

BAB II

TINJAUAN PUSTAKA

2.1 Kajian Teori

2.1.1 Regresi Logistik

Menurut widarjono, regresi logistik merupakan regresi yang digunakan untuk mempelajari hubungan antara variabel bebas dan variabel terikat yang bertujuan untuk prediksi nilai peluang terjadi pada sebuah peristiwa(roflin dkk., 2023).

Variabel regresi logistik dapat berupa kategorik ataupun numerik. Jika salah satu atau lebih adalah skala nominal bernilai lebih dari dua pada variabel independen maka harus membuat *dummy variable* yang merepresentasikan kategori dari variabel tersebut menggunakan *reference group*-nya dari salah satu kategori tersebut.

Rumus matematis umum regresi tunggal menurut (roflin dkk., 2023).dalam buku dengan judul regresi logistik biner dan multinomial sebagai berikut:

$$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \dots\dots\dots (2.1)$$

Keterangan:

Y = variabel bebas

X = variabel terikat

β_0 dan β_1 = parameter regresi atau koefisien dugaan nilai regresi berdasarkan data sampel

ε_i = galat atau error yang bersifat acak, yakni nilai besaran penyimpangan dari model regresi

Terdapat perbedaan antara regresi logistik dan regresi linier hal itu terletak pada variabel dependen, jika regresi linier menggunakan variabel terikat numerik, regresi logistik menggunakan variabel terikat kategorik. Berdasarkan jenis data atau variabel regresi logistik dibedakan menjadi tiga, yaitu:

a) Regresi Logistik Biner

Regresi logistik biner merupakan bentuk dari regresi non-linier yang memiliki variabel terikat berupa biner, yakni nol (0) dan satu (1). Regresi logistik ini menggunakan fungsi logit untuk menghubungkan model antara variabel terikat biner dan variabel bebas. Fungsi logit sendiri diartikan sebagai log natural dari variabel dependen, sehingga hanya memiliki dua kategori yakni sebagai nol dan satu.

b) Regresi Logistik Multinomial

Menurut agresti, regresi logistik multinomial digunakan apabila variabel terikat merupakan kategorik yang memiliki lebih dari dua klasifikasi(roflin dkk., 2023).

c) Regresi Logistik Ordinal

Regresi ini merupakan pemodelan dari regresi logistik untuk data kategorik ordinal dengan jumlah kategori nya lebih dari dua.

Tujuan regresi logistik adalah menghasilkan model terbaik dan sederhana, sehingga dapat memberikan gambaran hubungan antara variabel terikat dan bebas.

1) Regresi Logistik Biner

Menurut hosmer dan lemeshow, regresi logistik biner berguna untuk menganalisa hubungan antara variabel terikat dan variabel bebas yang bertujuan untuk memprediksi peluang kejadian suatu peristiwa (roflin dkk., 2023). Rumus matematis dari fungsi logit sebagai berikut:

$$g(X_{ki}) = \ln \left[\frac{P(Y|X_{ki})}{P(Y|X_{ki})} \right] = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} \dots\dots\dots (2.2)$$

Dimana rumus matematis regresi logistik biner umum sebagai berikut,

$$P(Y|X_{ki}) = \frac{e^{(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})}}{1 + e^{(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})}} \dots \dots \dots (2.3)$$

Keterangan:

P = peluang y (variabel dependen)

X_{ki} = nilai pengamatan ke- i pada perubahan ke- k

β_k = nilai koefisien peubah prediktor ke- k

K = indek perubahan prediktor

I = indeks untuk pengamatan

E = konstanta 2.71828

Menurut Ibnu, regresi logistik merupakan metode yang bersumber dari ilmu statistik yang diadopsi oleh *machine learning* yang digunakan untuk memprediksi probabilitas dengan memanfaatkan prinsip-prinsip statistika. Dalam implementasinya menggunakan *library scikit-learn*. (id, 2021) pada *machine learning* digunakan untuk membangun model untuk mempelajari pola-pola data latih yang kemudian diprediksi pada kelas label. Contoh *code* yang digunakan dalam machine learning dengan menggunakan bahasa pemrograman python sebagai berikut:

```
from sklearn.linear_model import LogisticRegression

model = LogisticRegression()

model.fit(X_train, y_train)
```

2) Estimasi Parameter

Estimasi parameter menggunakan metode *Maximum Likelihood Estimasi* (MLE) dengan parameter β yang memiliki dengan syarat harus mengikuti distribusi tertentu, pada regresi logistik biner menggunakan distribusi bernoulli. Apabila X_i dan Y_i adalah variabel independen dan dependen pada pengamatan ke- i

Fungsi *likelihood* sebagai berikut:

$$\pi(X_i) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_{1i}+\beta_2 X_{2i}+\dots+\beta_k X_{ki})}} \dots\dots\dots (2.4)$$

Keterangan:

$\pi(X_i)$ = fungsi *likelihood*

X_{ki} = nilai pengamatan ke-i pada perubahan ke-k

β_k = nilai koefisien peubah prediktor ke-k

K = indek perubahan prediktor

I = indeks untuk pengamatan

E= konstanta 2.71828

3) Signifikansi Parameter

Signifikansi parameter merupakan pengujian untuk mengetahui pengaruh variabel bebas pada model terhadap variabel terikat, maka dari itu menggunakan pengujian hipotesis statistik. Jika secara individu dengan uji *wald*, uji *wald* pengujian prediktor satu persatu terhadap hipotesis $H_0: \beta_j = 0$, rumus matematis uji *wald* sebagai berikut:

$$W = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim N(0,1) \dots\dots\dots (2.5)$$

Keterangan:

$\hat{\beta}_j$ = penduga β_j

$se(\hat{\beta}_j)$ = standar error penduga β_j

Hipotesis yang digunakan:

H_0 : diterima jika hasil dari p-value > 0.05

H_0 : ditolak jika hasil dari p-value <0.05

2.1.2 Desa Gogodeso

Berdasar pada dokumen dari pemerintah Desa Gogodeso, Desa Gogodeso adalah desa yang berada di kabupaten blitar tepatnya di kecamatan kanigoro. Luas dari wilayah Desa Gogodeso sekitar 300.390 ha/m² yang terdiri atas pemukiman warga, persawahan dan badan usaha milik desa. Tercatat pada tahun 2015 penduduk Desa Gogodeso terdiri dari 1683 KK, dengan 5500 jiwa orang yang tebagi dalam

jenis kelamin pria dan wanita dan juga usia. Desa ini terdiri atas 4 dusun seperti Dusun Dogong, Dusun Gogodeso, Dusun Ngade, Dan Dusun Serut.

2.1.3 Pola Distribusi Bantuan Beras Bulog Di Desa Gogodeso

Hasil wawancara yang sudah dilakukan peneliti dengan kepala seksi pelayanan dan kasun ngade bahwa pendistribusian bantuan beras Bulog di Desa Gogodeso dimulai di bulan agustus 2023 hal ini dikarenakan bantuan beras belum disalurkan dari bulan januari hingga juli, sehingga membuat pendistribusian bantuan beras Bulog untuk bulan januari hingga juli berjalan di bulan selanjutnya sehingga pendistribusian bantuan beras dalam satu bulan terjadi beberapa kali untuk mendistribusikan bantuan beras di januari hingga juli. Pada setiap pendistribusian beras penerima memperoleh 10 Kg beras yang sudah dimasukkan kedalam karung. Jika ditotal dalam satu tahun penerima beras mendapatkan 120 Kg beras.

2.1.4 Faktor-Faktor Yang Mempengaruhi Kelayakan Penerima Bantuan

Kelayakan seseorang menerima bantuan dari pemerintah dikutip dari file sidaksos terdapat beberapa kriteria keluarga mendapatkan bantuan, yakni:

- 1) Kondisi rumah luas keseluruhan kurang dari 30m², bangunan milik sendiri atau kontrak atau bebas sewa.
- 2) Kepemilikan aset seperti:
 - a) Sepeda motor roda dua atau tiga dengan tahun pembuatan dimulai dari 2015
 - b) Ternak unggas kurang dari sama dengan 100 ekor
 - c) Ternak sapi kurang dari sama dengan 2 ekor
 - d) Ternak kambing kurang dari sama dengan 4 ekor
 - e) Televisi kurang dari sama dengan 21 inch

- f) Luas lahan (sawah, kebun, tegal, dan lain-lain) kurang dari 140 M²
- 3) Sumber penerangan listrik pln 450 Watt atau tanpa meteran, termasuk 900 Watt subsidi atau bukan listrik yakni penerangan menggunakan petromak, pelita, sentir, dan lain-lain
- 4) Sumber air minum dengan volume penggunaan paling banyak digunakan menggunakan sumur
- 5) Penggunaan fasilitas bab (mandi, cuci, kakus) baik sendiri, bersama maupun umum individu per-bulan kurang dari
- 6) Pengeluaran individu per-bulan kurang dari Rp 500.000.-
- 7) Foto rumah
- 8) Lokasi verval

2.1.5 *Confusion Matrix*

Confusion matrix adalah alat yang digunakan untuk mengukur performa klasifikasi *machine learning*. Matriks terdiri atas empat kolom, yakni: *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), Dan *False Negative* (FN) (anggreany, 2020).

Keterangan elemen dalam *confusion matrix* sebagai berikut:

- a. *True Positive* (TP) merupakan jumlah sampel positif yang diprediksi benar
- b. *False Positive* (FP) merupakan jumlah sampel negatif yang diprediksi sebagai data positif
- c. *True Negative* (TN) merupakan jumlah sampel negatif yang diprediksi benar
- d. *False Negative* (FN) merupakan jumlah sampel positif yang diprediksi sebagai data negatif.

	Positif (1)	Negatif (0)
Positif (1)	TP <i>True Positive</i>	FP <i>False Positive</i>
Negatif (0)	FN <i>False Negative</i>	TN <i>True Negative</i>

Gambar 2.1 Confusion Matrix

2.1.6 One-Hot Encoding

Dalam pembuatan *machine learning* mengharuskan seluruh variabel masukan dan keluaran berbentuk numerik, jika terdapat data yang berbentuk kategorikal maka harus diubah menjadi biner haruslah melewati pengkodean *one-hot encoding*(galli, 2020).

Contoh penggunaan *one-hot encoding*:

```
import numpy as np
import pandas as pd
from sklearn.preprocessing import LabelEncoder

data ={"id":[1,2,3],
       "gender": ["M", "F", "F"],
       "age": [17,19,29]}

le=LabelEncoder()
data["gender"]=le.fit_transform(data["gender"])

df = pd.DataFrame(data)

print(df)
```

Gambar 2.2 Contoh Endcoding

Yang akan menghasilkan *Output* pada **gambar 2.3**.

	id	gender	age
0	1	1	17
1	2	0	19
2	3	0	29

Gambar 2.3 Hasil Encoding

2.2 Kajian Penelitian

Beberapa penelitian yang telah dilakukan sebelumnya terkait penggunaan regresi logistik biner. Terdapat 10 penelitian terkait dengan regresi logistik biner adapun penelitian tersebut terletak pada **tabel 2.1**.

Tabel 2.1. Kajian Penelitian

No	Nama Peneliti (Tahun)	Judul	Hasil
1	Safitri Dkk., (2019)	Model regresi logistik biner pada tingkat pengangguran terbuka di provinsi sulawesi barat tahun 2017	Persentase lapangan dan persentase laki-laki usaha dengan kategori pertanian pada perubah jenis kelamin p-value bernilai $0.017 < 5\%$ dan peubah lapangan usaha nilainya $0.006 < 5\%$ yang diartikan peubah tersebut berpengaruh pada tingkat pengangguran terbuka. Kelemahan pada penelitian ini terletak pada hasil disini hanya disajikan untuk jenis kelamin dengan kategori laki-laki saja sedangkan pada data terdapat jenis kelamin perempuan yang tidak disajikan oleh penulis
2	Hendayanti & Nurhidayati, (2020)	Regresi logistik biner dalam penentuan ketetapan klasifikasi tingkat kedalaman kemiskinan provinsi-provinsi di indonesia	Mys dan eys berpengaruh terhadap tingkat kedalaman kemiskinan dengan akurasi prediksi sebesar 85.3%. Kelemahan dari penelitian ini belum memperhatikan tingkat ekonomi yang ada di setiap provinsinya
3	Suhendra Dkk., (2020)	Ketepatan klasifikasi pemberian kartu keluarga sejahtera di kota semarang menggunakan metode regresi logistik biner dan metode chaid	Menggunakan regresi logistik biner menghasilkan ketepatan sebesar 88% dan kesalahan 12% sedangkan menggunakan chaid menghasilkan ketepatan klasifikasi sebesar 90.2% dan kesalahan 9.8%. Kelemahan terdapat beberapa variabel yang tidak dibutuhkan atau dirasa tidak terlalu penting seperti kepemilikan/penguasaan hp.
4	Amelia Dkk., (2020)	Analisis faktor yang mempengaruhi status penerimaan beras keluarga miskin	Hasil dari penelitian ini memiliki ketetapan klasifikasi sebesar 72.2% dan kesalahan sebesar 27.8% sehingga h_0 ditolak. Kelemahan pada penelitian ini terletak pada data sampel yang hanya 210 diujikan.

No	Nama Peneliti (Tahun)		Judul	Hasil
			menggunakan regresi logistik biner di kecamatan langsa barat	
5	Sari (2020)	Dkk.,	Analisis <i>financial distress</i> pada bank umum syariah melalui metode regresi logistik biner data panel	Hasil dari penelitian ini teori sinyal yang menunjukkan hubungan roa, roe, dan npf memiliki pengaruh yang cukup besar terhadap <i>financial distress</i> . Sedangkan dr, car, dkomin, dan ddrek kurang memiliki pengaruh terhadap <i>financial distress</i> . Kelemahan pada penelitian ini terdapat pada data yang digunakan dirasa kurang baik karena menggunakan data dengan kurun waktu yang cukup lama. Dalam waktu yang lama tersebut dirasa sudah mengalami perubahan yang cukup signifikan sehingga dirasa data yang digunakan tidak dapat mencerminkan kondisi yang terjadi saat ini
6	Marna (2021)	Dkk.,	Analisis regresi logistik biner untuk memprediksi faktor internal dan eksternal terhadap indeks prestasi	Menunjukkan bahwa terdapat pengaruh terhadap prestasi yang berkaitan tentang faktor internal dan eksternal sebesar 60%. Kelemahan pada saat pengambilan data menggunakan kuesioner karena terdapat beberapa pertanyaan yang cukup memusingkan pengisi kuisisioner karena tidak terukur baik. Lalu tidak dilakukan pengujian parameter dan
7	Salomo Parsaulian & Ispriyanti, (2021)		Analisa faktor-faktor yang mempengaruhi penerima beras raskin menggunakan regresi logistik biner dengan <i>gui r</i>	Hasil dari penelitian ini nilai akurasi bernilai 85,4% dengan nilai <i>aper</i> atau <i>error</i> 14,6%. Kelemahan pada penelitian ini adalah model mungkin bisa mengalami overfitting karena data dengan model dirasa sangat sesuai dengan data pelatihan dan dirasa tidak umum
8	Susetyoko (2022)	Dkk.,	Model klasifikasi pada seleksi mahasiswa baru penerima kip kuliah menggunakan regresi logistik biner	Menghasilkan rata-rata <i>f1-score</i> 92,40%, rata-rata <i>recall</i> 87,93%, <i>accuracy</i> 88,01%, <i>precision</i> 97,92%, dan <i>auc</i> 84,6%. Kelemahan pada penelitian ini adalah tidak dilakukannya <i>cross validation</i> dan juga tidak dilakukannya <i>best parameter</i> sehingga masih sangat dimungkinkan mengalami overfitting
9	Harris, (2021)		Primer on binary logistic regression	Hasil dari penelitian ini menunjukkan bahwa <i>or</i> dan <i>ci</i> pada setiap prediktor telah terpenuhi dan sesuai dengan model secara keseluruhan. Kelemahan pada penelitian ini belum dilakukan <i>cleansing data</i> untuk membersihkan dan menambahkan data

No	Nama Peneliti (Tahun)	Judul	Hasil
10	Hills & Eraso, (2021)	Factors associated with non-adherence to social distancing rules during the covid-19 pandemic: a logistic regression analysis	yang diperlukan atau tidak diperlukan sehingga mengurangi tingkat keaurasian Tidak patuhan terhadap peraturan memiliki hubungan yang lebih kuat dengan kerentanan dan pengendalian covid-19. 92.8% orang tidak mematuhi aturan pembatasan sosial dan 48.6% sengaja melanggar aturan tersebut. Kelemahan dari penelitian ini belum dilakukan pengujian signifikansi parameter sehingga dapat mengurangi tingkat akurasi.

Penelitian ini dilakukan dengan berpendoman pada penelitian sebelumnya yang telah dilakukan sebagai pembandingan dan kajian penelitian. Dengan penelitian terdahulu yang paling mendekati dengan penelitian yang peneliti lakukan yakni penelitian yang dilakukan amelia dan kawan-kawan dengan judul analisis faktor yang mempengaruhi status penerimaan beras keluarga miskin menggunakan regresi logistik biner di Kecamatan Langsa Barat dengan hasil prediksi sebesar 72,2% (amelia dkk., 2020). Sehingga hasil dari kajian penelitian yang dilakukan dapat digunakan sebagai perbandingan dari topik penelitian yakni penerapan regresi logistik biner.

Dengan mempelajari penelitian sebelumnya, peneliti dapat mengembangkan regresi logistik biner yang dapat mengatasi beberapa kondisi sehingga menghasilkan kontribusi baru dalam pemodelan evaluasi penerima bantuan beras Bulog. Penelitian ini menggabungkan wawasan dari penelitian sebelumnya dengan pendekatan baru, sehingga dapat memberikan kontribusi dalam penelitian tentang penerapan penelitian regresi logistik biner.

Perbedaan antara penelitian yang dilakukan penulis dengan penelitian terdahulu. Seperti, objek penelitian yang digunakan, variabel yang digunakan, penggunaan evaluasi model untuk menghasilkan akurasi model dengan data yang digunakan, dan pada penelitian ini dilakukan dengan menerapkan seluruh data untuk diolah menggunakan regresi logistik biner hingga menghasilkan probabilitas sebuah keputusan data tersebut dinyatakan sebagai penerima bantuan atau bukan penerima bantuan. Kelebihan dari penelitian yang dilakukan penulis dibandingkan dengan penelitian terdahulu yakni model dapat memaparkan data lebih baik dan dapat dipahami oleh pembaca, sehingga informasi dapat membantu pemerintah dalam meningkatkan kebijakan pendistribusian beras Bulog agar lebih efisien dan adil untuk seluruh warga yang benar benar memerlukan.