Statistika, Vol. 24 No. 1, 20 – 30
Mei 2024

Studi Komparasi Regresi Logistik Biner dan *K-Nearest Neighbor* Pada Kasus Prediksi Curah Hujan

FITRI RAHMAWATI[1], FITRI AMANAH[2], SEFRI IMANUEL FALLO[3]

[1)]Program Studi Matematika Fakultas PMIPA Universitas Pendidikan Indonesia, Indonesia
[2)]Program Studi Statistika Fakultas MIPA Universitas Islam Bandung, Indonesia
[3)]Program Studi Matematika Fakultas MIPA Universitas San Pedro, Indonesia

e-mail: fitrirahmawati@upi.edu

ABSTRAK

Perubahan iklim yang sedang terjadi di berbagai belahan dunia sebagai akibat dari pemanasan global telah menyebabkan ketidakpastian cuaca. Salah satu perubahan yang dirasakan adalah intensitas curah hujan. Hal ini mengakibatkan prediksi akan curah hujan menjadi penting untuk dilakukan. Ada beberapa teknik analisis data yang digunakan untuk prediksi curah hujan, diantaranya klasifikasi. Pada penelitian ini, dengan menggunakan variabel temperatur, kelembapan, lamanya penyinaran, dan kecepatan angin, akan dilakukan prediksi terhadap klasifikasi curah hujan di Kota Bogor. Model yang digunakan adalah Regresi Logistik Biner dan *K-Nearest Neighbor*. K yang digunakan pada model *K-Nearest Neighbor* yaitu 1 hingga 18. Untuk membandingkan kedua model, dibentuk *confusion matrix* yang selanjutnya digunakan untuk menghitung akurasi model. Akurasi model Regresi Logistik Biner sebesar 92,746%, adapun akurasi model *K-Nearest Neighbor* adalah sebesar 94,81865%. Dengan demikian, pada penelitian ini model *K-Nearest Neighbor* lebih baik dibandingkan model Regresi Logistik Biner.

Kata Kunci: Regresi Logistik Biner, KNN, Curah Hujan.

ABSTRACT

Climate change due to global warming occurring in all parts of the world makes the weather unpredictable. One of the changes felt is the intensity of rainfall. This makes it important to predict rainfall. There are several data analysis techniques used to predict rainfall, including classification. In this research, using the variables temperature, humidity, duration of sunlight, and wind speed, predictions will be made on the classification of rainfall in the city of Bogor. The models used are Binary Logistic Regression and K-Nearest Neighbor. The K used in the K-Nearest Neighbor model is 1 to 18. To compare the two models, a confusion matrix is formed and then used to calculate the model accuracy. The accuracy of the Binary Logistic Regression model is 92.746%, while the accuracy of the K-Nearest Neighbor model is 94,81865%. Thus, in this research the K-Nearest Neighbor model is better than the Binary Logistic Regression model.

Keywords: Binary Logistic Regression, KNN, Rainfall.

1. PENDAHULUAN

Isu mengenai pemanasan global sedang menjadi perbincangan hangat belakangan ini. Dampaknya sudah sangat terasa salah satunya dalam hal perubahan iklim. Dalam penelitian yang dilakukan oleh (Susilokarti *et al.*, 2015), variabel-variabel utama untuk menilai perubahan iklim meliputi suhu, pola musim (kemarau dan hujan), kelembapan, dan angin. Lebih lanjut berdasarkan (BMKG, 2011), fokus utama dalam mengukur perubahan iklim sering kali tertuju pada suhu dan curah hujan. Perubahan curah hujan yang tidak menentu akibat pemanasan global ini membuat prediksi akan curah hujan menjadi penting untuk dilakukan. Salah satu metode analisis yang dapat digunakan untuk meramalkan curah hujan adalah dengan menggunakan teknik klasifikasi.

Klasifikasi merupakan salah satu topik analisis data dalsam *machine learning*. Menurut (James *et al.*, 2017) klasifikasi adalah proses untuk memprediksi variabel respon yang berupa data kualitatif. Dengan memanfaatkan nilai-nilai yang diketahui pada data, dapat diprediksi kelas atau kategori dari data tersebut. Beberapa model dari klasifikasi diantaranya model Regresi Logistik, *K-Nearest Neighbor* (KNN), *Neural Network*, *Random Forest* dan sebagainya.

Menurut (Jayanti and Noeryanti, 2014), algoritma K-*Nearest Neighbor* (KNN) adalah metode klasifikasi yang mengidentifikasi kelompok K objek dari data *training* yang paling dekat dengan objek yang ada dalam data *testing* atau data baru. Model regresi logistik biner, sesuai dengan penelitian yang dilakukan oleh (Sepang, Komalig and Hatidja, 2012), dipergunakan untuk mengevaluasi hubungan di antara satu variabel respons dan beberapa variabel prediktor. Model ini dapat digunakan untuk situasi di mana variabel responsnya terdiri dari data kualitatif dikotomi, di mana nilai 1 mencerminkan keberadaan suatu karakteristik dan nilai 0 menggambarkan ketiadaannya.

Penelitian ini akan membandingkan kinerja dua teknik klasifikasi, yaitu Regresi Logistik Biner dan K-*Nearest Neighbor* (K-NN) dalam memprediksi curah hujan di Kota Bogor. Data yang digunakan adalah data pengamatan curah hujan di Kota Bogor yang diambil dari tahun 2018 hingga 2020.

2. METODE PENELITIAN

2.1 Gambar dan Tabel

Data yang dipergunakan dalam studi ini mencakup data curah hujan yang tercatat di Kota Bogor, yang diambil dari rentang tahun 2018 hingga 2020 yang sebelumnya telah digunakan dalam (Fallo, 2021). Variabel yang digunakan sebanyak 5 variabel, yang terdiri dari 4 variabel prediktor dan 1 variabel respon yang dijelaskan dalam tabel berikut.

Tabel 1. Variabel Penelitian

Nama Variabel	Simbol	Keterangan
Temperatur	X_1	Temperatur rata-rata (°C)
Kelembapan	X_2	Kelembapan rata-rata (%)
Penyinaran	X_3	Lamanya penyinaran matahari (jam)
Angin	X_4	Kecepatan angin rata-rata (m/s)
Cuaca	Y	Terdiri dari dua kelas: 0 berarti tidak hujan dan 1 berarti hujan

2.2 Regresi Logistik

Analisis regresi menurut (Rahmawati and Suratman, 2022) adalah suatu metode pengolahan data yang digunakan untuk memodelkan pengaruh variabel prediktor terhadap variabel respon. Regresi logistik merupakan jenis analisis regresi yang digunakan untuk mempelajari hubungan antara variabel prediktor dan variabel respons, khususnya ketika variabel responnya memiliki skala data nominal atau ordinal dan bersifat kategorik. (Tampil, Komaliq and Langi, 2017) menyatakan bahwa regresi logistik biner dimanfaatkan untuk mengkaji keterkaitan antara satu variabel respons dengan beberapa variabel prediktor, di mana variabel respons tersebut terdiri dari dua kategori.

Regresi logistik menurut (Rismia, Widiharih and Santoso, 2021) dapat dinyatakan sebagai:

$$\pi(x) = \frac{e^{(\beta_0+\beta_1 X_1+\cdots+\beta_p X_p)}}{1+e^{(\beta_0+\beta_1 X_1+\cdots+\beta_p X_p)}} \quad \ldots (1)$$

Dimana $\pi(x)$ adalah peluang sukses kejadian y diberikan nilai x, y adalah variabel respon yang bersifat kategorik, serta x adalah variabel prediktor. Selanjutnya dengan memanfaatkan transformasi logit, model tersebut dapat dinyatakan sebagai berikut:

$$g(x) = ln\left\{\frac{\pi(x)}{1-\pi(x)}\right\} = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \quad \ldots (2)$$

Dengan demikian model Regresi Logistik pada persamaan (1) dapat dinyatakan sebagai:

$$\pi(x) = \frac{e^{g(x)}}{1+e^{g(x)}} \quad \text{... (3)}$$

Uji parameter dan model regresi dilakukan dalam beberapa tahap yaitu:
1). Uji Koefisian Parameter secara Serentak
Hipotesis yang digunakan yaitu:
H_0: $\beta_1 = \cdots = \beta_p = 0$
H_1: $terdapat\ \beta_i \neq 0, i = 1,2,\ldots,p$
Statistik uji yang diterapkan adalah:

$$G = -2ln\frac{\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^{n}\pi_i^{y_i}(1-\pi_i)^{1-y_i}} \quad \text{... (4)}$$

Dengan
n_1 $= \sum_{i=1}^{n} y_i$
n_0 $= \sum_{i=1}^{n}(1-y_i)$
n $= n_1 + n_0$
Daerah penolakan:
Tolak H_0 jika nilai $G > \chi^2_{(db,\alpha)}$ atau p-value < α, dengan derajat bebas adalah banyak variabel prediktor dalam model tanpa β_0.
2). Uji Koefisian Parameter secara Parsial
Hipotesis yang digunakan yaitu:
H_0: $\beta_i = 0$,
H_1: $\beta_i \neq 0$.
Dengan i = 1,2, ..., p.
Statistik uji yang digunakan adalah:

$$W = \frac{\widehat{\beta_i}}{SE(\widehat{\beta_i})} \quad \text{... (5)}$$

dengan
$\widehat{\beta_i}$ = nilai dugaan untuk parameter β_i
$\frac{\widehat{\beta_i}}{SE(\widehat{\beta_i})}$ = dugaan standar error baku untuk koefisien β_i.
Tolak H_0 jika nilai $|W| > \chi^2_{(\alpha,1)}$ atau p-value < α.
3) Uji kesesuaian model
Hipotesis yang digunakan adalah:
H_0: *model yang dibangun sesuai*
H_1: *model yang dibangun tidak sesuai*
Statistik uji yang digunakan:

$$\hat{C} = \sum_{i=1}^{k}\frac{(O_i - n_i\overline{\pi_i})^2}{n_i\overline{\pi_i}(1-\overline{\pi_i})} \quad \text{... (6)}$$

dengan:
O_i = pengamatan pada kelompok ke-i
k = jumlah kelompok dalam model serentak
n_i = banyaknya pengamatan pada kelompok ke-i
$\overline{\pi_i}$ = rata-rata taksiran peluang pengamatan kelompok ke-i
Tolak H_0 jika nilai $\hat{C} > \chi^2_{(\alpha,db)}$, dengan derajat bebas db=(k-2).

2.3 K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* (KNN) merupakan salah satu dari algoritma *supervised learning* yang bertujuan untuk klasifikasi data. Menurut (Utami, Fadjryani and Daniaty, 2020), KNN adalah metode klasifikasi objek dengan mempertimbangkan kelas terdekat dari objek tersebut. Nilai K pada KNN merujuk pada banyaknya data terdekat dari data *testing* yang digunakan dalam penentuan kelas. Lebih lanjut, (James *et al.*, 2017) menyatakan jika diberikan K *integer* positif

dan data testing x_0, selanjutnya ditentukan K titik pada data *training* yang terdekat ke x_0, disimbolkan $\mathcal{N}_0$, maka probabilitas bersyarat untuk kelas j dapat dinyatakan:

$$P(Y = j|X = x_0) = \frac{1}{K}\sum_{i\epsilon\mathcal{N}_0} I(y_i = j). \quad \ldots (7)$$

Algoritma KNN yang digunakan yaitu:

1. Menentukan nilai K (jumlah tetangga terdekat) yang digunakan, dalam penelitian ini dipilih K dari 1 hingga 18.
2. Menentukan jarak *Euclid* masing-masing objek terhadap data *training*.
3. Menentukan titik-titik yang termuat dalam $\mathcal{N}_0$.
4. Menentukan data terklasifikasi pada kelas 0 atau 1 menggunakan kategori tetangga terdekat terbanyak.

2.4 K-Nearest Neighbor

Perhitungan ketepatan klasifikasi model dalam penelitian ini menggunakan *confusion matrix*. Berdasarkan Utami, Fadjryani and Daniaty, (2020), matriks konfusi adalah tabel yang digunakan untuk mencatat kinerja klasifikasi dan dinyatakan dalam tabel berikut:

Tabel 2. Matriks Konfusi

Aktual	Prediksi	
	M_1	M_2
M_1	A_{11}	A_{12}
M_2	A_{21}	A_{22}

Dengan:

A_{11} = jumlah observasi M_1 diklasifikasikan M_1
A_{12} = jumlah observasi M_1 diklasifikasikan M_2
A_{21} = jumlah observasi M_2 diklasifikasikan M_1
A_{22} = jumlah observasi M_2 diklasifikasikan M_2

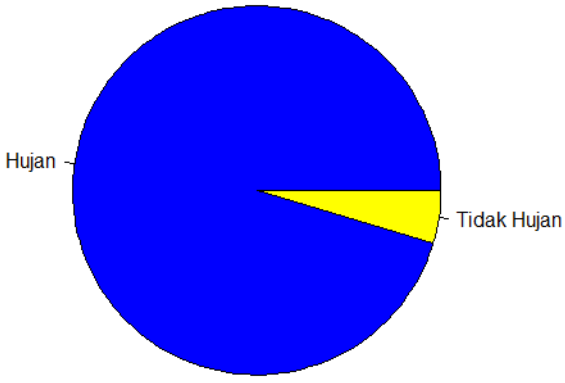
Akurasi model dapat dihitung dengan rumus:

$$\text{Akurasi} = \frac{A_{11}+A_{22}}{A_{11}+A_{12}+A_{21}+A_{22}} \quad \ldots (8)$$

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Data

Data diambil dari bulan Januari 2018 hingga Desember 2020 di kota Bogor, Jawa Barat dengan total 526 pengamatan. Dilakukan eksplorasi data untuk memperoleh informasi awal dari data.



Gambar 1. Pie Chart Variabel y (Curah hujan)

Berdasarkan Gambar 1, data pengamatan cuaca antara tahun 2018-2020 di Kota Bogor yang digunakan, sebesar 95,63%, terjadi hujan adapun sisanya sebesar 4,37% tidak terjadi hujan. Selanjutnya untuk deskripsi data variabel prediktor X_1 , X_2, X_3, X_4 dinyatakan dalam tabel berikut

Tabel 3. Deskripsi Data Variabel Prediktor

Statistika deskriptif	Variebel prediktor			
	X_1	X_2	X_3	X_4
Rata-rata	26.11	84.32	5.189	1.428
Standar deviasi	0.795	5.045	2.73	0.749
Nilai min	23.10	63	0	0
Nilai max	28.50	96	10.6	5

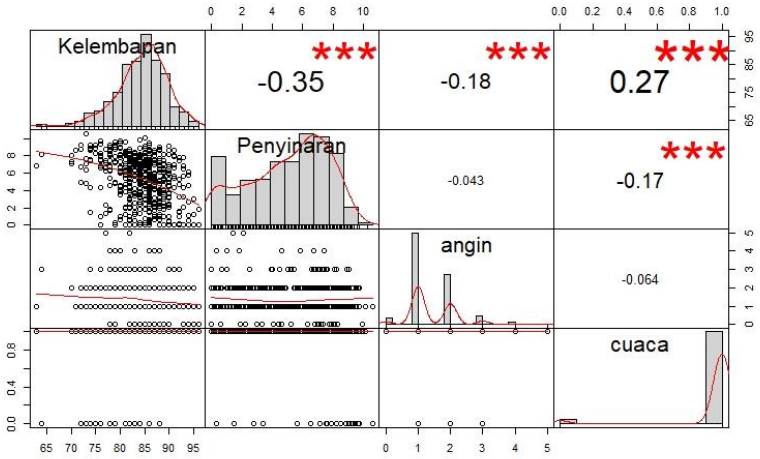
Multikolinearitas terjadi dalam model regresi ketika terdapat keterkaitan yang signifikan antara variabel prediktor. Agar tidak terjadi multikolinearitas, koefisien korelasi antar variabel menurut (Ridwan and Sunendiari, 2021) haruslah di bawah 0,8 dan lebih dari -0,8. Adapun perhitungan koefisien korelasi dinyatakan dalam persamaan berikut:

$$r_{X_pX_q} = \frac{n\sum_{i=1}^{n} X_{pi}X_{qi} - n\sum_{i=1}^{n} X_{pi}\sum_{i=1}^{n} X_{qi}}{\sqrt{\left(n\sum_{i=1}^{n} X_{pi}{}^2 - \left(\sum_{i=1}^{n} X_{pi}\right)^2\right)\left(n\sum_{i=1}^{n} X_{qi}{}^2 - \left(\sum_{i=1}^{n} X_{qi}\right)^2\right)}} \quad \text{... (9)}$$

Dengan:

X_p = variabel prediktor - p
X_q = variabel prediktor - q
n = banyak sampel yang digunakan

Koefisien korelasi variabel prediktor pada data penelitian ini dinyatakan dalam *chart* korelasi *output software* R pada gambar berikut.



Gambar 2. Chart Korelasi Variabel Prediktor.

Berdasarkan Gambar 2, tidak ada koefisien korelasi yang lebih dari 0,8 maupun kurang dari -0,8 antar variabel prediktornya. Hal ini mengindikasikan bahwa data tidak mengandung multikolinearitas sehingga dapat dilanjutkan ke tahap analisis data dengan model regresi logistik.

3.2 Analisis Data

Sebelum dianalisis, data dibagi menjadi dua kategori yakni *data training* dan *data testing*. Pengamatan di tahun 2019 dan 2020 sebanyak 333 data digunakan sebagai *data training*, adapun sisanya yaitu pengamatan di tahun 2018 sebanyak 193 data digunakan sebagai *data testing*. *Data training* yang dimiliki akan dipakai untuk membentuk model, adapun *data testing* akan digunakan untuk mengevaluasi kinerja model.

3.2.1 Analisis Pemodelan Regresi Logistik Biner

Koefisien regresi logistik biner diperoleh dengan mengestimasi parameter menggunakan *data training*. Selanjutnya dari hasil estimasi dapat disusun model regresi logistik binernya.

Model Tahap Pertama

Model regresi logistik biner yang diperoleh setelah dilakukan estimasi parameter yaitu:

$$\pi(x) = \frac{e^{f(x)}}{1+e^{f(x)}} \quad \text{... (10)}$$

Dengan

$$f(x) = -14.2 - 0.061\,X_1 + 0.32688X_2 - 0.87254X_3 - 0.75831X_4$$

1) Uji Koefisian Parameter secara Serentak

Uji serentak digunakan untuk menilai dampak variabel prediktor secara kolektif terhadap variabel respon.

Hipotesis yang digunakan:

H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

H_1: $terdapat\ \beta_i \neq 0, i = 1,2,3,4.$

Nilai G yang diperoleh yaitu G=30.594, sedangkan nilai $\chi^2_{(0,05,4)}$ = 9.4877. Karena nilai G>$\chi^2_{(0,05,4)}$ maka H_0 ditolak. Oleh karena itu, dapat disimpulkan bahwa secara bersama-sama, variabel prediktor memiliki pengaruh terhadap variabel respon.

2) Uji Koefisian Parameter secara Parsial

Dalam uji ini, hipotesis yang digunakan adalah:

H_0: $\beta_i = 0$,

H_1: $\beta_i \neq 0$, $dengan\ i = 1,2,3,4.$

Jika *p-value* < 5%, maka H_0 ditolak. Akibatnya dapat disimpulkan variabel berpengaruh signifikan terhadap model regresi logistik biner. Hasil ujil parsial dinyatakan dalam tabel berikut.

Tabel 4. Uji Parsial Model Pertama

Variabel	*p-value*	kesimpulan
X_1	0.90213	Tidak berpengaruh signifikan pada model
X_2	0.000097	Berpengaruh signifikan pada model
X_3	0.00394	Berpengaruh signifikan pada model
X_4	0.18277	Tidak berpengaruh signifikan pada model

Ditemukan bahwa hasil uji parsial menunjukkan bahwa variabel X_2 dan X_3 memiliki pengaruh signifikan pada model, sementara variabel X_1 dan X_4 tidak menunjukkan pengaruh yang signifikan. Selanjutnya, model regresi logistik biner kedua akan dibentuk dengan menghapus variabel yang tidak berpengaruh secara signifikan.

Model Tahap Kedua

Model regresi logistik biner yang diperoleh setelah dilakukan estimasi parameter yaitu:

$$\pi(x) = \frac{e^{f(x)}}{1+e^{f(x)}} \quad \text{... (11)}$$

Dengan

$$f(x) = -18.10582 + 0.33142X_2 - 0.76653X_3$$

1) Uji Koefisian Parameter secara Serentak
Uji serentak adalah teknik uji yang memungkinkan evaluasi pengaruh simultan dari beberapa variabel prediktor terhadap variabel respons.
Hipotesis yang digunakan:
$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$
$H_1: terdapat\ \beta_i \neq 0, i = 1,2,3,4.$
Nilai G yang diperoleh yaitu G=28.7125 sedangkan nilai $\chi^2_{(0,05,2)} = 5.991465$. Karena nilai G>$\chi^2_{(0,05,2)}$ maka H_0 ditolak. Dengan demikian dapat dikatakan variabel prediktor secara bersama-sama berpengaruh terhadap variabel respon.

2) Uji Koefisian Parameter secara Parsial
Dalam uji ini, hipotesis yang digunakan adalah:
$H_0: \beta_i = 0,$
$H_1: \beta_i \neq 0,\ dengan\ i = 1,2,3,4.$
Jika *p-value* < 5%, maka H_0 ditolak. Akibatnya dapat disimpulkan variabel berpengaruh signifikan terhadap model regresi logistik biner. Hasil ujil parsial dinyatakan dalam tabel berikut.

Tabel 5. Uji Parsial Model Pertama

Variabel	*p-value*	Kesimpulan
X_2	0.0000481	Berpengaruh signifikan pada model
X_3	0.00711	Berpengaruh signifikan pada model

Dari uji tahap kedua, diperoleh seluruh variabel yang digunakan yaitu X_2 dan X_3 berpengaruh signifikan pada model. Dengan demikian dilakukan dengan uji kesesuaian model.

3) Uji kesesuaian model
Hipotesis yang digunakan adalah:
$H_0: model\ yang\ dibangun\ sesuai$
$H_1: model\ yang\ dibangun\ tidak\ sesuai$
Hasil yang didapatkan berdasarkan uji kesesuaian model adalah sebagai berikut:

Tabel 6. Uji Kesesuaian Model

$\chi^2 hitung$	Df	$\chi^2 tabel$	*p-value*
2.749	8	15.5073 1	0.9491

Dari Tabel 6 diperoleh bahwa nilai $\chi^2 hitung < \chi^2 tabel$ serta p-value > 0,05. Selanjutnya dapat disimpulkan bahwa H_0 diterima, yang artinya model yang dibangun telah sesuai.

Model final dari regresi logistik yang diperoleh yaitu sesuai dengan persamaan (11):

$$\pi(x) = \frac{e^{f(x)}}{1+e^{f(x)}}$$

Dengan

$$f(x) = -18.10582 + 0.33142X_2 - 0.76653X_3$$

Dimana X_2 menyatakan kelembapan rata-rata dan X_3 lamanya penyinaran matahari. Berdasarkan model dapat dikatakan semakin tingginya kelembapan rata-rata akan meningkatkan peluang terjadinya hujan dan semakin lama penyinaran matahari akan menurunkan peluang terjadinya hujan di Kota Bogor.

Akurasi Model

Setelah mendapatkan model akhir dari regresi logistik biner sesuai dengan persamaan (9), langkah selanjutnya adalah menghitung akurasi model. Akurasi model dihitung dengan membandingkan hasil prediksi *data testing* dengan data aktual pada variabel respon. Hasilnya direpresentasikan dalam *confusion matrix* seperti yang dinyatakan pada tabel di bawah ini:

Tabel 7. *Confusion matrix* Model Regresi Logistik Biner

Aktual	Prediksi	
	Tidak Hujan	Hujan
Tidak Hujan	2	7
Hujan	7	177

Dari tabel di atas, ketepatan klasifikasi model regresi logistik biner yaitu:
Akurasi = $\frac{2+177}{2+7+7+177} = 92{,}746\%$,,

Dengan demikian dapat disimpulkan akurasi dari model regresi logistik biner yang dibentuk adalah 92,746%.

3.2.2 *K-Nearest Neighbors*

Tahapan pertama yang dilakukan pada pembentukan KNN yaitu dengan memisahkan data menjadi *data training* dan *data testing*. Pembagian yang dilakukan sama seperti pada model regresi logistik biner, yaitu sebanyak 333 data sebagai data *training* dan sisanya 193 data sebagai data *testing*.
Nilai K yang digunakan yaitu dari 1 sampai 18, dengan demikian ada 18 model KNN yang dilakukan dalam penelitian ini. Selanjutnya komputasi diselesaikan menggunakan *package Class* pada *software* R.
Confusion matrix pada tiap K dirangkum dalam tabel di bawah ini:

Tabel 8. Gabungan *Confusion matrix* Model KNN

Aktual	Prediksi	
	Tidak Hujan	Hujan
	K=1	
Tidak Hujan	1	12
Hujan	8	172
	K=2	
Tidak Hujan	2	9
Hujan	7	175
	K=3	
Tidak Hujan	1	6
Hujan	8	178
	K=4	
Tidak Hujan	1	8
Hujan	8	176
	K=5	
Tidak Hujan	1	6
Hujan	8	178
	K=6	
Tidak Hujan	2	8
Hujan	7	176
	K=7	
Tidak Hujan	1	7
Hujan	8	177
	K=8	
Tidak Hujan	2	7
Hujan	7	177

	K=9	
Tidak Hujan	1	11
Hujan	8	173
	K=10	
Tidak Hujan	2	8
Hujan	7	176
	K=11	
Tidak Hujan	1	8
Hujan	8	176
	K=12	
Tidak Hujan	1	8
Hujan	8	176
	K=13	
Tidak Hujan	1	7
Hujan	8	177
	K=14	
Tidak Hujan	0	5
Hujan	9	179
	K=15	
Tidak Hujan	1	4
Hujan	8	180
	K=16	
Tidak Hujan	0	6
Hujan	9	178
	K=17	
Tidak Hujan	0	3
Hujan	9	181
	K=18	
Tidak Hujan	1	2
Hujan	8	182

Nilai ketepatan klasifikasi model dinyatakan dalam tabel berikut.

Tabel 9. Ketepatan Klasifikasi

Nilai K	Akurasi
1	0.8963731
2	0.9170984
3	0.9274611
4	0.9170984
5	0.9274611
6	0.9222798
7	0.9222798
8	0.9274611
9	0.9015544
10	0.9222798
11	0.9222798
12	0.9170984
13	0.9222798
14	0.9274611
15	0.9378238
16	0.9222798
17	0.9378238
18	0.9481865

Berdasarkan Tabel 9, nilai akurasi terbaik diperoleh pada K=18 yaitu sebesar 94,81865%.

3.3 Perbandingan Ketepatan Klasifikasi Kedua Model

Hasil ketepatan atau akurasi klasifikasi pada model Regresi Logistik Biner yaitu sebesar 92,746%, adapun ketepatan klasifikasi *K-Nearest Neighbors* (KNN) pada K = 18 sebesar 94,81865%.. Berdasarkan hal tersebut, pada penelitian ini model KNN memiliki akurasi yang lebih baik daripada model Regresi Logistik Biner pada prediksi terjadinya hujan.

4. SIMPULAN DAN SARAN

Hasil analisis data serta pembahasan pada penelitian ini menunjukkan bahwa:

1) Model Regresi Logistik Biner menghasilkan ketepatan klasifikasi sebesar 92,746%,, pada kasus prediksi curah hujan di Kota Bogor.
2) Model K-Nearest Neighbors (KNN) menghasilkan ketepatan klasifikasi pada K=18 sebesar 94,81865% pada kasus prediksi curah hujan di Kota Bogor.
3) Dalam kasus prediksi curah hujan di Kota Bogor, model K-Nearest Neighbors (KNN) menunjukkan tingkat ketepatan klasifikasi yang lebih unggul dibandingkan dengan model Regresi Logistik Biner.

Untuk penelitian selanjutnya, dapat dilakukan studi komparasi *K-Nearest Neighbor* (K-NN) dan Regresi Logistik Biner pada studi kasus yang berbeda. Dapat pula dilakukan prediksi curah hujan di Kota Bogor dengan teknik klasifikasi lain seperti *decision tree, random forest, naive bayes*, dan sebagainya.

DAFTAR PUSTAKA

BMKG. (2011). Evaluasi cuaca dan sifat hujan Bulan Agustus 2011 serta prakiraan cuaca dan sifat hujan Bulan September 2011, *Bulletin Metereologi. Badan Metereologi Klimatologi dan Geofisika Stasiun Metereologi Otorita Batam*, 1, p. 39.

Fallo, S.I. (2021). *Support Vector Machine, Naive Bayes Classifier, dan Regresi Logistik Ordinal dalam Prediksi Cuaca*. Universitas Gadjah Mada.

James, G. *et al.* (2017). *An Introduction to Statistical Learning with Applications in R*. Available at: https://doi.org/10.1007/978-1-4614-7138-7_8.

Jayanti, R.D. and Noeryanti (2014) 'Apilkasi Metode K-Nearest Neighbor dan Analisis Diskriminan untuk Analisis Resiko Kredit pada Koperasi Simpan Pinjam di Kopinkra Sumber Rejeki', *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST)*, pp. 275–284.

Rahmawati, F. and Suratman, R.Y. (2022). Performa Regresi Ridge dan Regresi Lasso pada Data dengan Multikolinearitas, *Leibniz: Jurnal Matematika*, 2(2), pp. 1–10. Available at: https://doi.org/10.59632/leibniz.v2i2.176.

Ridwan, M. and Sunendiari, S. (2021). Mendeteksi dan Mengatasi Multikolinieritas pada Data Penelitian Diabetes Melitus Wanita Suku Indian Tahun 2018, *Prosiding Statistika*, pp. 64–70. Available at: https://karyailmiah.unisba.ac.id/index.php/statistika/article/view/25565.

Rismia, E.R., Widiharih, T. and Santoso, R. (2021). Klasifikasi Regresi Logistik Multinomial Dan Fuzzy K-Nearest Neighbor (Fk-Nn) Dalam Pemilihan Metode Kontrasepsi Di Kecamatan Bulakamba, Kabupaten Brebes, Jawa Tengah, *Jurnal Gaussian*, 10(4), pp. 476–487. Available at: https://doi.org/10.14710/j.gauss.v10i4.33095.

Sepang, F., Komalig, H. and Hatidja, D. (2012). Penerapan Regresi Logistik untuk Menentukan Faktor-Faktor yang Mempengaruhi Pemilihan Jenis Alat Kontrasepsi di Kecamatan Modayag Barat, *do. Jurnal MIPA Unsrat Online*, 1(1), pp. 1–5.

Susilokarti, D. *et al.* (2015). Identifikasi Perubahan Iklim Berdasarkan Data Curah Hujan di Wilayah Selatan Jatiluhur Kabupaten Subang, Jawa Barat, *Jurnal Agritech*, 35(01), p. 98. Available at: https://journal.ugm.ac.id/agritech/article/view/13038/15155.

Tampil, Y., Komaliq, H. and Langi, Y. (2017). Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado, *d'CARTESIAN*, 6(2), pp. 56–62. Available at: https://doi.org/10.35799/dc.6.2.2017.17023.

Utami, I.T., Fadjryani and Daniaty, F.F.D. (2020). Perbandingan Klasifikasi Status Pendonor Darah dengan Menggunakan Regresi Logistik dan K-Nearest Neighbor', *Jurnal.Stis.Ac.Id*, V.12.1.202. Available at:

https://jurnal.stis.ac.id/index.php/jurnalasks/article/view/217%0Ahttps://jurnal.stis.ac.id/index.php/jurnalasks/article/download/217/83.