

Klasifikasi *Waiting Time for Pilot* di Pelabuhan Tanjung Perak Menggunakan Metode Regresi Logistik – *Synthetic Minority Oversampling Technique* (SMOTE)

Cindy Meilany Fitri Andriani dan Destri Susilaningrum
Departemen Statistika Bisnis, Institut Teknologi Sepuluh Nopember (ITS)
e-mail: destri_s@statistika.its.ac.id

Abstrak—WTP didefinisikan sebagai selisih waktu antara waktu pandu naik ke atas kapal (*Pilot on Board* (POB)) dengan waktu penetapan pelayanan kapal. WTP pada Pelabuhan Tanjung Perak memiliki standar kinerja pelayanan operasional sebesar 108 menit. Namun, pada Pelabuhan Tanjung Perak masih terdapat beberapa kapal yang mengalami WTP yang melebihi standar yang ditetapkan. Hal ini menyebabkan kerugian bagi pihak pengguna jasa pelabuhan dan penurunan kinerja pelayanan operasional pelabuhan. Oleh karena itu, dilakukan penelitian klasifikasi kapal mengalami WTP ≥ 108 menit untuk mengurangi kerugian yang terjadi. Metode yang digunakan adalah regresi logistik - *Synthetic Minority Oversampling Technique* (SMOTE) dalam meng-klasifikasikan WTP karena proporsi sampel kelas minoritas sebesar 12%, dimana proporsi tersebut kurang dari 35% dari jumlah data yang artinya dikategorikan sebagai data tidak seimbang. Hasil penelitian ini menunjukkan diperoleh faktor-faktor yang berpengaruh signifikan terhadap WTP yaitu waktu kedatangan kapal, cuaca, jenis kapal, dan gerak kapal. Model regresi logistik-SMOTE memiliki nilai *accuracy* 86,67%, *missclassification rate* 0,133, *sensitivity* 0,873, *specificity* 0,800, *precision* 0,980, *AUC* 0,933.

Kata Kunci— Regresi Logistik Biner, SMOTE, *Waiting Time for Pilot*.

I. PENDAHULUAN

KONDISI kepadatan lalu lintas kapal di Pelabuhan Tanjung Perak memicu banyaknya permintaan pelayanan pemanduan. Pemanduan adalah kegiatan pandu dalam membantu, memberikan saran, dan informasi kepada nakhoda tentang keadaan perairan setempat yang penting agar navigasi-pelayaran dapat dilaksanakan dengan selamat, tertib, dan lancar demi keselamatan kapal dan lingkungan [1]. Pemanduan di Indonesia menilai kinerja jasa pemanduannya menggunakan *Approaching Time* (AT), *Waiting Time for Pilot* (WTP), dan *zero accident*. WTP didefinisikan sebagai selisih waktu antara waktu pandu naik ke atas kapal (*Pilot on Board* (POB)) dengan waktu penetapan pelayanan kapal.

WTP pada Pelabuhan Tanjung Perak memiliki standar kinerja pelayanan operasional yang telah ditetapkan, yaitu sebesar 108 menit. Namun, pada Pelabuhan Tanjung Perak masih terdapat beberapa kapal yang mengalami WTP yang melebihi standar yang ditetapkan. Hal ini tentunya menyebabkan kerugian dan penurunan kinerja pelayanan operasional pelabuhan yang menimbulkan meningkatnya biaya logistik yang akan berdampak pada harga barang di pasaran.

Banyaknya kerugian karena WTP selama di pelabuhan ini maka perlu dilakukan penelitian mengenai klasifikasi kapal mengalami WTP ≥ 108 menit untuk mengurangi kerugian

yang terjadi dan untuk meningkatkan kinerja pelayanan operasional Pelabuhan Tanjung Perak dengan metode regresi logistik- *Synthetic Minority Oversampling Technique* (SMOTE). Metode klasifikasi dengan nilai *specificity*, *precision*, dan *AUC* yang tinggi diharapkan memberi kemudahan dalam mengklasifikasikan kapal, menjadi gambaran dan pertimbangan pengguna jasa pelabuhan sebelum melakukan permohonan pelayanan pemanduan, dan pelabuhan dapat melakukan evaluasi pada kapal yang memiliki WTP lebih dari standar.

SMOTE dapat menangani ketidakseimbangan jumlah sampel dalam suatu kelas dengan mereplikasi data minoritas yang hasilnya dikenal sebagai data sintesis. Pada data kategorik dan numerik akan menggunakan *Synthetic Minority Oversampling Technique-Nominal Continuous* (SMOTE-NC). SMOTE digunakan karena proporsi sampel kelas minoritas sebesar 12%, dimana proporsi tersebut kurang dari 35% dari jumlah data yang artinya dikategorikan sebagai data tidak seimbang [2].

Penelitian terdahulu tentang WTP pernah dilakukan oleh Widyawati dan Yuliantini (2019) tentang faktor yang mempengaruhi WTP di Terminal Jamrud menggunakan regresi linear berganda menunjukkan waktu kedatangan kapal dan cuaca berpengaruh positif dan signifikan terhadap WTP [3]. Penelitian lain yang dilakukan oleh Damastuti dan Aisiyah (2015) tentang simulasi penjadwalan kapal tambat untuk mencapai *zero* WTP di Dermaga Jamrud Utara dengan optimasi algoritma *Sequential Searching* menunjukkan jenis kapal berhubungan positif dengan WTP [4].

II. STUDI LITERATUR

A. Tabel Kontingensi

Tabel 1 menunjukkan tabel kontingensi atau yang sering disebut tabulasi silang (*crosstabulation* atau *cross classification*) adalah tabel yang berisi data jumlah atau frekuensi atau beberapa klasifikasi [5]. Metode *cross tabulation* dapat menjawab hubungan antara dua atau lebih variabel penelitian tetapi bukan hubungan sebab akibat.

B. Uji Independensi

Uji independensi dilakukan untuk mengetahui hubungan antara dua variabel [5]. Setiap variabel yang digunakan harus memenuhi syarat homogen (setiap sel pada objek pengamatan harus merupakan objek yang sama), *Mutually Exclusive* (level satu dengan level lainnya harus saling lepas atau independen), *Mutually Exhaustive* (dekomposisi secara lengkap sampai unit terkecil, semua nilai harus masuk dalam klasifikasi yang dilakukan), skala nominal dan ordinal.

Tabel 1.
Tabel Kontingensi

Variabel X	Variabel Y				Total Baris
	1	2	...	j	
1	n_{11}	n_{12}	...	n_{1j}	$n_{1.}$
2	n_{21}	n_{22}	...	n_{2j}	$n_{2.}$
⋮	⋮	⋮	...	⋮	⋮
i	n_{i1}	n_{i2}	...	n_{ij}	$n_{i.}$
Total Kolom	$n_{.1}$	$n_{.2}$...	$n_{.j}$	$n_{..}$

Keterangan:

$i = 1, 2, \dots, r$ dan $j = 1, 2, \dots, c$.

n_{ij} = Jumlah pengamatan variabel X kategori ke- i dan variabel Y kategori ke- j .

$n_{i.}$ = Jumlah pengamatan variabel X kategori ke- i .

$n_{.j}$ = Jumlah pengamatan variabel Y kategori ke- j dengan.

$n_{..}$ = Jumlah keseluruhan pengamatan.

Tabel 2.
Nilai Confusion Matrix

Hasil Observasi	Hasil Klasifikasi		Total
	Positive Class (Y=0)	Negative Class (Y=1)	
Positive Class (Y=0)	True Positive (TP)	False Negative (FN)	TP + FN
Negative Class (Y=1)	False Positive (FP)	True Negative (TN)	FP + TN
Total	TP + FP	FN + TN	TP + FP + FN + TN = n

Keterangan :

TP = Jumlah kapal yang sesuai standar dan pada kenyataannya kapal sesuai standar.

TN = Jumlah kapal yang tidak sesuai standar dan pada kenyataannya kapal tidak sesuai standar.

FP = Jumlah kapal yang sesuai standar dan pada kenyataannya kapal tidak sesuai standar.

FN = Jumlah kapal yang tidak sesuai standar dan pada kenyataannya kapal sesuai standar.

Hipotesis untuk pengujian independensi antara variabel dependen dan variabel independen dengan $i = 1, 2, \dots, I$ dan $j = 1, 2, \dots, J$ dapat dituliskan sebagai berikut:

$H_0: P_{ij} = P_i \times P_j$ (Tidak ada hubungan antara variabel).

$H_1: P_{ij} \neq P_i \times P_j$ (Terdapat hubungan antara variabel).

Apabila ditetapkan tingkat signifikansi sebesar α maka H_0 ditolak jika nilai $\chi^2 > \chi^2_{(\alpha, (I-1)(J-1))}$ atau nilai $P\text{-value} < \alpha$.

Statistik Uji:

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad (1)$$

dengan $e_{ij} = \frac{n_{i.} \times n_{.j}}{n_{..}}$.

Keterangan:

χ^2 = Nilai peubah acak yang distribusi sampelnya didekati oleh distribusi *chi square*.

n_{ij} = Jumlah pengamatan pada baris ke- i dan kolom ke- j .

e_{ij} = Nilai ekspektasi pada baris ke- i dan kolom ke- j .

$n_{i.}$ = Jumlah pengamatan pada baris ke- i .

$n_{.j}$ = Jumlah pengamatan pada kolom ke- j .

C. Regresi Logistik Biner

Analisis regresi logistik biner digunakan untuk menjelaskan hubungan antara variabel dependen berupa data dikotomik atau biner dengan variabel independen berupa data berskala interval dan kategorik. Model regresi logistiknya dinyatakan dengan Persamaan 2.

Tabel 3.
Variabel Penelitian

Variabel	Keterangan	Kategori	Satuan	Skala
Y	WTP	1: < 108 2: ≥ 108	Menit	Nominal
X_1	GT	-	Ton	Rasio
X_2	LOA	-	Meter	Rasio
X_3	Waktu Kedatangan Kapal	1 : Malam 2 : Siang	-	Nominal
X_4	Cuaca	1 : Hujan 2 : Cerah	-	Nominal
X_5	Perusahaan Pelayaran	1 : Dalam Negeri 2 : Luar Negeri	-	Nominal
X_6	Jenis Kapal	1 : Kapal Penumpang 2 : Kapal Kargo 3 : Kapal Curah 4 : Kapal Petikemas	-	Nominal
X_7	Gerak Kapal	1 : Keluar 2 : Masuk 3 : Pindah	-	Nominal

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_p}} \quad (2)$$

Keterangan:

$\pi(x)$ = Peluang kejadian sukses dengan nilai probabilitas $0 \leq \pi \leq 1$.

β_i = Nilai parameter dengan $i = 0, 1, \dots, k$.

x_j = Nilai parameter dengan $j = 1, 2, \dots, p$.

Model regresi pada Persamaan 2 dapat diuraikan dengan menggunakan transformasi logit dari $\pi(x)$ untuk mempermudah pendugaan nilai parameter regresi. Model $g(x)$ ditunjukkan pada Persamaan 3.

$$g(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (3)$$

1) Estimasi Parameter

Estimasi parameter regresi logistik biner menggunakan metode *Maximum Likelihood* dimana estimasi parameter β mensyaratkan data mengikuti distribusi Bernoulli [6]. Setiap pasangan pengamatan diasumsikan independen sehingga fungsi *likelihood* merupakan gabungan fungsi distribusi masing masing sebagai berikut.

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \quad (4)$$

Fungsi *likelihood* tersebut lebih mudah dimaksimumkan dalam bentuk log $l(\beta)$ dan dinyatakan dengan $L(\beta)$.

$$L(\beta) = \ln l(\beta)$$

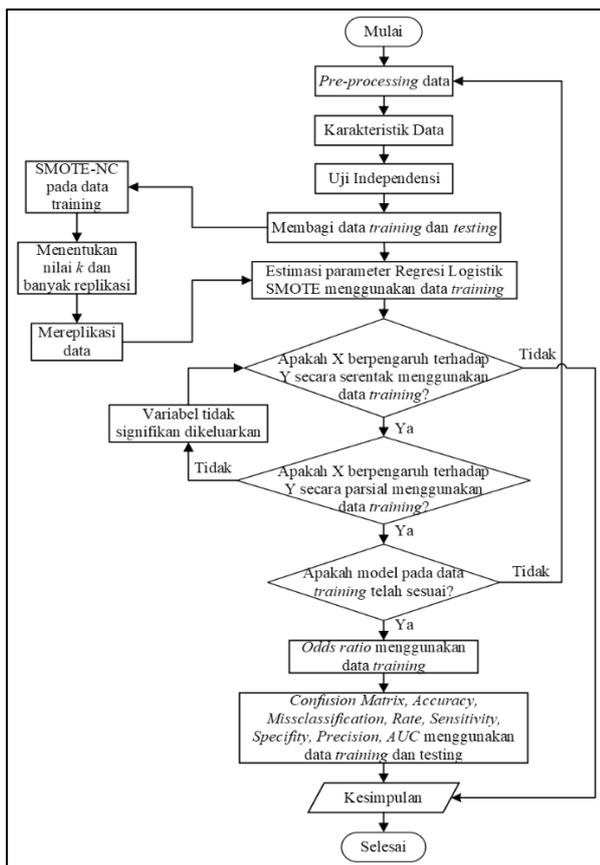
$$= \sum_{j=0}^p \left(\sum_{i=1}^n y_i x_{ij} \right) \beta_j - \sum_{i=1}^n \log \left(1 + e^{\sum_{j=0}^p \beta_j x_{ij}} \right) \quad (5)$$

a. Pengujian Estimasi Parameter Serentak

Pengujian signifikansi koefisien β secara serentak (multivariat) terhadap variabel dependen untuk menentukan variabel independen pada model berpengaruh signifikan secara bersama sama terhadap variabel dependen.

$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$ (Variabel yang diduga tidak berpengaruh signifikan secara bersama-sama terhadap variabel respon).

$H_1: \text{Minimal terdapat satu } \beta_l \neq 0; l = 1, 2, \dots, p$ (Minimal terdapat satu variabel yang diduga berpengaruh signifikan terhadap variabel respon).



Gambar 1. Diagram Alir.

Statistik Uji:

$$G = -2 \ln \frac{\binom{n_1}{n}^{n_1} \binom{n_0}{n}^{n_0}}{\sum_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{1-y_i}} \quad (6)$$

Keterangan:

- n_1 = Jumlah pengamatan berkategori $y = 1$.
- n_0 = Jumlah pengamatan berkategori $y = 0$.
- n = Jumlah pengamatan ($n_1 + n_0$).
- π_i = Rata rata taksiran peluang ke-i.

Statistik uji G adalah *Likelihood Ratio Test* dan mengikuti distribusi *Chi Squared* sehingga pada taraf signifikansi α maka H_0 ditolak jika $G > \chi^2_{(\alpha,df)}$ atau nilai *P-value* $< \alpha$. Derajat bebas merupakan banyaknya parameter pada model tanpa β_0 .

b. Pengujian Estimasi Parameter Parsial

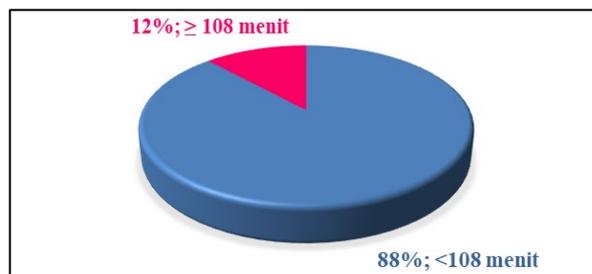
Pengujian estimasi parameter secara parsial digunakan untuk menguji pengaruh setiap koefisien β secara univariat dengan membandingkan parameter hasil *maximum likelihood* dugaan β dengan *standard error* parameter tersebut. Hasil pengujian akan menunjukkan variabel independen berpengaruh atau tidak terhadap variabel dependen dan layak masuk ke dalam model regresi atau tidak [6]. Hipotesis pengujian yang digunakan adalah sebagai berikut:

- $H_0: \beta_l = 0$.
- $H_1: \beta_l \neq 0; l = 1, 2, \dots, p$.

Statistik Uji:

$$W = \frac{\hat{\beta}_l}{SE(\hat{\beta}_l)} \quad (7)$$

Statistik uji W tersebut disebut sebagai statistik uji *Wald* dan mengikuti distribusi normal sehingga pada taraf signifikansi α maka Tolak H_0 jika $|Whitung| > Z_{\alpha/2}$ atau



Gambar 2. Pie Chart WTP.

Tabel 4. Karakteristik Waktu Kedatangan Kapal dan WTP

Variabel	Kategori	WTP		Total
		<108	≥108	
Waktu	Malam	93	20	113
		31,31%	6,73%	38,05%
Kedatangan Kapal	Siang	168	16	184
		56,57%	5,39%	61,95%
Total		261	36	297
		87,88%	12,12%	100,0%

nilai *P-value* $< \alpha$ (0,05). Statistik uji *Wald* dengan $SE(\hat{\beta}_l)$ merupakan taksiran *standar error* dari parameter.

2) Uji Kesesuaian Model

Pengujian kesesuaian model dilakukan untuk menguji apakah model regresi logistik yang dihasilkan sudah layak atau tidak terdapat perbedaan antara hasil pengamatan dan kemungkinan hasil prediksi. Hipotesis pengujian kesesuaian model adalah sebagai berikut:

- H_0 : Model sesuai (tidak terdapat perbedaan signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi).
 - H_1 : Model tidak sesuai (terdapat perbedaan signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi).
- Statistik uji:

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)} \quad (8)$$

Keterangan:

- O_k = Observasi grup ke- k ($\sum_{j=1}^{C_k} y_j$) dengan C_k respon (0,1).
- π_k = Rata-rata taksiran peluang ($\sum_{j=1}^{C_k} \frac{(m_j \hat{\pi}_j)^2}{n_k}$).
- g = Jumlah grup (kombinasi kategori dalam model serentak).
- n'_k = Banyak observasi pada grup ke- k .

Statistik uji \hat{C} mengikuti distribusi *Chi-Square* sehingga diperoleh keputusan tolak H_0 jika nilai $\hat{C} > \chi^2_{(\alpha,df)}$ atau nilai *P-value* $> \alpha$. Derajat bebas merupakan $g-2$ (g : jumlah kategori pada variabel independen). Saat gagal tolak H_0 maka model telah sesuai [6].

3) Odds Ratio

Odds ratio digunakan untuk menentukan kecenderungan hubungan fungsional lebih jauh antara variabel dependen dan variabel independen serta pengaruh perubahan nilai pada variabel. Keputusan tidak terdapat hubungan antara variabel independen dengan variabel dependen yaitu jika nilai *odds ratio* = 1. Jika nilai *odds ratio* < 1 maka hubungan variabel independen dengan variabel dependen memiliki hubungan negatif. Jika nilai *odds ratio* > 1 maka hubungan variabel independen dengan variabel dependen memiliki hubungan positif [5],

$$OR = e^{\beta_k} \quad (9)$$

Tabel 5.
Karakteristik Cuaca dan WTP

Variabel	Kategori	WTP		Total
		<108	≥108	
Cuaca	Cerah	165	10	175
		55,56%	3,37%	58,92%
	Hujan	96	26	122
		32,32%	8,75%	41,08%
Total	261	36	297	
		74,6%	25,4%	100,0%

Tabel 6.
Karakteristik Perusahaan Pelayaran dan WTP

Variabel	Kategori	WTP		Total
		<108	≥108	
Perusahaan Pelayaran	Dalam Negeri	216	26	242
		72,73%	8,75%	81,48%
	Luar Negeri	45	10	55
		15,15%	3,37%	18,52%
Total	261	36	297	
		87,88%	12,12%	100,0%

dengan β_k = Nilai parameter dengan $i = 0, 1, \dots, k$

D. Synthetic Minority Over-Sampling Technique (SMOTE)

SMOTE dapat menangani ketidakseimbangan jumlah sampel dalam suatu kelas dengan mereplikasi data kelas minoritas yang hasilnya dikenal sebagai data sintetis. Jika proporsi sampel kelas minoritas kurang dari 35% dari jumlah data, maka dikategorikan sebagai data tidak seimbang. Pada data numerik, SMOTE akan bekerja dengan mencari *k-nearest neighbors* (data tetangga terdekat sebanyak k) untuk setiap data di kelas minoritas menggunakan jarak *Euclidean* [2]. Misalkan terdapat data sebanyak q variabel yaitu $X^T = [X_1, X_2, \dots, X_q]$ dan $Z^T = [Z_1, Z_2, \dots, Z_q]$, maka jarak *Euclidean* $d(X, Z)$ yaitu:

$$d(X, Z) = \sqrt{(X_1 - Z_1)^2 + (X_2 - Z_2)^2 + \dots + (X_q - Z_q)^2} \tag{10}$$

Data sintetis akan dibangkitkan menggunakan persamaan sebagai berikut:

$$x_{syn} = x_m + (x_{knn} - x_m)\gamma \tag{11}$$

Keterangan:

x_{syn} = Data sintetis.

x_m = Data ke- i dari kelas minoritas.

x_{knn} = Data dari kelas minoritas yang memiliki jarak terdekat.

γ = Bilangan random antara nilai 0 dan 1.

SMOTE-NC merupakan perluasan dari metode SMOTE yang digunakan untuk menangani kumpulan data yang bersifat campuran nominal dan kontinyu. Algoritma SMOTE-NC dijelaskan sebagai berikut:

a. Perhitungan Median

Menghitung nilai median dari standar deviasi semua variabel kontinyu untuk kelas minoritas. Jika variabel nominal berbeda antara sampel dan nilai kemungkinan tetangga terdekatnya, maka median ini termasuk dalam jarak *Euclidean*.

b. Perhitungan *k Nearest Neighbours*

Menghitung jarak *euclidean* antara vektor variabel dimana k tetangga terdekat (sampel kelas minoritas) dengan variabel vektor lainnya (sampel kelas minoritas) diidentifikasi menggunakan variabel kontinyu. Setiap variabel nominal

Tabel 7.
Karakteristik Jenis Kapal dan WTP

Variabel	Kategori	WTP		Total
		<108	≥108	
Jenis Kapal	Kapal Kargo	78	14	92
		26,26%	4,71%	30,98%
	Kapal Curah	70	13	83
		23,57%	4,38%	27,95%
Kapal Penumpang	30	1	31	
	10,10%	0,34%	10,44%	
	Kapal Petikemas	83	8	91
27,95%		2,69%	30,64%	
Total	261	36	297	
		87,88%	12,12%	100,0%

yang berbeda antara variabel vektor dengan nilai tetangga terdekatnya, maka nilai median dari standar deviasi yang dihitung sebelumnya masuk dalam perhitungan jarak *euclidean*.

c. Membuat *Synthetic* Sampel

Variabel kontinyu pada data sintetis untuk kelas minoritas dibuat menggunakan pendekatan SMOTE seperti pada sub bab sebelumnya. Variabel nominal pada data sintetis merupakan nilai modus di sebagian besar k tetangga terdekat.

E. Pengukuran Performa Model Klasifikasi

Pengukuran performa model klasifikasi dapat menggunakan *Confusion Matrix* dan kurva *Receiver Operating Characteristics* (ROC). *Confusion Matrix* merupakan salah satu evaluasi prosedur klasifikasi untuk mengukur kinerja algoritma model klasifikasi [7]. Penentuan klasifikasi dapat diketahui melalui Tabel 2. Pengukuran performa model klasifikasi perlu dilakukan pengukuran menggunakan beberapa indikator yang dihasilkan dari *confusion matrix*, yaitu:

- Accuracy* = $[TP+TN / n]$ merupakan persentase model klasifikasi melakukan prediksi secara benar.
- Misclassification Rate* = $[FN+FP / n]$ merupakan persentase model klasifikasi melakukan kesalahan prediksi.
- Sensitivity* = $[TP / TP+FN]$ merupakan persentase kapal diprediksi sesuai standar dibanding seluruh kapal yang sebenarnya sesuai standar.
- Specificity* = $[TN / FP+TN]$ merupakan persentase kapal diprediksi tidak sesuai standar dibanding seluruh kapal yang sebenarnya tidak sesuai standar.
- Precision* = $[TP / FP+TP]$ merupakan persentase kapal yang sebenarnya sesuai standar dibanding seluruh kapal yang diprediksi sesuai standar.

Nilai *accuracy*, *sensitivity*, *specificity*, dan *precision* semakin baik jika mendekati 100%. Pada penelitian ini dengan kasus *imbalanced data* dipertimbangkan nilai *specificity* yang tinggi karena tidak ingin terjadinya *false positive* atau kapal yang diprediksi sesuai standar tetapi sebenarnya tidak sesuai. Nilai pengukuran dalam *confusion matrix* dapat ditafsirkan melalui kurva *Receiver Operation Characteristics* (ROC) yang menunjukkan hubungan antara *sensitivity* (*true positive rate*) dan $1-\textit{specificity}$ (*false positive rate*) pada sumbu X. Performa klasifikasi diukur dari luas

Tabel 8.
Karakteristik Gerak Kapal dan WTP

		WTP		Total
		<108	≥108	
Gerak Kapal	Keluar	116	8	124
		39,06%	2,69%	41,75%
	Masuk	75	27	102
		25,25%	9,09%	34,34%
	Pindah	70	1	71
		23,57%	0,34%	23,91%
Total		261	36	297
		87,88%	12,12%	100,0%

Tabel 9.
Karakteristik GT, LOA, dan WTP

Variabel	Karakteristik	WTP	
		<108	≥108
GT	Minimum	648	650
	Maksimum	43.256	36.259
	Rata-rata	8.195,79	10.306,08
	Standar Deviasi	9.172,16	9.974,32
	Minimum	50	30
LOA	Maksimum	261	206
	Rata-rata	115,12	122,01
	Standar Deviasi	41,56	45,93

area dibawah kurva ROC atau *Area Under Curve* (AUC). Kriteria interpretasi nilai AUC adalah sebagai berikut : >50% – 60% (sangat lemah), >60% – 70% (lemah), >70% – 80% (sedang), >80% – 90% (baik), >90% – 100% (sangat baik).

III. METODOLOGI PENELITIAN

A. Metode Pengumpulan Data

Data penelitian ini berupa data sekunder di PT Pelabuhan Indonesia Regional Jawa Timur yang menggunakan jasa pandu pada tanggal 1 Maret 2021 – 31 Maret 2021 di Pelabuhan Tanjung Perak. Jumlah data kapal menggunakan jasa pandu ada sebanyak 446 kapal. Dilakukan penghapusan data *double* sehingga menjadi 297 kapal. Data terbagi menjadi 80% (237 kapal) sebagai data *training* dan 20% (60 kapal) sebagai data *testing*. Pembagian data menjadi data *training* dan data *testing* diambil secara acak menggunakan *software R*.

B. Variabel Penelitian

Variabel penelitian yang digunakan dapat disajikan pada Tabel 3.

C. Metode Penelitian

Metode yang digunakan untuk mengklasifikasi WTP adalah regresi logistik biner karena variabel dependen (y) terdiri dari 2 kategori dan berskala nominal dengan variabel independen berupa data berskala rasio dan nominal dan regresi logistik-SMOTE karena proporsi sampel kelas minoritas sebesar 12%, yang kurang dari 35% dari jumlah data yang artinya dikategorikan sebagai data tidak seimbang.

D. Diagram Alir

Diagram alir pada penelitian ini dapat ditunjukkan oleh Gambar 1.

Tabel 10.
Uji Independensi

Variabel	χ^2	db	$\chi^2_{(0,05,db)}$	P-Value	Keputusan
Waktu Kedatangan Kapal	6,218	1	3,841	0,013	H ₀ ditolak
Cuaca	45,556	1	3,841	0,000	H ₀ ditolak
Perusahaan Pelayaran	0,069	1	3,841	0,794	H ₀ gagal ditolak
Jenis Kapal	18,423	3	7,815	0,000	H ₀ ditolak
Gerakan Kapal	69,195	2	5,991	0,000	H ₀ ditolak

Tabel 11.
Estimasi Parameter Logistik-SMOTE

Variabel	Keterangan	β
Constant	-	-2,612
GT (X ₁)	-	0,000
LOA (X ₂)	-	0,000
Waktu Kedatangan Kapal (X ₃)	2 = Siang	-0,960
Cuaca (X ₄)	2 = Cerah	-2,337
Perusahaan Pelayaran (X ₅)	2 = Luar Negeri	-0,806
Jenis Kapal (X ₆)	2 = Kapal Kargo	3,548
	3 = Kapal Curah	2,690
Gerak Kapal (X ₇)	4 = Kapal Petikemas	1,780
	2 = Masuk	2,095
	3 = Pindah	-2,890

IV. HASIL DAN PEMBAHASAN

A. Karakteristik WTP di Pelabuhan Tanjung Perak

Jumlah kapal di Pelabuhan Tanjung Perak yang menggunakan jasa pandu pada bulan Maret 2021 sebanyak 297 kapal, diantaranya terdapat 12% (36 kapal) dengan WTP ≥108 menit dan 88% (261 kapal) dengan WTP <108 menit. Persentase jumlah kapal disajikan pada Gambar 2.

1) Karakteristik Waktu Kedatangan Kapal dan WTP

Tabel 4 menunjukkan kapal dengan WTP ≥108 menit mayoritas waktu kedatangan kapalnya malam hari, kapal dengan WTP <108 menit mayoritas waktu kedatangan kapalnya siang hari.

2) Karakteristik Cuaca dan WTP

Tabel 5 menunjukkan kapal dengan WTP ≥108 menit mayoritas cuacanya adalah hujan, sedangkan kapal dengan WTP <108 menit mayoritas cuacanya adalah cerah.

3) Karakteristik Perusahaan Pelayaran dan WTP

Tabel 6 menunjukkan kapal dengan WTP ≥108 dan <108 menit mayoritas perusahaan pelayaran dalam negeri.

4) Karakteristik Jenis Kapal dan WTP

Tabel 7 menunjukkan bahwa persentase jenis kapal tertinggi yaitu kapal kargo sebesar 30,98% dan terendah kapal penumpang sebesar 10,44%. Kapal dengan WTP ≥108 mayoritas adalah kapal kargo, sedangkan kapal dengan WTP <108 mayoritas adalah kapal petikemas.

5) Karakteristik Gerak Kapal dan WTP

Tabel 8 menunjukkan bahwa persentase gerak kapal tertinggi yaitu saat keluar sebesar 41,75% dan terendah saat pindah sebesar 23,91%. Kapal dengan WTP ≥108 mayoritas adalah gerakannya adalah masuk, sedangkan kapal dengan WTP <108 mayoritas gerakannya adalah keluar.

Tabel 12.

Uji Serentak Regresi Logistik-SMOTE 7 Variabel	
G	$\chi^2_{(0,05;10)}$
163,996	18,307

Tabel 13.

Uji Parsial Regresi Logistik-SMOTE 7 Variabel				
Variabel	Kategori	β	$ W_{hitung} $	P-Value
Constant	-	2,612	-1,643	0,100
GT (X ₁)	-	0,000	1,011	0,312
LOA (X ₂)	-	0,000	-0,356	0,722
Waktu	-	-	-	-
Kedatangan Kapal (X ₃)	2 : Siang	0,960	-2,775	0,006
Cuaca (X ₄)	2 : Cerah	2,337	-6,210	0,000
Perusahaan Pelayaran (X ₅)	2 : Luar Negeri	0,806	-1,372	0,170
Jenis Kapal (X ₆)	2 : Kapal Kargo	3,548	2,891	0,004
	3 : Kapal Curah	2,690	2,222	0,026
	4 : Kapal Petikemas	1,780	1,483	0,138
Gerak Kapal (X ₇)	2 : Masuk	2,095	5,417	0,000
	3 : Pindah	2,890	-2,667	0,008

Tabel 14.

Uji Serentak Regresi Logistik-SMOTE 4 Variabel	
G	$\chi^2_{(0,05;7)}$
161,252	16,919

Tabel 15.

Uji Parsial Regresi Logistik-SMOTE 4 Variabel				
Variabel	Kategori	β	$ W_{hitung} $	P-Value
Constant	-	2,576	-2,187	0,029
Kedatangan Kapal (X ₃)	2 : Siang	0,922	-2,702	0,007
Cuaca (X ₄)	2 : Cerah	2,348	-6,316	0,000
	2 : Kapal Kargo	3,238	2,756	0,006
Jenis Kapal (X ₆)	3 : Kapal Curah	2,391	2,055	0,040
	4 : Kapal Petikemas	1,510	1,283	0,200
Gerak Kapal (X ₇)	2 : Masuk	2,164	5,612	0,000
	3 : Pindah	2,854	-2,636	0,008

6) Karakteristik GT, LOA, dan WTP

Tabel 9 menunjukkan rata-rata GT tertinggi terdapat pada kapal dengan WTP ≥108 menit yaitu sebesar 10.306,08 ton, maksimum GT tertinggi terdapat pada kapal dengan WTP <108 menit yaitu sebesar 43.256 ton. Sedangkan rata-rata LOA tertinggi terdapat pada kapal dengan WTP ≥108 menit yaitu sebesar 122,01 meter, maksimum LOA tertinggi terdapat pada kapal dengan WTP <108 menit yaitu sebesar 261 meter.

B. Uji Independensi

Uji Independensi dilakukan untuk mengetahui hubungan antara variabel independen. Hipotesis uji independensi adalah sebagai berikut.

H₀: $P_{ij} = P_i \times P_j$ (Tidak ada hubungan antara variabel).

H₁: $P_{ij} \neq P_i \times P_j$ (Terdapat hubungan antara variabel).

Pada taraf signifikansi $\alpha=0,05$ maka H₀ ditolak jika nilai $\chi^2 > \chi^2_{(\alpha,(I-1)(J-1))}$ atau nilai P-value < α . Hasil analisis uji independensi disajikan pada Tabel 10.

Tabel 10 menunjukkan bahwa terdapat 3 variabel yang memiliki nilai $\chi^2 > \chi^2_{(\alpha,(I-1)(J-1))}$ atau nilai P-value kurang dari $\alpha = 0,05$ sehingga H₀ ditolak yang berarti waktu kedatangan kapal, cuaca, jenis kapal, dan gerakan kapal berhubungan dengan WTP. Sedangkan perusahaan pelayaran tidak berhubungan dengan WTP.

C. Klasifikasi WTP di Pelabuhan Tanjung Perak Menggunakan Regresi Logistik-SMOTE

Data training memiliki jumlah kelas tidak seimbang antara kapal dengan WTP <108 menit sebanyak 87% (206 kapal) dan kapal dengan WTP ≥108 menit 13% (31 kapal). Salah satu cara mengatasi ketidakseimbangan jumlah data ialah menggunakan SMOTE-NC. Hasil model terbaik yaitu saat k=4 dengan persentase replikasi kelas minoritas sebanyak 50% dari kelas mayoritas. Sehingga, proporsi kelas setelah dilakukan SMOTE-NC pada WTP <108 menit sebanyak

67% (206 kapal) dan kapal dengan WTP lebih dari atau sama dengan 108 menit sebanyak 33% (103 kapal).

1) Estimasi Parameter Model Regresi Logistik-SMOTE

Estimasi parameter pada regresi logistik-SMOTE bertujuan mengetahui variabel yang berpengaruh signifikan terhadap WTP melalui pengujian parameter serentak dan pengujian parameter parsial. Estimasi parameter disajikan pada Tabel 11.

a. Pengujian Estimasi Parameter Serentak Model Regresi Logistik-SMOTE

Pengujian estimasi parameter secara serentak terhadap WTP bertujuan mengetahui pengaruh 7 variabel independen secara bersama-sama terhadap WTP.

H₀: $\beta_1 = \beta_2 = \dots = \beta_{10} = 0$ (Variabel tidak berpengaruh signifikan secara bersama-sama terhadap WTP).

H₁: Minimal terdapat satu $\beta_p \neq 0; p = 1, 2, \dots, 10$ (Minimal terdapat satu variabel berpengaruh signifikan terhadap WTP).

Pada taraf signifikansi $\alpha = 0,05$ maka H₀ ditolak jika $G > \chi^2_{(\alpha,db)}$ atau nilai P-value < α . Hasil analisis uji serentak disajikan pada Tabel 12. Hasil pengujian estimasi parameter serentak menunjukkan nilai G (163,996) lebih besar dari nilai $\chi^2_{(0,05;10)}$ (18,307) sehingga H₀ ditolak yang berarti minimal terdapat satu variabel yang diduga berpengaruh signifikan terhadap WTP.

b. Pengujian Estimasi Parameter Parsial Model Regresi Logistik-SMOTE

Pengujian estimasi parameter secara parsial digunakan untuk menguji setiap variabel independen terhadap WTP dan layak masuk ke dalam model regresi atau tidak.

H₀: $\beta_p = 0$ (Variabel tidak berpengaruh terhadap WTP).

H₁: $\beta_p \neq 0$ (Variabel berpengaruh terhadap WTP).

Pada taraf signifikansi $\alpha = 0,05$ maka H₀ ditolak jika $|W_{hitung}| > Z_{\alpha/2}$ atau nilai P-value < α . Hasil uji parsial regresi logistik-SMOTE 7 variabel disajikan oleh Tabel 13. Tabel 13 menunjukkan variabel waktu kedatangan kapal, cuaca, jenis kapal, dan gerak kapal memiliki nilai $|W_{hitung}| > Z_{\alpha/2}$ (0,96) atau nilai P-value lebih kecil dari nilai α (0,05) maka H₀ ditolak yang berarti variabel tersebut berpengaruh signifikan terhadap WTP. Variabel signifikan berhubungan dengan

Tabel 16.

Uji Kesesuaian Model Regresi Logistik-SMOTE		
χ^2	$\chi^2_{(0,05;8)}$	P-value
13,170	15,507	0,106

Tabel 17.

Odds Ratio Regresi Logistik-SMOTE			
Variabel	Keterangan	β	e^β
Waktu Kedatangan Kapal (X_3)	2 = Siang	-1,110	0,330
Cuaca (X_4)	2 = Cerah	-1,834	0,160
Gerak Kapal (X_7)	2 = Masuk	1,575	4,832
	3 = Pindah	-1,519	0,219

Tabel 18.

Confussion Matrix Regresi Logistik-SMOTE							
Observasi	Prediksi						
	Data Training			Data Testing			
	1	1	Total	1	2	Total	
1	186	186	206	48	7	55	
2	33	33	103	1	4	5	
Total	219	219	309	49	11	60	
Accuracy			82,85	Accuracy			86,67
Missclassification Rate			0,172	Missclassification Rate			0,133
Sensitivity			0,903	Sensitivity			0,873
Specificity			0,680	Specificity			0,800
Precision			0,850	Precision			0,980

WTP sesuai hasil uji independensi Tabel 10. Oleh karena itu, dilakukan estimasi parameter kembali dengan hanya menggunakan variabel yang signifikan.

c. Pengujian Estimasi Parameter Serentak Model Regresi Logistik-SMOTE

Pengujian estimasi parameter secara serentak terhadap WTP bertujuan mengetahui pengaruh 4 variabel independen secara bersama-sama terhadap WTP.

H_0 : $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ (Variabel tidak berpengaruh signifikan secara bersama-sama terhadap WTP).

H_1 : Minimal terdapat satu $\beta_p \neq 0; p = 1,2,3,4$ (Minimal terdapat satu variabel berpengaruh signifikan terhadap WTP).

Pada taraf signifikansi $\alpha = 0,05$ maka H_0 ditolak jika $G > \chi^2_{(\alpha,db)}$ atau nilai P-value < α . Hasil analisis uji serentak disajikan pada Tabel 14. Hasil pengujian estimasi parameter serentak menunjukkan nilai G (161,252) lebih besar dari nilai $\chi^2_{(0,05;7)}$ (16,919) sehingga H_0 ditolak yang berarti minimal terdapat satu variabel yang diduga berpengaruh signifikan terhadap WTP.

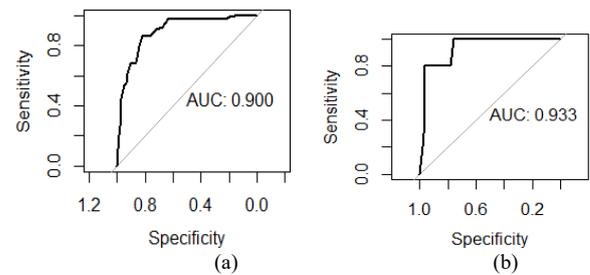
d. Pengujian Estimasi Parameter Parsial Model Regresi Logistik-SMOTE

Pengujian estimasi parameter secara parsial digunakan untuk menguji setiap variabel independen terhadap WTP.

H_0 : $\beta_p = 0$ (Variabel tidak berpengaruh terhadap WTP).

H_1 : $\beta_p \neq 0$ (Variabel berpengaruh terhadap WTP).

Pada taraf signifikansi $\alpha = 0,05$ maka H_0 ditolak jika $|W_{hitung}| > Z_{\alpha/2}$ atau nilai P-value < α . Hasil uji parsial regresi logistik-SMOTE 4 variabel ditunjukkan oleh Tabel 15. Tabel 15 menunjukkan variabel waktu kedatangan kapal, cuaca, jenis kapal, dan gerak kapal memiliki nilai $|W_{hitung}| > Z_{\alpha/2}$ (0,96) atau nilai P-value lebih kecil dari nilai α (0,05) maka H_0 ditolak yang berarti variabel tersebut berpengaruh signifikan terhadap WTP. Model menggunakan regresi logistik-SMOTE untuk prediksi klasifikasi WTP di Pelabuhan Tanjung Perak adalah sebagai berikut.



Gambar 3. Kurva ROC Regresi Logistik-SMOTE (a) Data Training. (b) Data Testing.

$$\hat{g}(x) = -2,576 - 0,922X_3(2) - 2,348X_4(2 + 3,238X_6(2) + 2,391X_6(3) + 1,510X_6(4) + 2,164X_7(2) - 2,854X_7(3) + 2,164X_7(2) - 2,854X_7(3)$$

2) Uji Kesesuaian Model Regresi Logistik-SMOTE

Uji kesesuaian model dilakukan untuk menguji apakah model WTP sudah layak.

H_0 : Model sesuai (tidak terdapat perbedaan signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi).

H_1 : Model tidak sesuai (terdapat perbedaan signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi).

Pada taraf signifikansi $\alpha = 0,05$ maka H_0 ditolak jika $\chi^2 > \chi^2_{(\alpha,db)}$ atau nilai P-value < α (0,05). Hasil analisis uji kesesuaian model menggunakan disajikan pada Tabel 16. Tabel 16 menunjukkan nilai χ^2 sebesar 13,170 lebih kecil dari nilai $\chi^2_{(0,05;8)}$ sebesar 15,507 atau nilai P-value 0,106 lebih besar dari α sebesar 0,05 sehingga gagal tolak H_0 yang berarti model sesuai atau tidak terdapat perbedaan signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi.

3) Odds Ratio Model Regresi Logistik-SMOTE

Odds ratio untuk menentukan kecenderungan hubungan fungsional lebih jauh antara WTP dengan waktu kedatangan kapal, cuaca, jenis kapal, dan gerak kapal. Odds ratio disajikan pada Tabel 17. Penjelasan nilai odds ratio adalah sebagai berikut.

a. Waktu kedatangan kapal (X_3)

Kapal yang dijadwalkan datang pada siang hari cenderung memiliki WTP ≥ 108 menit 0,398 lebih kecil dibandingkan kapal yang dijadwalkan datang pada saat malam hari.

b. Cuaca (X_4)

Kapal yang beroperasi saat cuaca cerah cenderung memiliki WTP ≥ 108 menit 0,096 lebih kecil dibandingkan kapal yang beroperasi saat cuaca hujan.

c. Jenis Kapal (X_6)

Kapal dengan jenis kapal kargo cenderung memiliki WTP ≥ 108 menit 25,476 kali lebih besar dibandingkan kapal dengan jenis kapal penumpang. Kapal dengan jenis kapal curah cenderung memiliki WTP ≥ 108 menit 10,922 kali lebih besar dibandingkan kapal dengan jenis kapal penumpang. Kapal dengan jenis kapal petikemas cenderung memiliki WTP ≥ 108 menit 4,525 kali lebih besar dibandingkan kapal dengan jenis kapal penumpang.

d. Gerak Kapal (X_7)

Kapal dengan gerak masuk cenderung memiliki WTP ≥ 108 menit 8,702 kali lebih besar dibandingkan kapal dengan gerak keluar. Kapal dengan gerak pindah cenderung memiliki WTP ≥ 108 menit 0,058 kali lebih kecil dibandingkan kapal dengan gerak keluar.

4) Confussion Matrix Model Regresi Logistik-SMOTE

Tabel 18 menunjukkan bahwa berdasarkan hasil prediksi model regresi logistik-SMOTE pada data *testing* diperoleh nilai *accuracy* ketepatan hasil prediksi sebesar 86,67, *misclassification rate* 0,133, *sensitivity* 0,873, *specificity* 0,800, dan *precision* 0,980. Nilai *specificity* bernilai 0,800 yang berarti kategori WTP yang diklasifikasi secara benar dapat diketahui sebesar 80%. Hal ini menandakan bahwa model regresi logistik-SMOTE sudah lebih baik dalam memprediksi kelas minoritas dan model tidak cenderung pada kelas mayoritas saja. Performa klasifikasi juga dapat diukur dari luas *Area Under Curve* (AUC). Kurva ROC dari model klasifikasi WTP pada data *training* dan data *testing* ditunjukkan pada Gambar 3.

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Hasil analisis regresi logistik-SMOTE diperoleh faktor-faktor yang berpengaruh signifikan terhadap WTP yaitu waktu kedatangan kapal, cuaca, jenis kapal, dan gerak kapal. Model regresi logistik-SMOTE memiliki nilai *accuracy* 86,67%, *missclassification rate* 0,133, *sensitivity* 0,873, *specificity* 0,800, *precision* 0,980, AUC 0,933 Model menggunakan regresi logistik-SMOTE untuk prediksi klasifikasi WTP di Pelabuhan Tanjung Perak adalah sebagai berikut.

$$\hat{g}(x) = -2,576 - 0,922X_3(2) - 2,348X_4(2) + 3,238X_6 \\ + 2,391X_6(3) + 1,510X_6(4) + 2,164X_7(2) \\ - 2,854X_7(3) + 2,164X_7(2) - 2,854X_7(3)$$

B. Saran

Saran yang dapat diberikan kepada perusahaan untuk

mengatasi WTP yang lama yaitu dilakukan sosialisasi terkait WTP kepada pandu dan perusahaan pelayaran dengan memperhatikan faktor-faktor yang diduga berpengaruh terhadap WTP yaitu ketika memiliki kapal dijadwalkan datang saat malam hari, kapal beroperasi saat hujan, berjenis kapal kargo, atau saat kapal melakukan gerakan masuk. Saran untuk penelitian selanjutnya adalah dapat dilakukan dengan menggunakan metode yang berbeda dengan tetap memperhatikan *imbalanced data* sehingga harapannya didapatkan hasil yang lebih baik, salah satunya bisa menggunakan metode *Decision Tree*, *Naïve Bayes*, dan *undersampling*.

DAFTAR PUSTAKA

- [1] Pemerintah Pusat, *Peraturan Pemerintah (PP) No. 17 Tahun 1988 tentang Penyelenggaraan dan Pengusahaan Angkutan Laut*. Jakarta: LN. 1988, No. 37, TLN No. 3378, LL Setkab : 8. HLM, JDH BPK RI, 1988, p. 8.
- [2] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.
- [3] N. Widyawati and B. E. Yuliantini, "Kesiapan alat bongkar muat, cuaca dan waktu kedatangan kapal curah kering terhadap waiting time berth di terminal Jamrud cabang Tanjung Perak," *J. Baruna Horiz.*, vol. 2, no. 1, pp. 1–10, 2019.
- [4] N. Damastuti and A. S. Aisjah, "Simulasi penjadwalan kapal untuk pencapaian zero waiting time di Dermaga Jamrud Utara dengan optimasi algoritma sequential searching," *e-NARODROID*, vol. 1, no. 1, pp. 33–40, Jan. 2015, doi: 10.31090/narodroid.v1i1.10.
- [5] A. Agresti, *An Introduction to Categorical Data Analysis Second Edition*. USA: John Wiley & Sons, Inc. ISBN: 978-0-471-22618-5, 2007.
- [6] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression Second Edition*. USA: John Wiley & Sons, Inc. ISBN: 0-471-35632-8, 2000.
- [7] M. R. Faisal and D. T. Nugrahadi, *Belajar Data Science: Klasifikasi dengan Bahasa Pemrograman R*. Banjarbaru: Scripta Cendekia, ISBN: 978-602-0950-59-4, 2019.