

Forecasting Modeling of Imported Goods Release Waiting Time in Importer Logistics Operations Using Multiple Linear Regression

Vivid Kristiani Alfad Zebua^{1*}, Rusdah^{2}**

* Magister Ilmu Komputer, Universitas Budi Luhur, Jakarta

** Teknik Multimedia Jaringan, Politeknik Negeri Batam

vividkazeuba12@gmail.com¹, rusdah@budiluhur.ac.id²

Article Info

Article history:

Received 2025-06-02

Revised 2025-07-05

Accepted 2025-07-09

Keyword:

*Import,
Waiting Time,
Multiple Linear Regression,
Machine Learning,
Logistics.*

ABSTRACT

Import activities play a critical role in international trade, directly affecting logistics efficiency and the competitiveness of importing companies. The process of releasing imported goods at ports often involves complex administrative procedures that can cause delays, leading to increased logistics costs. This study aims to predict the waiting time for the release of imported goods using a machine learning approach. A case study was conducted at PT. Sentra Sarana Logistic, a licensed customs broker responsible for import administration. The primary model applied was Multiple Linear Regression (MLR), and its performance was compared with Neural Network (NN) and Support Vector Machine (SVM) algorithms. Several influencing factors were considered, including tax payment time, inspection duration, and inspection status. Evaluation results indicate that the MLR model achieved the best performance, with an RMSE of 0.00653, MAE of 0.00544, and R-squared of 0.99999, demonstrating high prediction accuracy and a strong linear correlation. The SVM model yielded acceptable results (RMSE 0.74107, R-squared 0.98388) but underperformed compared to MLR. The NN model showed the lowest accuracy with RMSE 2.86599, MAE 2.38831, and R-squared 0.69510. The findings suggest that MLR, despite its simplicity, is highly effective for predicting waiting times in import logistics operations. This research not only offers a practical decision-support tool for importers but also contributes to the existing literature on machine learning applications in logistics operations and customs processing.



This is an open access article under the [CC-BY-SA](#) license.

I. PENDAHULUAN

Impor adalah kegiatan ekonomi yang melibatkan pembelian barang dari luar negeri untuk digunakan atau dipasarkan di dalam negeri [1]. Indonesia adalah negara yang selalu terlibat dalam perdagangan internasional dengan Negara China [2].

Para pengusaha tentunya harus memiliki pengetahuan yang cukup tentang prosedur impor agar bisnis mereka lebih mudah untuk dijalankan, terutama yang berkaitan dengan perdagangan internasional, kepabeanan, pengiriman, dan perbankan, dan masalah di lapangan yang sering terjadi [2].

Untuk mempercepat keluar masuk barang impor di Indonesia, prosedur jasa pelayanan barang diperlukan untuk pengiriman barang impor [1]. Dalam instansi pabean

Pengusaha Pengurusan Jasa Kepabeanan (PPJK) yang membantu Importir dan instansi pabean Direktorat Jenderal Bea dan Cukai dalam urusan impor [3]. Perusahaan Importir sebagai pengimpor barang akan menggunakan jasa kepabeanan untuk mengurus berbagai administrasi mulai dari orderan barang dari negara lain hingga masuk ke Indonesia. Selain itu, PPJK juga dapat didefinisikan sebagai usaha jasa pengantar yang membantu pemilik barang menyelesaikan pembayaran Bea Cukai (Custom Clearance) dan mengelola pengiriman dan penerimaan barang dengan perusahaan pelayaran [4].

Dalam studi kasus ini, proses bisnis impor dimulai dengan perjanjian antara importir Indonesia dan eksportir negara China untuk membeli bahan baku industri seperti bahan baku besi baja, tekstil, dan bahan baku lainnya. Karena banyak

importir tidak memiliki ijin impor dan modul pengajuan sendiri, Perusahaan pabean PPJK PT. Sentra Sarana Logistic mengurus administrasi impor, dan pengeluaran barang impor dari importir hingga barang tiba di Pelabuhan Tanjung Perak, Surabaya, dan kemudian dikirim ke gudang atau pabrik masing-masing importir.

Pada penelitian ini, terdapat urgensi yang harus segera diatasi dengan cepat dan tepat, dimana waktu yang ditargetkan perusahaan Importir untuk pengeluaran barang impor dengan target yang sudah ditentukan adalah maksimal 3 hari sesuai Standar Operasional Procedur (SOP) perusahaan.

Biaya logistik adalah komponen yang sangat menentukan daya saing perusahaan dan negara, biaya ini meningkatkan biaya bahan dan produk, dan seringkali menjadi masalah bagi baik importir maupun eksportir, permasalahannya adalah biaya logistik yang terkait dengan tetap berada di pelabuhan untuk menampung barang, yang pada akhirnya menghasilkan peningkatan biaya logistic, biaya logistik perusahaan dapat dipengaruhi jika pengaturan waktu tunggu dapat diterapkan [5]. Waktu tunggu yang dibutuhkan Importir untuk pengeluaran barang impor pada proses Custom Clearance dapat sangat bervariasi. Proses panjang yang dilalui sebuah kapal saat sudah tiba di pelabuhan menjadi salah satu hambatan bagi importir untuk melakukan pengeluaran barang secara cepat.

Faktor-faktor penyebab keterlambatan pengeluaran barang impor pada lapangan penumpukan kontainer pelabuhan, disebabkan oleh faktor eksternal dan internal, faktor internal yaitu : telatnya penyerahan PIB, dokumen impor yang masih belum lengkap, waktu pembayaran pajak untuk pengeluaran barang yang sering tidak tepat waktu, sedangkan faktor eksternal, yaitu: status jalur merah, dokumen sppb, barang impor sppb, trucking, traffic, sistem autogate dan jaringan, cuaca, dan biaya penumpukan [6]. Beberapa faktor internal dan eksternal adalah variabel dalam study kasus ini, Dimana faktor internal yaitu 7 atribut yang berasal dari dalam perusahaan atau pihak yang melakukan proses impor diantaranya, Nama Barang, Nama Perusahaan, Nomor Container, Jumlah Container, Tanggal Pemesanan Barang, Tanggal Kelengkapan Dokumen, Tanggal Pembayaran Pajak. Dan faktor eksternal yaitu 6 atribut yang berasal dari luar perusahaan, seperti instansi pemerintah, otoritas pelabuhan, atau kondisi logistik yang tidak bisa dikendalikan langsung oleh Perusahaan diantaranya, Tanggal Sandar Kapal, Status Periksa Barang, Tanggal Keluar Barang, Port Of Loading, Port Of Discharge, Waktu Tunggu.

Berbagai aspek dinamika peramalan atau prediksi waktu tunggu dari berbagai objek dan subjek telah dibahas dalam penelitian sebelumnya. Misalnya penelitian [7] membahas tentang Tingkat idle time di Terminal Petikemas Pelabuhan Tanjung Perak- Surabaya. periode Januari – Desember 2019 sebesar 0,47 yaitu pada bulan Desember 2019. Kemudian nilai Idle Time (IT) tertinggi sebesar 2,88 yaitu pada bulan Februari 2019 dengan menggunakan regresi sederhana.

Selanjutnya penelitian [8] menyebutkan atribut jam, kondisi cuaca, waktu tempuh, hasil dari penelitian ini menghasilkan prediksi dengan nilai error RMSE sebesar 3.80% dengan menggunakan Regresi Linear Decomposition untuk peramalan arus lalu lintas berdasarkan waktu tempuh dan cuaca.

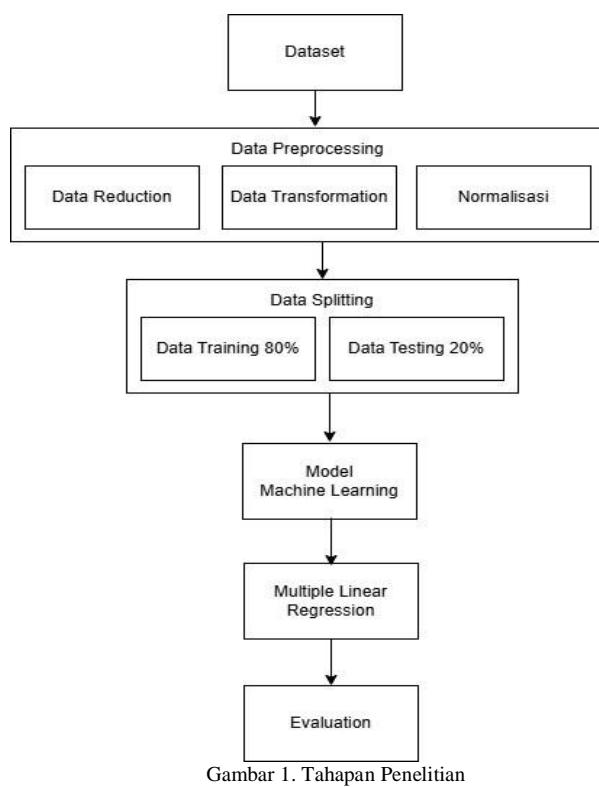
Kemudian penelitian [9] melakukan analisis menggunakan regresi berganda, untuk menganalisis penyebab waktu idle pada operasi bongkar muat di terminal kontainer pelabuhan internasional Indonesia, pada Tanjung Priok, Terdapat 6 kategori variabel kegagalan pada idle time yaitu kepala truk, sistem komputer, cuaca, pekerja bongkar muat, kegagalan peralatan dan aktivitas operator. Regresi linier berganda ini telah banyak digunakan dalam beberapa studi penelitian seperti peramalan atau prediksi dengan objek yang berbeda-beda [10]. Penelitian [11] menggunakan Backpropagation Neural Network dalam melakukan peramalan layanan kapal tunda dipelabuhan Tanjung Perak, beberapa variabel yang digunakan yaitu data permintaan pelayanan kapal tunda, bulan, jenis kapal, antrian, waktu tunggu hasil yang didapatkan yaitu menghasilkan nilai MAPE sangat baik dibawah 10%.

Beberapa penelitian tersebut menggunakan algoritma regresi, neural network dalam melakukan analisis maupun peramalan dalam memprediksi waktu tunggu. Namun sebagian fokus pada atribut dan pengaruh, sehingga belum ada yang melakukan peramalan waktu tunggu itu sendiri dengan machine learning lainnya dengan objek yang berbeda. Penelitian focus pada peramalan waktu tunggu pengeluaran barang impor pada logistic Perusahaan imporir dengan melakukan percobaan machine learning yang lebih dominan dan layak untuk penelitian ini. Karna banyak variabel independent atau faktor-faktor yang mempengaruhi peramalan ini, maka studi kasus ini menggunakan algoritma Multiple Regresi Linear, kemudian melakukan uji coba dengan algoritma Neural Network, dan menambahkan satu algoritma Support Vector Machine untuk melihat sejauh mana algoritma bekerja untuk menghasilkan model dan evaluasi yang terbaik dalam melakukan peramalan waktu tunggu. Sehingga permintaan peramalan waktu tunggu dimasa yang akan datang dapat memberikan kontribusi yang positif bagi perusahaan importir dalam meningkatkan efisiensi, baik dari segi waktu maupun biaya. Penelitian ini juga berkontribusi pada penambahan literatur akademik tentang penerapan machine learning dalam kasus logistik impor, Hasil dari penelitian ini diharapkan dapat mengidentifikasi algoritma peramalan yang paling sesuai untuk melakukan peramalan waktu tunggu pengeluaran barang impor menggunakan Multiple Linear Regression.

II. METODE

Penelitian ini menggunakan beberapa tahapan mulai dari pengumpulan dataset, data preprocessing yang meliputi data reduction, data transformation, normalisasi, selanjutnya data splitting meliputi data training dan testing, kemudian model machine learning dengan menggunakan ujicoba dari

algoritma Multiple Linear Regression, Neural Network, dan Support Vector Machine, hingga evaluasi dan hasil akhir dari penelitian, dapat dilihat pada Gambar 1.



A. Dataset

Data yang digunakan pada kasus ini adalah Dataset yang dari Januari tahun 2019 sampai Desember tahun 2023 dari PT. SS Logistic, sebanyak 3.005 record data dan 13 atribut data, diantaranya Nama Barang, Nama Perusahaan Impor, Nomor Container, Jumlah Container, Tanggal Pemesanan Barang, Tanggal Sandar Kapal, Tanggal Kelengkapan Dokumen, Tanggal Pembayaran Pajak, Status Pemeriksaan Barang, Tanggal Keluar Barang, Port Of Loading (POL), Port Of Discharge (POD), Waktu Tunggu. Atribut-atribut data ini menunjukkan variabel-variabel yang mempengaruhi waktu tunggu pengeluaran barang impor pada PT. SS Logistic.

B. Data Preprocessing

Prapeprosesan data merupakan langkah penting dalam pengembangan model sebelum pembuatan model text mining [12]. Proses yang dilakukan yaitu, data reduction, data transformation, dan normalisasi data.

1) *Data Reduction*: Reduksi data dengan mengumpulkan data kemudian memilih-milah dalam satuan konsep, kategori, dan tema tertentu [13]. Reduksi data mengurangi atribut-atribut atau dimensi pada data yang tidak digunakan dan tidak relevan dalam melakukan peramalan. Data yang dikumpulkan memiliki data kompleks dan cakupan yang cukup besar. Manfaat dari data reduction adalah memungkinkan analisis yang lebih efisien dan efektif terhadap dataset yang besar dan

kompleks. Dengan menghilangkan data yang tidak relevan serta mengurangi dimensi yang tidak perlu, proses data mining dapat lebih fokus pada pola dan informasi yang penting dalam pengambilan keputusan. Beberapa atribut yang direduksi ada 6 yaitu nama barang, nama perusahaan impor, nomor container, tanggal pemesanan barang, Port of Loading (POD), Port Of Discharge (POD).

2) *Data Transformation*: Setelah data direduksi, selanjutnya pengubahan bentuk data akan dilakukan. Data dalam bentuk kata akan diubah menjadi data dengan nilai numerik dengan menggunakan peta pengkodean dan pembentukan data baru. Ini dilakukan untuk pembelajaran mesin learning dapat memahami data [14]. Pada fase ini pembentukan tabel atribut baru menjadi 3 data atribut baru diantaranya yaitu : atribut lama Tanggal Kelengkapan Dokumen dan Tanggal Sandar Kapal menjadi atribut baru Waktu Pengurusan Dokumen, selanjutnya atribut lama Tanggal Kelengkapan Dokumen dan Tanggal Pembayaran Pajak menjadi atribut baru Waktu Pembayaran Pajak, kemudian atribut lama Tanggal Pembayaran Pajak dan Tanggal Keluar Barang menjadi Waktu Periksa Barang.

3) *Normalisasi*: Normalisasi fitur adalah langkah terakhir dalam preprocessing data, yang dilakukan untuk menyeimbangkan variasi nilai dari berbagai fitur dalam dataset dan meningkatkan kinerja model [15].

4) *Data Splitting*: Pembagian data dibagi menjadi dua bagian yaitu data training (data latih) dan data testing (data uji). Data pelatihan digunakan untuk melatih model machine learning, dan data pengujian digunakan untuk menguji kinerja model pada data yang belum diuji sebelumnya [16]. Data dibagi menjadi dua bagian yaitu data training (80%) dan data testing (20%) [15].

C. Model Machine Learning

Machine learning adalah cabang kecerdasan buatan, tentang cara komputer dapat belajar dari data untuk menjadi lebih cerdas. [17]

Pada tahap ini, model dibuat untuk mengolah data yang telah dipreprocessing. Tahap ini adalah untuk memaksimalkan kinerja model terbaik dari atribut/variabel yang digunakan. Pada pemodelan ini menggunakan machine learning dengan percobaan berbagai model algoritma untuk melakukan peramalan pada waktu tunggu yaitu Multiple Linear Regression (MLR), Neural Network (NN) dan Support Vector Machine (SVM), dengan menggunakan bahasa pemrograman python.

Multiple Linear Regression: Regresi linear berganda merupakan metode prediksi yang melibatkan dua atau lebih variabel-variabel pemberi pengaruh dan variabel terpengaruh. Ada hubungan sebab akibat antara variabel-variabel tersebut atau saling terkait. Bagaimana variabel tersebut berhubungan satu sama lain akan dijelaskan dengan menggunakan teknik regresi [18]. Dalam kasus ini, regresi linier berganda, yang melibatkan lebih dari satu variabel independen dan satu variabel dependen, yang digunakan. Persamaan regresi linier berganda adalah :

$$y = + b_1x_1 + b_2x_2 \dots + b_nx_n \quad (1)$$

Keterangan :

y = Variabel dependen (nilai yang diramalkan)
 x_1, x_2, \dots, x_n = Variabel independen
 a = Konstanta
 b_1, b_2, \dots, b_n = Koefisien regresi

D. Evaluation

Pada Teknik evaluasi digunakan teknik pengujian berupa Root Mean Square Error (RMSE) untuk mengukur nilai error yang dihasilkan dan R Square (R²) untuk mengukur akurasi nilai error terbaik yang dihasilkan dari perbandingan 3 algoritma yaitu, algoritma Multiple Linear Regression, Support Vector Machine dan Neural Network.

1) *Root Mean Square Error (RMSE)* : RMSE merupakan akar kuadrat dari kuadrat kesalahan rata-rata yang dihasilkan dari perhitungan. Jika hasil RMSE semakin rendah nilai error maka hasil prediksinya akan semakin baik [19]. Nilai RMSE yang rendah menunjukkan bahwa nilai yang dihasilkan suatu model perkiraan lebih dekat ke nilai aslinya, sementara nilai RMSE yang lebih besar menunjukkan bahwa nilai yang dihasilkan suatu model perkiraan lebih tidak akurat Persamaan RMSE adalah.

$$RMSE = \sqrt{\frac{1}{n} \sum (Y_t - \hat{Y}_t)^2}$$

Keterangan :

Y_t = data aktual periode t

\hat{Y}_t = data hasil peramalan periode t n = banyaknya data

2) *Mean Absolute Error (MAE)*: MAE adalah metrik evaluasi untuk mengukur selisih absolut rata-rata antara nilai aktual dan prediksi dalam regresi [20]. Berikut pada persamaan x merupakan rumus MAE.

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

Keterangan:

(\hat{y}_i) = rating yang actual

y_i = merupakan rating dari hasil prediksi

n = jumlah keseluruhan data.

3) *R Squara (R²)*: Kesesuaian suatu model dapat menerapkan persamaan koefisien determinasi R Square (R²). Dinamakan koefisien determinasi karena R² x 100% daripada variasi yang terjadi dalam variabel tak bebas Y dapat dijelaskan oleh variabel bebas X dengan adanya regresi linier Y atas X . Besarnya nilai koefisien determinasi adalah berkisar $0 < R^2 < 1$. Artinya jika R² mendekati 1 maka dapat dikatakan pengaruh variabel bebas terhadap variabel ikatan adalah sangat kuat. Berikut interpretasi koefisien determinasi [18]. Persamaan R² adalah.

$$Kd = R^2 \times 100\%$$

Keterangan :

Kd = besar atau jumlah koefisien determinasi

R² = nilai koefisien korelasi

TABEL I
INTERPRETASI KOEFISIEN DETERMINASI

Interval Koefisien	Tingkat Hubungan
0,80 – 1,000	Sangat Kuat
0,60 – 0,799	Kuat
0,40 – 0,599	Cukup Kuat
0,20 – 0,399	Rendah
0,00 – 0,199	Sangat Rendah

III. HASIL DAN PEMBAHASAN

Dalam kasus ini, dengan menggunakan eksperimen pelatihan dan uji dengan 6 parameter atribut data dari hasil preprocessing data, dan menggunakan percobaan dengan 3 algoritma yang berbeda. Kemudian hasil yang diperoleh dari setiap pemodelan dilakukan optimasi dengan Root Mean Square Error (RMSE), Mean Absolute Error (MAE) dan R Square (R²). Hasil error dengan nilai terbaik akan digunakan untuk peramalan waktu tunggu pengeluaran barang impor.

A. Dataset

Dataset yang digunakan dalam penelitian yaitu dari Januari tahun 2019 sampai Desember tahun 2023 dari importir PT. SS Logistic, sebanyak 3.005 record baris data dan 13 kolom atribut data. Dataset awal yang diperoleh memiliki 13 atribut data yang diantaranya, Nama Barang, Nama Perusahaan Impor, Nomor Container, Jumlah Container, Tanggal Pemesanan Barang, Tanggal Sandar Kapal, Tanggal Kelengkapan Dokumen, Tanggal Pembayaran Pajak, Status Pemeriksaan Barang, Tanggal Keluar Barang, Port Of Loading (POL), Port Of Discharge (POD), Waktu Tunggu. Contoh dataset yang digunakan dapat dilihat pada Tabel II.

TABEL II
DATASET AWAL

Nama Barang	Jumlah Container	Tanggal Pemesanan Barang	Tanggal Sandar Kapal	...	Waktu Tunggu
Stainless Steel Sheet	1	20/02/2019	10/03/2019	...	2
Stainless Steel Coil	3	01/03/2019	14/03/2019	...	4
Pvc Sheet	2	01/03/2019	20/03/2019	...	0
Roofing Nail	2	12/03/2019	26/03/2019	...	1
Electric Winch	1	10/03/2019	30/03/2019	...	2

B. Data Preprocessing

Proses selanjutnya adalah proses penyiapan data atau data preprocesing, data yang telah dilakukan seleksi dan disiapkan untuk proses tahapan selanjutnya sebelum dilakukan modeling. Tahapan ini dilakukan untuk memenuhi persyaratan proses data mining dengan melakukan langkah-langkah preprocessing. Kegiatan preprocessing dilakukan mulai dari data reduction sampai normalisasi data.

1) *Data Reduction:* Langkah yang dilakukan pada tahapan ini yaitu, melakukan penyaringan fitur atau atribut data yang tidak relevan. Pada Tabel III terdapat faktor internal yaitu 7 atribut yang berasal dari dalam perusahaan atau pihak yang melakukan proses impor diantaranya, Nama Barang, Nama Perusahaan, Nomor Container, Jumlah Container, Tanggal Pemesanan Barang, Tanggal Kelengkapan Dokumen, Tanggal Pembayaran Pajak. Dan faktor eksternal yaitu 6 atribut yang berasal dari luar perusahaan, seperti instansi pemerintah, otoritas pelabuhan, atau kondisi logistik yang tidak bisa dikendalikan langsung oleh Perusahaan diantaranya, Tanggal Sandar Kapal, Status Periksa Barang, Tanggal Keluar Barang, Port Of Loading, Port Of Discharge, Waktu Tunggu.

Selanjutnya dari kolom beberapa atribut yang direduksi atau dikurangi ada 6 yaitu Nama Barang, Nama Perusahaan Impor, Nomor Container, Tanggal Pemesanan Barang, Port of Loading (POD), Port Of Discharge (POD) dihilangkan untuk membuat model lebih cepat dilatih dan membantu mencegah masalah overfitting, yang sering terjadi dengan data yang sangat besar. Data tersebut dapat dilihat pada Tabel III.

TABEL III
ATRIBUT DATA

No.	Nama Atribut	Type Data	Faktor
1	Nama Barang	String	Internal
2	Nama Perusahaan	String	Internal
3	Nomor Container	String	Internal
4	Jumlah Container	Int	Internal
5	Tanggal Pemesanan Barang	String	Internal
6	Tanggal Kelengkapan Dokumen	String	Internal
7	Tanggal Pembayaran Pajak	String	Internal
8	Tanggal Sandar Kapal	String	Eksternal
9	Status Periksa	String	Eksternal
10	Tanggal Keluar Barang	String	Eksternal
11	Port Of Loading	String	Eksternal
12	Port Of Discharge	String	Eksternal
13	Waktu Tunggu	Int	Eksternal

2) *Data Transformation:* Setelah melakukan reduksi selanjutnya yaitu proses yang dilakukan pada tahap ini adalah transformasi data atau proses pembentukan atribut baru dan mengubah data menjadi type data integer dari data terpilih. Pada fase ini pembentukan tabel atribut baru menjadi 3 data atribut baru yaitu: atribut lama Tanggal Kelengkapan Dokumen dan Tanggal Sandar Kapal menjadi atribut baru Waktu Pengurusan Dokumen ditunjukkan pada Tabel IV, selanjutnya atribut lama Tanggal Kelengkapan Dokumen dan Tanggal Pembayaran Pajak menjadi atribut baru Waktu Pembayaran Pajak ditunjukkan pada Tabel V, kemudian atribut lama Tanggal Pembayaran Pajak dan Tanggal Keluar Barang menjadi Waktu Periksa Barang ditunjukkan pada Tabel VI. Dan transformasi status pemeriksaan barang, merah jadi 0 dan hijau jadi 1 dapat dilihat pada Tabel VII.

Selanjutnya contoh data yang diubah adalah beberapa data string diubah menjadi data integer yang mudah diolah. Hal ini

untuk memastikan keakuratan data yang akan digunakan dalam analisis dan peramalan waktu tunggu untuk memberikan gambaran yang spesifik tentang perbaikan dan kualitas data.

TABEL IV
CONTOH TRANSFORMASI WAKTU PENGURUSAN DOKUMEN

Tanggal Kelengkapan Dokumen	Tanggal Sandar Kapal	Waktu Pengurusan Dokumen
25/01/2019	23/01/2019	2
25/01/2019	25/01/2019	0

TABEL V
CONTOH TRANSFORMASI WAKTU PEMBAYARAN PAJAK

Tanggal Pembayaran Pajak	Tanggal Kelengkapan Dokumen	Waktu Pembayaran Pajak
14/12/2020	12/12/2020	2
14/12/2020	13/12/2020	1

TABEL VI
CONTOH TRANSFORMASI WAKTU PERIKSA BARANG

Tanggal Keluar Barang	Tanggal Pembayaran Pajak	Waktu Periksa Barang
15/12/2022	09/12/2022	6
09/12/2022	09/12/2022	0

TABEL VII
CONTOH TRANSFORMASI STATUS PEMERIKSAAN BARANG

Status Periksa Barang	Status
MERAH	0
HIJAU	1

Sehingga setelah data ditransformasi dan direduksi menghasilkan atribut-atribut data baru yang digunakan untuk melakukan pemodelan yaitu, terdapat faktor internal berasal dari aktivitas, keputusan, atau proses internal perusahaan importir, Jumlah Container, Waktu Pengurusan Dokumen, Waktu Pembayaran Pajak. Faktor eksternal yang berasal dari luar perusahaan, seperti otoritas pelabuhan, bea cukai, atau kondisi lapangan/logistik yang tidak bisa dikendalikan langsung oleh perusahaan. Yaitu, Status Periksa, Waktu Periksa Barang, dan Waktu Tunggu dapat dilihat pada Tabel VIII.

TABEL VIII
ATRIBUT DATA BARU

No.	Nama Atribut	Type Data	Faktor
1	Jumlah Container	Int	Internal
2	Waktu	Pengurusan Dokumen	Int
3	Waktu	Pembayaran Pajak	Int
4	Status Periksa	Int	Eksternal
5	Waktu	Periksa Barang	Int
6	Waktu Tunggu	Int	Eksternal

Selanjutnya hasil dari seluruh transformasi data dan pembentukan atribut baru dapat dilihat pada Tabel IX. Dengan 6 atribut data dan type data integer.

TABEL IX
CONTOH DATA SETELAH TRANSFORMASI

Jumlah Container	Waktu Pengurusan Dokumen	Waktu Pembayaran Pajak	Status Periksa	Waktu Periksa Barang	Waktu Tunggu
1	2	1	1	3	2
2	2	5	1	0	3
2	1	4	1	1	4
3	0	3	0	0	3
1	1	5	1	2	6

Setelah dilakukan transformasi data, selanjutnya data di normalisasikan untuk memaksimalkan hasil pemodelan yang baik yang mengarah pada peningkatan akurasi dan kualitas kinerja model yang dibangun dalam machine learning dan analisis data. Normalisasi data ditunjukkan pada Tabel X.

TABEL IX
CONTOH NORMALISASI DATA

Jumlah Container	Waktu Pengurusan Dokumen	Waktu Pembayaran Pajak	Status Periksa	Waktu Periksa Barang	Waktu Tunggu
0.000000	0.426471	0.229630	0.0	0.0465	4
0.000000	0.411765	0.229630	0.0	0.0465	5
0.057143	0.441176	0.222222	1.0	0.0465	2
0.028571	0.411765	0.222222	0.0	0.0465	4
0.028571	0.441176	0.214815	0.0	0.0465	1

C. Data Splitting

Jumlah data yang digunakan untuk melakukan eksperimen adalah 3.005 record data dan data dibagi menjadi dua bagian yaitu data training (80%) dan data testing (20%) [21]. Proporsi pelatihan data dapat dilihat pada Tabel XI.

TABEL XI
PROPORSI DATA TRAINING DAN TESTING

Keterangan	Training	Testing	Jumlah
Proporsi	80%	20%	100%
Jumlah	2.404	601	3005

D. Model Machine Learning

Pada tahap pemodelan dilakukan untuk meramalkan dan memahami hubungan setiap variabel. Proses modelling dilakukan dengan menggunakan bahasa pemrograman python. Modelling dilakukan dengan cara mengimpor dataset, kemudian import pandas as pd pada Library python from sklearn.model_selection import train_test_split. Memisahkan data training dan data testing dengan X_train, X_test, y_train, y_test =train_test_split(x,y,test_size=0.2,random_state=42). from sklearn.linear_model import Linear Regression. Data pelatihan yang dilakukan untuk membangun model algoritma dalam memprediksi kelas baru yaitu waktu tunggu pengeluaran barang impor.

Proses yang dilakukan pada tahap ini adalah dengan membagi data set training sebagai data pelatihan untuk melatih model regresi linear berganda yaitu dari 3005 data, data training yang di latih adalah 80 % yaitu 2404 data selanjutnya dilatih untuk menghasilkan prediksi peramalan data testing dari 20% yaitu 601 data. Data input yang digunakan pada tahap modelling yaitu dapat dilihat pada Tabel XII.

TABEL XII
CONTOH DATA ATRIBUT MODELLING

Nama Atribut	Type Data
Jumlah Container	Int
Waktu Pengurusan Dokumen	Pengurusan Dokumen
Waktu Pembayaran Pajak	Pembayaran Pajak
Status Periksa	Int
Waktu Periksa Barang	Periksa Barang
Waktu Tunggu	Int

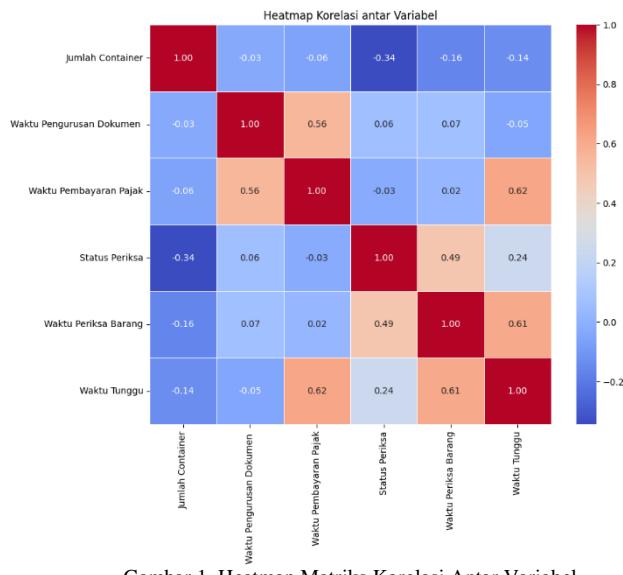
Hasil nilai yang didapatkan pada pemodelan 3 algoritma yaitu, dengan Multiple Linear Regression (MLR) memiliki nilai RMSE sebesar 0.08131 dan nilai MAE sebesar 0.06775, ini menunjukkan bahwa nilai error yang lebih rendah dengan ubungan antar variabel kemungkinan linear, dan MLR mampu memodelkannya dengan baik. Selanjutnya algoritma Support Vector Machine (SVM) memiliki nilai RMSE sebesar 0,09207, dan nilai MAE sebesar 0,07672, SVM dengan nilai error yang sedikit tinggi tidak seakurat MLR. Kemudian algoritma Neural Network (NN) memiliki nilai RMSE sebesar 5.12610 dan nilai MAE sebesar 4.27175 dengan nilai error sangat tinggi disbanding algoritma sebelumnya. Hasil nilai perbandingan algoritma dapat dilihat pada Tabel XIII.

TABEL XIII
MODELLING MACHINE LEARNING

Modelling	Multiple Linear Regression	Support Vector Machine	Neural Network
RMSE	0.08131	0.09207	5.12610
MAE	0.06775	0.07672	4.27175

Selanjutnya dilakukan dilakukan kontribusi heatmap korelasi matriks pada atribut-atribut karena hasil nilai error yang didapatkan masih kurang optimal dan beberapa ada yang overfitting.

Gambar 1. heatmap matriks korelasi antar variabel memberikan gambaran visual tentang hubungan antar variabel atau atribut untuk menghindari overfitting dari data yang tidak relevan dalam peralaman, kemudian hasil korelasi untuk menentukan tahap lanjutan evaluasi model terhadap variabel-variabel yang memengaruhi Waktu Tunggu.



Gambar 1. Heatmap Matriks Korelasi Antar Variabel

Seperti yang dijelaskan sebelumnya, fitur-fitur yang memiliki korelasi absolut yang lebih besar dari ambang batas 0,20 dipilih untuk analisis lanjutan. Dari hasil korelasi tersebut diperoleh atribut Waktu Tunggu, Waktu Pembayaran Pajak, Waktu Periksa Barang dan Status Periksa. Nilai korelasi antar atribut dapat dilihat pada Tabel XIV.

TABEL XIV
HASIL KORELASI

Nama Atribut	Nilai Korelasi	Tahap Lanjutan Evaluasi
Waktu Tunggu	1.00	Ya
Waktu	0.62	Ya
Waktu Periksa Barang	0.61	Ya
Status Periksa Barang	0.24	Ya
Waktu	-0.05	Tidak
Status Periksa	-0.14	Tidak

E. Evaluation Model

Pada machine learning, setelah dilakukan percobaan pemodelan algoritma lain dengan menemukan performa model terbaik untuk dilakukan peramalan pada tahap selanjutnya. Pelatihan pemodelan yang sudah dilakukan pada data training 80% proporsi data. Hasil dari percobaan 3 model algoritma menunjukkan bahwa algoritma Multiple Linear Regression (MLR), yang memiliki performa model terbaik dari percobaan algoritma lainnya, Neural Network (NN) dan Support Vector Machine (SVM).

Kemudian pada tahap evaluasi, data testing digunakan 20% untuk mengukur seberapa baik model mampu memprediksi nilai pada data yang tidak terlibat dalam proses pelatihan. Dengan memastikan bahwa model tidak hanya mengingat data training tetapi juga mampu menggeneralisasi pola yang ditemui dalam data baru.

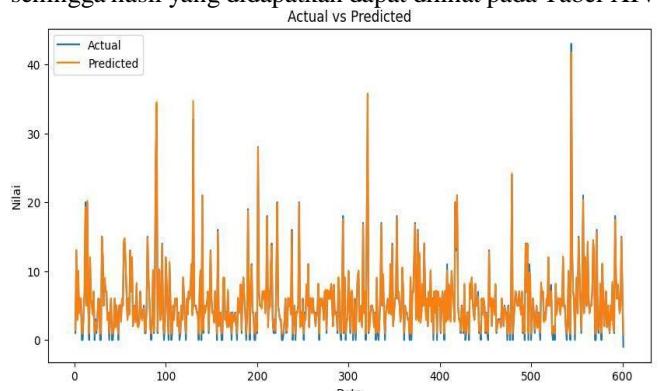
Hasil nilai yang didapatkan pada tahap evaluasi 3 algoritma yaitu, dengan Multiple Linear Regression (MLR) memiliki

nilai RMSE sebesar 0.00653 nilai error yang sangat kecil, dan MAE sebesar 0.00544 nilai sangat akurat dan nilai uji R-square : 0.99999 nilai yang sangat baik dan hampir sempurna. Artinya tahap evaluasi model ini adalah yang paling akurat karena prediksinya hampir identik dengan data asli dan bagus untuk situasi di mana ada hubungan linear yang kuat. Selanjutnya algoritma Support Vector Machine (SVM) memiliki nilai RMSE sebesar 0.74107, nilai MAE sebesar 0.61756, dan nilai uji R-square : 0.98388 dengan hasil masih sangat baik. Akurasinya cukup tinggi, tetapi tidak sebaik MLR. Bisa digunakan bila ada nonlinearitas ringan. Cocok untuk data kompleks, tetapi kurang optimal. Kemudian algoritma Neural Network (NN) memiliki dengan nilai error yang tetap masih tinggi yaitu RMSE sebesar 2.86599, nilai MAE sebesar 2.38832, dan nilai uji R-square : 0.69510 cukup rendah. Hasil nilai evaluasi model 3 algoritma dapat dilihat pada Tabel XV.

TABEL XV
EVALUASI MODEL

Evaluation	Multiple Linear Regression	Support Vectore Machine	Neural Network
RMSE	0.00653	0.74107	2.86599
MAE	0.00544	0.61756	2.38832
R-Squared	0.99999	0.98388	0.69510

Kemudian model MLR digunakan untuk dilakukan peramalan Waktu Tunggu pengeluaran barang impor, sehingga hasil yang didapatkan dapat dilihat pada Tabel XIV.



Gambar 2. Confusion Matrix Actual dan Predicted

TABEL XIV
HASIL PERBANDINGAN DATA ACTUAL DENGAN PREDICTED

No	Actual	Predicted
1	1	1.00156
2	13	12.99544
3	3	3.00095
4	10	9.99954
5	4	3.99976
****	****	****
596	6	5.99980
597	4	4.00089
598	5	5.00198
599	15	14.99653
600	7	7.00021

Pada Tabel XIV dan Gambar 2 adalah Confusion Matriks perbandingan hasil data prediksi dan actual dimana gap predicted dan actual hanya kurang lebih 0,01 sehingga titik pada gambar nya tidak akan terlalu terlihat. Artinya model yang dibangun sangat baik dan maksimal sehingga hasil peramalan dan actual hampir sama dengan nilai prediksinya.

V. KESIMPULAN

Penelitian ini berhasil mengembangkan dan mengevaluasi model prediksi waktu tunggu pengeluaran barang impor menggunakan pendekatan machine learning dengan studi kasus pada proses logistik impor di PT. Sentra Sarana Logistic. Tiga algoritma dibandingkan, yaitu Multiple Linear Regression (MLR), Support Vector Machine (SVM), dan Neural Network (NN), dengan mempertimbangkan sejumlah variabel seperti waktu pembayaran pajak, waktu pemeriksaan, status pemeriksaan, dan jenis layanan impor. Berdasarkan hasil evaluasi, MLR menunjukkan kinerja terbaik dengan nilai RMSE sebesar 0,00653, MAE 0,00544, dan R-squared sebesar 0,99999, mengindikasikan tingkat akurasi yang sangat tinggi dan korelasi linear yang kuat antara variabel. SVM memberikan hasil yang cukup baik, sedangkan NN menunjukkan performa yang paling rendah dalam konteks data yang dianalisis.

Hasil ini menegaskan bahwa model statistik yang sederhana namun interpretatif seperti MLR dapat menjadi alat prediktif yang efektif dalam mendukung pengambilan keputusan operasional di bidang logistik impor. Selain kontribusi praktis, penelitian ini juga memperkaya literatur terkait penerapan algoritma machine learning dalam optimalisasi proses logistik dan kepabeanan. Untuk kedepan, penelitian ini dapat dikembangkan lebih lanjut dengan melibatkan variabel eksternal yang lebih kompleks atau dengan pendekatan hybrid untuk meningkatkan generalisasi model pada skenario yang lebih dinamis..

DAFTAR PUSTAKA

- [1] S. Supartini, I. Iswanto, N. Astriawati, V. Dekanawati, and N. K. Hamzah Alfanzuri, "Pelayanan Jasa Impor Barang Dalam Masa Pandemi," *Dinamika Bahari*, vol. 3, no. 2, pp. 114–123, Oct. 2022, doi: 10.46484/db.v3i2.313.
- [2] I. Sani and S. Nanda Lubis, "Proses Pengurusan Dokumen Barang Ekspor Pada PT. Bahari Eka Nusantara Cabang Medan," *Journal of Maritime and Education (JME)*, vol. 4, no. 1, pp. 358–363, Feb. 2022, doi: 10.54196/jme.v4i1.70.
- [3] Hotrahelita Girsang and Dafid Ginting, "Peranan Divisi Operasional dalam Kegiatan Ekspor Impor pada Pt.Prima Indonesia Logistic Belawan," *Journal Of Social Science Research*, vol. 3, no. 6, pp. 9533–9551, 2023.
- [4] R. K. D. B. M. K. Karjono1, "Kesiapan Perusahaan Pengurusan Jasa Kepabeanan (Ppjk) Dalam Kegiatan Impor Barang Pada PT. Terminal Intimoda Utama Cabang Semarang", Refi, T.M. and Alhabsi, M., 2022. Prosedur Pengurusan Barang Impor Pada Kantor Pengawasan dan Pelayanan Bea Cukai Tipe Madya Pabean C Langsa Kota Langsa., 1(1), pp.21-25., pp. 21–25, 2022.
- [5] R. Sains, D. T. Kelautan, A. Ramadani, and C. Virandika, "Analisa Pengaruh Waktu Tunggu Kapal Terhadap Biaya Logistik Di Terminal Petikemas Makassar," 2022.
- [6] K. Verawati, H. Rahmayanti, W. Hadi, A. Costa, and S. Almira, "Faktor Penyebab Keterlambatan Pengeluaran Barang Impor Pada Lapangan Penumpukan Wilayah 2 Terminal Multipurpose PT Pelabuhan Tanjung Priok," *Jurnal Penelitian Transportasi Laut*, vol. 24, no. 1, pp. 21–32, Sep. 2022, doi: 10.25104/transla.v24i1.2022.
- [7] A. Lalita Ivana and H. Moetriono, "Optimasi Waktu Sandar Kapal Untuk Meningkatkan Kinerja Pelayanan Di Terminal Jamrud Pelabuhan Tanjung Perak Surabaya," 2021.
- [8] M. Hudzaifah and A. A. Rismayadi, "Peramalan Arus Lalu Lintas Berdasarkan Waktu Tempuh Dan Cuaca Menggunakan Metode Time Series Decomposition," *JURNAL RESPONSI*, vol. 3, no. 2, pp. 207–215, 2021. Available: <http://ejurnal.ars.ac.id/index.php/jti>
- [9] F. Suprata, C. Natalia, and A. Sugihoko, "Analysing the cause of idle time in loading and unloading operation at Indonesian international port container terminal: Port of Tanjung Priok case study," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, May 2020. doi: 10.1088/1757-899X/847/1/012090.
- [10] N. Nurdin, F. Fajriana, M. Maryana, and A. Zanati, "Information System for Predicting Fisheries Outcomes Using Regression Algorithm Multiple Linear," *Journal Of Informatics And Telecommunication Engineering*, vol. 5, no. 2, pp. 247–258, Jan. 2022, doi: 10.31289/jite.v5i2.6023.
- [11] U. Masruroh Kusman, A. Hamid, D. Candra Rini Novitasari, W. Dianita Utami, and I. Ariyanto Wijaya, "Optimasi Model Penugasan Berdasarkan Peramalan Layanan Kapal Tunda Di Pelabuhan Tanjung Perak Menggunakan Metode Backpropagation," 2023.
- [12] J. Setyanto and T. B. Sasongko, "Sentiment Analysis of Sirekap Application Users Using the Support Vector Machine Algorithm," 2024. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [13] A. D. A. E. S. F. E. R. Ahlan Syaeful Millah, "Analisis Data dalam Penelitian Tindakan Kelas," *Jurnal Kreativitas Mahasiswa*, vol. 1, no. 2, pp. 140–153, 2023.
- [14] Y. Christian, "Rental Price Prediction of Boarding Houses in Batam City Using Linear Regression and Random Forest Algorithms," 2023. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [15] E. Wijaya, M. Arief Soeleman, and N. Andono, "Comparative Performance Analysis of Optimization Algorithms in Artificial Neural Networks for Stock Price Prediction," 2025. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [16] R. F. Ramadhan and W. M. Ashari, "Performance Comparison of Random Forest and Decision Tree Algorithms for Anomaly Detection in Networks," 2024. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [17] B. J. Rana, N. A. Setiyanto, and M. Akrom, "Prediction of Corrosion Inhibitor Efficiency Based on Quinoxaline Compounds Using Polynomial Regression," 2025. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [18] A. Novebrian Maharadja, I. Maulana, and B. Arif Dermawan, "Penerapan Metode Regresi Linear Berganda untuk Prediksi Kerugian Negara Berdasarkan Kasus Tindak Pidana Korupsi," 2021. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [19] R. S. Ratih Yulia Hayuningtyas, "Implementasi Data Mining Dengan Algoritma Multiple Linear Regression Untuk Memprediksi Penyakit Diabetes," *Jurnal Teknik Komputer AMIK BSI*, vol. 8, no. 1, pp. 40–44, Jan. 2022.
- [20] M. M. Dewi, R. Andriani, and M. Nuraminudin, "Performance Analysis of the Item-Based Collaborative Filtering Model in Yogyakarta Tourism Recommendations," 2025. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [21] A. Dewantoro and T. B. Sasongko, "Comparison of LSTM Model Performance with Classical Regression in Predicting Gaming Laptop Prices in Indonesia," 2024. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>