

PENGUNAAN DATA MINING DALAM HIT RATE IMPORTASI JALUR MERAH DENGAN MODEL DECISION TREE

Alfin Yudistira¹⁾, Muh Nurkhamid²⁾

¹⁾ Kantor Pelayanan Utama Bea dan Cukai Tipe A Tanjung Priok, Jakarta Utara, 14310

²⁾ Dosen Jurusan Kepabeanan dan Cukai, Politeknik Keuangan Negara STAN, Bintaro Sektor V, Tangerang Selatan, Indonesia, 15222

E-mail: alfinyudistira@gmail.com

INFORMASI ARTIKEL

Tanggal masuk
[11-08-2021]

Revisi
[28-10-2021]

Tanggal terima
[27-11-2021]

ABSTRACT:

Customs and Excise faces a big challenge to be able to increase the hit rate of red line imports by 40% in accordance with the Blueprint for the 2014-2025 Ministry of Finance Institutional Transformation Program and international benchmarks. Through a qualitative study, this study aims to determine the use of data mining that is applied to the risk engine based on import data, people's experiences, and research results of customs institutions of other countries. The data mining method used is CRISP-DM, classification method, and decision tree model, using data imported from the red line KPU BC Type A Tanjung Priok for the period September – December 2019 and January 2020. The results show that the use of data mining can increase the hit rate of red line importation. The most relevant attribute in classifying data is the sending country which is categorized as a root node, while the import duty tariff attribute does not provide information on data classification. This research is expected to provide a new perspective for the KPU BC Type A Tanjung Priok in an effort to improve the risk engine targeting and risk engine routing of Customs and Excise.

Keywords: CRISP-DM, data mining, decision tree, hit rate, the red line import.

ABSTRAK:

Bea dan Cukai menghadapi tantangan besar untuk dapat meningkatkan capaian *hit rate* importasi jalur merah sebesar 40% sesuai dengan Cetak Biru Program Transformasi Kelembagaan Kementerian Keuangan Tahun 2014 – 2025 dan *benchmark* internasional. Melalui studi kualitatif, penelitian ini bertujuan untuk mengetahui penggunaan *data mining* yang diterapkan dalam *risk engine* berdasarkan data importasi, pengalaman orang, dan data hasil penelitian institusi kepabeanan negara lain. Metode *data mining* yang digunakan adalah CRISP-DM, metode klasifikasi, dan model *decision tree*, dengan menggunakan data importasi jalur merah Kantor Pelayanan Utama (KPU) Bea dan Cukai (BC) Tipe A Tanjung Priok periode September – Desember 2019 dan Januari 2020. Hasil penelitian menunjukkan bahwa penggunaan *data mining* dapat meningkatkan capaian *hit rate* importasi jalur merah. Atribut yang paling relevan dalam mengklasifikasikan data adalah negara pengirim yang dikategorikan sebagai *root node* (akar), sedangkan atribut tarif bea masuk tidak memberikan informasi dalam klasifikasi data. Penelitian ini diharapkan dapat memberikan pandangan baru bagi KPU BC Tipe A Tanjung Priok dalam upaya perbaikan *risk engine targeting* dan *risk engine* penjaluran Bea dan Cukai.

Kata Kunci: CRISP-DM, data mining, decision tree, hit rate, importasi jalur merah.

1. PENDAHULUAN

Pengawasan kepabeanan berupa pemeriksaan fisik barang, penelitian dokumen, audit, dan patroli dilakukan untuk menemukan adanya pelanggaran kepabeanan atau tidak (Semedi, 2013). Secara umum di Indonesia masih terdapat pelanggaran kepabeanan yang menyebabkan kekurangan atau potensi kehilangan penerimaan negara.

Sepanjang tahun 2019 DJBC menemukan 67 kasus penyelundupan kendaraan (Priasmoro, 2019). Cara penyelundupan dilakukan mulai dari pemberitahuan kepabeanan yang salah, pengeluaran dan pembongkaran barang impor tanpa izin, bahkan tanpa pemberitahuan kepabeanan (Fauzia, 2019). Dihitung secara nasional dari tahun 2016 sampai 2019, penyelundupan telah merugikan negara sekitar Rp650-an miliar (Jatmiko, 2019).

Sebagai upaya mengimbangi risiko penambahan volume dan kompleksitas perdagangan internasional, Bea dan Cukai melakukan manajemen risiko berupa pemeriksaan Bea dan Cukai. Penerapan manajemen risiko juga diimplementasikan pada Sistem Komputer Pelayanan (SKP)/*risk engine*

untuk penentuan kriteria jalur impor. *Output risk engine* berupa kriteria jalur impor merah, kuning, hijau, Mitra Utama (MITA), dan MITA Prioritas (Firdiansyah & Nugroho, 2017).

DJBC memiliki indikator *hit rate* untuk mendeteksi dan menemukan pelanggaran melalui penjaluran barang importasi (Pramudyo & Arimbhi, 2018). *Hit rate* dihitung pada Pemberitahuan Impor Barang (PIB) jalur merah berisiko *medium* dan *high*. Perhitungan secara nasional hanya dilakukan pada empat kantor besar, salah satunya adalah KPU BC Tipe A Tanjung Priok.

Perhitungan *hit rate* adalah salah satu Indikator Kinerja Utama (IKU) dari Direktorat Jenderal (Dirjen) Bea dan Cukai sejak tahun 2014. Dari tahun ke tahun, DJBC berhasil mencapai angka yang ditargetkan dalam IKU, tetapi belum menunjukkan hasil yang ideal berdasarkan *best practice* internasional yaitu sebesar 40% (Pramudyo & Arimbhi, 2018). KPU BC Tipe A Tanjung Priok yang menyumbang pelayanan dokumen impor sekitar 60 - 70% di Indonesia menurut Nurhidayati & Cahyani (2020) juga belum sepenuhnya memenuhi angka target DJBC seperti data capaian yang tersaji pada tabel 1.1.

Tabel 1.1 Capaian Hit Rate DJBC dan KPU BC Tipe A Tanjung Priok Periode 2014 sampai dengan 2019

Tahun	PIB Merah	SPTNP	SPBL	Temuan Lain	Hit Rate Tj. Priok (%)	Target DJBC (%)	Realisasi DJBC (%)
(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
2014	41926	5.965	-	-	14,23	10	18,00
2015	37901	4.170	-	-	11	12	25,2
2016	32870	5.034	401	-	16,53	25	34,70
2017	46330	10.596	322	-	23,57	25	32,27
2018	50499	11.420	460	942	25,39	25	28,84
2019	48094	11.241	275	776	25,56	26	30,42

Sumber: Diolah dari Laporan Kinerja DJBC Tahun 2014 – 2019

Berdasarkan tabel 1.1, rata-rata capaian *hit rate* DJBC pada tahun 2014-2019 hanya 28,23% dan capaian KPU BC Tipe A Tanjung Priok sebesar rata-rata 19,38%. *Risk engine* hanya mampu mendeteksi pelanggaran rata-rata 28,23% per tahun secara nasional, dan rata-rata 19,38% di KPU BC Tipe A Tanjung Priok. Capaian *hit rate* KPU BC Tipe A Tanjung Priok dengan volume dokumen impor tertinggi di Indonesia yang masih rendah akan berdampak pada capaian DJBC secara keseluruhan.

Penelitian yang dilakukan oleh Mubarak dalam Nurhidayati dan Cahyani (2020) mengungkapkan bahwa *risk profiling system* pada DJBC belum bekerja efektif. Hasil ini diperkuat dengan penelitian Dlava (2012) dan Pramudyo & Arimbhi (2018) bahwa *updating* data importir belum optimal sehingga mengganggu kinerja *risk*

engine. Berkaitan dengan tingkat kepatuhan kepabeanan di Indonesia, penelitian Syaifullah & Ramdany (2020) mengungkapkan bahwa kepatuhan kepabeanan eksportir dan importir Indonesia pada tingkatan *moderate*. Tinggi pelanggaran juga dapat dilihat pada tabel 1.1, kolom c, d, dan e yang menunjukkan jumlah pelanggaran dari seluruh PIB jalur merah pada kolom b. Oleh karena itu, sebagai upaya mengoptimalkan capaian *hit rate*, DJBC melakukan penyempurnaan *risk engine* melalui Instruksi Dirjen Bea dan Cukai Nomor INS-01/BC/2018 terkait pemutakhiran profil importir dan profil komoditas. Mulai 1 November 2019 DJBC melakukan optimalisasi *risk engine* yang akan diimplementasikan kepada seluruh kantor.

Bea dan Cukai semakin mengandalkan teknologi untuk

transparansi dan prediksi aliran barang, jasa, dan orang (Jonathan, 2018). Aktivitas untuk menemukan sebuah proses, menganalisis klaster data dengan perspektif berbeda hingga menemukan informasi yang berguna dikenal dengan *data mining* (Okazaki, 2017). Sejak dahulu manusia melakukan aktivitas intelijen secara sistematis, baik dengan membangun model proses ataupun membuat pohon keputusan secara manual (Jonathan, 2018). *Decision tree* dapat digunakan untuk menghitung kemungkinan-kemungkinan yang kompleks untuk digunakan sebagai rencana kontijensi dan penentuan strategi mitigasi (Firdiansyah, 2019).

Penelitian manajemen risiko kepabeanan yang dilakukan oleh (Al-Shbail (2020) pada *Jordan Customs* menyebutkan bahwa rekam jejak operator (importir atau *consignee*), eksportir di negara asal barang, asal pelabuhan muat, negara asal, tipe komoditas, dan informasi intelijen berpengaruh dalam sistem manajemen risiko. Penelitian yang dilakukan terhadap *South African Customs*, *consignor* (entitas yang menyerahkan pemberitahuan pabean), negara asal, dan *cargo type* yang diinterpretasikan pada

kode *Harmonized System* (HS) berkorelasi rendah dengan pelanggaran kepabeanan sehingga tidak menghasilkan *hit rate* yang tinggi (Hoffman et al., 2018). Penelitian dengan *data mining* untuk meningkatkan *hit rate* importasi jalur merah belum pernah dilakukan di Indonesia, tetapi metode ini pernah digunakan dalam penelitian Lestari & Silaban (2018) untuk mencari korelasi antaratribut-atribut dalam penetapan SPTNP (Surat Penetapan Tarif dan/atau Nilai Pabean).

Rumusan masalah dalam penelitian ini adalah bagaimana merancang *data mining* untuk *hit rate* importasi jalur merah pada KPU BC Tipe A Tanjung Priok dan apakah hasil *data mining* tersebut dapat meningkatkan capaian *hit rate* importasi jalur merah pada KPU BC Tipe A Tanjung Priok.

Penelitian ini bertujuan untuk mengetahui penggunaan *data mining* yang dapat diterapkan dalam *risk engine* berdasarkan data importasi, pengalaman orang, dan perspektif data hasil penelitian institusi kepabeanan negara lain sehingga dapat meningkatkan capaian *hit rate* importasi jalur merah pada KPU BC Tipe A Tanjung Priok.

2. KAJIAN LITERATUR

2.1 Agency Theory

Jensen & Meckling (1976) mendefinisikan hubungan keagenan sebagai kontrak di mana ada satu pihak atau lebih yang disebut *principal*, dengan pihak lain yang disebut *agent*. Dalam hubungan tersebut *principal* diwakili oleh *agent* yang mendapatkan delegasi atas otorisasi untuk mengambil keputusan. Namun, adanya asimetri informasi, memberi celah bagi *agent* untuk bertindak oportunistik (Kasper & Streit, 1999).

Prinsip *self assessment* dalam Undang-Undang Nomor 17 tahun 2006 dimaksudkan untuk kelancaran pelayanan atas aliran barang, orang, serta dokumen. Namun, masih ada pengguna jasa yang memanfaatkan *self assessment* untuk melakukan pelanggaran kepabeanan (Herlinawati *et al.*, 2016). Pengguna jasa mencari celah dengan memberikan keterangan pada dokumen kepabeanan yang dapat menguntungkan dirinya dengan memberikan informasi yang dapat menurunkan jumlah kewajiban pajak dalam rangka impor dengan tarif yang tidak sesuai, memberikan keterangan kuantitas yang lebih sedikit, atau cara lainnya. Di lain sisi, DJBC berupaya untuk

mengamankan penerimaan negara. Adanya perbedaan kepentingan tersebut mengharuskan DJBC untuk melakukan pengawasan yang lebih optimal.

2.2 Signal Detection Theory

Signal Detection Theory berhubungan dengan kemampuan klasifikasi *observer* (penerima) untuk mendeteksi *signal* di tengah keberadaan *noise* (Spackman, 1989). Teori ini berperan dalam proses pengambilan keputusan dengan mengkarakteristikan penerima ketika mengambil keputusan dengan cara memisahkan informasi yang berguna dengan *noise*. Saat ini penggunaannya umum untuk menilai memori, persepsi, ataupun tugas-tugas kategorisasi (Lynn & Barrett, 2014). Teori ini juga berguna dalam situasi yang tidak tentu atau mengandung risiko. Hasil penjaluran importasi pada *risk engine* DJBC dapat dijelaskan dalam empat kemungkinan hasil, seperti pada gambar 2.1 berikut.

	Pelanggaran	Tidak Pelanggaran
Merah	((1) <i>Hit</i>	(2) <i>False Alarm</i>
Tidak Merah	(3) <i>Miss</i>	(4) <i>Correct Identification</i>
	<i>Sensitivity</i>	<i>Specificity</i>

Sumber : Diolah dari Karim & Siegel
(1998)

Gambar 2.1 *Confusion Matrix Signal
Detection Theory*

2.3 Manajemen Risiko Kepabeanan

Dalam kegiatan administrasi kepabeanan, manajemen risiko diperlukan untuk mengatur pergerakan lalu lintas orang, barang, dan transportasi (WCO, 2011). Instistusi kepabeanan harus dapat mengimbangi perkembangan dengan menerapkan manajemen risiko untuk mengendalikan dan meminimalisasi tantangan. Secara umum, pengguna jasa kepabeanan dibagi menjadi empat kategori umum, yaitu sukarela patuh, berusaha patuh tetapi tidak selalu berhasil dalam usahanya, berusaha menghindari kepatuhan apabila memungkinkan, dan sengaja tidak mematuhi peraturan (WCO, 2011). Dengan kategori tersebut diperlukan perlakuan yang berbeda berdasarkan masing-masing tingkat risiko. Salah satu *treatment* yang dilakukan adalah manajemen penjaluran importasi.

DJBC melakukan aktivitas intelijen dengan mengembangkan *risk engine* penentuan jalur importasi dalam Sistem Analisis Risiko Importasi berdasarkan Instruksi Dirjen Bea dan Cukai Nomor

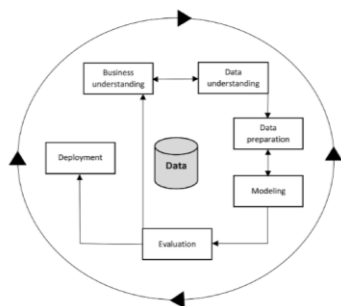
INS-03/BC/2019 yang diterapkan dalam skala 192rowt maupun nasional dengan parameter *targeting*. Implementasinya, diperlukan peran unit 192rowth192n dalam 192rowth192n, mengevaluasi, serta memutakhirkan sistem analisis risiko importasi yang bersifat 192rowt.

2.4 Data Mining

Berdasarkan 192rowth dalam Chang & Chen (2006) dijelaskan bahwa *data mining* adalah proses menemukan *trends* dan pola pada data. *Data mining* diperlukan guna mengungkap informasi di balik data yang jumlahnya sangat banyak sehingga penggunaan analisis data secara tradisional tidak dimungkinkan. Saat ini berkembang metodologi yang digunakan untuk proses *data mining*, salah satunya adalah *Cross-Reference Industry Process for Data Mining* (CRISP-DM) yang dikembangkan pada pertengahan tahun 1990an (Xia & Gong, 2014).

2.5 Cross-Reference Industry Process for Data Mining (CRISP-DM)

CRISP-DM merupakan salah satu metodologi yang digunakan dalam proses *data mining* yang terdiri dari tahapan-tahapan untuk proses *data mining*. CRISP-DM terdiri dari enam tahapan proses sebagai berikut.



Sumber : Wirth & Hipp dalam Jaggia *et al.* (2020)

Gambar 2.2 Tahapan CRISP-DM

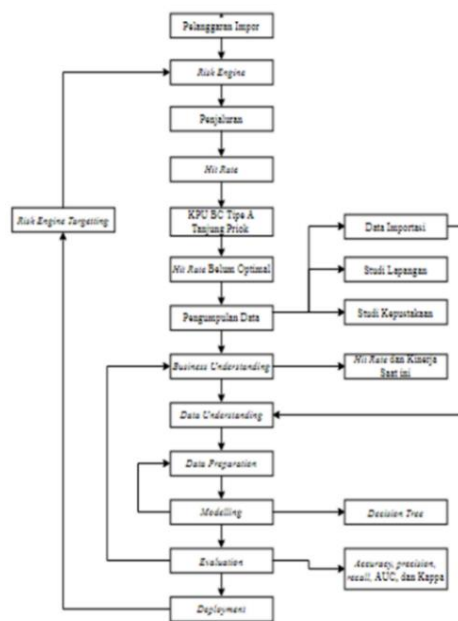
2.6 Decision Tree

Decision Tree adalah model yang paling sederhana dan cara tercepat yang dapat digunakan dalam metode klasifikasi dalam *data mining* (Sohrabi *et al.*, 2019). Penggunaan *decision tree* dipilih karena menghasilkan model tidak rumit dan mudah diinterpretasikan (Vulandari, 2017, p. 14). *Decision tree* terdiri dari akar dan cabang. Akar merupakan atribut yang menjadi fokus atensi utama yang didapatkan berdasarkan *information gain* Du *et al.* (2002), menggunakan *gain ratio* sebagai perhitungan *splitting* pohon algoritma C4.5 White & Liu (1994) dan untuk menentukan fitur yang paling penting/atribut yang relevan (Karegowda *et al.*, 2010). Dalam penelitian ini *decision tree* digunakan dalam penentuan atribut yang paling relevan untuk mengklasifikasi data

importasi jalur merah sehingga dapat digunakan sebagai informasi dalam *risk engine targeting* dan penjaluran.

3. METODE PENELITIAN

Metodologi penelitian yang digunakan dalam penelitian ini adalah penelitian kualitatif berupa studi kasus. Model penelitian ditunjukkan pada gambar 3.1. Penulis menganalisis dua sumber data yang berbeda yaitu hasil wawancara kepada pegawai DJBC sebagai narasumber dan data importasi. Penelitian pemanfaatan *data mining* pada data kepabeanan dengan model *decision tree* dilakukan oleh Chermiti (2019). Penelitian dilakukan dengan membangun *risk profile* untuk memprediksi pelanggaran dalam pemberitahuan pabean. Hasilnya menunjukkan bahwa negara pengiriman akhir (*country of last shipment*) dan *reporting firm kode* (kode entitas yang menyampaikan dokumen kepabeanan) yang paling relevan dengan faktor risiko dalam deklarasi impor.



Sumber: Penulis

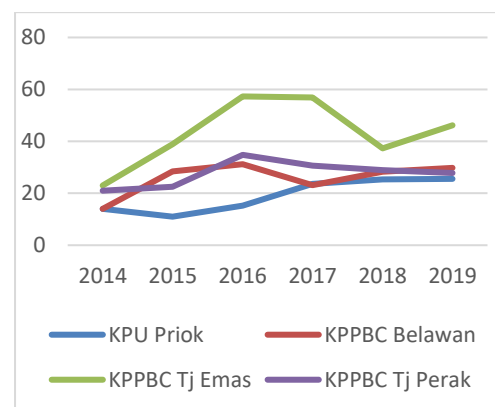
Gambar 3.1 Model Penelitian

Jenis data yang digunakan dalam penelitian ini adalah data primer dan data sekunder. Data primer adalah informasi wawancara yang dilakukan kepada narasumber pegawai DJBC pada Direktorat Penindakan dan Penyidikan (P2) dan KPU BC Tipe A Tanjung Priok. Data sekunder berupa informasi *database* impor KPU BC Tipe A Tanjung Priok periode September 2019 sampai dengan Desember 2019 dan Januari 2020. Pengolahan *data mining* dengan metodologi CRISP-DM. Lingkup penelitian dibatasi terkait *hit rate* importasi jalur merah pada KPU BC Tipe A Tanjung Priok.

4. HASIL DAN PEMBAHASAN

4.1 Business Understanding

Hit rate telah dijadikan IKU oleh DJBC sejak 2014 hingga saat ini. Variabelnya dengan SPTNP, Surat Penetapan Barang Larangan/Pembatasan (SPBL), perubahan HS, dan jumlah PIB jalur merah. Berikut disampaikan data capaian *hit rate* 4 kantor pelayanan yang menjadi objek perhitungan *hit rate* periode tahun 2014 sampai dengan 2019.



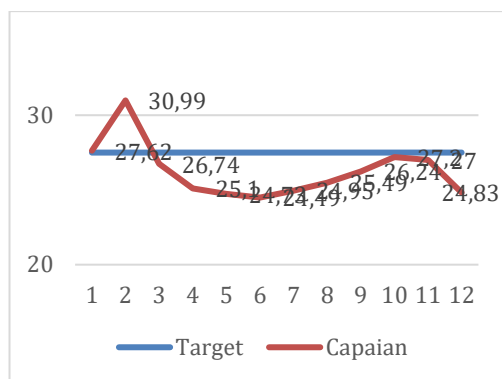
Sumber: Diolah dari Laporan Kinerja DJBC Tahun 2014 – 2019

Gambar 4.1 Capaian *Hit Rate* 4 Kantor Besar Tahun 2014 Sampai Dengan 2019

Terkait capaian dari 4 kantor besar, Kantor Pengawasan dan Pelayanan (KPPBC) Tanjung Emas merupakan kantor dengan capaian rata-rata tertinggi dibandingkan dengan 3 lainnya.

Dari pergerakan grafik gambar 4.1, KPU BC Tipe A Tanjung Priok merupakan satu-satunya yang memiliki

trend kecenderungan capaian naik dibandingkan dengan 3 kantor lainnya. Gambar 4.2 menunjukkan capaian *hit rate* KPU BC Tipe A Tanjung Priok Tahun 2020.



Sumber : Diolah dari Direktorat P2
(2021)

Gambar 4.2 Capaian *Hit Rate* KPU BC Tipe A Tanjung Priok Tahun 2020

Berdasarkan perhitungan IKU *hit rate* dengan *take last known value* (TLKV), capaian *hit rate* KPU BC Tipe A Tanjung Priok tahun 2020 sebesar 24,83% atau lebih rendah 0,73% dibandingkan capaian tahun 2019 sebesar 25,56%.

4.2 Data Understanding

Terdapat sembilan atribut data yaitu: cara pembayaran, pelabuhan muat, negara pemasok (pasok), negara pengirim (kirim), *Harmonized System Code* atau kode HS, negara asal (asal), tarif bea masuk (tarif BM), tarif pajak

penghasilan impor (tarif PPh), dan HIT MISS sebagai atribut label.

4.3 Data Preparation

Pengkategorian atribut dan label awal dijelaskan sebagai berikut:

1. **Cara pembayaran** terdiri dari tiga kategori, yaitu biasa, dengan jaminan, dan lainnya;
2. **Pelabuhan muat** terdiri dari empat kategori, yaitu NAN (Nansha), NING (Ningbo), SGP (Singapura), dan OTH (lainnya);
3. **Pasok** adalah negara pemasok yang terdiri dari empat kategori, yaitu CHN (China), JPN (Jepang), SGP (Singapura), dan OTH (lainnya);
4. **Kirim** adalah negara pengirim barang yang terdiri dari lima kategori, yaitu CHN (China), JPN (Jepang), MLY (Malaysia), SGP (Singapura), dan OTH (Lainnya);
5. **HS** merupakan *Harmonized System Code* atau kode HS. HS dibagi menjadi lima kategori, yaitu 64,73, 94, 95, dan OTH. OTH merupakan HS dengan pos selain yang disebutkan diawal. Dalam penelitian ini, kode HS yang digunakan adalah menggunakan dua digit pertama yang menunjukkan bab;
6. **Asal** adalah negara asal barang.

Terdiri dari empat kategori, yaitu CHN (China), JPN (Jepang), TUR (Turki), dan OTH (lainnya);

7. **Tarif BM** adalah tarif bea masuk. Tarif bea masuk dibagi menjadi tiga kategori, yaitu *maximum* (MAX), *middle* (MID), dan *minimum* (MIN) dengan perhitungan yang disajikan pada gambar 4.3 berikut ini:

$$\frac{\text{Tarif BM terbesar} - \text{Tarif BM terkecil}}{3}$$

Sumber : diolah penulis dari Ananda & Fadhli (2018)

Gambar 4.3 Kategorisasi Tarif BM

Berdasarkan data importasi September sampai dengan Desember 2019.

Tarif BM terkecil adalah 0% dan yang terbesar adalah 450% sehingga menghasilkan kategori MAX apabila lebih dari 300%, MID lebih dari 150% sampai dengan 300%, dan MIN adalah 150% dan kurang dari itu;

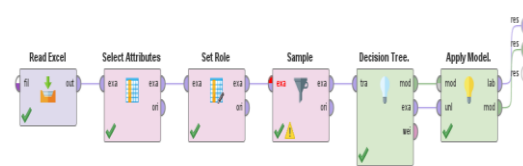
8. **Tarif PPh** adalah tarif PPh impor. Tarif PPh dibagi menjadi tiga kategori yaitu MAX, MID, dan MIN dengan perhitungan seperti gambar 4.3. Berdasarkan data importasi September sampai dengan Desember 2019. Tarif PPh terkecil adalah 1% dan yang terbesar adalah 10% sehingga menghasilkan kategori MAX apabila

lebih dari 7%, MID lebih dari 4% sampai dengan 7%, dan MIN adalah 1% hingga 4%;

9. **HIT (H) dan MISS (M)** H adalah atribut yang merupakan sebuah label yang berarti menunjukkan adanya temuan, sedangkan M adalah *miss* yang menunjukkan tidak adanya temuan.

4.4 Modelling

Pada tahap ini akan dilakukan *modelling decision tree*. Proses *modelling* menggunakan beberapa operator pada gambar 4.4.



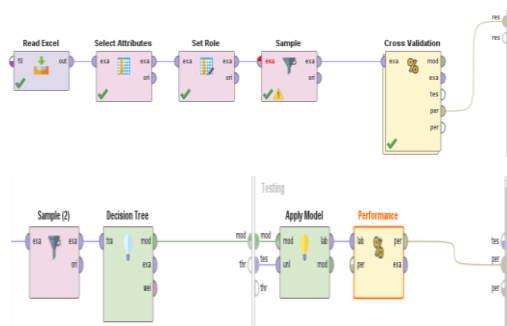
Sumber: Diolah dari RapidMiner

Gambar 4.4 Modelling Decision Tree

Terdapat *imbalance* pada set data. Untuk mengatasi *imbalance* data pada *machine learning*, berdasarkan penelitian yang dilakukan Johnson & Khoshgoftaar (2019), *data-level methods* dapat dilakukan pada *data training* dengan melakukan *under-sampling*. Hal serupa juga dijelaskan dalam penelitian Japkowicz (2000) bahwa *over-sampling* dan *down-sizing* adalah metode efektif untuk *imbalance* data. Japkowicz juga menyebutkan *down-sizing* bekerja lebih baik

dibandingkan *over-sampling*. Dalam penelitian ini dilakukan *under-sampling/down-sizing* label M pada atribut label HIT MISS.

Setelah model terbentuk, kemudian dilakukan penilaian atas model dengan melakukan *10-fold cross validation*. Proses validasi merupakan tes yang dilakukan kepada *data testing* dari *data training*. Hasil validasi model sebelum dan sesudah *balancing* data ditunjukkan pada tabel 4.1.



Sumber: Diolah dari RapidMiner

Gambar 4.5 Proses Validasi Model

Model yang dibentuk sebanyak 37 model dengan percobaan *unpruned* dan *pruned* dengan *depth* antara 5 – 10 dan *confidence* 0,1 – 0,5. Model terbaik yang dihasilkan adalah Model 29 *pruned* dengan *depth* 9 dan *confidence* 0,2. Model 29 memiliki *accuracy* 70,92%, *precision* 68,11%, *recall* 78,67%, AUC 0,778, dan kappa 0,418.

Tabel 4.1 *Confusion Matrix* Model 29
Sebelum dan Sesudah *Balancing* Data

sebelum *balancing*

	<i>True M</i>	<i>True H</i>	<i>Class Precision</i>
<i>Pred M</i>	288358	80034	78,27%
<i>Pred H</i>	29685	42687	58,98%
<i>Class Recall</i>	90,67%	34,78%	

setelah *balancing*

	<i>True M</i>	<i>True H</i>	<i>Class Precision</i>
<i>Pred M</i>	77519	26182	74,75%
<i>Pred H</i>	45202	96539	68,11%
<i>Class Recall</i>	63,17%	78,67%	

Sumber: Diolah dari RapidMiner

Tabel 4.1 merupakan perbandingan *confusion matrix* pada model 29 *decision tree pruned* dengan *depth* 9 dan *confidence* 0,2 yang belum dilakukan *balancing* (atas) dan sesudah dilakukan *balancing* (bawah). *Confusion matriks* yang berisi data aktual dan prediksi digunakan untuk menilai performa algoritma klasifikasi (Márquez-Vera et al., 2013).

4.5 Evaluation

Model 29 merupakan model terbaik yang dihasilkan dengan nilai *accuracy*

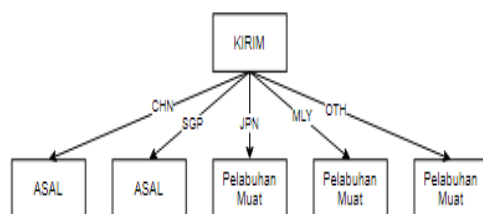
70,92%, *precision* 68,11%, *recall* 78,67%, AUC 0,778, dan kappa 0,418. Terkait dengan kategori nilai AUC dan kappa model 29, dijelaskan pada tabel 4.2.

Tabel 4.2 *Evaluation Model 29*

Nilai	Angka	Kategori
AUC	0,778	<i>Fair Classification</i>
Kappa	0,418	<i>Moderate Agreement</i>

Sumber : Diolah dari Gorunescu dalam Lestari & Silaban (2018) dan Viera & Garrett (2005)

Berdasarkan hasil *decision tree* model 29, negara pengirim (**Kirim**) merupakan atribut *root node* (akar). Negara pengirim merupakan atribut yang paling kuat untuk mendiferensiasi kelas dalam grup label H dan M. Berikut adalah potongan *decision tree* model 29 digambarkan pada gambar 4.6.



Sumber: Diolah dari RapidMiner

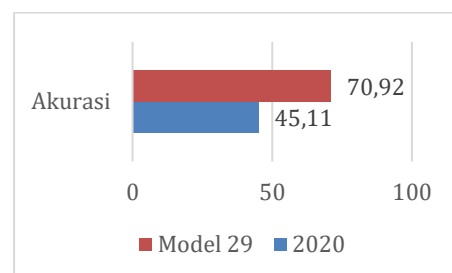
Gambar 4.6 Potongan *Decision Tree* Model 29

Nilai *information gain ratio* masing-masing atribut adalah **Kirim** 0,084, **Pasok** 0,084, **Asal** 0,082, **Pelabuhan** 0,082, **Cara Pembayaran** 0,082, **Tarif PPh** 0,003, **Tarif BM** 0,010, **HS** 0,012, dan **Cara pembayaran** 0,010. Tarif BM tidak dapat memberikan informasi tentang kelas H dan M.

muat 0.061, **Tarif PPh** 0,003, **Tarif BM** 0, **HS** 0,012, dan **Cara pembayaran** 0,010. Tarif BM tidak dapat memberikan informasi tentang kelas H dan M.

4.6 Deployment

Akurasi *risk engine targeting* KPU BC Tipe A Tanjung Priok adalah sebesar 45,11%, sedangkan akurasi *targeting* dengan model *decision tree* dalam penelitian ini adalah sebesar 70,92% seperti yang ditunjukkan pada gambar 4.7. Pada tahap *deployment*, dilakukan dengan menyelesaikan seluruh rangkaian proses *data mining* mulai dari *business understanding* hingga *deployment*.



Sumber: Diolah dari Direktorat P2 (2021)

Gambar 4.7 Akurasi *Risk Engine Targeting* KPU BC Tipe A Tanjung Priok Tahun 2020 dan Model 29

Model 29 kemudian digunakan untuk mengklasifikasi data importasi Januari 2020. Penggunaan model 29 dilakukan pada 60.000 baris data. Hasil

klasifikasi ditunjukkan pada tabel 4.3 menunjukkan model 29 dapat mengklasifikasikan data temuan pelanggaran 71,77% secara tepat.

Tabel 4.3 Penggunaan Model 29 Pada Data Importasi KPU BC Tipe A Tanjung Priok Bulan Januari 2020

Keterangan	Jumlah Data	%
Klasifikasi Tepat	43.062	71,77
Klasifikasi Tidak Tepat	16.938	28,38
Total	60.000	100

Sumber: Diolah dari RapidMiner

Penelitian ini menghasilkan model *decision tree* yang dapat menentukan atribut-atribut yang relevan dalam mengklasifikasi data importasi jalur merah KPU BC Tipe A Tanjung Priok.

5. SIMPULAN

Proses *data mining* dengan metodologi CRISP-DM, metode klasifikasi, dan model *decision tree* dalam *hit rate* importasi jalur merah pada KPU BC Tipe A Tanjung Priok dapat dilakukan menggunakan atribut-atribut pada PIB dan hasil penelitian dari instansi kepabeanan negara lain. Negara pengirim yang dikategorikan sebagai *root node* (akar) adalah atribut yang paling relevan memberikan informasi dalam mengklasifikasikan

kelas HIT dan MISS dengan nilai *information gain ratio* 0,084. Atribut lain yang relevan adalah negara pemasok 0,084, negara asal 0,082, pelabuhan muat 0,061, kode HS 0,012, cara pembayaran 0,010, dan tarif PPh 0,003. Tarif BM dengan nilai *information gain ratio* 0 tidak memberikan informasi mengenai kelas HIT dan MISS. Model terbaik yang dibentuk memiliki *accuracy*, *precision*, *recall*, AUC, dan *kappa* masing-masing sebesar 70,92%, 68,11%, 78,67%, 0,778, dan 0,418. Model tersebut memiliki akurasi yang lebih tinggi sebesar 25,81% dibandingkan dengan akurasi *risk engine argetting* KPU BC Tipe A Tanjung Priok tahun 2020 yaitu 45,11%.

6. SARAN

KPU BC Tipe A Tanjung Priok bersama dengan Direktorat P2 dapat mempertimbangkan indikator angka evaluasi tambahan selain *accuracy* untuk menilai model seperti *precision*, *recall*, AUC, dan *kappa* untuk menilai efektivitas *risk engine targeting*. Atribut negara pengirim dapat dijadikan informasi tambahan bagi KPU BC Tipe A Tanjung Priok sebagai usulan parameter kepada Direktorat P2 dalam *risk engine targeting* sesuai dengan

ketentuan dalam Instruksi Dirjen Bea dan Cukai nomor INS- 03/BC/2019.

DAFTAR PUSTAKA

- Al-Shbail, T. (2020). The impact of risk management on revenue protection: empirical evidence from Jordan customs. *Transforming Government: People, Process and Policy*. <https://doi.org/10.1108/TG-02-2020-0025>.
- Ananda, R., & Fadhli, M. (2018). *Statistik Pendidikan Teori dan Praktik Dalam Pendidikan* (S. Saleh (Peny.)). CV. Widya Puspita.
- Chang, C. C., & Chen, R. S. (2006). Using data mining technology to solve classification problems: A case study of campus digital library. *Electronic Library*, 24(3), 307–321. <https://doi.org/10.1108/02640470610671178>.
- Chermiti, B. (2019). Establishing risk and targeting profiles using data mining: Decision trees. *World Customs Journal*, 13(2), 39–58.
- Dlava, D. (2012). *Implementasi manajemen risiko dalam bidang impor skripsi*. Universitas Indonesia.
- Du, W., Du, W., Zhan, Z., & Zhan, Z. (2002). Building decision tree classifier on private data. *Proceedings of the IEEE International Conference on Privacy, Security and Data Mining*, 14, 1–8. <http://portal.acm.org/citation.cfm?id=850784>.
- Fauzia, M. (17 Desember 2019). Bea dan Cukai Ungkap Penyelundupan Puluhan Mobil dan Motor Mewah di Tanjung Priok. www.kompas.com. <https://money.kompas.com/read/2019/12/17/174039826/bea-dan-cukai-ungkap-penyelundupan-puluhan-mobil-dan-motor-mewah-di-tanjung>.
- Firdiansyah, A. (2019). Tinjauan Terhadap Identifikasi Risiko Penetapan Tarif Kepabeanan pada Kantor Pelayanan Utama Bea dan Cukai Tanjung Priok. *Jurnal Perspektif Bea Dan Cukai*, 3(1), 133–151.
- Firdiansyah, A., & Nugroho, A. S. (2017). Evaluasi Kebijakan Pemeriksaan Fisik Barang Pada Direktorat Jenderal Bea Dan Cukai. *Jurnal Perspektif Bea Dan Cukai*, 1(1). <https://doi.org/10.31092/jpbc.v1i1.121>.
- Herlinawati, Y., Hidayat, K., & Setyawan, A. (2016). Analisis Implementasi Pengawasan Ekspor Impor Barang pada KPPBC Tipe Madya Pabean Juanda. *Jurnal Perpajakan (JEJAK)*, 10(1), 1–6.
- Hoffman, A. J., Grater, S., Venter, W. C., Maree, J., & Liebenberg, D. (2018). An explorative study into the effectiveness of a customs operation and its impact on trade. *World Customs Journal*, 12(2), 63–86.
- Jaggia, S., Kelly, A., Lertwachara, K., & Chen, L. (2020). Applying the CRISP-DM Framework for Teaching Business Analytics. *Decision Sciences Journal of Innovative Education*, 18(4), 612–634.
- Japkowicz, N. (2000). Learning from imbalanced data sets: a comparison of various strategies. *AAAI Workshop on Learning from Imbalanced Data Sets*, 0–5.
- Jatmiko, B. P. (17 Desember 2019). Penyelundupan Mobil dan Motor

- Mewah Sepanjang 2019 Berpotensi Rugikan Negara Rp 647,5 Miliar. www.kompas.com.
<https://money.kompas.com/read/2019/12/17/204501026/penyelundupan-mobil-dan-motor-mewah-sepanjang-2019-berpotensi-rugikan-negara>.
- Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of financial economics*, 3(4), 305-360.
- Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1), 1-54.
<https://doi.org/10.1186/s40537-019-0192-5>.
- Jonathan, B. (2018). Artificial Intelligence and The Future of Customs. In I. Muscat (Ed.), *The EU Customs Union @ 50 Concept to Continuum* (pp. 122-125). Malta Customs.
- Karegowda, A. G., Manjunath, A. S., Ratio, G., & Evaluation, C. F. (2010). Comparative study of Attribute Selection Using Gain Ratio. *International Journal of Information Technology and Knowledge Management*, 2(2), 271-277.
[https://pdfs.semanticscholar.org/3555/1bc9ec8b6ee3c97c524f9c9ceee798c2026e.pdf%0Ahttp://csjournals.com/IJITKM/PDF 3-1/19.pdf](https://pdfs.semanticscholar.org/3555/1bc9ec8b6ee3c97c524f9c9ceee798c2026e.pdf%0Ahttp://csjournals.com/IJITKM/PDF%203-1/19.pdf).
- Karim, K. E., & Siegel, P. H. (1998). A signal detection theory approach to analyzing the efficiency and effectiveness of auditing to detect management fraud. *Managerial Auditing Journal*, 13(6), 367-375.
<https://doi.org/10.1108/02686909810222384>.
- Kasper, W., & Streit, M. E. (1999). Institutional economics. *Books*.
- Lestari, S., & Silaban, H. A. (2018). IMPLEMENTASI DATA MINING DALAM PENERBITAN SURAT PENETAPAN TARIF DAN NILAI PABEAN MENGGUNAKAN METODE CLASSIFICATION PADA DIREKTORAT JENDERAL BEA DAN CUKAI. *CKI ON SPOT*, 11(2).
- Lynn, S. K., & Barrett, L. F. (2014). "Utilizing" Signal Detection Theory. *Psychological Science*, 25(9), 1663-1673.
<https://doi.org/10.1177/0956797614541991>.
- Márquez-Vera, C., Cano, A., Romero, C., & Ventura, S. (2013). Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. *Applied intelligence*, 38(3), 315-330.
- Nurhidayati, & Cahyani, P. (2020). Pengaruh Kebijakan Penurunan Jalur Merah terhadap Penerimaan Negara The Impact of Reducing The Red Line Policy on Government Revenue Pendahuluan. *Jurnal Ekonomi Dan Pembangunan Indonesia*, 20(1), 79-93.
- Okazaki, Y. (2017). Implications of Big Data for Customs-How It Can Support Risk Management Capabilities. *WCO Research Paper*, (39).
- Pramudyo, J., & Arimbhi, P. (2018). Implementasi Kebijakan Penetapan Jalur Pengeluaran Barang Impor Pada Kantor Pelayanan Bea Cukai Tipe C Soekarno Hatta Tahun 2016. *Jurnal Ilmiah Untuk Mewujudkan Masyarakat Madani*, 5(1), 51-65.

- Priyasmo, M. R. (17 Desember 2019). *Bea Cukai Bongkar Penyelundupan Mobil dan Motor Mewah di Tanjung Priok*. liputan6.com. <https://www.liputan6.com/news/read/4136492/bea-cukai-bongkar-penyelundupan-mobil-dan-motor-mewah-di-tanjung-priok>.
- Semedi, B. (2013). Pengawasan kepabeanan. *Pusdiklat Bea Cukai*, 1–11.
- Sohrabi, B., Raeesi Vanani, I., Nikaein, N., & Kakavand, S. (2019). Predictive analytics of physician's prescription and pharmacies sales correlation using data mining. *International Journal of Pharmaceutical and Healthcare Marketing*, 13(3), 346–363. <https://doi.org/10.1108/IJPHM-11-2017-0066>.
- Spackman, K. A. (1989). Signal Detection Theory: Valuable Tools for Evaluating Inductive Learning. *Proceedings of the Sixth International Workshop on Machine Learning*, 160–163. <https://doi.org/10.1016/b978-1-55860-036-2.50047-3>.
- Syaifullah, S., & Ramdany, R. (2020). Mengukur Tingkat Kepatuhan Kepabeanan Perusahaan Ekspor Dan Import Di Indonesia. *Jurnal Akuntansi*, 9(1), 69–89. <https://doi.org/10.37932/ja.v9i1.89>.
- Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam med*, 37(5), 360–363.
- Vulandari, R. T. (2017). *Data Mining Teori dan Aplikasi Rapidminer* (1st ed.). Penerbit Gava Media.
- WCO. (2011). *WCO Customs Risk Management Compendium* (Vol. 1).
- White, A. P., & Liu, W. Z. (1994). Technical Note: Bias in Information-Based Measures in Decision Tree Induction. *Machine Learning*, 15(3), 321–329. <https://doi.org/10.1023/A:1022694010754>.
- Xia, B. S., & Gong, P. (2014). Review of business intelligence through data analysis. *Benchmarking*, 21(2), 300–311. <https://doi.org/10.1108/BIJ-08-2012-0050>.