

Algoritma Machine Learning Dalam Melakukan Prediksi Pemilihan Konfigurasi Kapal Tunda di Pelabuhan Tanjung Priok

Budi Tri Yulianto^{1*}, Raden Muhammad Atok²

¹Sekolah Interdisiplin Manajemen Teknologi, Institut Teknologi Sepuluh Nopember, Surabaya, Jawa Timur

²Departemen Aktuaria, Institut Teknologi Sepuluh Nopember, Surabaya, Jawa Timur

Email: ^{1*}6032211150@student.its.ac.id, ²moh_atok@statistika.its.ac.id

(Naskah masuk: 30 Mei 2024, direvisi: 24 Jun 2024, 2 Jul 2024, diterima: 3 Jul 2024)

Abstrak

Pengoperasian pelabuhan secara dasar meliputi berbagai kegiatan pelayanan, salah satu proses pelayanan kegiatan di pelabuhan yaitu pelayanan pemanduan dan penundaan kapal. Salah satu langkah yang dibutuhkan dalam proses penetapan kapal tunda dalam pelayanan pemanduan dan penundaan kapal yaitu pemilihan konfigurasi Kapal Tunda. Penelitian ini menguji performa klasifikasi *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier* (NBC) pada data Laporan Harian Gerakan Kapal (LHGK) di Pelabuhan Tanjung Priok selama periode 2021 untuk proses pemodelan dan evaluasi. Penelitian ini bertujuan untuk membuat modelan prediksi dalam penentuan konfigurasi Kapal Tunda, evaluasi hasil model prediksi untuk memilih konfigurasi kapal tunda di Pelabuhan Tanjung Priok. Dengan menerapkan model klasifikasi NBC dan SVM yang ditingkatkan dengan *kernel Linier* dan RBF, termasuk juga pemilihan fitur baik untuk SVM dan *Naïve Bayes*. Hasil uji perbandingan model prediksi antara SVM dan NBC menunjukan bahwa klasifikasi SVM memberikan hasil yang paling optimal, yaitu menggunakan *kernel linier* pada nilai $C=10$, diperoleh akurasi sebesar 84,7%, *recall* sebesar 84,7%, *F1-score* sebesar 88,7%, dan akurasi sebesar 88,7%. Penelitian ini dimasa yang akan datang dapat dimanfaatkan dalam proses pengambilan keputusan dalam menentukan susunan konfigurasi Kapal Tunda oleh petugas pelabuhan.

Kata Kunci: Pelabuhan, Kapal Tunda, Prediksi, *Support Vector Machine*, *Naïve Bayes*

Machine Learning Algorithm For Predicting Tugboat Configuration Selection at Tanjung Priok Port

Abstract

The basic operation of the port includes service activities related to pilotage and providing towage services for sea-going vessels. One of the activities at the port is pilotage and towage services. One of the necessary steps in the process is selection of Tugboat configuration. The dataset for this research uses data from the Daily Ship Movement Report (LHGK) at Tanjung Priok Port during the 2021 period for modeling and evaluation. This research aims to create a prediction model in determining the Tugboat configuration using *Support Vector Machine* (SVM) and *Naïve Bayes Classifier* classification methods, evaluating the prediction model results to select the tugboat configuration at Tanjung Priok Port by comparing the accuracy of the prediction models. Various enhancements, including feature selection for SVM and *Naïve Bayes*, are applied. The comparison test results of the prediction models between SVM and NBC indicate that SVM classification provides the most optimal results, using a linear kernel with a value of $C=10$, achieving an accuracy of 84.7%, a recall of 84.7%, an *F1-score* of 88.7%, and an accuracy of 88.7%.

Keywords: Port, Tug Boat, Prediction, *Support Vector Machine*, *Naïve Bayes*

I. PENDAHULUAN

Pengoperasian pelabuhan secara dasar meliputi kegiatan pelayanan jasa yang berhubungan dengan pemanduan kapal dan pemberian jasa kapal tunda untuk kapal-kapal laut [1]. Terlepas dari struktur kepemilikannya dalam suatu Pelabuhan harus menyediakan serangkaian fasilitas dan layanan tertentu antara lain pelayanan jasa kapal diantaranya Pemanduan dan Penundaan [2].

Keterbatasan area Pelabuhan dan pertimbangan faktor keselamatan menjadi perhatian utama dalam regulasi dalam hal pelayanan penundaan yang diatur dalam aturan pemerintah melalui Peraturan Menteri Perhubungan Republik Indonesia Nomor PM 57 Tahun 2015 tentang Pemanduan dan Penundaan Kapal. Dengan pertimbangan keselamatan dan keamanan serta kelancaran pelaksanaan pemanduan, pandu wajib menggunakan sarana bantu dan prasarana pemanduan yang dipersyaratkan sesuai ketentuan untuk membantu olah gerak kapal.

Penggunaan kapal tunda sebagai sarana bantu pemanduan untuk membantu olah gerak kapal diatur dengan ketentuan sebagai berikut:

1. Panjang kapal 70 (tujuh puluh) meter sampai dengan 150 (seratus lima puluh) meter menggunakan 1 (satu) unit kapal tunda yang dengan jumlah daya paling rendah 2000 (dua ribu) DK dengan jumlah gaya tarik paling rendah 24 ton *bollard pull*.
2. Panjang kapal diatas 150 (seratus lima puluh) meter sampai dengan 250 (dua ratus lima puluh) meter menggunakan minimal 2 (dua) unit kapal tunda dengan jumlah daya paling rendah 6000 (enam ribu DK) dengan jumlah gaya tarik paling rendah 65 ton *bollard pull*.
3. Panjang kapal 250 (dua ratus lima puluh) meter ke atas minimal 3 (tiga) unit kapal tunda dengan jumlah daya paling rendah 11.000 (sebelas ribu) DK dengan jumlah gaya tarik rendah 125 ton *bollard pull*.

Perencanaan dan pengaturan transportasi merupakan isu yang tidak pernah berhenti. Karakter lingkungan yang terus berubah, menjadi penyebab permasalahan transportasi menjadi lebih kompleks. Sektor maritim sering dianggap konservatif dalam hal penerapan suatu teknologi baru atau mengembangkan solusi pasar baru. Asumsi ini sering digarisbawahi oleh makalah kebijakan terkait, misalnya industri pelabuhan yang diperingkat sebagai salah satu sektor inovatif yang lebih rendah dalam penelitian yang dipromosikan oleh Forum Transportasi Antar-nasional tentang inovasi *Transformation Forum on Innovation* dan penelitian akademis [3].

Pada era *Big Data* saat ini industri kepelabuhanan mempunyai peluang yang sangat besar dalam mendukung pengimplementasian teknologi baru dengan pendekatan *machine learning* pada era industri 4.0. Industri 4.0 adalah gelombang transformasi terbaru, berdasarkan sistem *cyber-fisik* yang mencakup berbagai teknologi baru elektrik yaitu *Internet of Things* (IoT), teknologi *Blockchain*, *Big Data*, Robotika, *Augmented Reality* (AR), dan *Artificial Intelligence* (AI) [4].

Meskipun implementasi Industri 4.0 mendapatkan momentum dalam beberapa tahun terakhir di antara para pemangku kepentingan pelabuhan, masih ada banyak upaya yang harus dilakukan untuk mencapai potensi akhir dari ini yang disebut "*blue ocean*" [5], dengan menjalankan teknologi semacam itu di pelabuhan, jumlah aliran data yang sangat luas dan besar menjadi lebih mudah tersedia di pelabuhan. Berangkat dari hal tersebut salah satu upaya dalam pemanfaatan data yang dihasilkan di pelabuhan untuk mendukung efektivitas proses bisnis dalam pelayanan jasa pemanduan dan penundaan menggunakan algoritma *machine learning*. Klasifikasi adalah proses umum untuk menemukan model (kemampuan) yang menggambarkan dan mengidentifikasi kelas-kelas data dengan tujuan dapat memprediksi kelas dari objek yang kelasnya tidak diketahui [6].

Pada penelitian sebelumnya [7] dan [8], penelitian ilmiah tentang algoritma SVM dengan kernel linier dan algoritma *Naïve Bayes* yang dilakukan, didapatkan kesimpulan bahwa performa SVM lebih akurat dibandingkan dengan *Naïve Bayes* dalam penelitian klasifikasi yang dilakukan [7]. Sedangkan penelitian ilmiah lainnya yang dilakukan didapatkan akurasi model sebelum dilakukan *tunning* sebesar 89,93 % dan setelah dilakukan *tunning* didapatkan akurasi sebesar 100 % dengan kernel *Polynomial* dan *K-fold validation* sebanyak 10 kali iterasi [8].

Model klasifikasi yang paling optimal dalam penelitian ini dapat dimanfaatkan di masa mendatang untuk membantu permasalahan dalam proses pengambilan keputusan dalam menentukan susunan konfigurasi kapal tunda oleh petugas pelabuhan. Hasil prediksi dapat menjadi pertimbangan dan dasar pengembangan sistem untuk menentukan kapal tunda saat petugas pelabuhan menetapkan konfigurasi kapal tunda dalam pelayanan jasa kapal, khususnya dalam pemanduan dan penundaan di Pelabuhan Tanjung Priok.

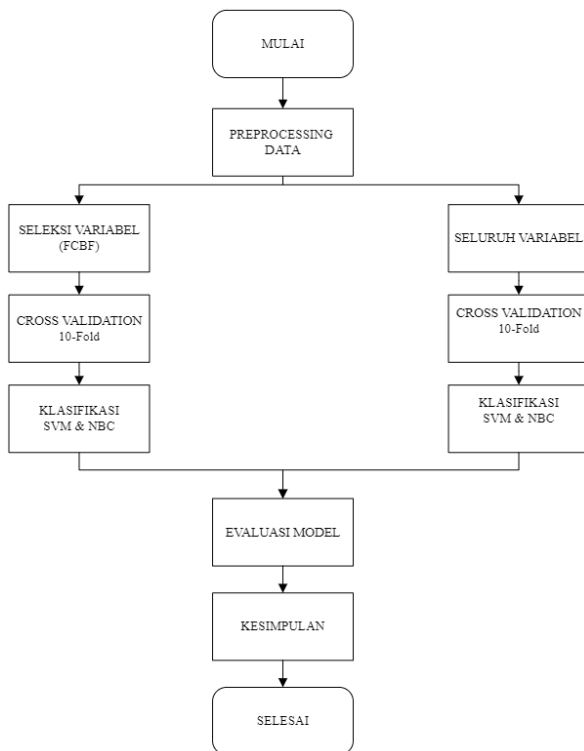
Tujuan dalam penelitian ini adalah membahas mengenai hasil prediksi dalam penentuan konfigurasi kapal tunda menggunakan algoritma *machine learning* yaitu metode klasifikasi SVM dengan kernel linier dan kernel *Radial Based Function* (RBF), serta model klasifikasi *Naïve Bayes* pada data Laporan Harian Gerakan Kapal di Pelabuhan Tanjung Priok, serta mengetahui ketepatan klasifikasi pada metode SVM dan NBC. Dengan pendekatan berbagai metode klasifikasi, penelitian ini diharapkan dapat mendukung proses pengambilan keputusan dalam menentukan susunan konfigurasi Kapal Tunda secara efektif dalam proses pelayanan jasa kapal di Pelabuhan.

II. METODOLOGI

Data dan Informasi utama yang digunakan dalam melakukan penelitian ini berasal dari Laporan Harian Gerakan Kapal (LHGK) di Pelabuhan Tanjung Priok selama periode tahun 2021, sesuai dengan data mengenai pelayanan jasa Pemanduan dan Penundaan Kapal. Dimana variabel tunda ke-1, tunda ke-2 dan tunda ke-3 sebagai variabel respon, sedangkan variabel Jam, GT, LOA, Total HP, Bendera Kapal,

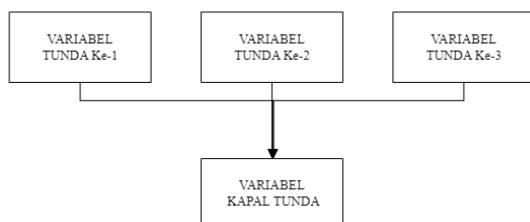
Asal (Dermaga Awal), KE (Tujuan Dermaga), Kegiatan (Masuk/Keluar/Pindah) dan Pandu sebagai variabel Prediktor.

Langkah-langkah dalam proses klasifikasi pada model algoritma *machine learning* dalam penentuan konfigurasi kapal tunda dapat divisualisasikan dalam bentuk *flowchart*. *Flowchart* ini memberikan gambaran umum tentang bagaimana melakukan prediksi menggunakan *Support Vector Machine (SVM)* dan *Naïve Bayes* yang dikombinasikan dengan beberapa tahapan dalam penentuan variabel yang paling optimal dan tahapan dalam penentuan data testing dan training menggunakan *Cross Validation* seperti pada Gambar 1 sebagai berikut.



Gambar 1. *Flowchart* Penelitian

Langkah *preprocessing* yaitu pada variabel kapal tunda ke-1, ke-2 dan ke-3 sesuai dengan kriteria kapal tunda dikategorikan berdasarkan daya kapal dan sistem propulsi di setiap kapal tunda yang tercatat. Pengkategorian 3 variabel kapal tunda yang telah terbentuk, dilakukan penggabungan maka akan terbentuk 1 variabel konfigurasi kapal tunda di setiap pelayanan kegiatan yang selanjutnya akan digunakan sebagai variabel respon (Gambar 2).



Gambar 2. Variabel Konfigurasi Tunda

A. Fast Correlation Based Filter (FCBF)

Melakukan seleksi variabel menggunakan metode FCBF Sehingga terdapat 2 kelompok data yaitu data menggunakan seluruh variabel prediktor dan data hanya menggunakan variabel yang terseleksi. Selanjutnya 2 kelompok data tersebut masing-masing dibagi menjadi data training dan data testing. FCBF adalah salah satu algoritma pemilihan variabel. Sebuah variabel dapat dianggap baik jika relevan dengan konsep kelas namun tidak redundan dengan variabel lainnya, Pendekatan lain yang pernah dilakukan didasarkan pada metode entropi teoritis informasi [9]. Persamaan (1) menetapkan definisi entropi dari variabel (X).

$$H(X) = - \sum_{i=1}^n P(x_i) \log_2(p(x_i)) \tag{1}$$

Entropi dari variabel (X) yang diketahui variabel (Y) didefinisikan dalam Persamaan (2).

$$H(X|Y) = - \sum_{i=1}^n P(y_i) \sum_{i=1}^n P(x_i|y_i) \log_2(P(x_i|y_i)) \tag{2}$$

$P(X_i)$ adalah *posterior probabilities* untuk semua nilai X dan $P(x_i|y_i)$ *posterior probabilities* dari X jika Y diketahui. Berdasarkan entropi tersebut dapat diperoleh *Information Gain* pada Persamaan (3).

$$IG(X|Y) = H(X) - H(X|Y) \tag{3}$$

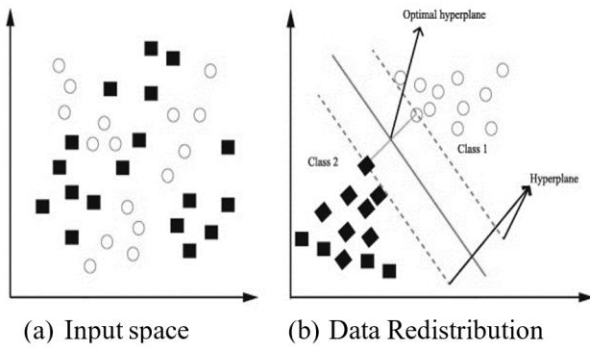
B. K-Fold Cross Validation

Partisi dataset pada penelitian ini akan dilakukan dengan pembagian dataset menjadi dua kelompok dataset training dan testing. *Dataset training* merupakan kumpulan data sebagai pembentuk model prediktif. *Data testing* digunakan sebagai pengukuran kinerja model prediktif. Pembagian *data testing* dan *data training* menggunakan *K-fold cross validation* yaitu sebanyak 10 *fold*.

K-fold cross validation adalah sebuah metode yang digunakan untuk memastikan ketidakteraturan dalam menentukan data sebagai data latih dan uji [10]. Salah satu metode untuk membagi data set menjadi *data testing* dan *data training*, dengan teknik menggunakan seluruh *dataset* sebagai *data training* dan *testing*. Metode ini berguna untuk mencapai tingkat kepercayaan yang tinggi karena semua dataset digunakan sebagai *data testing* dan *data training* [11].

C. Klasifikasi SVM Linier

SVM linear melakukan klasifikasi terhadap dua kelas secara linear yang dapat dipisahkan, Ilustrasi *hyperplane* pemisah dan margin SVM pada data yang dapat dipisahkan terdapat pada Gambar 3.



Gambar 3. Konsep *Hyperplane* Pada SVM Linier

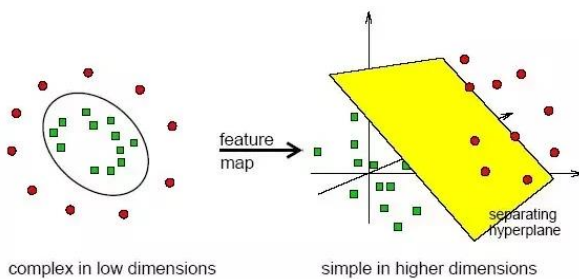
Misalkan diberikan suatu himpunan $X_i = \{X_1, X_2, \dots, X_n\}$ dimana $x_i \in R^n, i = 1, 2, \dots, n$ diketahui X_i berpola tertentu yaitu Ketika X_i termasuk dalam suatu class (kelas), maka X_i diberikan label (target) $Y_i = +1$ dan jika tidak maka diberi label $Y_i = -1$. Apabila data terpisah secara linier, maka fungsi pemisah atau *hyperplane* didefinisikan oleh [12] seperti terlihat pada Persamaan (4).

$$D(x) = w^T x + b \tag{4}$$

dimana w adalah vektor berdimensi m dan b adalah suatu konstanta yang disebut bias. Data X_i termasuk dalam $Y_i = +1$ bila $w^T x_i + b > 0$ dan data X_i termasuk dalam $Y_i = -1$ apabila $w^T x_i + b < 0$.

D. Klasifikasi SVM Non Linier.

Memilih fungsi kernel yang tepat sangat penting karena akan berpengaruh dalam menentukan ruang fitur dimana fungsi klasifikasi akan dicari. Oleh karena itu, vektor input dipetakan oleh fungsi $\Phi(x)$ ke dalam ruang vektor baru dengan dimensi yang lebih tinggi (3 dimensi).



Gambar 4. Konsep *Hyperplane* Non Linier

Pada ilustrasi Gambar 4, jenis pemetaan non linier tidak perlu diketahui secara eksplisit. Pada SVM non linier, hal itu diubah menjadi bentuk $\phi(x_i)^T \phi(x_j)$. Fungsi transformasi dalam SVM menggunakan "Trick Kernel" (Scholopf & Simola, 2002). Trik kernel dirumuskan dalam Persamaan (5).

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \tag{5}$$

Fungsi kernel memberikan berbagai kemudahan karena menggunakan kernel untuk menentukan *support vector* dalam SVM, tidak memerlukan pengetahuan tentang bentuk pemetaan sebenarnya $\phi(x)$ Fungsi kernel yang umum digunakan dalam metode SVM [13] adalah

1. Kernel Linear
 $K(x_i, x_j) = x_i^T x_j$
2. Kernel Polynomial
 $K(x_i, x_j) = (-\gamma x_i^T x_j + r)^p, \gamma > 0$
3. Kernel Radial Basis Function (RBF)
 $K(x_i, x_j) = exp(-\gamma \| x_i - x_j \|^2)$
4. Sigmoid Kernel
 $K(x_i, x_j) = tanh(-\gamma x_i^T x_j + r)$

Persamaan *dual* pada klasifikasi non linier dihasilkan dengan pada Persamaan (6).

$$L_D(\alpha) = \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j K(x_i y_j) \tag{6}$$

Fungsi keputusan pada SVM non linier diperoleh melalui dengan Persamaan (7) sebagai berikut:

$$D(x) = \sum_{i \in S} \alpha_i y_i K(x_i, x_j) + b \tag{7}$$

Dengan $K(x_i, x)$ adalah fungsi kernel. Selanjutnya data testing diklasifikasikan menggunakan fungsi keputusan pada Persamaan (8).

$$X \in \begin{cases} \text{Kelas 1, jika } D(x) > 0 \\ \text{Kelas 2, jika } D(x) < 0 \end{cases} \tag{8}$$

Membangun model prediksi dengan SVM melalui 2 tahap yaitu menggunakan semua variabel prediktor dan menggunakan seleksi variabel prediktor baik SVM Linier dan SVM non linier.

E. Naïve Bayes Classifier

Naïve Bayes termasuk dalam kategori pembelajaran *supervised*, sehingga memerlukan data pelatihan pada tahap pembelajaran untuk membuat keputusan. Secara umum, persamaan Naïve Bayes dapat dinotasikan sebagai Persamaan (9):

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \tag{9}$$

$P(A|B)$ merupakan *posterior probability* karena nilainya bergantung pada nilai B, sedangkan $P(A)$ merupakan *prior probability* karena nilainya tidak bergantung pada nilai B.

Dalam klasifikasi menggunakan *Naïve Bayes Classifier* dilakukan melalui 2 tahapan yaitu menggunakan semua variabel prediktor dan menggunakan seleksi variabel dengan metode FCBF.

III. HASIL DAN PEMBAHASAN

Variabel respon yang digunakan dalam penelitian ini adalah Tunda ke-1, Tunda ke-2 dan Tunda ke-3. Setiap kapal tunda yang beroperasi di Pelabuhan Tanjung Priok akan dikelompokkan menjadi data kategorikal berdasarkan kriteria daya mesin dan jenis propulsi sebagaimana Tabel 1 berikut.

Tabel 1. Kriteria Variabel Tunda

Daya Mesin (Hp)	Type Propulsi	Kategori Tug Boat
1200	Stand Tug - FIX	Type A – Tunda Kecil
3200	ASD - SRP	
2400	ASD - SRP	
3200	ATD - SRP	Type B – Tunda Sedang
2400	ATD - SRP	
3600	ATD - SRP	Type C – Tunda Besar
4000	ATD - SRP	
4400	ATD - SRP	
0	Tanpa Tunda	Type D – Tanpa Tunda

Selanjutnya dilakukan penyusunan acuan konfigurasi terhadap kondisi Kapal Tunda pada dataset LHGK, sehingga dari 3 variabel kapal Tunda yang tercatat dilakukan penggabungan menjadi 1 variabel disesuaikan berdasarkan acuan konfigurasi kapal Tunda. Tabel 2 menyajikan variabel respon sebagai acuan konfigurasi kapal tunda berdasarkan dengan daya mesin dan type propulsinya.

Tabel 2. Variabel Respon

Konfigurasi	Kategori Kapal Tunda	Keterangan
1	AAA	3 Kapal Tunda Kecil
2	AAB	2 Kapal Tunda Kecil dan 1 Kapal Tunda sedang
3	AAC	2 Kapal Tunda kecil dan 1 Kapal Tunda Besar
4	ACC	1 Kapal Tunda Kecil dan 2 Kapal Tunda Besar
5	AAD	2 Kapal Tunda Kecil
6	ABB	1 Kapal Tunda Kecil dan 2 Kapal Tunda Sedang
7	ABC	1 Kapal Tunda Kecil, 1 Kapal Tunda Sedang dan 1 Kapal Tunda Besar
8	ABD	1 Kapal Tunda Kecil dan 1 Kapal Tunda Sedang
9	ACD	1 Kapal Tunda Kecil dan 1 Kapal Tunda Besar
10	ADD	1 Kapal Tunda Kecil
11	BBB	3 Kapal Tunda Sedang

Konfigurasi	Kategori Kapal Tunda	Keterangan
12	BBC	2 Kapal Tunda Sedang dan 1 Kapal Tunda Besar
13	BBD	2 Kapal Tunda Sedang
14	BCC	1 Kapal Tunda Sedang dan 2 Kapal Tunda Besar
15	BCD	1 Kapal Tunda Sedang dan 1 Kapal Tunda Besar
16	BDD	1 Kapal Tunda Sedang
17	CCC	3 Kapal Tunda Besar
18	CCD	2 Kapal Tunda Besar
19	DDD	Tanpa Kapal Tunda
20	CDD	1 Kapal Tunda Besar

Sebagai bagian dari tahap *pre-processing* dilakukan proses konversi menjadi data numerik pada variabel LHGK tahun 2021, agar dapat diolah untuk kebutuhan penelitian ini. Variabel prediktor penelitian yang digunakan dalam analisis penelitian ditampilkan pada Tabel 3.

Tabel 3. Variabel Prediktor

Variabel	Nama Variabel	Skala Data
Y ₁	Jam Pemakain Start	Ordinal
Y ₂	<i>Grt/GT</i>	Interval
Y ₃	<i>LOA</i>	Rasio
Y ₄	Total <i>HP</i>	Interval
Y ₅	(<i>R/D</i>))	Nominal
Y ₆	Dari	Nominal
Y ₇	Ke	Nominal
Y ₈	Kegiatan	Nominal
Y ₉	Pandu	Nominal

Beberapa variabel respon merupakan variabel yang sudah bersifat numerik sehingga tidak perlu dilakukan konversi data. Variabel yang tidak perlu dilakukan konversi data disajikan pada Tabel 4. Variabel tersebut merupakan jenis variable yang kuantitatif dan memiliki besaran angka yang jelas sebagai berikut:

Tabel 4. Variabel Respon Numerik

Variabel	Nama Variabel
Y ₂	<i>Grt/GT</i>
Y ₃	<i>LOA</i>
Y ₄	Total <i>HP</i>

Adapun variabel yang perlu dilakukan proses konversi menjadi data kategorik yang akan digunakan sebagai variabel prediktor yaitu Jam Pemakaian Start, Domestik = D / Asing = R, Dari, Ke, kegiatan dan Pandu. Ke enam variabel tersebut perlu dilakukan dilakukan proses konversi menjadi data numerik terlebih dahulu sebagaimana Tabel 5.

Tabel 5. Konversi Variabel Respon

Variabel	Nama Variabel	Konversi Data Numerik						
		1	2	3	4	5	6	7
Y_1	Jam Pemakai an Start	05:00-10:59	11:00-15:00	15:01-18:00	18:01-04:59			
Y_5	(R/D)	D	R	-	-	-		
Y_6	Dari	Nusantara	Pelabuhan I	Pelabuhan II	Pelabuhan III	Tpk Koja Ke Timur	Laut	Labuh
Y_7	Ke	Nusantara	Pelabuhan I	Pelabuhan II	Pelabuhan III	Tpk Koja Ke Timur	Laut	Labuh
Y_8	Kegiatan	Keluar	Masuk	Pindah	-	-	-	-
Y_9	Pandu	Dengan Pandu	Tanpa Pandu	-	-	-	-	-

Hasil perbandingan variabel prediktor, antara seluruh variabel dan hasil variabel setelah dilakukan seleksi fitur (FCBF) pada penelitian ini ditampilkan Tabel 6.

Tabel 6. Perbandingan Variabel Prediktor

Variabel	Nama Variabel	Variabel Prediktor	
		Seluruh Variabel	Seleksi Fitur (FCBF)
Y_1	Jam Pemakain Start	√	X
Y_2	<i>Grt/GT</i>	√	√
Y_3	<i>LOA</i>	√	√
Y_4	Total <i>HP</i>	√	√
Y_5	(R/D))	√	X
Y_6	Dari	√	√
Y_7	Ke	√	√
Y_8	Kegiatan	√	√
Y_9	Pandu	√	X

Selanjutnya klasifikasi akan dilakukan dan diulang pada data latih 10-fold dengan berbagai kombinasi nilai parameter C dalam SVM linear. Setiap fold data latih dievaluasi menggunakan kombinasi nilai C sebanyak tiga kali, menghasilkan total 30 kombinasi yang berbeda. Setiap percobaan pada data latih 10-fold menghasilkan berbagai nilai akurasi yang berbeda. Parameter optimal yang dipilih adalah klasifikasi dengan hasil rata-rata akurasi dan recall tertinggi.

Dalam penelitian ini, klasifikasi akan dilakukan dalam dua tahap klasifikasi menggunakan semua prediktor dan variabel yang dipilih oleh FCBF. Metode SVM akan dilakukan dengan dua jenis kernel, kernel linier dan RBF. Penggunaan kernel linier memerlukan penyetelan parameter Cost (C), sementara penggunaan kernel RBF memerlukan penyetelan kedua parameter Cost (C) dan Gamma (γ). Untuk model dengan kernel linear yang menggunakan seluruh fitur, hasil dari Tabel 7 menunjukkan bahwa nilai Cost (C) menggunakan C=10, menghasilkan rata-rata akurasi mencapai 88,52%, dengan rata-rata recall yang mencapai 82,22%.

Tabel 7. Klasifikasi SVM Kernel Linier Dengan Seluruh Variabel

Kernel	Cost (C)	Precision	Recall	F1 Score	Akurasi
Linier	0,1	0,7430	0,7430	0,8099	0,8067
Linier	1	0,8084	0,8084	0,8849	0,8829
Linier	10	0,8222	0,8222	0,8862	0,8852

Setelah dilakukan seleksi fitur, dilakukan evaluasi kinerja model dengan kernel linier pada tiga nilai Cost (C) yang berbeda sebagaimana Tabel 8 Diperoleh nilai C =10, didapatkan hasil akurasi mencapai 88,70% dan recall mencapai 84,77%.

Tabel 8. Klasifikasi SVM Kernel Linier Dengan Seleksi Fitur

Kernel	Cost (C)	Precision	Recall	F1 Score	Akurasi
Linier	0,1	0,7440	0,7440	0,8104	0,8081
Linier	1	0,8091	0,8091	0,8858	0,8838
Linier	10	0,8477	0,8477	0,8870	0,8870

Peningkatan nilai cost pada SVM Linier secara signifikan dapat meningkatkan akurasi dan recall pada model, Seleksi fitur juga berkontribusi positif terhadap peningkatan kinerja model, dengan akurasi tertinggi mencapai 88,70% dan recall 84,77% pada nilai C=10 setelah seleksi fitur. Nilai C=10 memberikan hasil yang optimal, baik sebelum maupun setelah seleksi fitur.

Hasil klasifikasi SVM dengan kernel RBF pada semua variabel sesuai pada Tabel 9, kesimpulannya adalah kombinasi optimal untuk nilai C dan γ adalah C = 10 dan γ = 0,1, yang menghasilkan rata-rata akurasi tertinggi mencapai 87%.

Tabel 9. Klasifikasi SVM Kernel RBF Seluruh Variabel

Run	Cost (C)	Gamma (Y)	Precision	Recall	F1 Score	Akurasi
1	0,1	0,1	0,1167	0,1167	0,2892	0,1532
2	0,1	1	0,3822	0,3882	0,5478	0,4671
3	0,1	10	0,1200	0,1200	0,2458	0,1586
4	1	0,1	0,6250	0,6250	0,7805	0,7302
5	1	1	0,6652	0,6652	0,7373	0,7333
6	1	10	0,5367	0,5367	0,6520	0,5996
7	10	0,1	0,8133	0,8133	0,8724	0,8734
8	10	1	0,7163	0,7163	0,7844	0,7843
9	10	10	0,5478	0,5478	0,6613	0,6126

Hasil dari kernel RBF pada Tabel 10, dengan melakukan seleksi fitur terlihat bahwa menggunakan kombinasi parameter C=10 dan Y=1 mampu menghasilkan nilai rata-rata akurasi paling tinggi, mencapai 88%.

Tabel 10. Klasifikasi SVM Kernel RBF Dengan Seleksi Fitur

Run	Cost (C)	Gama (Y)	Precision	Recall	F1 Score	Akurasi
1	0,1	0,1	0,6144	0,6144	0,7602	0,7184
2	0,1	1	0,6250	0,6250	0,7552	0,7279
3	0,1	10	0,5463	0,5463	0,6889	0,6437
4	1	0,1	0,7730	0,7730	0,8445	0,8419
5	1	1	0,7833	0,7833	0,8584	0,8590
6	1	10	0,8130	0,8130	0,8605	0,8590
7	10	0,1	0,7730	0,7730	0,8445	0,8419
8	10	1	0,8377	0,8377	0,8798	0,8811
9	10	10	0,8409	0,8409	0,8638	0,8734

Peningkatan nilai C dan γ pada SVM dengan kernel RBF dapat berpengaruh meningkatkan akurasi dan recall model. Seleksi fitur juga berkontribusi positif terhadap peningkatan kinerja model, dengan akurasi tertinggi mencapai 88.11% pada nilai C=10 dan γ =1 setelah seleksi fitur. Nilai C=10 memberikan hasil terbaik baik sebelum maupun setelah seleksi fitur, menjadikannya pilihan yang optimal dalam analisis ini.

Hasil klasifikasi NBC dengan seluruh prediktor didapatkan hasil sebagaimana Tabel 11 dan klasifikasi NBC dengan seleksi variabel didapatkan hasil performa klasifikasi sebagaimana Tabel 12 berikut.

Tabel 11. Klasifikasi Naïve Bayes Classifier Seluruh Variabel

	Precision	Recall	F1_Score	Accuracy
Naïve Bayes	0,7869	0,7869	0,8231	0,8284

Tabel 12. Klasifikasi Naïve Bayes Classifier Seleksi Variabel FCBF

	Precision	Recall	F1_Score	Accuracy
Naïve Bayes	0,8177	0,8177	0,8377	0,8401

Penggunaan Naïve Bayes Classifier (NBC) dengan semua prediktor menghasilkan akurasi 82,84%, precision dan recall sebesar 78,69%, serta F1 score sebesar 82,31% sebagaimana Tabel 11. Setelah dilakukan seleksi variabel menggunakan metode Fast Correlation-Based Filter (FCBF), performa model meningkat. Hasil klasifikasi menunjukkan akurasi sebesar 84,01%, precision dan recall sebesar 81,77%, serta F1 score sebesar 83,77% sebagaimana Tabel 12. Seleksi variabel terbukti berkontribusi positif terhadap peningkatan kinerja model NBC, dengan peningkatan akurasi menjadi 84,01% setelah seleksi variabel menggunakan metode FCBF.

Performa klasifikasi yang dihasilkan dari model klasifikasi dengan membandingkan hasil nilai akurasi SVM dan NBC yang disimpulkan sesuai dengan Tabel 13 berikut:

Tabel 13. Perbandingan Hasil Akurasi

Metode	Precision	Recall	F1 Score	Accuracy
SVM Linier	0,8222	0,8222	0,8862	0,8852
SVM RBF	0,8133	0,8133	0,8724	0,8734
SVM Linier Seleksi Fitur	0,8477	0,8477	0,8870	0,8870
SVM RBF Seleksi Fitur	0,8130	0,8130	0,8605	0,8590
NBC	0,7869	0,7869	0,8231	0,8284
NBC Seleksi Fitur	0,8177	0,8177	0,8377	0,8401

Sebagaimana data hasil perbandingan pada Tabel 13 didapatkan bahwa hasil dari pengukuran nilai precision, Recall, F1 Score dan Akurasi, Penggunaan SVM dengan kernel linier dan seleksi variabel memberikan hasil yang paling optimal.

IV. KESIMPULAN

Pada penelitian ini sebelum dilakukan proses klasifikasi, didapatkan bahwa komposisi konfigurasi kapal tunda yang dominan di Pelabuhan Tanjung priok pada data LHGK Periode 2021, didapatkan sebanyak 35% pelayanan penundaan dilakukan untuk kapal dengan panjang 70-150 meter dan sebanyak 45% pelayanan penundaan dilakukan untuk kapal dengan penjang 150-250 meter. Sedangkan untuk proses hasil uji performa ketepatan klasifikasi terbaik, didapatkan pada model klasifikasi SVM Linier dengan seleksi fitur, dalam menentukan konfigurasi kapal Tunda di Pelabuhan Tanjung Priok pada data LHGK Periode 2021 diperoleh nilai rata-rata precision sebesar 84,7 %, recall 84,7 %, F1-score 88,7% dan akurasi yang baik sebesar 88,7%.

Secara umum keseluruhan perbandingan kinerja metode klasifikasi SVM dan NBC menunjukkan bahwa hasil performa model SVM dengan kernel Linier lebih baik dalam mengklasifikasikan data LHGK di Pelabuhan Tanjung Priok periode tahun 2021 dibandingkan metode klasifikasi NBC. Berdasarkan nilai evaluasi model yang diperoleh, dapat disimpulkan bahwa algoritma klasifikasi dalam menentukan konfigurasi kapal tunda dapat dikategorikan sebagai hasil yang baik. Dalam pelayanan jasa pemanduan dan penundaan berikutnya dapat menggunakan metode atau algoritma yang sama, dengan harapan hasil yang baik dan akurat seperti yang diperoleh untuk penentuan konfigurasi pada LHGK tahun 2021.

REFERENSI

- [1] R. O. S. Gurning and E. Hariyadi, *Manajemen Bisnis Pelabuhan*, Surabaya: PT Andhika Prasetya Ekawahana, 2007.
- [2] W. S. Dong and P. M. Panayides, *Maritime Logistic*, 2 ed., Croydon: Kogan Page, 2015, p. 252.
- [3] T. Koukaki and A. Tei, "Innovation and maritime transport: A systematic review," *Case Studies on Transport Policy*, vol. 8, pp. 700-710, 2020.
- [4] S. Filom, A. M. Amiri and S. Razavi, "Applications of machine learning methods in port operations – A systematic literature review," *Transportation Research Part E*, vol. 161, pp. 1-30, May 2022.
- [5] I. d. I. P. Zarzuelo, M. J. F. Soeane and B. . L. Bermúdez, "Industry 4.0 in the port and maritime industry: A literature review," *Journal of Industrial Information Integration*, vol. 20, pp. 1-18, 2020.
- [6] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, vol. 2, Illionis: Morgan Kaufman, 2006.
- [7] Q. n. Li and T. h. Li, "Research on The Application of Naive Bayes and Suoport Vector Machine Alhorithm on exercises Classification," *Juornal of Physics*, pp. 1-6, 2020.
- [8] M. Asrol, P. Papilo and F. E. Gunawan, "Support Vector Machine with K-fold Validation to Improve the Industry's Sustainability Performance Classification," *International Conference on Computer Science and Computational intelligence*, pp. 854-862, 2020.
- [9] L. Yu and H. Liu, "Feature Selection for High-Dimensional Data:," *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, pp. 856-863, 2003.
- [10] A. Arabameri, S. Saha, J. Roy, J. Tiefenbacher, A. Cerda, T. Biggs, B. Pradhan, P. T. T. Ngo and A. L. Collins, "A Novel Ensemble Computational Intelligence Approach For The Spatial," *Science of the Total Environment*, vol. 726, pp. 1-18, 2020.
- [11] Y. Bengio, "No Unbiased Estimator of the Variance of K-Fold Cross-Validation," *Journal of Machine Learning Research* 5, vol. 5, pp. 1089-1105, 2004.
- [12] C. Cortes and V. Vapnik, "Support-vector networks," *Journal of Machine Learning Research*, pp. 273-297, 1995.
- [13] Schölkopf, Bernhard, Smola and A.J, *Learning with Kernels : Support Vector Machines, Regularization, Optimization and Beyond*, Cambridge, MA: MIT Press, 2002.