

## Analisis Kecenderungan Keterlambatan Pembayaran Pengecekan Kapal di Pelabuhan Regional Riau

Hespri Yomeldi  
School of Electrical Engineering  
and Informatics  
Institut Teknologi Bandung  
Bandung, Indonesia  
hespri@s.itb.ac.id

Moh. Roufiq Azmy  
School of Electrical Engineering  
and Informatics  
Institut Teknologi Bandung  
Bandung, Indonesia  
roufiq@s.itb.ac.id

Ryche Pranita  
School of Electrical Engineering  
and Informatics  
Institut Teknologi Bandung  
Bandung, Indonesia  
ryche@s.itb.ac.id

*Abstract— Ship health checks must be carried out which function to provide a sailing permit. The implementation of ship health checks is carried out in collaboration with the ministry of health and transportation. The implementation of the activity, commonly known as Port Health Quarantine Clearance (PHQC) requires time to check and the ship makes a payment check to be able to issue a sailing permit. The problem that arises in the field is that the ship delays PHQC payments and then impacts on the buildup of ships in the port, besides that officers also need longer time to process the issuance of sailing permits. This of course has an impact on other port services such as dwelling time and scheduled departures that can be delayed. In overcoming this problem, an in-depth study is needed to analyze the trend of late payment of ship health checks, what variables influence it and how treatment is done to overcome these problems. Using logistic regression and decision tree with Classification and Regression Tree algorithm, a model is then developed that determines the variables that affect the delay of the ship making PHQC payments*

*Keywords— data analytic, logistic regression, decision tree, port health quarantine clearance*

### I. PENDAHULUAN

Pelabuhan Laut Regional Riau merupakan salah satu pelabuhan utama di wilayah Sumatera. Pelabuhan ini dikelola oleh PT. Pelindo I. Pelabuhan ini merupakan pintu gerbang di bagian utara Provinsi Riau, dan melayani hinterland di bagian selatan Provinsi Sumatera Utara serta bagian utara Provinsi Sumatera Barat. Fungsi pelabuhan ini adalah sebagai pelabuhan internasional (terbuka bagi perdagangan luar negeri) yang melayani minyak sawit mentah (CPO) Cargo, General Cargo, Peti Kemas, Dry Bulk Cargo dan juga penumpang. Pelabuhan Regional Riau mencatat jumlah kapal yang datang mencapai 50.000 lebih selama 2 tahun terakhir dengan Pelabuhan Kota Dumai memiliki jumlah kapal masuk tertinggi, lebih dari 80 persen total kapal yang datang atau berlayar. Peningkatan tersebut, terutama berasal kapal dari mancanegara menjadikan Pelabuhan Regional Riau sangat berpotensi untuk dikembangkan sebagai kawasan pelabuhan yang strategis dan bernilai ekonomi tinggi.

Kawasan pelabuhan yang bernilai ekonomis, pemerintah perlu menentukan standar pelayanan pelabuhan yang baik. Efisiensi dan efektivitas seluruh aktivitas di pelabuhan harus dapat dikelola secara tepat dan komprehensif. Pengelolaan bongkar muat, pengelolaan waktu kedatangan dan keberangkatan kapal, pengelolaan izin berlayar, pemeriksaan kesehatan kapal harus dapat terintegrasi dengan baik. Salah satu kendala yang sering dihadapi dalam pelayanan pelabuhan terhadap kapal yang bersandar dan berlayar adalah permasalahan pemeriksaan kesehatan kapal sesuai dengan klasifikasi yang telah ditetapkan [1]. Pemeriksaan kesehatan kapal wajib dilakukan sebagai salah satu syarat pemberian sertifikat ijin berlayar suatu kapal. Pelaksanaan pemeriksaan kesehatan kapal dilakukan dengan kerjasama Kementerian Kesehatan dan Kementerian Perhubungan. Pada pelaksanaannya kegiatan yang biasa dikenal sebagai *Port Health Quarantine Clearance* (PHQC) membutuhkan waktu untuk pengecekan dan pihak kapal melakukan pembayaran pengecekan untuk dapat menerbitkan surat izin berlayar. Permasalahan yang muncul di lapangan adalah pihak kapal melakukan penundaan pembayaran PHQC yang berimbas

pada penumpukan kapal di pelabuhan. Selain itu petugas juga membutuhkan waktu yang lebih lama dalam memproses penerbitan surat izin berlayar. Hal ini tentu saja berimbas pada pelayanan pelabuhan lain seperti *dwelling time* [2] dan jadwal keberangkatan yang bisa tertunda. Guna mengatasi permasalahan ini diperlukan suatu penelitian mendalam untuk menganalisis kecenderungan keterlambatan pembayaran pengecekan kesehatan kapal, variabel apa saja yang mempengaruhinya dan bagaimana treatment yang dilakukan untuk mengatasi permasalahan tersebut.

Kontribusi dari penelitian ini adalah bagaimana membangun model prediksi yang tepat untuk mengetahui jenis kapal yang sering melakukan penundaan pembayaran berdasarkan data PHQC di Pelabuhan Regional Dumai. Penelitian ini juga mengidentifikasi variabel apa yang mempengaruhi kecenderungan keterlambatan pembayaran pengecekan kesehatan kapal menggunakan *decision tree*. Pemegang kebijakan dalam pelaksanaan PHQC selanjutnya dapat melakukan *treatment* berdasarkan variabel tersebut untuk menekan jumlah kapal yang mengalami keterlambatan pembayaran, sehingga pelabuhan dapat beroperasi secara maksimal.

## II. METODE PENELITIAN

Metode Penelitian yang digunakan pada penelitian ini adalah sebagai berikut:

### A. Regresi Logistik

Regresi logistik adalah sebuah pendekatan untuk membuat model prediksi seperti halnya regresi linear atau yang biasa disebut dengan istilah *Ordinary Least Squares* (OLS) regression. Regresi logistik memprediksi variabel terikat yang berskala dikotomi. Skala dikotomi yang dimaksud adalah skala data nominal dengan dua kategori. Model regresi logistik (LR) dapat digunakan untuk memprediksi probabilitas kelas berdasarkan fitur input, setelah peringkat mereka sesuai dengan kepentingan relatif mereka [3].

### B. Decision Tree

Metodologi *Decision Tree* adalah metode *data mining* yang umum digunakan untuk membangun sistem klasifikasi berdasarkan beberapa kovariat atau untuk mengembangkan algoritma prediksi untuk variabel target. *Decision tree* memiliki struktur pohon seperti bagan aliran, di mana setiap node internal menunjukkan tes pada atribut, setiap cabang mewakili hasil dari tes, dan node daun mewakili kelas atau distribusi kelas [4]. *Decision Tree* pada penelitian ini menggunakan algoritma *Classification and Regression Tree* (CART) Algoritma dapat secara efisien menangani kumpulan data yang besar, baik data tersebut berupa variabel numerik maupun variabel kategorik [5]. CART akan menghasilkan *classification tree* jika variabel output berupa kategori, dan akan menghasilkan *regression tree* jika variabel output berupa numerik [6]. Ketika ukuran sampel cukup besar, data studi dapat dibagi menjadi set data pelatihan (training) dan validasi (testing). Dataset training digunakan untuk membangun model *Decision Tree*, sedangkan dataset testing untuk memutuskan ukuran *tree* yang sesuai yang diperlukan untuk mencapai model akhir yang optimal. Penggunaan umum metode *Decision Tree* sebagai berikut:

1. Pemilihan variabel. Banyak dari variabel ini memiliki relevansi marjinal, oleh karena itu mungkin tidak boleh dimasukkan dalam training Data Mining. Seperti pemilihan variabel bertahap dalam analisis regresi, metode *Decision Tree* dapat digunakan untuk memilih variabel input paling relevan yang harus digunakan untuk membentuk model pohon keputusan, yang selanjutnya dapat digunakan untuk merumuskan hipotesis dan menginformasikan penelitian selanjutnya.
2. Menilai kepentingan relatif variabel. Variabel penting dihitung berdasarkan pengurangan akurasi model (atau dalam kemurnian simpul di pohon) ketika

variabel dihapus. Dalam sebagian besar keadaan, semakin banyak catatan variabel berpengaruh, semakin besar pentingnya variabel.

3. Penanganan nilai yang hilang. Metode umum dalam menangani data yang hilang adalah untuk mengecualikan kasus dengan nilai yang hilang; ini tidak efisien dan berisiko menimbulkan bias dalam analisis. Analisis *Decision Tree* dapat menangani data yang hilang dengan dua cara: itu dapat mengklasifikasikan nilai yang hilang sebagai kategori terpisah yang dapat dianalisis dengan kategori lain atau menggunakan model *Decision Tree* yang dibangun dengan mengatur variabel dengan banyak nilai yang hilang sebagai variabel target untuk membuat prediksi dan mengganti yang hilang ini dengan nilai prediksi.
4. Prediksi. Ini adalah salah satu penggunaan paling penting dari model *Decision Tree*. Menggunakan model tree yang berasal dari data historis, mudah untuk memprediksi hasil untuk catatan di masa mendatang.
5. Manipulasi data. Banyaknya kategori dari satu *categorical variable* atau data kontinu adalah umum dalam penelitian. Dalam keadaan ini, model *Decision Tree* dapat membantu dalam memutuskan bagaimana cara terbaik menentukan variabel kategori menjadi jumlah kategori yang lebih mudah dikelola atau bagaimana membagi variabel dalam rentang angka tertentu

### C. Analisis Dataset

Data PHQC diperoleh dari KKP Riau. Data tersebut merupakan data transaksi PHQC selama tahun 2017-2019 (hingga April 2019) berisi 50.497 record data.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
PelabuhanLabel	String	29	0		None	None	29	Left	Nominal	Input
JenisKapal	String	9	0		None	None	9	Left	Nominal	Input
BenderaKapal	String	22	0		None	None	22	Left	Nominal	Input
Berat	Numeric	12	1		None	None	12	Right	Scale	Input
Klasifikasi	Numeric	12	1		None	None	12	Right	Nominal	Input
TARIF	Numeric	12	1		None	None	12	Right	Scale	Input
PelabuhanKedatangan	String	31	0		None	None	31	Left	Nominal	Input
PelabuhanBerikutnya	String	30	0		None	None	30	Left	Nominal	Input
TanggalTerbit	Date	11	0		None	None	11	Right	Scale	Input
JamTerbit	Date	11	2		None	None	11	Right	Scale	Input
cat_time	Numeric	8	2		None	None	8	Right	Nominal	Input
TanggalPublished	Date	20	0		None	None	20	Right	Scale	Input
STATUSBIAYA	String	9	0		None	None	9	Left	Nominal	Input
BULAN	Numeric	12	0		None	None	12	Right	Nominal	Input
Hari	Numeric	8	2		None	None	8	Right	Nominal	Input
Time	String	9	0		None	None	9	Left	Nominal	Input

Figure 1. Daftar Variabel Dataset

## III. HASIL PENELITIAN

Pada tahapan ini dilakukan analisis data, pembentukan model, uji coba dan evaluasi model. Model yang diperoleh dapat diubah berdasarkan hasil ujicoba dan evaluasi apakah fit atau tidak.

### A. Analisis Data

Pada tahapan ini dibentuk satu variabel baru yang digunakan sebagai variabel dependent/predictor. Variabel yang dibentuk diberi inisialisasi “**status\_biaya**”, variabel ini bersifat dikotomi berupa “**Terlambat**” dan “**Lunas**”. Variabel ini diperoleh dari perbedaan antara waktu terbit dan waktu publish pada raw data. Perbedaan ini menunjukkan keterlambatan atau penundaan pembayaran pengecekan kesehatan kapal.

Tahapan selanjutnya dilakukan pengkodean data. Pengkodean atau encode merupakan pemberian coding pada data sehingga format data nya seragam dan bersifat kategorik. Pada

umumnya pengkodean menggunakan format numeric, sehingga lebih mudah di analisis. Pengkodean harus terdokumentasikan dengan baik untuk memudahkan interpretasi data.

## B. Pembangunan Model

### 1) Model Regresi Logistik

Dengan variabel dependent bersifat dikotomi dan variabel bebas kategorikal, metode yang dapat digunakan adalah Regresi Logistik. Pemodelan menggunakan Tools SPSS dengan tingkat kepercayaan = 95% dan  $\alpha = 5\%$ .

**Model Summary**

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	29981.183 <sup>a</sup>	.032	.070

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Figure 2. Ringkasan Model

**Variables in the Equation**

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>						
PelabuhanLabel(1)	-1.584	.047	1127.771	1	.000	.205
JenisKapal(1)	-1.010	.036	802.482	1	.000	.364
Klasifikasi	.070	.008	68.178	1	.000	1.072
BULAN	-.207	.031	44.750	1	.000	.813
Constant	3.055	.063	2375.473	1	.000	21.218

a. Variable(s) entered on step 1: PelabuhanLabel, JenisKapal, Klasifikasi, BULAN.

Figure 3. Korelasi Variabel Prediktor

Berdasarkan hasil di atas dapat diperoleh persamaan model:

$$Y = \beta + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4$$

$$Y = 3,055 - 1,58x_1 - 1,01x_2 + 0,07x_3 - 0,207x_4$$

Y = Status Keterlambatan  
 $X_1$  = Pelabuhan  
 $X_2$  = Jenis Kapal  
 $X_3$  = Klasifikasi Kapal  
 $X_4$  = Bulan

### 2) Model Decision Tree

Pemodelan *decision tree* menggunakan *RapidMiner* sebagai *tools* analisis. Variabel **klasifikasi kapal** berdasarkan berat menjadi variabel terikat/variabel yang diukur. **Tarif, kategori waktu kedatangan, pelabuhan, klasifikasi bendera, hari dan bulan** menjadi variabel prediktor. Sebagian model decision yang dihasilkan dapat dilihat pada Figure 4.

```

Tree
TARIF = 10000
| cat_time = afternoon
| | Pelabuhan Label = Pelabuhan Laut Bengkalis: 3 {11=0, 1=0, 10=0, 9=0, 8=0, 7=0, 6=0, 2=0, 5=0, 4=7, 3=145}
| | Pelabuhan Label = Pelabuhan Laut Dumai
| | | Hari = Friday: 3 {11=0, 1=0, 10=0, 9=0, 8=0, 7=0, 6=0, 2=0, 5=0, 4=0, 3=3}
| | | Hari = Monday: 3 {11=0, 1=0, 10=0, 9=0, 8=0, 7=0, 6=0, 2=0, 5=0, 4=8, 3=15}
| | | Hari = Thursday
| | | | BULAN = 1: 3 {11=0, 1=0, 10=0, 9=0, 8=0, 7=0, 6=0, 2=0, 5=0, 4=0, 3=2}
| | | | BULAN = 2: 3 {11=0, 1=0, 10=0, 9=0, 8=0, 7=0, 6=0, 2=0, 5=0, 4=1, 3=2}
| | | | BULAN = 4: 4 {11=0, 1=0, 10=0, 9=0, 8=0, 7=0, 6=0, 2=0, 5=0, 4=2, 3=0}
| | | | BULAN = 6: 4 {11=0, 1=0, 10=0, 9=0, 8=0, 7=0, 6=0, 2=0, 5=0, 4=2, 3=2}
| | | | Hari = Tuesday: 3 {11=0, 1=0, 10=0, 9=0, 8=0, 7=0, 6=0, 2=0, 5=0, 4=2, 3=4}
| | | | Hari = Wednesday: 3 {11=0, 1=0, 10=0, 9=0, 8=0, 7=0, 6=0, 2=0, 5=0, 4=0, 3=5}
| | Pelabuhan Label = Pelabuhan Laut Sungai Pakning: 4 {11=0, 1=0, 10=0, 9=0, 8=0, 7=0, 6=0, 2=0, 5=0, 4=3, 3=1}
| cat_time = evening
| | Pelabuhan Label = Pelabuhan Laut Bengkalis: 4 {11=0, 1=0, 10=0, 9=0, 8=0, 7=0, 6=0, 2=0, 5=1, 4=4, 3=0}
| | Pelabuhan Label = Pelabuhan Laut Dumai: 3 {11=0, 1=0, 10=0, 9=0, 8=0, 7=0, 6=0, 2=0, 5=0, 4=3, 3=5}
| cat_time = morning: 4 {11=0, 1=0, 10=0, 9=0, 8=0, 7=0, 6=0, 2=0, 5=1, 4=1114, 3=268}
TARIF = 100000: 9 {11=0, 1=0, 10=0, 9=50, 8=0, 7=0, 6=0, 2=0, 5=0, 4=0, 3=0}
TARIF = 125000: 10 {11=0, 1=0, 10=19, 9=0, 8=0, 7=0, 6=0, 2=0, 5=0, 4=0, 3=0}
TARIF = 150000: 11 {11=101, 1=0, 10=0, 9=0, 8=0, 7=0, 6=0, 2=0, 5=0, 4=0, 3=0}
TARIF = 200000: 1 {11=0, 1=654, 10=0, 9=0, 8=0, 7=0, 6=0, 2=0, 5=0, 4=0, 3=0}
TARIF = 250000: 2 {11=0, 1=0, 10=0, 9=0, 8=0, 7=0, 6=0, 2=311, 5=0, 4=0, 3=0}
TARIF = 300000: 3 {11=0, 1=0, 10=0, 9=0, 8=0, 7=0, 6=0, 2=0, 5=0, 4=0, 3=141}
TARIF = 350000: 4 {11=0, 1=0, 10=0, 9=0, 8=0, 7=0, 6=0, 2=0, 5=0, 4=88, 3=0}
TARIF = 5000
| Klas Bendera = Asing: 2 {11=0, 1=0, 10=0, 9=0, 8=0, 7=0, 6=0, 2=2, 5=0, 4=0, 3=0}
| Klas Bendera = Indonesia

```

Figure 4. Model Decision Tree

### C. Evaluasi Model

#### 1) Evaluasi Model Regresi Logistik

Evaluasi Model Regresi digunakan *The Hosmer-Lemeshow goodness of fit test* [7] dengan menggunakan pendekatan Chi-Square. Uji ini melibatkan pengelompokan pengamatan berdasarkan probabilitas yang diharapkan dan kemudian menguji hipotesis bahwa perbedaan antara peristiwa yang diamati dan yang diharapkan adalah nol untuk semua kelompok secara bersamaan.

$$\hat{\pi} = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_p X_p)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_p X_p)}$$

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	126.615	8	.000

Figure 5. Hasil Hosmer and Lemeshow Test

Berdasarkan hasil evaluasi di atas menggunakan *The Hosmer-Lemeshow goodness of fit test*, model yang dibentuk belum fit karena signifikan menolak  $H_0$  (Hipotesis awal). Hal ini dikarenakan angka observasi dan ekspektasi yang berbeda cukup jauh dan signifikan.

#### 2) Evaluasi Model Decision Tree

Evaluasi decision tree menggunakan *Rapid Miner* ditunjukkan dengan *confusion matrix* sebagai berikut, dimana model yang dibentuk memiliki tingkat akurasi sebesar 87,21 %

accuracy: 87.21% +/- 1.83% (micro average: 87.21%)

	true 11	true 1	true 10	true 9	true 8	true 7	true 6	true 2	true 5	true 4	true 3	class pre...
pred. 11	101	0	0	0	0	0	0	0	0	0	0	100.00%
pred. 1	0	1078	0	0	0	0	0	71	0	0	0	93.82%
pred. 10	0	0	19	0	0	0	0	0	0	0	0	100.00%
pred. 9	0	0	0	50	0	0	0	0	0	0	0	100.00%
pred. 8	0	0	0	0	45	0	0	0	0	0	0	100.00%
pred. 7	0	0	0	0	0	75	0	0	0	0	0	100.00%
pred. 6	0	0	0	0	0	0	213	0	0	0	0	100.00%
pred. 2	0	241	0	0	0	0	0	834	0	0	0	77.58%
pred. 5	0	0	0	0	0	0	0	0	221	0	0	100.00%
pred. 4	0	0	0	0	0	0	0	0	2	1206	269	81.65%
pred. 3	0	0	0	0	0	0	0	0	0	28	324	92.05%
class rec...	100.00%	81.73%	100.00%	100.00%	100.00%	100.00%	100.00%	92.15%	99.10%	97.73%	54.64%	

Figure 6. Evaluasi Decision Tree

#### IV. KESIMPULAN

Keterlambatan Pembayaran pengecekan kapal dipengaruhi oleh variabel Pelabuhan, Jenis, Klasifikasi dan Bulan dan Tarif. Hasil model regresi belum fit untuk analisis kecenderungan keterlambatan pembayaran, sehingga dilakukan analisis lanjutan menggunakan model *Decision Tree* untuk memprediksi kapal yang memiliki kemungkinan terlambat melakukan pembayaran. Berdasarkan hasil penelitian, variabel tarif yang merupakan turunan dari variabel berat kapal menjadi variabel prediktor utama yang menentukan kecenderungan keterlambatan pembayaran. KKP dapat melakukan antisipasi terhadap kemungkinan keterlambatan pembayaran berdasarkan tipe klasifikasi berat kapal yang melakukan keterlambatan pembayaran. Sebagian besar kapal yang mengalami keterlambatan pembayaran berasal dari kapal non-ferry yaitu kapal membawa muatan ekspor-impor, sehingga diperlukan treatment berbeda untuk kapal-kapal tersebut untuk mempercepat proses pembayaran

#### REFERENSI

- [1] Kementerian Kesehatan RI, *Klasifikasi Kantor Kesehatan Pelabuhan*. Indonesia, 2014.
- [2] S. Rahmadika, P. H. Rusmin, H. Hindersah, and K. H. Rhee, "Providing data integrity for container dwelling time in the seaport," in *Proceedings - 2016 6th International Annual Engineering Seminar, InAES 2016*, 2017.
- [3] Q. Cheng, P. K. Varshney, and M. K. Arora, "Logistic regression for feature selection and soft classification of remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, 2006.
- [4] Lanshen Guo, Minglu Zhang, Lixin Sun, and Zhong Wang, "Churn Analysis Model of Securities Business Based on the Decision Tree," in *2006 6th World Congress on Intelligent Control and Automation*, 2006, pp. 6048–6051.
- [5] A. F. A. Pinem and E. B. Setiawan, "Implementation of classification and regression Tree (CART) and fuzzy logic algorithm for intrusion detection system," in *2015 3rd International Conference on Information and Communication Technology, ICoICT 2015*, 2015.
- [6] G. Vagliasindi, L. Fortuna, A. Murari, G. Mazzitelli, and A. Gallo, "An Automatic Identifier of Confinement Regimes at JET combining Fuzzy Logic and Classification Trees," *Analysis*, 2008.

- [7] B. V. Shah and B. G. Barnwell, “Hosmer-Lemeshow goodness of fit test for Survey data,” *Proc. Jt. Stat. Meet. Surv. Res. Methods*, no. 3, pp. 3778–3781, 2003.