada konsep *true positive (TP)*, *true negative (TN)*, *false positive (FP)*, dan *false negative (FN)*. Ditampilkan dalam tabel *Confusion Matrix* Konsep berikut.

Tabel *Confusion Matrix* Konsep. Sumber:

		Kenyataan	
		Positif	Negatif
Prediksi	Positif	TP	FP
	Negatif	FN	TN

Kurniawan, 2021

Tabel ini sering disebut *confusion matrix* karena istilah yang digunakan bisa membuat kebingungan bagi orang yang membacanya. Secara singkat bisa diingat dengan :

- 1) TP: model sukses memprediksi positif (ya), karena kenyataannya memang positif (ya)
- 2) TN: model sukses memprediksi negatif (tidak), karena kenyataannya memang negatif (tidak)
- 3) FP: model memprediksi positif (ya), namun salah karena kenyataannya negatif (tidak)
- 4) FN: model memprediksi negatif (ya), namun salah karena kenyataannya positif (ya)

Setelah memahami konsep pada tabel 1 statistik kinerja model sebelumnya yang telah dibuat bisa dilihat dengan *confusion matrix* nya. Langkah awal dengan impor terlebih dahulu paket *metrics* dari *Scikit-learn*, dan masukkan *test dataset* yang ditunjukkan dalam gambar 15 dibawah ini.

```
In [12]: import sklearn.metrics as met
confusionmatrix = met.confusion_matrix(y_test,y_prediksi)
In [13]: print(confusionmatrix)
[[0 2]
 [0 6]]
```

Gambar *Confusion Matrix*

Dari tabel *Confusion Matrix* Konsep dapat dilihat bahwa model regresi logistik dari dataset menghasilkan TP = 0, FP = 2, FN = 0, dan TN = 6. Angka TP=0 dan FP=2 artinya model memprediksi 0 kasus positif yang tepat, dan 2 kasus positif yang tidak tepat. Demikian dengan FN=0 dan TN=6 yang berarti model memprediksi 0 kasus negatif yang tepat dan 6 kasus negatif yang keliru. Hasil dari *confusion matrix* ini bisa digunakan untuk mengukur akurasi model. Dengan rumusnya adalah

jumlah prediksi yang benar dibagi dengan total seluruh populasi prediksi:

$$\begin{aligned} \text{accuracy} &= (TP + TN)/(TP + TN + \\ & FP + FN) \\ \text{accuracy} &= (0 + 6)/(0 + 6 + 2 + 0) \\ &= \frac{6}{8} \\ &= 0,75 \end{aligned}$$

Dengan rumus ini dapat dihitung secara manual akurasi model, yaitu 0+6 dibagi 0+6+2+2 hasilnya 0,75. Dapat juga dengan memanggil fungsi `score()` pada gambar di bawah ini untuk menghitungnya.

```
In [14]: score = model.score(X_test, y_test)
print(score)

0.75
```

Gambar Akurasi model

Hasil dari fungsi `score()` dan perhitungan manual sama yaitu 0,75 sehingga akurasi model bernilai 75% yang berarti sudah cukup baik. Secara umum di bidang *Data Science*, model dengan akurasi di atas 70% sudah digolongkan sebagai model yang berkinerja cukup baik.

Feature Selection

Feature selection adalah proses pemilihan *feature* atau atribut data yang memberikan kontribusi paling besar terhadap keluaran prediksi model yang diharapkan. Metode *Recursive Feature Elimination* (RFE) dapat dipakai untuk memilih *feature* yang terbaik dalam pembuatan model.

```
In [18]: import sklearn.feature_selection as fs
rfe = fs.RFE(model, 3)
rfe = rfe.fit(X_train, y_train)
print('Support=', rfe.support_)
print('Ranking=', rfe.ranking_)

Support= [ True True True]
Ranking= [1 1 1]
```

Gambar Feature selection

Dalam Scikit-learn sudah terdapat fungsi RFE sehingga tidak perlu melakukan proses manual dengan menghilangkan *feature* secara satu-per-satu. Untuk mengidentifikasi *feature* yang berpengaruh terhadap akurasi model, selanjutnya `import module feature_selection` seperti pada gambar 17 dan panggil fungsi RFE dengan menyertakan model yang sudah dilatih.

Dari gambar di atas didapat hasil bahwa semua *feature* berpengaruh dalam penelitian dan sudah cukup untuk membuat model regresi logistik menghasilkan akurasi yang tinggi, sehingga bila salah satu *feature* diabaikan tetap dapat memberikan hasil prediksi yang cukup baik.

Analisis Regresi Logistik

Analisis regresi memungkinkan kita untuk memprediksi nilai dari satu variabel berdasarkan nilai variabel lain. Dalam analisis regresi, variabel yang kita prediksi disebut variabel dependen, dan variabel yang digunakan untuk prediksi disebut variabel independen. Dalam gambar di bawah ini ada data sebanyak 36 yang diamati hal ini membuktikan semua data terbaca dalam program.

```
In [19]: print(log_reg.summary())
```

Logit Regression Results					
Dep. Variable:	inflasi	No. Observations:	36		
Model:	Logit	Df Residuals:	32		
Method:	MLE	Df Model:	3		
Date:	Wed, 09 Aug 2023	Pseudo R-squ.:	0.81297		
Time:	13:37:35	Log-Likelihood:	-18.822		
converged:	True	LL-Null:	-19.069		
		LLR p-value:	0.9200		
	coef	std err	z	P> z	[0.025 0.975]
intercept	-16.8044	33.650	-0.499	0.618	-82.757 49.148
X1	0.2523	0.447	0.564	0.572	-0.624 1.129
X2	-0.1020	0.892	-0.114	0.909	-1.851 1.647
X3	-1.197e-06	1.88e-06	-0.636	0.525	-4.88e-06 2.49e-06

Gambar Analisis Regresi Logistik

Regresi logistik adalah model klasifikasi yang digunakan untuk memprediksi respon kualitatif dengan dua kemungkinan hasil atau kategori. Model regresi logistik memprediksi peluang setiap amatan untuk masuk ke salah satu dari dua kemungkinan tersebut. Pada gambar 18 dapat kita cari persamaan regresi logistiknya dengan rumus analisis berikut:

$$\begin{aligned} P(X) &= (\exp(\beta_0 + \beta_1 X)) / (1 + \exp(\beta_0 + \beta_1 X)) \\ \text{logit}[P(x)] &= g(x) \\ &= \ln(P(x)/(1 - P(x))) \\ &= \beta_0 \\ &\quad + \beta_1 x_1 + \dots + \beta_n x_n \end{aligned}$$

Persamaan pada rumus (4.3) merupakan fungsi logistik (kurvanya berbentuk sigmoid) dimana β_0 dan β_1 adalah suatu koefisien yang diketahui. Fungsi $P(X)$ untuk sembarang X akan bernilai antara 0 sampai 1. Semakin kecil nominator-nya maka $P(X)$ akan menuju 0 dan sebaliknya maka $P(X)$ mendekati nilai 1. Bentuk ini cocok untuk merepresentasikan nilai peluang yang berada pada kisaran 0 sampai 1.

Selanjutnya, dari persamaan rumus (4.4) merupakan bentuk regresi logistik. Pada sisi kiri terdapat pembagian antara $P(x)$ dan $1 - P(x)$, dimana $P(x)$ sendiri bernilai antara 0 sampai 1. Karena total nilai peluang adalah 1, maka jika kita nyatakan $P(x)$ tersebut sebagai peluang “sukses” suatu kejadian, maka $1 - P(x)$ adalah peluang “gagal“-nya. Perbandingan antara peluang “sukses” suatu kejadian terhadap peluang “gagal“-nya disebut *odds*. Nilai *odds* menunjukkan seberapa besar peluang “sukses” dibandingkan peluang

“gagal”. Contohnya, jika nilai *odds* adalah 2, artinya peluang sukses dua kali lipat dari peluang gagal dan jika nilai *odds* adalah 0,5, peluang sukses hanya 0,5 kali peluang gagal (peluang gagalnya 2 kali lebih besar dari peluang sukses).

Pada hasil keluaran dalam gambar 18 didapati nilai $\beta_0 = -16,804$, $\beta_1 = 0,252$ dan $\beta_2 = -0,102$ sehingga model regresi logistik dari hasil keluaran pada gambar adalah sebagai berikut

$$P(X) = \frac{\exp(-16,804 + 0,252 X_1 - 0,102 X_2 + 0,0 X_3)}{1 + \exp(-16,804 + 0,252 X_1 - 0,102 X_2 + 0,0 X_3)}$$

atau

$$\ln\left(\frac{P(x)}{1 - P(x)}\right) = -16,804 + 0,252 X_1 - 0,102 X_2$$

Selanjutnya adalah melakukan prediksi berdasarkan hasil model regresi logistik. Misalnya suatu waktu pada bulan tertentu mempunyai indeks harga konsumen 100 ribu, nilai tukar uang 3, dan uang beredar 5 juta. Probabilitas naiknya nilai inflasi dapat dihitung sebagai berikut. Dengan mencari regresi logistiknya terlebih dahulu lalu,

$$\begin{aligned} \ln\left(\frac{P(x)}{1 - P(x)}\right) &= -16,804 + 0,252 X_1 - 0,102 X_2 \\ &= -16,804 + 0,252(100) - 0,102(3) + 0,0(5) \\ &= 8,090 \end{aligned}$$

Dilanjutkan mencari probabilitas dengan memasukkan hasil regresi logistik sebelumnya ke dalam rumus probabilitas regresi logistik.

$$\begin{aligned} p &= \frac{e^{(\beta_0 + \beta_1 X)}}{1 + e^{(\beta_0 + \beta_1 X)}} \\ &= \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}} \\ &= \frac{1}{1 + e^{-8,090}} \\ &= 0,999 \end{aligned}$$

Dengan demikian, nilai prediksi probabilitas suatu bulan tertentu tersebut memiliki nilai inflasi naik sebesar 0,999 sedangkan probabilitas nilai inflasi turun sebesar $1 - 0,999 = 0,001$.

4. SIMPULAN

Simpulan

Berdasarkan hasil penelitian analisis data inflasi yang ada Indonesia memanfaatkan *machine learning* dengan metode regresi logistik sebagai salah satu tahap analisisnya dapat ditarik kesimpulan sebagai berikut:

1. Dalam proses analisis data inflasi yang dilakukan dengan menggunakan bahasa pemrograman python didapatkan hasil yang jelas bahwa data inflasi dapat dilihat dari indeks harga konsumen, nilai tukar uang, dan jumlah uang yang beredar setelah diperhitungkan memanfaatkan algoritma regresi logistik sebagai acuan perhitungan analisis.
2. Data sampel sebanyak 36 sampel adalah data gabungan per bulan selama 3 tahun yang diteliti dan menghasilkan persamaan regresi logistik yaitu $\ln\left(\frac{P(x)}{1 - P(x)}\right) = -16,804 + 0,252 X_1 - 0,102 X_2$ yang artinya data sesuai dan analisis yang dilakukan berhasil dengan tahapan beberapa uji data yang dilakukan.
3. Hasil analisis data inflasi dengan metode regresi logistik ini membuktikan bahwa dalam prosesnya inflasi dipengaruhi beberapa faktor dalam memutuskan apakah inflasi pada suatu negara itu nilainya naik atau turun dan dalam prosesnya faktor tersebut adalah indeks harga konsumen, nilai tukar uang, dan jumlah uang yang beredar. Pada intinya inflasi dipengaruhi dengan uang dalam konteks bidang yang berbeda. Dengan nilai akurasi yang cukup tinggi sebesar 75%.
4. Kecanggihan teknologi masa kini membuat banyak kemudahan seperti pemanfaatan *machine learning* dengan menggunakan beberapa simbol pemrograman, sebuah analisis bisa dilakukan dengan bahasa komputer itu. Di dalam penelitian ini python digunakan sebagai salah satu acuan untuk memperkirakan naik turunnya inflasi dan pada akhirnya membuktikan inflasi selama tahun 2020 sampai dengan 2022 mengalami kenaikan nilai inflasi dengan nilai probabilitas sebesar 0,999 yang datanya diambil dari sampel per bulan selama 3 tahun tersebut.

Saran

Pada proses penelitian didapati sumber hanya terpaku garis besar pada data resmi yang dikeluarkan pemerintah Indonesia saja untuk memperkembangkan penelitian lebih baik lagi alangkah baiknya untuk meneliti lebih spesifik dalam menentukan atau mengambil suatu kasus kejadian.

Dalam proses penelitian penulis sadar masih banyak kekurangan dikarenakan terbatasnya ilmu yang dimiliki penulis sehingga dalam penyusunan masih menggunakan kata dan kalimat tidak baku dan tidak sesuai. Keterbatasan penulis menyebabkan dalam analisis yang digunakan adalah metode regresi logistik saja, diharapkan untuk penelitian selanjutnya bisa mempertimbangkan penggunaan metode analisis dan algoritma lain dalam proses penelitian.

5. DAFTAR PUSTAKA

- [1] Ahmad. (2021). *Pengertian, Penyebab, Dampak dan Peran Bank Sentral dalam Menghadapi Inflasi*. Gramedia Literasi Dikutip Februari 8, 2023, dari <https://www.gramedia.com/literasi/inflasi>
- [2] Ananta, A. D., & Widodo, P. (2021). Analisis Determinasi Inflasi Di Indonesia Tahun 2015-2019. *Prosiding Seminar Nasional & Call for Paper STIE AAS*, 4(1), 162-167. <https://prosiding.stie-aas.ac.id/index.php/prosenas/article/view/108>
- [3] Anastavia, A. (2018). *Pengaruh Inflasi Terhadap Kestabilan Dan Eksistensi Usaha Ekonomi Mikro (Studi Kasus Usaha Keripik Pisang Tunas Kota Metro)*. Skripsi Sarjana, Institut Agama Islam Negeri Metro. Dikutip Februari 17, 2023, dari <https://repository.metrouniv.ac.id/id/eprint/1420>
- [4] Astuti, A. B., Efendi, A., Astutik, S., & Sumarningsih, E. (2020). *Analisis Data Kategorik Menggunakan R Teori dan Aplikasinya pada Berbagai Bidang* (1st ed.). Malang: UB Press.
- [5] Ayunda, P. L. (2018). *Analisis Perbandingan Regresi Logistik Model Logit Dan Probit Untuk Menentukan Variabel Yang Mempengaruhi Fluktuasi Harga Beras*. Skripsi Sarjana, Institut Teknologi Sepuluh Nopember. Dikutip Desember 6, 2022, dari <https://repository.its.ac.id/49444>
- [6] Badan Pusat Statistik. (2022). *Badan Pusat Statistik*. Badan Pusat Statistik. Dikutip Oktober 8, 2022, dari <https://www.bps.go.id>
- [7] Budiharto, W. (2018). *Pemrograman Python Untuk Ilmu Komputer Dan Teknik* (1st ed.). Yogyakarta: ANDI.
- [8] Chandrasekaran, M. (2021). *Logistic Regression for Machine Learning*. Capital One. Dikutip November 4, 2022, dari <https://www.capitalone.com/tech/machine-learning/what-is-logistic-regression>
- [9] Fernando, J. (2022). *Inflation: What It Is, How It Can Be Controlled, and Extreme Examples*. Investopedia. Dikutip Desember 8, 2022, dari <https://www.investopedia.com/terms/i/inflation.asp>
- [10] Hadjar, I. (2019). *Statistik: Untuk Ilmu Pendidikan, Sosial, dan Humaniora*. Bandung: Remaja Rosdakarya.
- [11] Heryadi, Y., & Wahyono, T. (2020). *Machine Learning (Konsep dan Implementasi)* (1st ed.). Yogyakarta: Gava Media.
- [12] Jupyter. (2023). Project Jupyter. Dikutip September 5, 2023, dari <https://jupyter.org>
- [13] Kurniawan, D. (2021). *Pengenalan Machine Learning dengan Python* (2nd ed.). Jakarta: Elex Media Komputindo.
- [14] Malona, O. L. (2020). Penerapan Metode Fuzzy Inference System (FIS) Sugeno Dalam Menentukan Nilai Inflasi (Studi Kasus Pada Data Inflasi Medan). *KARISMATIKA*, 6(3), 33-38. <https://doi.org/10.24114/jmk.v6i3.22215>
- [15] Mulachela, H. (2022). *Jenis-jenis Inflasi Beserta Dampaknya*. Katadata. Dikutip Februari 8, 2023, dari <https://katadata.co.id/safrezi/berita/62134806c5205/jenis-jenis-inflasi-beserta-dampaknya>
- [16] Nafalana, R. D. (2022). *Identifikasi Kondisi Pasar Saham Dengan Algoritma Bry-Boschan Dan Prediksinya Menggunakan Regresi Logistik Biner (Studi Kasus: Indeks Harga Saham Gabungan (Ihsg) Januari 2003 – Mei 2022)*. Skripsi Sarjana, Universitas Islam Indonesia. Dikutip Februari 22, 2023, dari <https://dspace.uii.ac.id/handle/123456789/40256>
- [17] Nursyafina. (2020). *Pengaruh Inflasi Dan Pertumbuhan Ekonomi Terhadap Tingkat Pengangguran Di Indonesia*. Skripsi Sarjana, Universitas Islam Riau. Dikutip Februari 17, 2023, dari <https://repository.uir.ac.id/10443>
- [18] Pamungkas, F. S., Prasetya, B. D., & Kharisudin, I. (2020). Perbandingan Metode Klasifikasi Supervised Learning pada Data Bank Customers Menggunakan

- Python. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 689-694. <https://journal.unnes.ac.id/sju/index.php/prisma/article/view/37875>
- [19] Pangaribuan, J. J., Tanjaya, H., & Kenichi, K. (2021). Mendeteksi Penyakit Jantung Menggunakan Machine Learning Dengan Algoritma Logistic Regression. *Journal Information System Development (ISD)*, 6(2), 1-10. <https://ejournal-medan.uph.edu/index.php/isd/article/view/468>
- [20] Pratama, S. R., & Mirza, A. H. (2021). Penerapan Data Mining Untuk Memprediksi Tingkat Inflasi Menggunakan Metode Regresi Linier Berganda Pada BPS Kota Palembang. *Bina Darma Conference on Computer Science (BDCCS)*, 3(2), 245-255. <https://conference.binadarma.ac.id/index.php/BDCCS/article/view/2148>
- [21] Seldon. (2021). *Machine Learning Regression Explained*. Seldon. Dikutip Desember 6, 2022, dari <https://www.seldon.io/machine-learning-regression-explained>
- [22] Stojiljković, M. (2022). *Logistic Regression in Python*. Real Python. Dikutip Maret 9, 2023, dari <https://realpython.com/logistic-regression-python>
- [23] Suprpto, & Ni'mah, Y. L. (2019). *Pengantar Analisis Data Menggunakan Python* (1st ed.). Surabaya: DEEPUBLISH.
- [24] Syahza, A. (2021). *Metodologi Penelitian* (Revisi ed.). Pekanbaru: UR Press. https://www.researchgate.net/publication/354697863_Buku_Metodologi_Penelitian_Edisi_Revisi_Tahun_2021
- [25] Tyasnurita, R., & Pamungkas, A. Y. M. (2020). Deteksi Diabetik Retinopati menggunakan Regresi Logistik. *ILKOM Jurnal Ilmiah*, 12(2), 130-135. <https://doi.org/10.33096/ilkom.v12i2.578>
- [26] Widarjono, A. (2018). *Analisis Regresi Dengan SPSS* (1st ed.). Yogyakarta: UPP STIM YKPN.
- [27] Wulandari, A., Faruk, F. M., Doven, F. S., & Budyanra. (2019). Penerapan Metode Regresi Logistik Biner Untuk Mengetahui Determinan Kesiapsiagaan Rumah Tangga Dalam Menghadapi Bencana Alam. *Seminar Nasional Official Statistics 2019: Pengembangan Official Statistics dalam Mendukung Implementasi SDG's*, 379-389. <https://prosiding.stis.ac.id>
- [28] Yanuar, A. (2018). *Statistical Machine Learning*. Universitas Gadjah Mada Menara Ilmu Machine Learning. Dikutip November 9, 2022, dari <https://machinelearning.mipa.ugm.ac.id/2018/05/20/statistical-machine-learning>
- [29] Zarman, W., & Wicaksono, M. F. (2020). *Implementasi Algoritma Dalam Bahasa Python*. Bandung: Informatika.