

PREDIKSI STATUS PENGIRIMAN BARANG MENGGUNAKAN METODE *MACHINE LEARNING*

Hardian Kokoh Pambudi¹, Putu Giri Artha Kusuma², Femi Yulianti³, Kevin Ahessa Julian⁴

Program Studi Teknik Logistik, Fakultas Rekayasa Industri^{1,2,3,4}

Telkom University

Jl. Telekomunikasi no.1, Terusan Buah Batu, Dayeuhkolot, Kab. Bandung

hkpambudi@telkomuniversity.ac.id¹, putugiriak@telkomuniversity.ac.id², femiyulianti@telkomuniversity.ac.id³,

kevinahessajulian@student.telkomuniversity.ac.id⁴

Abstrak

Salah satu ukuran kinerja bagi industri logistik, khususnya perusahaan pengiriman barang, adalah ketepatan waktu penghantaran. Hal ini masih menjadi tantangan bagi perusahaan guna menjamin tingkat kepuasan pelanggan dan menurunkan biaya transportasi. Di sisi lain, perkembangan teknologi informasi saat ini memungkinkan organisasi atau perusahaan dapat mengumpulkan data dalam jumlah besar secara otomatis. Metode yang cukup andal yang dapat digunakan dalam melakukan analisis data prediksi adalah *machine learning*, yaitu metode ekstraksi data menjadi sebuah pola informasi tertentu. Penelitian ini bertujuan untuk mengaplikasikan tiga metode *machine learning* untuk memperkirakan status pengiriman barang. Metodologi yang digunakan pada penelitian ini mengikuti proses *machine learning* yang dirilis oleh the *Cross Industry Standard Process for Data Mining* (CRISP-DM), yaitu; memahami proses bisnis, memahami data, persiapan data, pengembangan model, evaluasi, dan implementasi. Berdasarkan hasil penelitian, metode *random forest* menghasilkan nilai akurasi yang lebih baik jika dibandingkan dengan metode regresi logistik dan *artificial neural network* (ANN), yaitu sebesar 76,6%, sedangkan metode ANN dan regresi logistik sebesar 73,81% dan 72,84%.

Kata kunci:

analisis data prediksi, *machine learning*, waktu pengiriman, transportasi dan logistik

Abstract

One of the key performance indicators for the logistics industry, especially freight forwarder company (cargo), is the delivery time. This is still a challenge in this industry in terms of ensuring the customer service level and reducing transportation costs. On the other hand, the development of information technology now allows an organization or company to collect large amounts of data automatically. A decent method that can be used to analyze the data for prediction purposes is machine learning, which is a method of extracting data into a certain pattern of information. This research aims to apply three machine learning methods to estimate the status of shipping goods. The method used in this study follows the machine learning process published by the Cross Industry Standard Process for Data Mining (CRISP-DM), namely; business processes understanding, data understanding, data preparation, model development, evaluation, and implementation. Based on the results of the study, the random forest method produces better accuracy than the logistic regression and artificial neural network (ANN) methods, which is 76.6%, while the results of ANN and logistic regression methods are 73.81% and 72.84% respectively.

Keywords:

predictive data analytics, machine learning, delivery time, transportation and logistics

I. PENDAHULUAN

Perkembangan teknologi informasi saat ini memungkinkan sebuah organisasi atau perusahaan mempunyai kapabilitas dalam mengumpulkan data yang berhubungan dengan proses bisnisnya dalam jumlah yang besar dan waktu yang singkat. Hanya saja, diperlukan tahapan tertentu agar data yang ada dapat dijadikan sumber informasi atau pengetahuan baru sebagai dasar pengambilan keputusan. Analisis data yang bertujuan untuk memprediksi suatu nilai yang belum diketahui berdasarkan data yang ada disebut analisis data prediksi atau *predictive data analytics* (Kelleher dkk., 2015).

Metode yang cukup andal digunakan dalam melakukan analisis data prediksi adalah *machine learning*, yaitu algoritma ekstraksi data yang merupakan irisan dari ilmu statistik, kecerdasan buatan (*artificial intelligence*), dan ilmu komputer (*computer science*) (Müller dan Guido, 2016). Aplikasi dari *machine learning* untuk membangun model prediksi juga sudah cukup massif, mulai dari sektor keuangan (Lin dkk., 2011; Soni, 2011; Yoo dkk., 2005), teknologi telekomunikasi (Mozer dkk., 2000; Qureshi dkk., 2013), energi (Amasyali & El-Gohary, 2018; Grolinger dkk., 2016; Y. Guo dkk., 2018), atau bahkan pada industri jasa seperti pariwisata (Li & Cao, 2018; Nilashi dkk., 2017).

Pada bidang logistik dan rantai pasok, *machine learning* untuk analisis prediksi juga sudah diterapkan di beberapa aktivitas logistik, seperti peramalan permintaan (Carbonneau dkk., 2008; Ferreira dkk., 2016), penentuan kebijakan persediaan (de Santis dkk., 2017; X. Guo dkk., 2014), dan bidang-bidang lain yang membutuhkan prediksi data sehingga proses pengambilan keputusan bisa dilakukan lebih tepat dan cepat.

Analisis prediksi berbasis *machine learning* juga diaplikasikan pada aktivitas transportasi logistik, misalnya digunakan untuk memperkirakan waktu perjalanan kendaraan (Simroth & Zahle, 2010) keterlambatan pengantaran bahan-bahan konstruksi (Asadi dkk., 2015), dan keterlambatan keberangkatan transportasi udara (Nigam & Govinda, 2017). Beberapa penelitian ini erat kaitannya dengan ketepatan waktu penghantaran yang merupakan kinerja utama bidang logistik. Ketepatan waktu penghantaran ini sangat penting guna meningkatkan kepuasan pelanggan (*service level*) dan menurunkan biaya logistik.

Penelitian ini bertujuan untuk memprediksi status pengiriman barang yang dilakukan oleh perusahaan ekspedisi pengiriman barang melalui pesawat udara. Data yang tercatat oleh perusahaan, terjadi rata-rata keterlambatan sebesar 27% dari total aktivitas bisnisnya. Adapun metode yang dipilih adalah analisis prediksi berbasis *machine learning*, yaitu *logistic regression* (regresi logistik), *random forest*, dan *artificial neural network* (ANN).

II. KAJIAN LITERATUR

II.1 *Machine learning*

Machine learning adalah serangkaian algoritma komputer yang digunakan untuk mengoptimalkan kinerja komputer atau sistem berdasarkan data sampel yang ada (Alpaydin, 2020). Kemampuan utama dari *machine learning* adalah modifikasi dan adaptasi keputusan dalam merespon perubahan (Marsland, 2015). Kegunaan dari *machine learning* antara lain sebagai berikut (Provost & Fawcett, 2013)

- *Classification* (Klasifikasi) adalah metode *machine learning* yang digunakan untuk memprediksi suatu nilai/kelas sebuah individu dalam sebuah populasi.
- *Similarity matching* (Pencocokan kemiripan) adalah metode *machine learning* yang digunakan untuk mengidentifikasi kemiripan antar individu berdasarkan data yang ada.
- *Clustering* (Pengklasteran) adalah metode *machine learning* yang digunakan untuk mengelompokkan individu dalam grup yang sama berdasarkan kesamaan yang dimiliki.

Dalam penelitian ini, kegunaan metode *machine learning* mengacu pada poin pertama, yaitu digunakan untuk memprediksi suatu nilai/kelas dari suatu individu.

II.2 *Logistic Regression* (Regresi Logistik)

Model regresi, termasuk regresi logistik (*logistic regression*), merupakan metode analisis data yang menjelaskan korelasi antara sebuah variabel dependen dan satu atau beberapa variabel independen. Tujuan pengembangan model regresi adalah menemukan hubungan yang paling bisa menjelaskan dengan baik antara variabel-variabel tersebut. Hal yang membedakan model regresi logistik dengan model regresi linier sederhana adalah pada regresi logistik,

variabel keluaran bersifat biner atau dikotomis, sehingga, nilai variabel dependen dari hasil perhitungan di dalam model, akan dipetakan dengan sebuah fungsi menjadi nilai yang sifatnya biner (Hosmer Jr dkk., 2013).

Berdasarkan jumlah variabel masukannya, regresi logistik dibedakan menjadi dua, yaitu regresi logistik tunggal (*single logistic regression*), di mana hanya terdapat satu jenis variabel masukan, dan regresi logistik jamak (*multiple logistic regression*), di mana terdapat dua atau lebih variabel masukan, untuk memprediksi nilai variabel keluarannya. Dalam penelitian ini, model regresi yang digunakan adalah regresi logistik jamak.

Jika terdapat n buah variabel masukan yang disimbolkan oleh $x = (x_1, x_2, \dots, x_n)$, dan Y adalah variabel keluaran di mana peluang $P(Y = 1 | x) = \pi(x)$, maka persamaan regresi logistik dituliskan sebagai berikut:

$$y(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

di mana

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

II.3 Random Forest

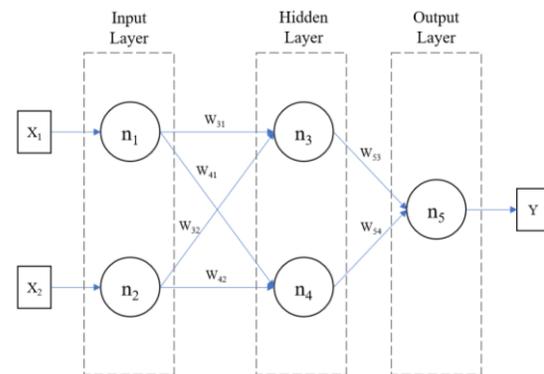
Salah satu kekurangan dari algoritma pohon keputusan (*decision tree*) adalah strukturnya akan sangat kompleks untuk data dengan jumlah variabel yang banyak, sehingga risiko yang muncul adalah *overfitting*. Beberapa cara dapat ditempuh untuk mengurangi risiko tersebut. Selain membatasi jumlah variabel, cara lain yang dapat dilakukan adalah merestrukturisasi *tree* dari algoritma sebelumnya. Pendekatan ini memunculkan algoritma baru yang disebut *random forest* (Shalev-Shwartz & Ben-David, 2014).

Random forest adalah algoritma prediksi yang merupakan gabungan dari beberapa *decision tree*. Algoritma ini cukup efisien secara komputasional karena setiap pohon keputusan memilih secara acak variabel yang digunakan dan tidak menggunakan semua variabel (Oshiro dkk., 2012). Model *random forest* dibangun dengan menggunakan konsep *bagging* (*bootstrapp aggregation*) yaitu pengumpulan sampel observasi secara acak ke dalam suatu

penampung atau disebut *bag*. Beberapa *bag* tersusun dari observasi yang dipilih secara acak dari data observasi orisinal dalam *training dataset*. *Dataset* pada *bag* terbentuk dari pengambilan dengan pengembalian yang memungkinkan sebuah hasil observasi memiliki peluang untuk terpilih kembali.

II.4 Artificial Neural Network (ANN)

Artificial neural network (ANN) merupakan metode klasifikasi dari *machine learning* yang terinspirasi dari struktur sistem saraf makhluk hidup (*neural systems*). Seperti halnya sistem saraf pada otak manusia ANN mempunyai struktur dasar yang serupa, yaitu terdiri dari beberapa simpul (*nodes*), yang bisa sebagai variabel masukan (*input*), keluaran (*output*), atau antara, yang antar simpulnya dihubungkan oleh penghubung atau *link* (Tan dkk., 2016). Masing-masing penghubung terdapat bobot (*weight*) yang mengindikasikan kekuatan amplifikasi yang diterima oleh simpul masukan menuju simpul tujuan, sehingga, ANN dapat didefinisikan sebagai sebuah kesatuan fungsi yang memetakan nilai variabel masukan yang ada di lapisan masukan (*input layer*) menuju lapisan keluaran (*output layer*) dengan menggunakan bobot sebagai parameter penghubungnya (Aggarwal, 2018). Adapaun struktur dasar dari ANN digambarkan oleh Gambar 1.



Gambar 1. Topologi Dasar ANN (Tan dkk., 2016)

II.5 Penggunaan Model Prediksi pada Proses Bisnis

Penggunaan model prediksi untuk memperkirakan waktu proses bisnis tertentu telah menjadi perhatian beberapa peneliti. Bidang bisnisnya pun beragam, mulai dari industri manufaktur, logistik, maupun jasa. Van der Aalst dkk. (2011) misalnya

mengembangkan model prediksi yang digunakan untuk memperkirakan durasi waktu pelayanan pada industri jasa asuransi menggunakan metode *process mining*. Model tersebut juga telah diimplementasikan di kasus nyata dengan tingkat akurasi yang baik (Van der Aalst dkk., 2011).

Kang dkk. (2012) mengembangkan model prediksi menggunakan metode *machine learning*, yaitu *support vector machines* (SVM) untuk memantau proses bisnis secara waktu nyata (*real-time*). Model ini juga menyediakan informasi berupa status proses yang sedang berjalan. Hasil penelitian ini menunjukkan bahwa pengembangan dari algoritma SVM dapat digunakan untuk menghasilkan prediksi secara waktu nyata dan dapat digunakan oleh industri (Kang dkk., 2012).

Evermann dkk. (2016) menggunakan metode *deep learning*, salah satu jenis metode *machine learning*, untuk memprediksi kondisi akhir dari proses bisnis yang sedang berjalan. Hal ini dilakukan dengan tujuan untuk mengidentifikasi proses-proses yang berisiko tidak memenuhi indikator kinerja utama (KPI), misalnya ketepatan waktu. Metode yang digunakan adalah *recurrent neural network* (RNN). Hasil dari penelitian ini menunjukkan tingkat akurasi yang tinggi, yaitu lebih dari 90% untuk dua *dataset* yang digunakan (Evermann dkk., 2016).

Bevacqua dkk. (2013) mengembangkan model prediksi waktu proses (*run-time*) dengan menggunakan kombinasi dari beberapa teknik penambangan (*data mining*), mulai dari *pattern mining*, *non-parametric regression*, dan *predictive clustering*. Hasil dari penelitian ini adalah model yang dikembangkan mampu memprediksi dengan andal dan akurat. Saat dipadukan dengan variabel lain seperti beban kerja dan kalender, metode *predictive clustering* menunjukkan hasil terbaik jika dibandingkan dengan metode lain (Bevacqua dkk., 2013).

Feldmann dkk. (2013) mengembangkan model yang disebut *proactive event processing*, yang merupakan pengembangan dari *complex event processing*. Model ini bertujuan untuk mendeteksi adanya kemungkinan munculnya masalah dalam rangkaian bisnis proses. Penelitian ini menerapkan model prediksinya di industri transportasi untuk memprediksi berat muatan kargo. Menurut penelitian ini, karena industri kargo membutuhkan biaya yang mahal sehingga perlu ada sistem yang dapat

mendeteksi potensi masalah. Hasil dari penelitian ini adalah kemampuan model dalam memprediksi status pengiriman antara *overload* dan *underload* dari kapasitas (Feldman dkk., 2013).

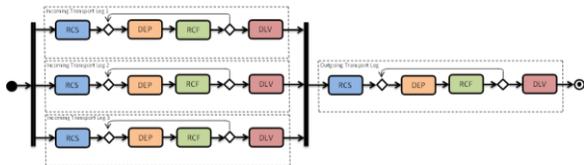
Adapun Metzger dkk. (2014) mengembangkan dan membandingkan model prediksi untuk memperkirakan masalah yang mungkin akan terjadi selama aktivitas proses bisnis pada industri logistik. Adapun metode yang digunakan antara lain *machine learning*, *constraint satisfaction*, dan *quality of service* (QoS) *aggregation*. Model prediksi yang dikembangkan kemudian digunakan untuk menganalisis sebuah *dataset* pada perusahaan kargo/jasa pengiriman barang udara. Hasil dari penelitian ini didapatkan bahwa dengan menggunakan ketiga metode di atas, akurasi prediksi yang dihasilkan minimal sebesar 70%. Adapun kombinasi dari dua metode, yaitu *machine learning* dan *constraint satisfaction* mampu meningkatkan tingkat akurasi prediksi sebesar 23% (Metzger dkk., 2014).

Penelitian ini merupakan pengembangan dari dua penelitian sebelumnya (Feldman dkk., 2013; Metzger dkk., 2014) dan menggunakan data dari sumber yang sama. Tujuan utama dari penelitian ini adalah adanya perbandingan tingkat akurasi model prediksi yang dikembangkan dengan menggunakan tiga jenis metode *machine learning* yang dikenal memiliki tingkat akurasi yang tinggi, yaitu regresi logistik (*logistic regression*), *random forest*, dan *artificial neural network* (ANN).

II.6 Proses Bisnis Pengiriman Barang

Gambar 1 menggambarkan proses bisnis pengiriman barang yang menjadi kasus studi pada penelitian ini. Model tersebut mewakili struktur proses bisnis pada *freight forwarding company* atau perusahaan kargo, khususnya jasa ekspedisi pesawat udara. Beberapa pengiriman yang kuantitasnya relatif kecil dari pengirim dikonsolidasikan, kemudian dikirim bersama kepada pelanggan agar lebih aman dan murah. Oleh karena itu, proses bisnis pada perusahaan ini dibagi menjadi dua bagian, yaitu *incoming transport* dan *outgoing transport*, yang secara bersama bertujuan untuk memastikan bahwa pengiriman dilakukan secara tepat waktu sampai ke pelanggan. Setiap bagian dari proses bisnis melibatkan layanan transportasi fisik berikut (Metzger dkk., 2014):

- 1) RCS (*Freight reception*): Barang diterima oleh maskapai, kemudian diperiksa di gudang keberangkatan.
- 2) DEP (*Freight departure*): Barang dikirim ke pesawat dan dikonfirmasi di atas pesawat, pesawat berangkat.
- 3) RCF (*Freight arrival*): Barang sampai di bandara tujuan dan disimpa di gudang kedatangan.
- 4) DLV (*Freight delivery*): Barang dikirim dari gudang bandara ke pelanggan.

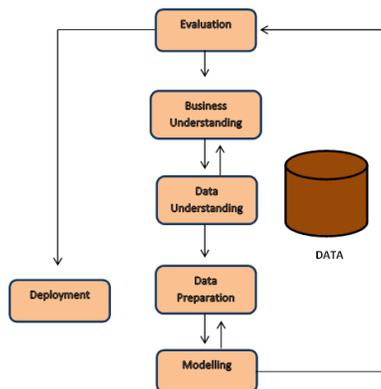


Gambar 1. Proses Bisnis Pengiriman Barang (Metzger dkk., 2014)

III. ANALISIS DAN PERANCANGAN

III.1 Metodologi

Metodologi dari penelitian ini mengacu pada proses *machine learning* yang mengacu pada proses *data mining* oleh the *Cross Industry Standard Process for Data Mining* (CRISP-DM) yang dijelaskan oleh Gambar 2.



Gambar 2. Proses Penggunaan Metode *Machine learning* oleh CRISP-DM (Provost & Fawcett, 2013)

Adapun penjelasan secara lebih detail mengenai metodologi yang dilakukan dalam penelitian ini adalah sebagai berikut:

III.1.1 Business Understanding (Memahami Proses Bisnis)

Dalam penelitian ini, tahap pertama yang dilakukan oleh peneliti adalah memahami proses bisnis dari pengiriman barang oleh perusahaan kargo. Seperti yang telah dijelaskan pada bagian II.2, bahwa proses pengiriman barang melalui empat kegiatan utama, yaitu RCS, DEP, RCF, dan DLV.

III.1.2 Data Understanding (Memahami Data)

Setelah memahami proses bisnis, tahap selanjutnya adalah memahami data yang terkait dengan proses bisnisnya, baik yang sudah tersedia maupun yang masih perlu dilakukan pencarian lebih lanjut. Pada penelitian ini, data yang digunakan adalah data yang dirilis oleh the *International Air Transport Association* (IATA).

Seperti yang sudah disebutkan pada bagian II.6 bahwa proses pengiriman barang melalui empat kegiatan utama. Masing-masing dari setiap kegiatan dalam proses bisnis tersebut tercatat waktu yang direncanakan (*planned time*) dan waktu yang sebenarnya (*actual time*). Selain berupa atribut durasi waktu proses per bagian, terdapat juga atribut lain seperti identitas pengiriman dan bandara yang dilalui oleh layanan transportasinya.

III.1.3 Data Preparation (Mempersiapkan Data)

Setelah proses memahami selesai, selanjutnya dilakukan tahap mempersiapkan data agar siap digunakan untuk pengembangan model prediksi. Tahapan ini dilakukan antara lain dengan mengubah format yang sesuai dengan kebutuhan (*tabular format*), mereduksi variabel yang tidak diperlukan, dan menghilangkan data dengan nilai yang hilang (*missing value*), dan penambahan atau pengubahan nilai pada salah satu atau beberapa variabel agar sesuai dengan tujuan penelitian dan pendekatan pengembangan model yang akan digunakan.

Pada penelitian ini, nilai variabel yang disesuaikan adalah penambahan variabel "*remark*" yang mempunyai dua nilai kategorikal, yaitu "*late*" jika nilai "*actual time*" lebih besar daripada "*planned time*", dan nilai "*fast*", jika nilai "*actual time*" lebih besar daripada "*planned time*".

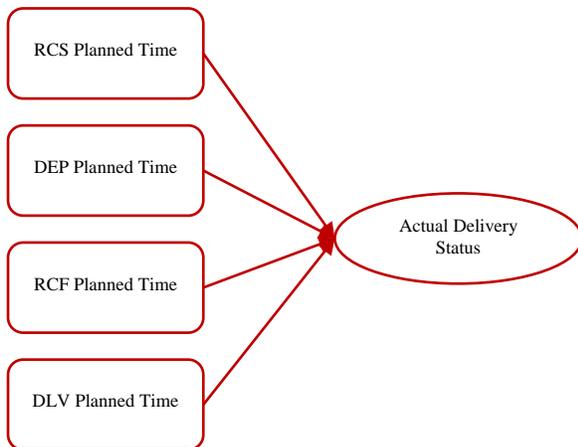
Selanjutnya, dilakukan penentuan variabel independen dan variabel dependen dari beberapa variabel yang ada. Berdasarkan studi literatur yang dilakukan, ditentukan bahwa berdasarkan seluruh

proses bisnis yang dilalui, terdapat data RCS dan DLV sebanyak satu kali, serta DEP dan RCF sebanyak dua kali. Keenam data tersebut dalam pengembangan model sebagai variabel independen. Adapun variabel dependen adalah status pengiriman dari seluruh proses yang ada, yaitu “late” jika penghantarannya terlambat, dan “fast” jika penghantarannya lebih cepat dibandingkan waktu yang direncanakan.

Selanjutnya, dipilih beberapa catatan transaksi yang terhitung lengkap, atau menghilangkan *data instance* yang tidak lengkap. Dari proses tersebut, didapatlah 1195 data transaksi dari total 3269 data. Langkah ketiga adalah membagi dari total data transaksi yang ada menjadi dua bagian, yaitu data yang digunakan untuk *training* dan *testing*. Berdasarkan studi literatur yang ada, digunakan 80% data atau sebanyak 856 data untuk *training*, dan 20% data atau sebanyak 239 data digunakan sebagai data *testing*.

III.1.4 Modelling (Pengembangan Model)

Setelah data sudah melalui tahap pre-processing, selanjutnya dilakukan pengembangan model prediksi dengan variabel masukan berupa durasi waktu yang direncanakan (*planned time*) untuk keempat aktivitas pengiriman barang, guna memprediksi status pengiriman barang (*actual time*) yang menjadi variabel keluaran. Secara umum, diagram yang menunjukkan variabel masukan dan keluaran, serta menggambarkan model prediksi yang dibangun digambarkan dalam Gambar 3.



Gambar 3. Model Prediksi Waktu Pengiriman Barang

Pengembangan model prediksi pada penelitian ini menggunakan tiga metode yang berbeda, yaitu regresi

logistik, *random forrest*, dan *artificial neural network*, dan (ANN). Perangkat lunak yang digunakan adalah *Weka 3.8*, sebuah *open-source software* pengolahan data berbasis *machine learning* yang cukup menyediakan antar muka yang ramah pengguna.

Walaupun hubungan antara variabel masukan dan keluaran relatif sama dan mengacu pada Gambar 3, terdapat dua bagian pada tahap pengembangan model ini, yaitu model yang menggambarkan pengiriman barang dari tempat asal menuju gudang konsolidasi atau *incoming transport*, dan model yang menggambarkan pengiriman barang dari gudang konsolidasi menuju tempat penerima atau *outgoing transport*, sehingga proses pengembangan model dan uji akurasi akan dibedakan berdasarkan kedua proses tersebut.

III.1.5 Evaluation (Evaluasi)

Pada tahap ini, dilakukan evaluasi berupa perbandingan tingkat akurasi dari ketiga metode *machine learning* yang digunakan pada tahap pengembangan model, baik pada tahap *incoming transport* maupun *outgoing transport*. Hasil akurasi rata-rata dari kedua proses tersebut untuk ketiga metode juga diperbandingkan untuk mengetahui metode yang memiliki kinerja terbaik. Adapun hasil dan pembahasan secara lebih detail dijelaskan pada bagian IV.

III.1.6 Deployment (Implementasi)

Setelah model sudah dievaluasi dan dinyatakan layak untuk digunakan untuk memprediksi sebuah proses bisnis, maka proses selanjutnya adalah implementasi, yaitu penggunaan model di kondisi nyata. Pada penelitian ini, tahapan implementasi tidak dilakukan dan merupakan batasan dari penelitian.

III.2 Hasil Penelitian

III.2.1 Incoming Transport

Incoming transportation adalah tahapan pengiriman barang dari tempat asal menuju lokasi konsolidasi. Metode pertama yang digunakan untuk pengembangan model adalah regresi logistik. Adapun hasil perbandingan akurasi untuk lima kali percobaan disajikan oleh Tabel 1.

Tabel 1. Tingkat Akurasi Metode Regresi Logistik untuk Incoming Transport

Percobaan ke-	Akurasi	Akurasi Rata-Rata
1	75,25	76,56
2	76,15	
3	77,23	
4	76,81	
5	77,40	

Selanjutnya, tingkat akurasi dengan metode *random forest* disajikan oleh Tabel 2. Metode ini dijalankan dengan melakukan perubahan nilai iterasi dari 100, 500, 1000, dan 2000.

percobaan disajikan oleh Tabel 1.

Tabel 1. Tingkat Akurasi Metode Random Forest untuk Incoming Transport

Jumlah Iterasi	Akurasi	Akurasi Rata-Rata
100	79,89	79,70
500	79,50	
1000	79,91	
2000	79,49	

Untuk metode ANN, percobaan dilakukan dengan melakukan perubahan pada *learning rate* dengan kombinasi 0,1, 0,05, dan 0,01, serta jumlah iterasinya yaitu 100, 500, dan 1000 kali. Adapun hasil perbandingan akurasi disajikan dalam Tabel 3.

Tabel 3. Tingkat Akurasi Metode ANN untuk Incoming Transport

Learning Rate	Iterasi	Akurasi	Akurasi Rata-Rata
0.1	100	74,74	76,24
	500	76,57	
	1000	77,41	
0.05	100	74,90	76,57
	500	76,99	
	1000	77,82	
0.01	100	75,31	75,03
	500	74,89	
	1000	74,89	

Berdasarkan nilai akurasi rata-rata, metode ANN memiliki tingkat akurasi terbaik sebesar 76,57% saat menggunakan *learning rate* 0.05.

Berdasarkan hasil yang didapat untuk ketiga metode, dapat disimpulkan bahwa pada proses *incoming transport*, hasil akurasi terbaik dicapai dengan menerapkan metode *random forest* dengan capaian tingkat akurasi sebesar 79,70 %, sedangkan nilai akurasi yang dihasilkan oleh metode regresi logistik dan ANN bernilai hampir sama, yaitu 76,56% dan 76,57%

III.2.2 Outgoing Transport

Outgoing transportation adalah tahapan pengiriman barang dari gudang konsolidasi menuju lokasi pelanggan atau penerima. Sama halnya dengan proses *incoming transport*, metode pertama yang digunakan untuk pengembangan model adalah regresi logistik. Adapun hasil perbandingan akurasi untuk lima kali percobaan disajikan oleh Tabel 4.

Tabel 1. Tingkat Akurasi Metode Regresi Logistik untuk Incoming Transport

Percobaan ke-	Akurasi	Akurasi Rata-Rata
1	69,72	71,06
2	69,52	
3	70,27	
4	73,23	
5	72,57	

Metode kedua yang digunakan untuk menganalisis data *outgoing transport* adalah metode *random forest*, yang disajikan oleh Tabel 5. Metode ini dijalankan dengan melakukan perubahan nilai iterasi dari 100, 500, 1000, dan 2000.

Tabel 5. Tingkat Akurasi Metode Random Forest untuk Outgoing Transport

Jumlah Iterasi	Akurasi	Akurasi Rata-Rata
100	73,33	73,50
500	72,00	
1000	74,66	
2000	74,00	

Selanjutnya, untuk metode ANN, percobaan dilakukan dengan melakukan perubahan pada *learning rate* dengan kombinasi 0,1, 0,05, dan 0,01, serta jumlah iterasinya yaitu 100, 500, dan 1000 kali. Adapun hasil perbandingan akurasi disajikan dalam Tabel 6.

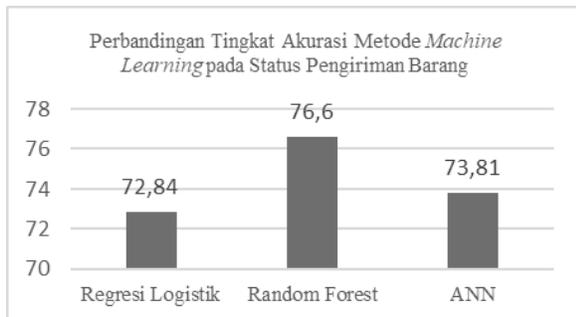
Tabel 6. Tingkat Akurasi Metode ANN untuk Outgoing Transport

Learning Rate	Iterasi	Akurasi	Akurasi Rata-Rata
0.1	100	68,66	68,88
	500	68,66	
	1000	69,33	
0.05	100	68,67	69,11
	500	69,33	
	1000	69,33	
0.01	100	64,00	66,89
	500	68,00	
	1000	68,67	

Berdasarkan nilai akurasi rata-rata, metode ANN memiliki tingkat akurasi terbaik sebesar 69,11% saat

menggunakan *learning rate* 0.05. Adapun untuk semua metode pada proses *outgoing transport*, metode *random forest* masih memiliki performansi terbaik, yaitu mencapai 73,50%. Berikutnya regresi logistik dengan 71,06% dan ANN dengan 69,11%.

Jika dibandingkan hasil tingkat akurasi antara *incoming transport* dan *outgoing transport*, untuk semua metode yang digunakan, terlihat bahwa nilai akurasi untuk proses *incoming transport* lebih tinggi dibandingkan nilai akurasi pada *outgoing transport*. Gambar 4 menyajikan data perbandingan tingkat akurasi dari ketiga metode untuk keseluruhan proses pengiriman barang. Nilai akurasi yang ditampilkan merupakan rata-rata dari nilai akurasi pada proses *incoming transport* dan *outgoing transport* untuk masing-masing metode.



Gambar 4. Perbandingan Tingkat Akurasi Semua Metode

Secara umum, baik untuk proses *incoming transport* maupun *outgoing transport*, metode *random forest* memiliki tingkat akurasi yang paling dibandingkan metode yang lain, yaitu 76,6%. Selanjutnya, metode ANN memiliki tingkat akurasi sebesar 73,81% dan regresi logistik sebesar 72,84%. Seperti halnya penelitian pada bidang *machine learning* lainnya, tingkat akurasi sebuah model prediksi yang dikembangkan dengan metode tertentu akan bergantung kepada jenis data yang digunakan sebagai kasus studi (*study case*), sehingga akan sangat dimungkinkan jika metode yang memiliki tingkat akurasi tertentu, tidak sepenuhnya akan mencapai tingkat akurasi yang sama jika diaplikasikan untuk jenis data yang berbeda.

IV. KESIMPULAN DAN SARAN

Model prediksi berbasis metode *machine learning* dapat diaplikasikan dalam dunia logistik, khususnya dalam memprediksi status pengiriman barang pada perusahaan kargo. Data-data berupa

durasi waktu yang direncanakan dapat menjadi variabel masukan untuk memprediksi status pengiriman yang bertindak sebagai variabel keluaran. Dalam penelitian ini, digunakan tiga metode yang berbeda, yaitu regresi logistik, *random forest*, dan *artificial neural network* (ANN).

Berdasarkan hasil percobaan yang dilakukan, metode *random forest* menghasilkan nilai akurasi yang lebih baik jika dibandingkan dengan metode regresi logistik dan ANN. Pada aktivitas *incoming transport* maupun *outgoing transport* menunjukkan bahwa metode *random forest* dapat menghasilkan akurasi terbaik sebesar 76,6%, sedangkan metode ANN dan regresi logistik sebesar 73,81% dan 72,84%.

Penelitian ini memiliki potensi untuk dikembangkan kedepan. Selain perlu menggunakan data yang lebih lengkap, misalnya data kondisi ekonomi dan lingkungan yang mempengaruhi aktivitas logistik, peneliti merekomendasikan untuk menggunakan metode *machine learning* yang lain, atau gabungan dari beberapa metode *machine learning*, guna meningkatkan tingkat akurasi prediksi.

REFERENSI

- Aggarwal, C. C. (2018). Neural networks and deep learning. *Springer, 10*, 973–978.
- Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.
- Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews, 81*, 1192–1205.
- Asadi, A., Alsubaey, M., & Makatsoris, C. (2015). A machine learning approach for predicting delays in construction logistics. *International Journal of Advanced Logistics, 4*(2), 115–130.
- Bevacqua, A., Carnuccio, M., Folino, F., Guarascio, M., & Pontieri, L. (2013). A Data-adaptive Trace Abstraction Approach to the Prediction of Business Process Performances. *ICEIS (1)*, 56–65.
- Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research,*

- 184(3), 1140–1154.
- de Santis, R. B., de Aguiar, E. P., & Goliatt, L. (2017). Predicting material backorders in inventory management using machine learning. *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, 1–6.
- Evermann, J., Rehse, J.-R., & Fettke, P. (2016). A deep learning approach for predicting process behaviour at runtime. *International Conference on Business Process Management*, 327–338.
- Feldman, Z., Fournier, F., Franklin, R., & Metzger, A. (2013). Proactive event processing in action: a case study on the proactive management of transport processes (industry article). *Proceedings of the 7th ACM International Conference on Distributed Event-Based Systems*, 97–106.
- Ferreira, K. J., Lee, B. H. A., & Simchi-Levi, D. (2016). Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management*, 18(1), 69–88.
- Grolinger, K., L’Heureux, A., Capretz, M. A. M., & Seewald, L. (2016). Energy forecasting for event venues: Big data and prediction accuracy. *Energy and Buildings*, 112, 222–233.
- Guo, X., Liu, C., Xu, W., Yuan, H., & Wang, M. (2014). A prediction-based inventory optimization using data mining models. *2014 Seventh International Joint Conference on Computational Sciences and Optimization*, 611–615.
- Guo, Y., Wang, J., Chen, H., Li, G., Liu, J., Xu, C., Huang, R., & Huang, Y. (2018). Machine learning-based thermal response time ahead energy demand prediction for building heating systems. *Applied Energy*, 221, 16–27.
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.
- Kang, B., Kim, D., & Kang, S. (2012). Periodic performance prediction for real-time business process monitoring. *Industrial Management & Data Systems*.
- Kelleher, J. D., Mac Namee, B., & D’arcy, A. (2015). *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. MIT press.
- Li, Y., & Cao, H. (2018). Prediction for tourism flow based on LSTM neural network. *Procedia Computer Science*, 129, 277–283.
- Lin, W.-Y., Hu, Y.-H., & Tsai, C.-F. (2011). Machine learning in financial crisis prediction: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(4), 421–436.
- Marsland, S. (2015). *Machine learning: an algorithmic perspective*. CRC press.
- Metzger, A., Leitner, P., Ivanović, D., Schmieders, E., Franklin, R., Carro, M., Dustdar, S., & Pohl, K. (2014). Comparing and combining predictive business process monitoring techniques. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(2), 276–290.
- Mozer, M. C., Wolniewicz, R., Grimes, D. B., Johnson, E., & Kaushansky, H. (2000). Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Transactions on Neural Networks*, 11(3), 690–696.
- Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python: a guide for data scientists*. “O’Reilly Media, Inc.”
- Nigam, R., & Govinda, K. (2017). Cloud based flight delay prediction using logistic regression. *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, 662–667.
- Nilashi, M., Bagherifard, K., Rahmani, M., & Rafe, V. (2017). A recommender system for tourism industry using cluster ensemble and prediction machine learning techniques. *Computers & Industrial Engineering*, 109, 357–368.
- Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How many trees in a random forest? *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, 154–168.
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. “O’Reilly Media, Inc.”

- Qureshi, S. A., Rehman, A. S., Qamar, A. M., Kamal, A., & Rehman, A. (2013). Telecommunication subscribers' churn prediction model using machine learning. *Eighth International Conference on Digital Information Management (ICDIM 2013)*, 131–136.
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge university press.
- Simroth, A., & Zahle, H. (2010). Travel time prediction using floating car data applied to logistics planning. *IEEE Transactions on Intelligent Transportation Systems*, 12(1), 243–253.
- Soni, S. (2011). Applications of ANNs in stock market prediction: a survey. *International Journal of Computer Science & Engineering Technology*, 2(3), 71–83.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2016). *Introduction to data mining*. Pearson Education India.
- Van der Aalst, W. M. P., Schonenberg, M. H., & Song, M. (2011). Time prediction based on process mining. *Information Systems*, 36(2), 450–475.
- Yoo, P. D., Kim, M. H., & Jan, T. (2005). Machine learning techniques and use of event information for stock market prediction: A survey and evaluation. *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, 2, 835–841.