

# Analisis Perbandingan Metode *Decision Tree*, *Naïve Bayes*, dan *Radial Basis Function Neural Network* pada Klasifikasi Tinggi Muka Air Laut Marina Ancol

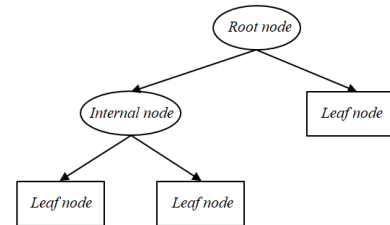
Qatrunnada Gusti Nadhifa, Ulil Azmi, dan Prilyandari Dina Saputri  
 Departemen Aktuaria, Institut Teknologi Sepuluh Nopember (ITS)  
 e-mail: ulil\_azmi@its.ac.id

**Abstrak**—Secara geografis, Indonesia merupakan negara kepulauan dengan luas lautan yang mencapai dua per tiga dibandingkan dengan luas daratannya. Berdasarkan hal tersebut, salah satu potensi yang dapat dikembangkan pada bidang perekonomian Indonesia adalah wisata bahari. Pantai Marina Ancol merupakan salah satu kawasan wisata bahari yang terletak di wilayah Jakarta Utara. Untuk menciptakan rasa keamanan dan kenyamanan bagi para wisatawan Pantai Marina Ancol, analisis klasifikasi tinggi muka air laut perlu dilakukan sebagai upaya antisipasi terhadap terjadinya bencana banjir rob yang merupakan peristiwa penggenangan daratan akibat terjadinya kenaikan permukaan air laut. Pada penelitian ini akan dilakukan analisis klasifikasi tinggi muka air laut Marina Ancol menggunakan metode *Decision Tree*, *Naïve Bayes*, dan *Radial Basis Function Neural Network* untuk mengetahui metode terbaik yang dapat digunakan dalam mengklasifikasikan tinggi muka air ke dalam empat kategori level air yaitu Siaga 1, Siaga 2, Siaga 3, dan Siaga 4 berdasarkan *accuracy score* dan *false negative rate*. Dalam memprediksikan kategori level air laut Marina Ancol, model yang didapat dengan metode *Decision Tree* menghasilkan nilai akurasi sebesar 85,19%, nilai *false negative rate* Siaga 1 sebesar 16,67%, dan nilai *false negative rate* Siaga 2 sebesar 12,24%. Dengan menggunakan metode *Naïve Bayes*, nilai akurasi yang dihasilkan dari model adalah sebesar 73,37%, nilai *false negative rate* Siaga 1 sebesar 16,67%, dan nilai *false negative rate* Siaga 2 sebesar 24,27%. Arsitektur model *Radial Basis Function Neural Network* yang terbaik adalah model dengan jumlah *input layer* sebanyak tiga, *hidden neuron* sebanyak lima belas, dan *output neuron* sebanyak empat. Nilai akurasi model RBFNN (3-15-4) menunjukkan nilai sebesar 86,64% nilai *false negative rate* Siaga 1 sebesar 100%, dan nilai *false negative rate* Siaga 2 sebesar 14,73%. Berdasarkan hasil tersebut, *Decision Tree* merupakan metode yang menghasilkan model lebih baik dibandingkan *Radial Basis Function Neural Network* dan *Naïve Bayes* dalam melakukan prediksi kategori level air.

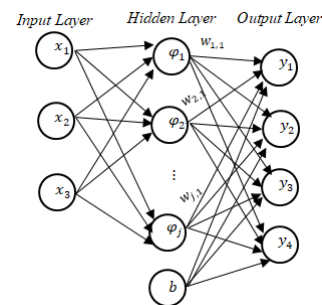
**Kata Kunci**—*Decision Tree*, *Klasifikasi*, *Naïve Bayes*, *Radial Basis Function Neural Network*, *Tinggi Muka Air*.

## I. PENDAHULUAN

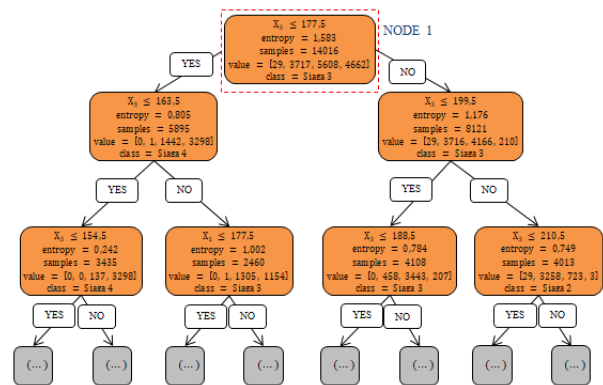
Secara geografis, Indonesia merupakan negara kepulauan dengan luas lautan yang mencapai dua per tiga dibandingkan dengan luas daratannya. Indonesia memiliki garis pantai lebih dari 99.000 km sehingga memiliki potensi yang besar pada bidang kelautan. Salah satu potensi yang dapat dikembangkan berdasarkan hal ini adalah perekonomian Indonesia melalui bidang wisata Bahari. Pantai Marina Ancol merupakan salah satu kawasan wisata bahari yang terletak di wilayah Jakarta Utara. Dilansir dari situs resmi Ancol (2023), kegiatan yang dapat dilakukan di daerah Taman Impian Jaya Ancol diantaranya adalah wisata



Gambar 1. Struktur *decision tree*.



Gambar 2. Arsitektur *radial basic function neural network*.



Gambar 3. Struktur model *decision tree*.

pantai, hiburan dan pendidikan dunia laut, wisata kuliner makanan laut, hingga pusat hiburan *theme park*. Beragamnya kegiatan ini menjadikan Pantai Marina Ancol menjadi kawasan wisata andalan di Jakarta yang diminati oleh wisatawan.

Pasang surut merupakan suatu fenomena yang terjadi di laut yaitu pergerakan naik turunnya permukaan air laut secara berkala yang diakibatkan oleh gaya tarik menarik benda-benda langit terutama matahari dan bulan terhadap massa air laut di bumi [1]. Peristiwa pasang surut ini mempengaruhi perubahan tinggi muka air dan dapat menyebabkan bencana yang sering terjadi di daerah pesisir pantai yaitu banjir rob. Banjir rob merupakan peristiwa penggenangan daratan akibat terjadinya kenaikan permukaan air laut [2]. Untuk

Tabel 4.

Kategori tinggi muka air pada pintu air marina ancol

Kategori Level Air	Status Tinggi Air	Tinggi Muka Air (cm)
Siaga 1	Bahaya	>250
Siaga 2	Siaga	200 – 250
Siaga 3	Waspada	170 – 200
Siaga 4	Normal	<170

Tabel 5.

Confusion matrix 4x4 dalam sudut pandang kelas A

Actual Values	Predicted Values			
	Classes			
	A	B	C	D
Classes				
A	TP	FN	FN	FN
B	FP	TP	TN	TN
C	FP	TN	TP	TN
D	FP	TN	TN	TP

Tabel 6.

Statistik deskriptif tinggi muka air marina ancol periode 2021-2022

Tinggi Muka Air (cm)	
Minimum	105
1 <sup>st</sup> Quartile	164
Median	183
3 <sup>rd</sup> Quartile	202
Maximum	262
Mean	183

menciptakan rasa keamanan dan kenyamanan bagi para wisatawan Pantai Marina Ancol, analisis klasifikasi tinggi muka air laut perlu dilakukan sebagai upaya antisipasi terhadap terjadinya peristiwa luapan air laut ini yang dapat membahayakan dan mengganggu kenyamanan wisatawan. Tinggi muka air dapat dikategorikan ke dalam empat kategori level air yaitu Siaga 1 atau berstatus bahaya, Siaga 2 atau berstatus siaga, Siaga 3 atau berstatus waspada, dan Siaga 4 atau berstatus normal.

Klasifikasi tinggi muka air laut dapat dilakukan dengan berbagai metode. Penelitian terkait perkiraan tinggi gelombang air laut menggunakan metode *Association Rule* dan *Classification Tree* dan didapatkan bahwa metode *classification tree* menghasilkan tingkat akurasi yang lebih tinggi yaitu sebesar 88% jika dibandingkan *association rule* dalam mendapatkan model prediksi [3]. Penelitian analisis prediksi gelombang ekstrim air laut di Pantai Ancol dengan kesimpulan yang didapatkan adalah algoritma *Iterative Dichotomiser 3* (ID3) memberikan nilai akurasi yang tinggi yaitu sebesar 88% terhadap prediksi tinggi muka air laut di Pantai Ancol [4]. Studi komparasi metode algoritma ID3 dan algoritma *Naïve Bayes* untuk klasifikasi penyakit diabetes mellitus. Pada penelitian ini didapatkan hasil nilai akurasi algoritma *Naïve Bayes* yang lebih baik yaitu sebesar 76% dibanding pada algoritma ID3 sebesar 74% [5]. Penelitian terkait pengklasifikasian daerah tertinggal di Indonesia dengan metode *Radial Basis Function Neural Network* (RBFNN). Pada penelitian ini digunakan struktur RBFNN (27-34-2) yaitu 27 neuron pada *input layer*, 34 neuron pada *hidden layer*, dan 2 neuron pada *output layer* [6]. Berdasarkan evaluasi performa klasifikasi, didapatkan nilai akurasi sebesar 93,48% dan nilai *F-Measure* sebesar 85,36%.

Pada penelitian ini akan dilakukan analisis klasifikasi tinggi muka air pada Pintu Air Marina Ancol menggunakan metode *Decision Tree*, *Naïve Bayes*, dan *Radial Basis*

Tabel 1.

Rincian jumlah data per kategori level air

Kategori Level Air	Status Tinggi Air	Total Data
Siaga 1	Bahaya	35
Siaga 2	Siaga	4640
Siaga 3	Waspada	7012
Siaga 4	Normal	5833

Tabel 2.

Nilai entropi dan *information gain* pada node pertama

NODE 1	Partisi	Entropi	Information Gain
X <sub>1</sub>	<177,5 cm	1,157342	0,229864
	≥177,5 cm	1,495679	
X <sub>2</sub>	<177,5 cm	0,955112	0,387019
	≥177,5 cm	1,370935	
X <sub>3</sub>	<177,5 cm	0,804930	0,563597
	≥177,5 cm	1,175503	

Tabel 3.

Contoh data testing

X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	Y
178	172	166	?

*Function Neural Network* untuk mengetahui metode terbaik yang dapat digunakan dalam mengklasifikasikan tinggi muka air ke dalam empat kategori level air yaitu Siaga 1, Siaga 2, Siaga 3, dan Siaga 4 berdasarkan *accuracy score* dan *false negative rate*. Data tinggi muka air yang digunakan merupakan tinggi muka air pada Pintu Air Marina Ancol periode Januari 2021 hingga Desember 2022 per jam setiap harinya.

## II. TINJAUAN PUSTAKA

### A. Tinggi Muka Air dan Pintu Air Marina Ancol

Pintu air merupakan suatu bangunan yang berfungsi sebagai pengatur lalu lintas air. Pintu air merupakan salah satu infrastruktur yang dapat digunakan dalam mengatasi masalah luapan air. Pintu air berfungsi untuk mengendalikan debit air dengan mengalirkannya menuju sungai ataupun lautan. DKI Jakarta sendiri memiliki beberapa pintu air di berbagai tempat yang salah satunya yaitu pintu air Marina Ancol yang letaknya bersisian langsung dengan Taman Impian Jaya Ancol yang merupakan salah satu destinasi wisata terkenal di DKI Jakarta. Menurut Dinas SDA DKI Jakarta, berdasarkan tinggi muka air, ketinggian air di setiap pintu air dikelompokkan ke dalam empat kategori yaitu Siaga 1, Siaga 2, Siaga 3, dan Siaga 4 dengan batas tinggi muka air pada pintu air Marina Ancol dijabarkan seperti pada Tabel 1.

### B. Decision Tree ID3 Algorithm

*Decision Tree* merupakan salah satu algoritma dalam *data mining* yang dapat digunakan dalam proses klasifikasi. Algoritma ini membuat model prediksi terhadap suatu keputusan menggunakan struktur hirarki atau pohon dimana setiap pohon memiliki cabang yang mewakili atribut tertentu yang harus dipenuhi sebelum menuju cabang berikutnya sampai berakhir tidak ada cabang lagi [7]. Contoh struktur pohon dari *Decision Tree* disajikan pada Gambar 1.

Pada struktur pohon *Decision Tree* terdiri atas tiga jenis simpul (*node*) yang memiliki fungsinya masing-masing. Simpul pertama yaitu simpul akar (*root node*) merupakan titik awal dari suatu *Decision Tree*. Selanjutnya yaitu *internal node* yang merupakan percabangan pada struktur *Decision*

Tabel 9. Hasil perhitungan nilai *mean*, standar deviasi, dan *varians*

	Y	$\mu$	$\sigma$	$\sigma^2$
$X_1$	Siaga 1	234,2413	17,7193	313,9753
	Siaga 2	205,3949	18,4140	339,0775
	Siaga 3	184,0872	20,5758	423,3662
	Siaga 4	164,1540	20,6022	424,4542
$X_2$	Siaga 1	242,5517	12,5059	156,3990
	Siaga 2	210,2119	15,8106	249,9765
	Siaga 3	184,9830	16,2719	264,7757
	Siaga 4	159,2537	16,7418	280,2906
$X_3$	Siaga 1	249,2068	6,58124	43,31280
	Siaga 2	213,5082	13,0206	169,5363
	Siaga 3	185,5356	11,7739	138,6258
	Siaga 4	155,9519	13,1768	173,6282

Tabel 10. Hasil perhitungan nilai *likelihood*

	Y	Likelihood
$X_1$	Siaga 1	0,0006
	Siaga 2	0,0307
	Siaga 3	0,0841
	Siaga 4	0,0701
$X_2$	Siaga 1	1,3849 E-08
	Siaga 2	0,0054
	Siaga 3	0,0719
	Siaga 4	0,0729
$X_3$	Siaga 1	3,0311 E-36
	Siaga 2	0,0001
	Siaga 3	0,0293
	Siaga 4	0,0821

*Tree* ini. Simpul terakhir yaitu *leaf node* merupakan titik akhir yang memuat keputusan akhir untuk suatu *decision tree* dan menunjukkan hasil akhir dari klasifikasi.

*Iterative Dichotomiser 3* (ID3) merupakan salah satu algoritma paling dasar dari *Decision Tree*. Adapun tahapan dari pengujian ini adalah sebagai berikut:

1. Menghitung nilai entropi yang merupakan informasi yang menyatakan ukuran ketidakpastian dari atribut data dengan persamaan berikut:

$$\text{Entropy}(S) = -\sum_{i=1}^n p_i \times \log_2 p_i, \quad (1)$$

dengan  $S$  merupakan himpunan kasus,  $n$  adalah jumlah partisi pada  $S$ , dan  $p_i$  merupakan proporsi kejadian yang termasuk ke dalam kelas ke- $i$ .

2. Menghitung nilai *Information Gain* yang mengukur pengurangan yang diharapkan dari *Entropy* jika diketahui nilai pada fitur  $A$  dengan persamaan:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times \text{Entropy}(S_i) \quad (2)$$

dengan  $S$  merupakan himpunan kasus,  $A$  merupakan atribut atau fitur,  $n$  adalah jumlah partisi atau kelas,  $S_i$  merupakan *subset*  $S$ ,  $|S_i|$  menunjukkan jumlah kasus pada pada *subset*  $S_i$ , dan  $|S|$  merupakan jumlah kasus dalam  $S$ .

Atribut dengan nilai *Information Gain* tertinggi akan menjadi akar dari *Decision Tree* ini dan atribut dengan nilai *Information Gain* terbesar setelah itu akan menjadi *node* selanjutnya hingga terbentuklah suatu struktur pohon keputusan.

### C. Naïve Bayes Classifier

*Naïve Bayes* merupakan suatu metode klasifikasi berdasarkan pada teorema *Bayes*. Teroema ini diungkapkan oleh ilmuwan Inggris Thomas Bayes yaitu mendapatkan

Tabel 7. Nilai bobot *input layer* menuju *hidden layer* pada struktur RBFNN

Input Layer	Hidden Layer			
	H(1)	H(2)	...	H(15)
$X_1$	154,941	140,956	...	184,468
$X_2$	153,642	139,180	...	187,281
$X_3$	154,610	140,574	...	189,391
Spread	7,532	8,713	...	4,662

Tabel 8. Nilai bobot *hidden layer* menuju *output layer* pada struktur RBFNN

Hidden Layer	Output Layer			
	SIAGA 1	SIAGA 2	SIAGA 3	SIAGA 4
Intercept	0,013	0,306	0,319	0,362
H(1)	-0,018	-0,420	-0,442	0,880
H(2)	-0,019	-0,467	-0,474	0,960
H(3)	-0,020	-0,474	-0,529	1,023
H(4)	-0,016	-0,388	0,899	-0,495
H(5)	-0,016	-0,344	0,204	0,156
H(6)	-0,019	-0,476	0,884	-0,389
H(7)	-0,017	-0,429	-0,371	0,817
H(8)	-0,019	-0,352	0,988	-0,617
H(9)	-0,011	-0,346	0,920	-0,562
H(10)	-0,013	1,103	-0,488	-0,602
H(11)	-0,017	-0,240	0,739	-0,481
H(12)	-0,015	1,139	-0,704	-0,420
H(13)	-0,018	0,977	-0,463	-0,496
H(14)	-0,018	0,776	-0,281	-0,477
H(15)	-0,005	-0,469	0,876	-0,402

peluang terjadinya suatu peristiwa berdasarkan informasi dari kejadian yang sudah terjadi. Naive bayes termasuk ke dalam algoritma *supervised* sehingga pada tahapan pembelajaran dibutuhkan data awal berupa data pelatihan (*data training*) untuk dapat mengambil Keputusan [7]. Tahapan dari metode ini adalah sebagai berikut:

1. Menghitung peluang prior yang merupakan informasi dari kejadian masa lalu dengan persamaan:

$$p(c_i) = \frac{n_{c_i}}{N}, \quad (3)$$

dengan  $N$  merupakan total keseluruhan pengamatan,  $n_{c_i}$  adalah total pengamatan yang tergabung dalam kelas  $c_i$ , dan  $p(c_i)$  merepresentasikan peluang suatu objek termasuk ke dalam kelas  $c_i$  tanpa informasi lainnya yang diketahui mengenai objek tersebut.

2. Menghitung peluang posterior yang merupakan peluang suatu pengamatan termasuk ke dalam  $c_i$  saat diketahui fitur  $x$  dengan persamaan:

$$P(c_i|x) = \frac{P(x|c_i) \times P(c_i)}{P(x)} \quad (4)$$

Keterangan:

- $x$  : kriteria suatu kasus berdasarkan fitur
- $c_i$  : kelas solusi pola ke- $i$ ,  $i$  adalah jumlah label kelas
- $P(c_i|x)$  : peluang muncul kelas  $c_i$  dengan kriteria  $x$
- $P(x|c_i)$  : peluang kriteria masukan  $x$  dengan label kelas  $c_i$
- $P(c_i)$  : peluang label kelas  $c_i$ .

Suatu algoritma *Naïve Bayes Classifier* akan menetapkan suatu objek ke dalam kelas  $c_i$  berdasarkan atas objek yang memiliki nilai peluang posterior  $P(c_i|x)$  terbesar ketika diberikan suatu fitur  $x$  [8].

Tabel 11.

Perbandingan kinerja evaluasi model pada *decision tree*, *naïve bayes*, dan RBFNN

Parameter Evaluasi	<i>Decision Tree</i>	<i>Naïve Bayes</i>	RBFNN
Nilai Akurasi	85,22%	73,37%	86,42%
<i>False Negative Rate</i> Siaga 1	16,67%	16,67%	100%
<i>False Negative Rate</i> Siaga 2	12,24%	24,27%	14,73%

Tabel 12.

Hasil prediksi model RBFNN untuk periode 12 jam berikutnya

Tanggal	Jam	$X_1$	$X_2$	$X_3$	Y Prediksi
1 Januari 2023	00.00	158	159	156	SIAGA 4
	01.00	159	156	159	SIAGA 4
	02.00	156	159	163	SIAGA 4
	03.00	159	163	176	SIAGA 3
	04.00	163	176	187	SIAGA 3
	05.00	176	187	194	SIAGA 3
	06.00	187	194	199	SIAGA 2
	07.00	194	199	208	SIAGA 2
	08.00	199	208	204	SIAGA 2
	09.00	208	204	213	SIAGA 2
	10.00	204	213	202	SIAGA 3
	11.00	213	202	205	SIAGA 2

D. *Radial Basis Function Neural Network (RBFNN)*

*Radial Basis Function Neural Network (RBFNN)* merupakan arsitektur dari *Artificial Neural Network (ANN)* yang bersifat *feed forward* [6]. RBFNN menggunakan fungsi aktivasi *Gaussian* yang memperhitungkan jarak antara data dengan pusat data. RBFNN termasuk tipe *multi layer network* yang terdiri atas tiga lapisan yaitu *input layer*, *hidden layer*, dan *output layer*. Pada penelitian ini, arsitektur dari RBFNN dengan tiga buah *input neuron* dan empat buah *output neuron* dapat dilihat pada Gambar 2.

Adapun tahapan dari metode ini adalah sebagai berikut:

1. Inisialisasi nilai *centroid* dan *spread*.
2. Menghitung fungsi aktivasi *Gaussian* dengan persamaan berikut:

$$\varphi_{ij} = \exp\left(-\frac{\|x_i - c_j\|^2}{2\sigma_j^2}\right) \tag{5}$$

dimana  $\varphi_{ij}$  merupakan fungsi aktivasi *Gaussian* pada *hidden neuron* ke- $j$ ,  $x_i$  adalah data *input* ke- $i$ ,  $c_j$  adalah nilai *centroid* dari *hidden neuron* ke- $j$ ,  $\|x_i - c_j\|^2$  merupakan jarak antara data ke- $i$  dengan nilai *centroid* dari *hidden neuron* ke- $j$ , dan  $\sigma_j^2$  merupakan nilai *spread hidden neuron* ke- $j$ .

3. Mendapatkan nilai bobot dan bias melalui pendekatan perhitungan matriks *Gaussian* dengan persamaan:

$$w = (G^T G)^{-1} G^T d \tag{6}$$

dimana  $w$  merupakan matriks bobot dari *hidden neuron* ke *output neuron*,  $G^T$  adalah *transpose* dari matriks *Gaussian*,  $(G^T G)^{-1}$  merupakan *inverse* dari perkalian  $G^T$  dan  $G$ , serta  $d$  merupakan matriks target.

4. Perhitungan *output* dengan persamaan:

$$y_k = \sum_{k=1}^j \varphi_{ij} w_{jk} + b_k \tag{7}$$

dimana  $y_k$  merupakan nilai keluaran *output* ke- $k$ ,  $w_{jk}$  adalah bobot *hidden neuron* ke- $j$  ke *output neuron* ke- $k$ ,  $\varphi_{ij}$  merupakan fungsi aktivasi *Gaussian* data ke- $i$  *hidden*

*neuron* ke- $j$ , dan  $b_k$  adalah bobot bias *output neuron* ke- $k$ .

5. Menentukan kelas tiap data berdasar indeks *output neuron* yang memiliki  $y_k$  maksimum untuk data tersebut.

E. *Confusion Matrix*

*Confusion Matrix* merupakan metode yang dilakukan untuk mengevaluasi hasil dari suatu algoritma klasifikasi. Hal ini merupakan sebuah matrix  $n \times n$  dimana  $n$  merupakan banyaknya kelas yang digunakan untuk menilai performa suatu algoritma klasifikasi [9]. Dalam *Confusion Matrix*, dikenal beberapa istilah yaitu *True Positive (TP)* yang berarti benar diprediksi sebagai kelas  $c_i$ , *True Negative (TN)* yang berarti benar diprediksi sebagai bukan kelas  $c_i$ , *False Positive (FP)* yang salah mengklasifikasikannya menjadi kelas  $c_i$  dan *False Negative (FN)* yang salah mengklasifikasikannya sebagai bukan kelas  $c_i$ . Visualisasi dari *Confusion Matrix* yang pada kasus ini terdapat 4 kelas dengan penempatan TP, TN, FP, dan FN jika dilihat dari sudut pandang kelas A sebagai contoh, dapat dilihat pada Tabel 2.

Pada penelitian ini, untuk mengukur kinerja pada *Confusion Matrix*, digunakan parameter nilai akurasi dan *false negative rate*. Nilai akurasi dapat didefinisikan dalam persamaan sebagai berikut:

$$\text{Akurasi} = \frac{\text{Total TP} + \text{Total TN}}{\text{Total TP} + \text{Total TN} + \text{Total FP} + \text{Total FN}} \times 100\% \tag{8}$$

Sedangkan *flase negative rate* dapat didefinisikan dalam persamaan berikut:

$$\text{FNR} = \left(1 - \frac{\text{Total TP}}{\text{Total TP} + \text{Total FN}}\right) \times 100\% \tag{9}$$

III. METODOLOGI PENELITIAN

A. *Sumber Data dan Variabel Penelitian*

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh Dinas Sumber Daya Air Provinsi DKI Jakarta berupa data tinggi muka air di pintu air Marina Ancol untuk periode 1 Januari 2021 – 31 Desember 2022 per jam untuk setiap harinya. Total observasi yang digunakan berjumlah 17.520 observasi. Metode *Decision Tree*, *Naïve Bayes*, dan *Radial Basis Function Neural Network* digunakan untuk mengetahui prediksi klasifikasi Tinggi Muka Air di laut Marina Ancol.

B. *Langkah Analitis*

Langkah-langkah analisis pada penelitian ini diantaranya yaitu sebagai berikut:

1. Pengkajian literatur dan pengumpulan data yang akan digunakan dalam penelitian;
2. Melakukan *preprocessing* data untuk mengatasi *missing value*;
3. Melakukan analisis statistika deskriptif untuk mendapatkan gambaran umum mengenai data yang akan digunakan pada penelitian;
4. Membagi data menjadi data *training* dan data *testing* dengan rincian 80% data *training* dan 20% data *testing*;
5. Melakukan pemodelan dengan metode *Decision Tree* dan mengevaluasi pada data *testing*;
6. Melakukan pemodelan dengan metode *Naïve Bayes* dan mengevaluasi model pada data *testing*;

7. Melakukan pemodelan dengan metode *Radial Basis Function Neural Network* dan mengevaluasi model pada data *testing*;
8. Pemilihan model terbaik berdasarkan parameter evaluasi;
9. Penarikan kesimpulan dan saran.

#### IV. HASIL DAN PEMBAHASAN

##### A. Preprocessing Data Tinggi Muka Air

Data Tinggi Muka Air yang digunakan pada penelitian ini didapatkan dari hasil pencatatan yang dilakukan oleh Dinas Sumber Daya Air (SDA) Provinsi DKI Jakarta untuk setiap jam. Proses *preprocessing* data digunakan untuk mengatasi data yang tidak terekam atau dapat dikatakan sebagai *missing value*. *Missing value* diatasi dengan melakukan imputasi berdasarkan data yang tersedia secara *online* pada situs Pusat Data dan Informasi Sumber Daya Air Dinas Sumber Daya Air Provinsi DKI Jakarta. Setelah itu, *missing value* yang juga tidak tercatat pada situs *online* diatasi dengan melakukan input data menggunakan data pada satu periode sebelumnya.

##### B. Analisis Deskriptif

Setelah dilakukan *preprocessing* untuk mengatasi *missing value* pada data, selanjutnya dilakukan analisis deskriptif dari data tinggi muka air yang digunakan. Analisis deskriptif tersebut disajikan pada Tabel 3.

Tabel 3 menunjukkan bahwa rata-rata tinggi muka air pada pintu Marina Ancol pada periode 2021-2022 adalah 183 cm dan termasuk ke dalam status Siaga 3. Tinggi air minimum yaitu 105 cm terjadi pada 1 November 2022 jam 04.00 WIB. Sedangkan tinggi air maksimum sebesar 262 cm terjadi pada 25 Desember 2022 jam 11.00 WIB. Rincian untuk jumlah data masing-masing kategori level air dapat dilihat pada Tabel 4.

##### C. Analisis Decision Tree

*Decision Tree* merupakan salah satu algoritma klasifikasi yang membuat model prediksi terhadap suatu keputusan menggunakan struktur hirarki atau pohon [9]. Algoritma ini membentuk suatu pohon keputusan melalui dua langkah yaitu perhitungan nilai *entropy* dan nilai *information gain*. Sebelum dilakukan analisis, dilakukan pembagian data menjadi data *training* dan data *testing* terlebih dahulu dengan proporsi 80% untuk data *training* dan 20% untuk data *testing*.

Perhitungan nilai entropi akan dilakukan pada data *traininig* melalui dua tahap yaitu nilai entropi keseluruhan data dan nilai entropi masing-masing variabel. Dengan menggunakan persamaan 1, didapatkan nilai entropi untuk keseluruhan data sebagai berikut [10]:

$$\begin{aligned} \text{Entropy}(S) &= \left( -\frac{29}{14016} \times \log_2 \left( \frac{29}{14016} \right) \right) \\ &+ \left( -\frac{3717}{14016} \times \log_2 \left( \frac{3717}{14016} \right) \right) \\ &+ \left( -\frac{5608}{14016} \times \log_2 \left( \frac{5608}{14016} \right) \right) \\ &+ \left( -\frac{4662}{14016} \times \log_2 \left( \frac{4662}{14016} \right) \right) \\ &= 1,583241 \end{aligned}$$

Nilai entropi yang sudah didapatkan ini akan digunakan dala perhitungan *information gain*. Selanjutnya, dilakukan

perhitungan nilai entropi dan *information gain* untuk masing-masing variabelnya. Diketahui bahwa variabel prediktor pada penelitian ini merupakan data numerik sehingga diperlukan suatu nilai pembagi yang ditentukan dengan bantuan *software Python* untuk mendapat nilai terbaik. Dengan menggunakan persamaan 1 dan 2, Tabel 5 menunjukkan hasil perhitungan dari nilai entropi dan *information gain* yang digunakan pada *node* pertama.

Tabel 5 menunjukkan bahwa nilai *information gain* terbesar dimiliki oleh variabel  $X_3$  yaitu sebesar 0,563597. Dikarenakan hal tersebut, variabel  $X_3$  akan menjadi *node* pertama atau biasa disebut dengan *root node*. Perhitungan pada *node* berikutnya juga dilakukan dengan langkah yang sama hingga terbentuk suatu struktur pohon keputusan. Struktur pohon yang didapatkan dari model *Decision Tree* ini dapat dilihat pada Gambar 3.

Struktur model *Decision Tree* pada Gambar 3 akan diinterpretasikan dengan menggunakan contoh data *testing* seperti pada Tabel 6.

Tabel 6 menunjukkan data observasi pertama pada data *testing*. Pada *node* pertama, yang akan diperhatikan adalah nilai dari variabel  $X_3$ . Variabel  $X_3$  bernilai 166 cm yang berarti kondisi ini memenuhi syarat pada *node* pertama yaitu bernilai lebih kecil dari 177 cm sehingga pada *node* selanjutnya akan berjalan ke arah kiri atau mengikuti jalan *YES* pada struktur pohon. Pada *node* kedua ini nilai dari variabel  $X_3$  tidak memenuhi syarat bernilai lebih kecil dari 163,5 cm sehingga akan dilanjutkan dengan berjalan ke arah kanan atau mengikuti jalan *NO* pada struktur pohon. Pada *node* ketiga ini, yang diperhatikan selanjutnya adalah nilai dari variabel  $X_1$ . Variabel  $X_1$  bernilai lebih besar dari 177,5 cm yang berarti tidak memenuhi syarat pada *node* ini sehingga akan berjalan ke kanan melalui jalur *NO* pada struktur pohon. Hal ini akan terus dilakukan hingga mencapai *leafnode* atau cabang terakhir yang akan menentukan di kelas manakah data pengamatan ini akan dikategorikan.

##### D. Analisis Naïve Bayes

*Naïve Bayes* merupakan suatu metode klasifikasi yang berdasarkan pada teorema *Bayes* dimana peluang terjadinya suatu peristiwa diperoleh berdasarkan informasi dari kejadian yang sudah terjadi. Algoritma ini membentuk model prediksi berdasarkan perhitungan nilai peluang prior dan nilai peluang posterior. Peluang prior dihitung menggunakan rumus pada persamaan 3 dengan rincian perhitungan seperti berikut ini. Peluang suatu data termasuk ke dalam kelas Siaga 1 adalah sebagai berikut:

$$p(\text{Siaga 1}) = \frac{29}{14016} = 0,00206$$

Selanjutnya, peluang dari suatu pengamatan termasuk ke dalam kelas Siaga 2 ditunjukkan sebagai berikut:

$$p(\text{Siaga 2}) = \frac{3717}{14016} = 0,2651$$

Peluang suatu pengamatan termasuk ke dalam kelas Siaga 3 adalah sebagai berikut:

$$p(\text{Siaga 3}) = \frac{5608}{14016} = 0,4001$$

Selanjutnya, peluang dari suatu pengamatan termasuk ke

dalam kelas Siaga 4 ditunjukkan sebagai berikut:

$$p(\text{Siaga 4}) = \frac{4662}{14016} = 0,3326$$

Nilai yang sudah didapatkan dari perhitungan peluang prior ini, akan digunakan pada tahap selanjutnya yaitu perhitungan peluang posterior.

Perhitungan peluang posterior ini akan dilakukan pada data *testing* dengan menggunakan rumus pada persamaan 4. Pada penelitian ini, variabel independen yang digunakan merupakan variabel numerik sehingga dibutuhkan beberapa perhitungan terlebih dahulu sebelum dapat dilakukan perhitungan peluang posterior. Nilai *mean* dan standar deviasi diperlukan agar dapat dilakukan perhitungan nilai *likelihood* yang nantinya akan digunakan pada perhitungan nilai peluang posterior. Tabel 7 menunjukkan hasil dari perhitungan nilai *mean* dan standar deviasi yang telah dilakukan pada variabel  $X_1$ ,  $X_2$ , dan  $X_3$ .

Nilai-nilai pada Tabel 7 akan digunakan dalam perhitungan nilai *likelihood*. Selanjutnya, dengan menggunakan contoh data *testing* yang ada pada Tabel 6, dilakukan perhitungan nilai *likelihood* pada setiap variabel. Tabel 8 menunjukkan hasil dari perhitungan nilai *likelihood* yang telah dilakukan pada variabel  $X_1$ ,  $X_2$ , dan  $X_3$ .

Selanjutnya, nilai-nilai *likelihood* yang terdapat pada Tabel 8 akan digunakan dalam perhitungan nilai peluang posterior. Peluang posterior didapatkan dengan menggunakan persamaan 4. Dengan:

$$P(x_1, x_2, x_3 | c_i) = p(x_1 | c_i) \times p(x_2 | c_i) \times p(x_3 | c_i)$$

maka:

$$P(\text{Siaga 1} | x) = p(x_1 | \text{Siaga 1}) \times p(x_2 | \text{Siaga 1}) \times p(x_3 | \text{Siaga 1}) \times p(\text{Siaga 1})$$

$$P(\text{Siaga 1} | x) = 0,0006 \times 1,3849E - 08 \times 3,0311E - 36 \times 0,00206 = 5,3441E - 50$$

Menggunakan langkah yang sama, perhitungan juga dilakukan untuk mendapatkan peluang dari Siaga 2, Siaga 3, dan Siaga 4 sebagai berikut:

$$P(\text{Siaga 2} | x) = p(x_1 | \text{Siaga 2}) \times p(x_2 | \text{Siaga 2}) \times p(x_3 | \text{Siaga 2}) \times p(\text{Siaga 2})$$

$$P(\text{Siaga 2} | x) = 0,0307 \times 0,0054 \times 0,0001 \times 0,2651 = 6,2669E - 09$$

$$P(\text{Siaga 3} | x) = p(x_1 | \text{Siaga 3}) \times p(x_2 | \text{Siaga 3}) \times p(x_3 | \text{Siaga 3}) \times p(\text{Siaga 3})$$

$$P(\text{Siaga 3} | x) = 0,0841 \times 0,0719 \times 0,0293 \times 0,4001 = 7,1121E - 05$$

$$P(\text{Siaga 4} | x) = p(x_1 | \text{Siaga 4}) \times p(x_2 | \text{Siaga 4}) \times p(x_3 | \text{Siaga 4}) \times p(\text{Siaga 4})$$

$$P(\text{Siaga 4} | x) = 0,0701 \times 0,0729 \times 0,0821 \times 0,3326 = 0,0001$$

Berdasarkan nilai akhir peluang posterior pada perhitungan peluang akhir, didapatkan bahwa kelas Siaga 4 memiliki nilai peluang yang terbesar. Oleh karena itu, contoh data *testing* pada Tabel 6 akan diklasifikasikan sebagai Siaga 4. Untuk

pengujian data *testing* lainnya akan dibantu dengan penggunaan *software Jupyter Notebook*.

Selanjutnya, akan dilakukan evaluasi model berdasarkan model yang sudah didapatkan dengan menggunakan persamaan 8 dan 9. Didapatkan nilai akurasi sebesar 73,37%, FNR Siaga 1 sebesar 16,67%, dan FNR siaga 2 sebesar 24,47% berdasarkan metode *Naïve Bayes*.

#### E. Analisis Radial Basis Function Neural Network

*Radial Basis Function Neural Network* (RBFNN) merupakan arsitektur dari *Artificial Neural Network* (ANN) yang bersifat *feed forward*. RBFNN menggunakan fungsi aktivasi *Gaussian* yang memperhitungkan jarak antara data dengan pusat data. Pada analisis ini jumlah *neuron* pada lapisan *input* berjumlah tiga buah dimana ini merepresentasikan jumlah variabel independen. Pada lapisan *output* terdapat empat buah *neuron* yang merepresentasikan jumlah kelas pada klasifikasi yaitu Siaga 1, Siaga 2, dan Siaga 3. Jumlah *hidden neuron* ditentukan melalui percobaan *trial error* dengan memperhatikan hasil percobaan dengan nilai akurasi data *testing* terbesar sebelum terjadinya *overfitting* pada data. Pada penelitian ini, stuktur yang digunakan 15 buah unit *hidden neuron* sehingga struktur yang digunakan adalah model RBFNN (3-15-4).

Setelah ditentukan rancangan arsitektur yang akan digunakan, selanjutnya akan dilakukan perhitungan untuk mendapatkan nilai bobot *input layer* terhadap *hidden layer*. Pada RBFNN, nilai bobot ini berupa nilai *centroid* dan juga nilai *spread* dari setiap *hidden unit*. Nilai bobot ini ditunjukkan pada Tabel 9.

Tabel 9 menunjukkan nilai bobot *input layer* menuju *hidden layer*. Bobot yang menghubungkan *input*  $X_1$  dengan *hidden* pertama adalah sebesar 154,91. Bobot yang menghubungkan *input*  $X_3$  dengan *hidden* kelimabelas adalah sebesar 189,391. Nilai bobot ini selanjutnya akan digunakan dalam perhitungan nilai aktivasi *gaussian*. Nilai aktivasi pada RBFNN didapatkan dengan menggunakan fungsi aktivasi *gaussian*. Nilai fungsi *gaussian* dari data *training* akan dihitung dengan menggunakan rumus yang ada pada persamaan (5). Nilai ini selanjutnya akan digunakan untuk membangun matriks *gaussian* dimana matriks ini akan digunakan untuk mendapatkan nilai bobot *hidden layer* terhadap *output layer*. Tabel 10 menampilkan rincian dari nilai bobot ini.

Tabel 10 menunjukan nilai bobot *hidden layer* menuju *output layer*. Bobot yang menghubungkan *hidden* pertama dengan *output* Siaga 1 adalah sebesar -0,018. Bobot yang menghubungkan *hidden* kelimabelas dengan *output* Siaga 4 adalah sebesar -0,402. Bobot ini selanjutnya akan digunakan dalam pembentukan model RBFNN dan juga perhitungan nilai *output* yang berupa nilai *pseudo probability*. Selanjutnya, dihitung nilai aktivasi pada data *testing* dengan menggunakan persamaan 5. Contoh rincian perhitungan untuk contoh data *testing* pada Tabel 6 adalah sebagai berikut. Nilai aktivasi pada data *testing* pengamatan pertama dan *hidden neuron* pertama adalah sebagai berikut:

$$\varphi_{1,1} = \exp \left( - \frac{|178 - 154,941|^2}{2 \times 7,532^2} + \frac{|172 - 153,642|^2}{2 \times 7,532^2} + \frac{|166 - 154,610|^2}{2 \times 7,532^2} \right)$$

$$\varphi_{1,15} = 1,507 \times 10^{-4}$$

Sedangkan nilai aktivasi pada data *testing* pengamatan pertama untuk *hidden neuron* kedua ditunjukkan sebagai berikut:

$$\varphi_{1,2} = \exp\left(-\frac{|178 - 140,956|^2}{2 \times 8,713^2} + \frac{|172 - 138,18|^2}{2 \times 8,713^2} + \frac{|166 - 140,574|^2}{2 \times 8,713^2}\right)$$

$$\varphi_{1,15} = 1,39 \times 10^{-9}$$

Nilai aktivasi pada data *testing* pengamatan pertama untuk *hidden neuron* kelimabelas ditunjukkan sebagai berikut:

$$\varphi_{1,15} = \exp\left(-\frac{|178 - 184,468|^2}{2 \times 4,662^2} + \frac{|172 - 187,281|^2}{2 \times 4,662^2} + \frac{|166 - 189,391|^2}{2 \times 4,662^2}\right)$$

$$\varphi_{1,15} = 6,05 \times 10^{-9}$$

Perhitungan nilai aktivasi ini juga dilakukan pada *hidden neuron* ketiga hingga keempatbelas dan untuk seluruh data *testing* dengan langkah yang sama. Setelah didapatkan nilai aktivasi gaussiannya, selanjutnya akan dilakukan perhitungan nilai *output*. Pada RBFNN dalam hal klasifikasi, *output* yang dihasilkan akan berupa suatu nilai *pseudo probability*, dimana nilai ini nantinya akan digunakan sebagai rujukan suatu data pengamatan akan termasuk ke dalam kelas yang mana. Model RBFNN yang didapatkan berdasar atas persamaan 7 adalah sebagai berikut:

$$y_{k,1} = 0,013 - 0,18\varphi_{k,1} - 0,19\varphi_{k,2} - 0,20\varphi_{k,3} - 0,16\varphi_{k,4} - 0,16\varphi_{k,5} - 0,19\varphi_{k,6} - 0,17\varphi_{k,7} - 0,19\varphi_{k,8} - 0,11\varphi_{k,9} - 0,13\varphi_{k,10} - 0,17\varphi_{k,11} - 0,15\varphi_{k,12} - 0,18\varphi_{k,13} - 0,18\varphi_{k,14} - 0,005\varphi_{k,15}$$

$$y_{k,2} = 0,306 - 0,42\varphi_{k,1} - 0,467\varphi_{k,2} - 0,474\varphi_{k,3} - 0,388\varphi_{k,4} - 0,344\varphi_{k,5} - 0,476\varphi_{k,6} - 0,429\varphi_{k,7} - 0,352\varphi_{k,8} - 0,346\varphi_{k,9} + 1,103\varphi_{k,10} - 0,24\varphi_{k,11} + 1,139\varphi_{k,12} + 0,977\varphi_{k,13} + 0,776\varphi_{k,14} - 0,469\varphi_{k,15}$$

$$y_{k,3} = 0,319 - 0,442\varphi_{k,1} - 0,474\varphi_{k,2} - 0,529\varphi_{k,3} + 0,899\varphi_{k,4} + 0,204\varphi_{k,5} + 0,884\varphi_{k,6} - 0,371\varphi_{k,7} + 0,988\varphi_{k,8} + 0,92\varphi_{k,9} - 0,488\varphi_{k,10} + 0,739\varphi_{k,11} - 0,704\varphi_{k,12} - 0,463\varphi_{k,13} - 0,281\varphi_{k,14} + 0,876\varphi_{k,15}$$

$$y_{k,4} = 0,362 + 0,88\varphi_{k,1} + 0,96\varphi_{k,2} + 1,023\varphi_{k,3} - 0,495\varphi_{k,4} + 0,156\varphi_{k,5} - 0,389\varphi_{k,6} + 0,817\varphi_{k,7} - 0,617\varphi_{k,8} - 0,562\varphi_{k,9} - 0,602\varphi_{k,10} - 0,481\varphi_{k,11} - 0,42\varphi_{k,12} - 0,496\varphi_{k,13} - 0,477\varphi_{k,14} - 0,402\varphi_{k,15}$$

Berdasarkan model tersebut, nilai dari *pseudo probability* dapat ditentukan dengan menggunakan nilai aktivasi *gaussian* yang sudah didapatkan sebelumnya. Rincian

perhitungan pada contoh data *testing* dapat dilihat sebagai berikut:

Dengan persamaan menggunakan persamaan 7 dan nilai fungsi aktivasi *gaussian* didefinisikan sebagai:

$$\varphi_{1,1} = \exp\left(-\frac{|x_{1,1}-154,941|^2}{2 \times 7,532^2} + \frac{|x_{1,2}-153,642|^2}{2 \times 7,532^2} + \frac{|x_{1,3}-154,610|^2}{2 \times 7,532^2}\right) + \vdots$$

$$\varphi_{1,15} = \exp\left(-\frac{|x_{1,1}-184,468|^2}{2 \times 4,662^2} + \frac{|x_{1,2}-187,281|^2}{2 \times 4,662^2} + \frac{|x_{1,3}-189,391|^2}{2 \times 4,662^2}\right)$$

Nilai *pseudo probability* dari contoh data *testing* adalah sebagai berikut:

$$y_{1,1} = 0,013 - 0,18 \times 1,5 \times 10^{-4} + \dots - 0,005 \times 6,05 \times 10^{-9} = 0,003$$

$$y_{1,2} = 0,306 - 0,42 \times 1,5 \times 10^{-4} + \dots - 0,469 \times 6,05 \times 10^{-9} = 0,065$$

$$y_{1,3} = 0,319 - 0,442 \times 1,5 \times 10^{-4} + \dots - 0,876 \times 6,05 \times 10^{-9} = 0,137$$

$$y_{1,4} = 0,3620,88 \times 1,5 \times 10^{-4} + \dots - 0,402 \times 6,05 \times 10^{-9} = 0,795$$

Berdasarkan hasil perhitungan pada nilai peluang akhir, nilai *pseudo probability* terbesar ditunjukkan pada  $y_{1,4}$  yang merepresentasikan kelas Siaga 4. Oleh karena itu, prediksi yang dihasilkan untuk data *testing* yang ditunjukkan pada Tabel 6 adalah termasuk ke dalam kelas Siaga 4.

Selanjutnya, akan dilakukan perhitungan nilai akurasi berdasarkan model yang sudah didapatkan dengan menggunakan persamaan 8 dan 9. Didapatkan nilai akurasi sebesar 86,42%, FNR Siaga 1 sebesar 16,67%, dan FNR siaga 2 sebesar 100% berdasarkan metode *Radial Basis Function Neural Network* adalah 14,73%.

#### F. Pemilihan Metode Terbaik

Pemilihan metode terbaik dalam memprediksikan kategori level air laut pada Marina Ancol ini akan didasarkan pada nilai akurasi, *false negative rate* Siaga 1, dan *false negative rate* Siaga 2 yang didapatkan dari masing-masing metode.. Hasil pengukuran parameter evaluasi dari ketiga metode *Decision Tree*, *Naïve Bayes*, dan *Radial Basis Function Neural Network* dapat dilihat pada Tabel 11.

Tabel 11 menunjukkan bahwa metode *Radial Basis Function Neural Network* menghasilkan nilai akurasi yang paling tinggi pada data *testing* yaitu sebesar 86,42%. Parameter *false negative rate* pada metode *Decision Tree* menghasilkan nilai yang lebih kecil jika dibandingkan metode lainnya yaitu masing-masing sebesar 16,67% pada kelas Siaga 1 dan 12,24% pada kelas Siaga 2. Pada kasus klasifikasi TMA ini, salah mengklasifikasikan status bahaya ataupun siaga sebagai keadaan yang normal lebih berbahaya jika dibandingkan sebaliknya. Hal ini dapat menyebabkan kurangnya persiapan bagi masyarakat dan pemerintah dalam menghadapi terjadinya luapan air laut akibat prediksi yang

salah tersebut. Untuk mencegah hal tersebut, meminimalkan *false negative rate* menjadi salah satu parameter yang penting pada penelitian ini. Berdasarkan ketiga nilai parameter evaluasi model pada Tabel 11, *Decision Tree* merupakan metode yang terbaik dalam memprediksikan kategori level air laut Marina Ancol jika dibandingkan dengan *Naive Bayes* dan *Radial Basis Function Neural Network*.

#### G. Prediksi Status Level Air untuk Duabelas Periode Selanjutnya

Hasil analisis yang telah dilakukan menunjukkan hasil bahwa *Decision Tree* merupakan metode yang terbaik dalam memprediksikan kategori level air laut Marina Ancol. Selanjutnya, akan dilakukan prediksi status level air untuk duabelas periode kedepan menggunakan model *Decision Tree* yang telah didapatkan. Hasil dari prediksi ini ditunjukkan pada Tabel 12.

Tabel 12 menunjukkan bahwa level air mencapai status Siaga pada tanggal 1 Januari 2023 pukul 06.00 WIB hingga pukul 09.00 WIB serta pada pukul 11.00 WIB. Selain itu, ketinggian air menunjukkan status Waspada pada pukul 03.00 WIB hingga pukul 05.00 WIB serta pada pukul 10.00 WIB. Status level air Normal berlangsung selama tiga jam yaitu pada pukul 00.00 WIB hingga pukul 02.00 WIB.

## V. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis yang telah dilakukan pada penelitian tugas akhir ini, didapatkan beberapa kesimpulan dan saran yang dijabarkan seperti berikut ini: (1) Metode *Decision Tree* dapat digunakan dalam memprediksikan kategori level air laut Marina Ancol dengan nilai akurasi sebesar 85,19%, nilai *false negative rate* Siaga 1 sebesar 16,67%, dan nilai *false negative rate* Siaga 2 sebesar 12,24%. (2) Metode *Naive Bayes* dapat digunakan dalam memprediksikan kategori level air laut Marina Ancol dengan nilai akurasi sebesar 73,37%, nilai *false negative rate* Siaga 1 sebesar 16,67%, dan nilai *false negative rate* Siaga 2 sebesar 24,27%. (3) Arsitektur model *Radial Basis Function Neural Network* yang terbaik ditinjau atas nilai akurasi terbesar pada data *testing* adalah arsitektur dengan jumlah *input neuron* sebanyak tiga, *hidden neuron* sebanyak lima belas, dan *output neuron* sebanyak empat. Nilai akurasi model RBFNN (3-15-4) menunjukkan nilai sebesar 86,64% nilai *false negative rate* Siaga 1 sebesar 100%, dan nilai *false negative rate* Siaga 2 sebesar 14,73%. (4) Dalam melakukan prediksi klasifikasi kategori level air laut Marina Ancol, ditinjau berdasarkan nilai akurasi model dan *false negative rate* dari setiap metode, *Decision Tree* merupakan metode yang menghasilkan model

lebih baik dibandingkan *Naive Bayes* dan RBFNN. Berdasarkan evaluasi kinerja klasifikasi, model *Decision Tree* mencapai nilai akurasi sebesar 85,19%, nilai *false negative rate* Siaga 1 sebesar 16,67%, dan nilai *false negative rate* Siaga 2 sebesar 12,24% sehingga model ini dikatakan baik dalam melakukan prediksi kategori level air laut Marina Ancol.

Saran yang dapat diberikan bagi masyarakat yaitu berdasarkan hasil prediksi yang sudah dilakukan, tidak disarankan untuk berkunjung ke pantai Marina Ancol pada kisaran pukul 06.00 WIB hingga pukul 09.00 WIB karena level air yang menunjukkan status Siaga 2 yaitu status siaga. Sedangkan terkait pengembangan penelitian selanjutnya, dapat mempertimbangkan faktor eksternal seperti pasang surut air laut dalam memprediksi tinggi muka air laut, serta pengembangan terkait pemilihan *input neuron* terbaik dikarenakan pada penelitian ini lebih difokuskan pada pencarian jumlah *hidden neuron* terbaik.

## DAFTAR PUSTAKA

- [1] W. A. Arifin, I. Ariawan, A. A. Rosalia, A. S. Sasongko, M. R. Apriansyah, and A. Satibi, "Model prediksi pasang surut air laut pada stasiun pushidrosal bakauheni lampung menggunakan support vector regression," *Jurnal Kemaritiman: Indonesian Journal of Maritime*, vol. 2, no. 2, pp. 139–148, Dec. 2021, doi: 10.17509/ijom.v2i2.35149.
- [2] R. C. Karana and R. D. Supriharjo, "Mitigasi bencana banjir rob di Jakarta Utara," *Jurnal Teknik POMITS*, vol. 2, no. 1, p. 25, 2013, doi: 10.12962/j23373539.v2i1.2465.
- [3] L. A. Hermanto, "Prakiraan tinggi gelombang air laut menggunakan data mining," *Jurnal IPTEK*, vol. 22, no. 1, p. 37, May 2018, doi: 10.31284/j.iptek.2018.v22i1.232.
- [4] A. B. Harti, D. H. Christianto, O. Pearl, R. Nabillah, and S. I. Kelautan, "Prediction of extreme sea water waves at ancil beach using ID3 algoritma algorithm," *Journal of Intelligent Decision Support System (IDSS)*, vol. 5, no. 2, pp. 64–72, Jun. 2022.
- [5] N. Nurdiana and A. Algifari, "Studi komparasi algoritma ID3 dan algoritma naive bayes untuk klasifikasi penyakit diabetes mellitus," *Infotech Journal*, vol. 6, no. 2, pp. 18–23, Dec. 2020.
- [6] V. Wahyuningrum, "Penerapan radial basis function neural network dalam pengklasifikasian daerah tertinggal di Indonesia," *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. 12, no. 1, pp. 37–54, Jun. 2020, doi: 10.34123/jurnalasks.v12i1.250.
- [7] D. Sartika and D. I. Sensuse, "Perbandingan algoritma klasifikasi naive bayes, nearest neighbour, dan decision tree pada studi kasus pengambilan keputusan pemilihan pola pakaian," *Jurnal Jatisi*, vol. 1, no. 2, pp. 151–161, Mar. 2017.
- [8] R. W. S. W. P. L. A. K. Krzysztow J. Cios, *Data Mining A Knowledge Discovery Approach*. Boston, MA: Springer US, 2007. doi: 10.1007/978-0-387-36795-8. ISBN: 978-0-387-33333-5.
- [9] B. P. Salmon, W. Kleynhans, C. P. Schwegmann, and J. C. Olivier, "Proper comparison among methods using a confusion matrix," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE, Jul. 2015, pp. 3057–3060. doi: 10.1109/IGARSS.2015.7326461.
- [10] F. Gorunescu, *Data Mining*, vol. 12. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. doi: 10.1007/978-3-642-19721-5. ISBN: 978-3-642-19720-8.