

## Klasifikasi Profil Risiko Pengusaha Cukai: Studi Kasus Direktorat Jenderal Bea dan Cukai

**Budi Dwi Oktianto dan Indra Budi**  
Fakultas Ilmu Komputer Universitas Indonesia  
Kampus UI Depok  
e-mail: budi.dwi@ui.ac.id, indra@cs.ui.ac.id

### Abstrak

*Rencana Strategis Direktorat Jenderal Bea dan Cukai (DJBC) Tahun 2015-2019 menyebutkan bahwa penerapan teknologi informasi diharapkan dapat mendukung pengambilan keputusan dalam memberikan pelayanan di bidang kepabeanan dan cukai yang berbasis manajemen risiko. Penggolongan profil risiko pengusaha cukai pada penerapannya mengalami kendala ketika klasifikasi profil risiko pengusaha cukai tidak tersedia sehingga berdampak pada terganggunya pengambilan keputusan dalam pemberian fasilitas penundaan pembayaran dan dalam pelayanan permohonan penyediaan pita cukai. Sistem manajemen risiko di lingkungan administrasi kepabeanan dan cukai dapat diterapkan menggunakan metode klasifikasi pada teknologi data mining. Penelitian ini termasuk kategori klasifikasi multi kelas karena atribut output atau target terdiri dari tiga kelas yaitu risiko rendah (L), sedang (M), dan tinggi (H). Metode klasifikasi yang digunakan untuk membangun model yaitu multinomial logistic regression dan naïve bayes, kemudian dilakukan evaluasi terhadap model dengan teknik k-fold cross. Hasil pengujian memperlihatkan bahwa multinomial logistic regression memiliki akurasi yang lebih baik dibanding naïve bayes dalam mengklasifikasi profil risiko pengusaha cukai.*

**Kata kunci:** data mining, classification, multinomial logistic regression, naïve bayes

### 1. Pendahuluan

Berdasarkan Peraturan Menteri Keuangan Nomor 234/PMK.01/2015 tentang Organisasi dan Tata Kerja Kementerian Keuangan [1], Direktorat Jenderal Bea dan Cukai (DJBC) merupakan salah satu unit eselon I di Kementerian Keuangan yang bertugas menyelenggarakan perumusan dan pelaksanaan kebijakan di bidang pengawasan, penegakan hukum, pelayanan dan optimalisasi penerimaan negara di bidang kepabeanan dan cukai sesuai dengan ketentuan peraturan perundang-undangan. Fungsi utama DJBC selaras dengan yang diamanatkan dalam Keputusan Menteri Keuangan Nomor 36/KMK.1/2014 tentang Cetak Biru Transformasi Kelembagaan Kementerian Keuangan 2014-2025 [2], menyebutkan bahwa salah satu inisiatif yang perlu diperhatikan oleh DJBC yaitu dalam hal integrasi manajemen risiko dalam memberikan pelayanan dan melakukan pengawasan di bidang kepabeanan dan cukai. Implementasi manajemen risiko merupakan kunci utama bagi organisasi kepabeanan dan cukai dalam melakukan pengawasan dan pelayanan di bidang kepabeanan dan cukai [3].

DJBC mengimplementasi manajemen risiko untuk melakukan pengawasan dan pelayanan di bidang cukai mulai tahun 2011 melalui Instruksi Direktur Jenderal Bea dan Cukai Nomor INS-6/BC/2011 tentang Profil Pengusaha Pabrik dan Importir Hasil Tembakau [4]. Berdasarkan instruksi tersebut, penilaian tingkat risiko atau profil dibagi menjadi tiga komponen yaitu registrasi, operasional, dan lokal. Ketiga komponen tersebut digunakan oleh Direktorat Penindakan dan Penyidikan di Kantor Pusat untuk menentukan atau memutakhirkan profil risiko perusahaan cukai. Proses penilaian dan pemutakhiran ini dapat dilakukan sewaktu-waktu ataupun secara periodik. Penggolongan profil risiko pengusaha di bidang cukai dibagi menjadi tiga kelompok yaitu profil risiko tinggi, menengah, dan rendah. Sebelum penentuan atau pemutakhiran profil risiko, dilakukan penyaringan (*filter*) oleh Direktorat Penindakan dan Penyidikan mengenai status Nomor Pokok Pengusaha Barang Kena Cukai (NPPBKC), eksistensi perusahaan, kegiatan perusahaan di bidang cukai dalam 6 bulan terakhir, pencatatan atau pembukuan perusahaan, pelanggaran administrasi di bidang cukai, dan Pelanggaran pidana di bidang cukai. Penggolongan profil risiko pengusaha cukai hasil tembakau ini digunakan untuk menentukan tingkat perlakuan dalam pelaksanaan pemberian fasilitas penundaan pembayaran dan dalam pelayanan Permohonan Penyediaan Pita Cukai (P3C) tambahan.

Berdasarkan Peraturan Menteri Keuangan Nomor 169/PMK.04/2011 tentang Penundaan Pembayaran Cukai untuk Pengusaha Pabrik atau Importir Barang Kena Cukai [5], pada pasal 5 disebutkan bahwa untuk mendapatkan penundaan dengan jaminan perusahaan, pengusaha pabrik harus memenuhi persyaratan salah satunya yaitu merupakan pengusaha dengan profil risiko rendah. Bentuk perlakuan layanan sesuai profil risiko pengusaha yang lain yaitu berdasarkan Peraturan Direktur Jenderal Bea dan Cukai Nomor PER-08/BC/2014 tentang Penyediaan dan Pemesanan Pita Cukai [6], pada pasal 8 dan pasal 9 memuat perbedaan tingkat perlakuan layanan permohonan penyediaan pita cukai (P3C) berdasarkan profil risiko pengusaha cukai hasil tembakau.

Pelayanan cukai berbasis manajemen risiko ini pada penerapannya mengalami kendala apabila profil risiko pengusaha cukai tidak tersedia yang berdampak pada terganggunya pengambilan keputusan dalam pemberian fasilitas penundaan pembayaran dan layanan Permohonan Penyediaan Pita Cukai (P3C). Proses penilaian profil risiko pengusaha cukai tersebut juga dapat menyebabkan inkonsistensi atau ketidaksesuaian apabila pada saat pemutakhiran profil risiko menggunakan data lama, sedangkan data komponen penilaian terbaru sudah tersedia belum digunakan.

Rencana Strategis DJBC Tahun 2015-2019 [7] menyebutkan penerapan teknologi informasi diharapkan dapat mendukung pengambilan keputusan dalam memberikan pelayanan di bidang kepabeanan dan cukai yang berbasis manajemen risiko. Sistem manajemen risiko di lingkungan administrasi kepabeanan dan cukai dapat diterapkan menggunakan teknologi *data mining* dengan teknik-teknik statistik tertentu [8].

Laporte [8] membahas pemanfaatan *data mining* dalam memprediksi ada atau tidaknya pelanggaran terhadap pemberitahuan pabean. *Data mining* juga digunakan oleh Arundina, Omar, dan Kartiwi [9] dalam memprediksi peringkat Sukuk. Hoe dan Dhillon [10] menggunakan *data mining* untuk memprediksi kesanggupan konsumen membayar atas layanan yang diberikan, dan juga Cleary [11] memanfaatkannya untuk membantu dalam menganalisis pemilihan target audit di sektor publik.

Menurut Han, Kamber, dan Pei [12], *data mining* merupakan bagian dari proses KDD (*Knowledge Discovery in Databases*) yang terdiri dari beberapa tahapan seperti pemilihan data, pra pengolahan, transformasi, *data mining*, dan evaluasi hasil. *Data mining* adalah suatu konsep yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam *database*.

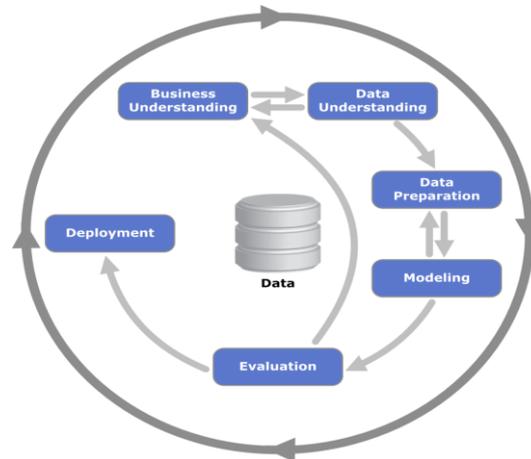
Secara umum, *data mining* dapat dikelompokkan menjadi dua kategori [12], yaitu *Descriptive* dan *Predictive mining*. *Descriptive mining* yaitu proses untuk menemukan karakteristik penting dari data dalam suatu basis data. Teknik atau metode yang dikategorikan dalam kategori ini adalah *clustering*, dan *association*. *Predictive*, yaitu proses untuk menemukan pola dari data dengan menggunakan beberapa variabel lain di masa depan. Metode yang termasuk dalam kategori ini adalah *classification*. Menurut Han, Kamber, dan Pei [12], metode klasifikasi adalah proses pencarian sebuah model (fungsi) yang menjelaskan dan membedakan kelas-kelas data atau konsep-konsep. Model diperoleh dari hasil analisis sekumpulan data *training*. Model tersebut digunakan untuk memprediksi objek yang kelasnya belum diketahui.

Penelitian ini bertujuan untuk mengklasifikasi profil risiko pengusaha di bidang cukai, sehingga dapat diketahui tingkat risiko suatu kegiatan usaha di bidang cukai. Klasifikasi profil risiko pengusaha cukai dengan *data mining* ini diharapkan diharapkan dapat meningkatkan efisiensi melalui otomasi proses penentuan profil risiko pengusaha cukai dengan memanfaatkan teknologi *data mining* berdasarkan *database* yang sudah ada di DJBC.

Penelitian ini dilakukan dengan pengelompokan tingkat profil risiko pengusaha cukai menjadi tiga kategori risiko yaitu tinggi, sedang, dan rendah yang merupakan klasifikasi multi kelas (lebih dari dua kelas). Oleh karena itu metode/algorithm klasifikasi yang digunakan yaitu *multinomial logistic regression* sesuai dengan yang dikemukakan oleh Ledolter [13] dan sesuai dengan penelitian sebelumnya yang dilakukan oleh Arundina, Omar, dan Kartiwi [9] dalam memprediksi peringkat Sukuk. Penelitian ini menggunakan Weka dalam menerapkan tahapan metodologi *data mining*.

## **2. Metode Penelitian**

Tahapan pada penelitian ini dilakukan sesuai dengan tahapan dalam metodologi CRISP-DM seperti terlihat pada Gambar 1.



Gambar 1. Tahapan proses CRISP [14]

Pemahaman proses bisnis mencakup pemahaman terhadap bagaimana proses penentuan tingkat atau profil risiko pengusaha cukai berdasarkan variabel penelitian. Tahap ini mengumpulkan informasi yang berkaitan dengan penelitian, tinjauan pustaka atas dokumen organisasi (Peraturan / Keputusan Menteri Keuangan, Peraturan/Instruksi Dirjen Bea dan Cukai), jurnal, dan penelitian terdahulu yang digunakan untuk membantu penelitian.

Pemahaman data mempelajari data-data yang berhubungan dengan proses penentuan tingkat atau profil risiko pengusaha cukai. Data yang diobservasi yaitu data profil risiko pengusaha cukai hasil tembakau periode tahun 2012 sampai 2015 pada basis data Sistem Aplikasi Cukai Sentralisasi (SACS) terkait fitur profiling pengusaha pabrik dan importir hasil tembakau.

Penyiapan data meliputi pemilihan variabel atau atribut yang disusun menjadi himpunan data untuk digunakan dalam pemodelan. Penelitian ini menggunakan data hasil observasi dari basis data aplikasi cukai di DJBC. Data yang diobservasi yaitu data profil risiko pengusaha hasil tembakau periode tahun 2012 sampai 2015. Data yang diobservasi yaitu sebanyak 9360 tingkat atau profil risiko pengusaha hasil tembakau. Variabel penelitian yang digunakan diperoleh dari variabel penilaian di Instruksi Direktur Jenderal Bea dan Cukai Nomor INS-6/BC/2011 tentang Profil Pengusaha Pabrik dan Importir Hasil Tembakau dalam Rangka Pelayanan dan Pengawasan di Bidang Cukai. Variabel atau atribut yang menjadi dataset pada penelitian ini terdiri dari enam variabel filter penilaian, tiga variabel penilaian, dan satu variabel kelas atau target seperti pada Tabel 1.

Tabel 1. Variabel Penelitian

No	Variabel	Keterangan	Type Data	Input/Output
1	F1	Variabel filter penilaian ke-1 (status NPPBKC)	Biner	Input
2	F2	Variabel filter penilaian ke-2 (eksistensi)	Biner	Input
3	F3	Variabel filter penilaian ke-3 (kegiatan cukai)	Biner	Input
4	F4	Variabel filter penilaian ke-4 (pembukuan)	Biner	Input
5	F5	Variabel filter penilaian ke-5 (pelanggaran adm.)	Biner	Input
6	F6	Variabel filter penilaian ke-6 (pelanggaran pidana)	Biner	Input
7	NR	Nilai variabel registrasi	Numerik	Input
8	NO	Nilai variabel operasional	Numerik	Input
9	NL	Nilai variabel lokal	Numerik	Input
10	RISK	Tingkat/profil risiko pengusaha cukai, berupa kategori risiko rendah (L), sedang (M), dan tinggi (H)	Nominal	Output

Atribut atau variabel F1, F2, F3, F4, F5, dan F6 merupakan variabel penanda sebagai filter dalam penilaian profil risiko pengusaha cukai. F1 merupakan atribut untuk mengetahui status Nomor Pokok Pengusaha Barang Kena Cukai (NPPBKC) apakah aktif atau tidak. Atribut F2 untuk mengetahui status eksistensi pengusaha. Atribut F3 untuk mengetahui apakah pengusaha melakukan kegiatan cukai atau tidak. Atribut F4 untuk mengetahui apakah ada mekanisme pembukuan atau tidak. Atribut F5 untuk mengetahui apakah pernah melakukan pelanggaran administrasi atau tidak. Atribut F6 untuk mengetahui apakah pernah

melakukan pelanggaran pidana atau tidak. NR, NO, dan NL merupakan indikator penilaian yang juga digunakan untuk menentukan profil risiko pengusaha cukai. Atribut NR merupakan hasil penilaian data registrasi perusahaan. Atribut NO merupakan hasil penilaian data operasional cukai yang meliputi aspek kewajaran transaksional dan pelanggaran di bidang cukai. Atribut NL merupakan hasil penilaian lokal yang berasal dari penelitian lapangan perusahaan. RISK adalah variabel output (target) yang merupakan variabel dependen (kelas) yaitu tingkat/profil risiko pengusaha cukai, berupa kategori risiko rendah (L), sedang (M), dan tinggi (H) yang akan diprediksi pada penelitian ini melalui metode klasifikasi menggunakan teknik data mining.

Penyiapan atau *preprocessing* data dilakukan dengan menggunakan aplikasi Weka. Dataset di Weka dipersiapkan dalam format .arff sebelum diproses lebih lanjut. Atribut F1, F2, F3, F4, F5, dan F6 dikonversi kedalam format yang sesuai pada Weka. Penghilangan outlier di Weka menggunakan filter RemoveMisclassified. Filter ini mereduksi instance atau objek data dari 9360 menjadi 9210. Data kelas yang tersebar secara tidak merata (imbalanced) dapat mempengaruhi hasil atau kinerja algoritma yang digunakan dalam membangun model. Sebagai perbandingan untuk bahan evaluasi, disiapkan dataset tambahan yaitu dataset yang datanya ditambah (oversampling) supaya lebih seimbang (balanced) berdasarkan variabel kelas (RISK). Balancing data juga menggunakan aplikasi Weka dengan metode Synthetic Minority Oversampling TEchnique (SMOTE). Hasil proses balancing data dapat terlihat pada Tabel 2.

Tabel 2. Dataset yang digunakan

Dataset	Klasifikasi Risiko			Jumlah
	Rendah (L)	Sedang (M)	Tinggi (H)	
A (original)	666	4839	3705	9210
B (balanced, oversampling SMOTE)	3330	4839	3705	11874

Tahap pemodelan memproses data sehingga terbentuk model, menggunakan teknik atau metode data mining. Metode data mining yang digunakan untuk mengklasifikasi profil risiko pengusaha cukai adalah klasifikasi regresi logistik multinomial (multinomial logistic regression) dan naïve bayes sebagai pembandingan.

Evaluasi terhadap model dilakukan dengan teknik k-fold cross validation terhadap dataset yang sudah disiapkan. Teknik ini memvalidasi model untuk menguji hasil secara independen terhadap dataset. Pada tahap evaluasi model dilakukan pengukuran akurasi untuk mengetahui kinerja model klasifikasi yang dihasilkan. Tahapan deployment dalam CRISP-DM yaitu bagaimana model yang dihasilkan diaplikasikan, tidak dibahas dalam penelitian ini.

### 3. Hasil dan Pembahasan

Hasil uji coba *cross validation* dengan menggunakan metode *multinomial logistic regression* dan *naïve bayes* dapat dilihat pada Tabel 3. Berdasarkan Tabel 3 terlihat bahwa akurasi yang tertinggi diperoleh sebesar 89,62% pada *dataset B* (hasil *oversampling-SMOTE*) dengan *multinomial logistic regression*. Pada *dataset A* akurasi terbaik diperoleh sebesar 88,62% yang juga menggunakan *multinomial logistic regression*. Model klasifikasi *naïve bayes* berdasarkan hasil uji pada *dataset A* memperoleh akurasi sebesar 88,39% dan pada *dataset B* sebesar 89,44%.

Hasil uji coba yang diperlihatkan pada Tabel 3 memperlihatkan bahwa secara umum metode Multinomial Logistic Regression menghasilkan akurasi yang lebih baik dibandingkan dengan Naïve Bayes baik pada dataset yang balanced ataupun yang tidak balanced. Sedangkan jika dilihat mengenai keseimbangan data, terlihat bahwa data yang seimbang (balanced) menghasilkan akurasi yang lebih baik dibandingkan dengan data yang tidak balanced. Disamping itu juga dapat terlihat bahwa perbedaan nilai *fold (k)* pada *cross validation* terhadap model menunjukkan hasil yang hampir sama atau tidak ada perbedaan yang signifikan terhadap pengukuran akurasi model yang dihasilkan pada kedua *dataset*.

Tabel 3. Akurasi hasil uji coba

Cross Validation	Multinomial Logistic Regression		Naïve Bayes	
	Dataset A	Dataset B	Dataset A	Dataset B
6-fold	88,62%	<b>89,62%</b>	88,38%	89,44%
7-fold	88,61%	<b>89,62%</b>	88,38%	89,44%
8-fold	88,60%	<b>89,62%</b>	88,38%	89,44%
9-fold	88,61%	<b>89,62%</b>	88,39%	89,44%
10-fold	88,62%	89,62%	88,38%	89,44%

Berdasarkan pengamatan terhadap *dataset* diperoleh pengaruh variabel F1, F2, F3, F4, F5, dan F6 terhadap RISK yaitu jika variabel bernilai 0 maka kelas risikonya adalah risiko tinggi (H), sedangkan jika bernilai 1 maka kelas risikonya dapat bervariasi. pengaruh variabel NR, NO, dan NL terhadap RISK yaitu semakin tinggi nilai variabel maka kelas risikonya cenderung rendah (L), sedangkan kelas risiko sedang (M) dan tinggi (H) memiliki nilai yang cenderung seimbang tersebar secara merata.

Kinerja model yang dihasilkan pada penelitian ini memiliki akurasi yang tergolong baik yaitu 89,62%, namun untuk implementasi aplikasi lebih lanjut di DJBC diperlukan pengujian menggunakan data-data (*instance*) profil risiko baru yang belum pernah digunakan dalam proses *training* untuk menguji kehandalan model. Model klasifikasi yang dihasilkan dapat diaplikasikan lebih lanjut sehingga dapat digunakan oleh aplikasi yang dimiliki organisasi (DJBC). Model yang dihasilkan dapat digunakan oleh Aplikasi Cukai yang berbasis Java dengan bantuan API (*application programming interface*) atau *library* yang disediakan oleh Weka untuk melakukan klasifikasi terhadap *instance* data profil risiko pengusaha cukai yang baru.

#### 4. Simpulan

Hasil pengujian menunjukkan bahwa model menggunakan algoritma *multinomial logistic regression* dan *naïve bayes* terhadap dua *dataset* yang disiapkan yaitu *dataset* dengan *oversampling* (SMOTE) dan tanpa *oversampling* menunjukkan bahwa pemodelan dengan *dataset* hasil *balancing* dengan *oversampling* memiliki akurasi yang sedikit lebih baik dibanding *dataset* awal tanpa *oversampling*. Hasil pengujian dengan *cross validation* juga memperlihatkan bahwa algoritma klasifikasi *multinomial logistic regression* memiliki akurasi yang lebih baik dibanding *naïve bayes* dalam mengklasifikasi profil risiko pengusaha cukai. *Multinomial logistic regression* memperoleh akurasi terbaik sebesar 89,62%, sedangkan *naïve bayes* memperoleh akurasi terbaik sebesar 89,44%.

Algoritma klasifikasi yang digunakan untuk pemodelan pada penelitian ini hanya menggunakan *multinomial logistic regression* dan *naïve bayes*. Untuk penelitian selanjutnya algoritma klasifikasi yang lain yang mendukung multi kelas dapat digunakan sebagai pembandingan. Karena keterbatasan jumlah data profil risiko pengusaha cukai pada saat ini, diharapkan dengan data yang lebih lengkap di masa mendatang dapat digunakan untuk melakukan *training* kembali terhadap model klasifikasi untuk meningkatkan kinerja model.

#### Daftar Pustaka

- [1] Kementerian Keuangan. (2015). Peraturan menteri keuangan nomor 234/PMK.01/2015 tentang organisasi dan tata kerja kementerian keuangan. Jakarta: Kementerian Keuangan.
- [2] Kementerian Keuangan. (2014). Keputusan menteri keuangan nomor 36/KMK.1/2014 tentang cetak biru transformasi kelembagaan kementerian keuangan 2014-2025. Jakarta: Kementerian Keuangan.
- [3] World Customs Organization. (2011). WCO customs risk management compendium. Brussels: World Customs Organization.
- [4] Direktorat Jenderal Bea dan Cukai. (2011). Instruksi direktur jenderal bea dan cukai nomor INS-6/BC/2011 tentang profil pengusaha pabrik dan importir hasil tembakau dalam rangka pelayanan dan pengawasan di bidang cukai. Jakarta: Direktorat Jenderal Bea dan Cukai.
- [5] Kementerian Keuangan. (2011). Peraturan menteri keuangan nomor 169/PMK.04/2011 tentang penundaan pembayaran cukai untuk pengusaha pabrik atau importir barang kena cukai. Jakarta: Kementerian Keuangan.
- [6] Direktorat Jenderal Bea dan Cukai. (2014). Peraturan direktur jenderal bea dan cukai nomor PER-08/BC/2014 tentang penyediaan dan pemesanan pita cukai. Jakarta: Direktorat Jenderal Bea dan Cukai.
- [7] Direktorat Jenderal Bea dan Cukai. (2015). Rencana strategis direktur jenderal bea dan cukai 2015-2019. Jakarta: Direktorat Jenderal Bea dan Cukai.

- [8] Laporte, B. (2011). Risk management systems: using data mining in developing countries customs administrations. *World Customs Journal- Volume 5 No 1 - 2011*.
- [9] Arundina, T., Omar, M., & Kartiwi, M. (2015). The predictive accuracy of sukuk ratings. *Pacific-Basin Finance Journal 34* , 273-292.
- [10]Hoe, A. C., & Dhillon, J. S. (2014). Predicting bad utility consumers in malaysia. *International Conference on IT and Multimedia – ICIMU* , 234-237.
- [11] Cleary, D. (2011). Predictive analytics in the public sector: using data mining to assist better target selection for audit. *Electronic Journal of e-Government Volume 9 Issue 2* , 132-140.
- [12]Han, J., Kamber, M., & Pei, J. (2012). *Data mining concepts and techniques third edition*. San Fransisco: Morgan Kaufmann.
- [13]Ledolter, J. (2013). *Data mining and business analytics with R*. New Jersey: Wiley.
- [14]Azevedo, A., & Santos, M. F. (2008). KDD, SEMMA and CRISP-DM: a parallel overview. In *Proceedings of the IADIS European Conference on Data Mining* , 182-185.