Jurnal Matematika UNAND
Vol. **15** No. **4** Hal. 495 – 505
Edisi Oktober 2026
ISSN : 2303–291X
e-ISSN : 2721–9410
©Departemen Matematika dan Sains Data FMIPA UNAND
https://doi.org/10.25077/jmua.15.4.495-505.2026

EVALUATION AND DEVELOPMENT OF MACHINE-LEARNING-BASED COMPLIANCE SCORING MODELS FOR ALCOHOLIC BEVERAGES EXCISE TAXPAYER IN INDONESIA WITH INNOVATIVE ENCODING STRATEGY (EVALUASI DAN PENGEMBANGAN MODEL PENILAIAN KEPATUHAN BERBASIS PEMBELAJARAN MESIN UNTUK WAJIB PAJAK CUKAI MINUMAN BERALKOHOL DI INDONESIA DENGAN STRATEGI PENGKODEAN INOVATIF)

BOY RIANSYAH*, BAGUS SARTONO, FARIT M AFENDI

Sekolah Sains Data, Matematika dan Informatika, IPB University
Kampus IPB Dramaga, Bogor 16680
email : boyriansyahboy@apps.ipb.ac.id, bagusco@apps.ipb.ac.id, fmafendi@apps.ipb.ac.id

Diajukan 24 November 2025 Direvisi 7 Juli 2026
Diterima 4 Agustus 2026 Diterbitkan 22 Oktober 2026

Abstrak. Transformasi variabel kategorik sebagai bagian dari proses *feature engineering* merupakan tahapan penting dalam pengembangan model klasifikasi karena menentukan kualitas representasi informasi yang digunakan selama proses pemodelan. Kondisi tersebut dijumpai pada pengembangan model skor kepatuhan pengusaha barang kena cukai Minuman Mengandung Etil Alkohol (MMEA) di Indonesia yang seluruh variabel prediktornya berbentuk kategorik. Model skor kepatuhan yang saat ini digunakan oleh Direktorat Jenderal Bea dan Cukai (DJBC) masih berbasis *scorecard* manual dengan penetapan struktur kategori variabel prediktor yang bersifat subjektif. Model tersebut juga belum memiliki validasi empiris dalam merepresentasikan peluang pelanggaran yang dilakukan oleh pengusaha. Penelitian ini bertujuan untuk mengevaluasi model skor kepatuhan cukai yang berlaku sekaligus mengembangkan model alternatif berbasis data menggunakan pendekatan regresi logistik dan algoritma *Categorical Boosting* (CatBoost) dengan strategi *encoding* yang inovatif. Dua metode *encoding* variabel kategorik yang dibandingkan adalah *Weight of Evidence* (WoE) sebagai representasi metode statistika klasik dan *Ordered Target* (OT) yang diadaptasi dari mekanisme CatBoost. Seluruh model divalidasi menggunakan *repeated k-fold cross-validation* dan dievaluasi melalui metrik *Area Under Curve* (AUC), Kolmogorov–Smirnov (KS) serta beberapa metrik evaluasi lainnya. Hasil penelitian menunjukkan bahwa penerapan *encoding* OT menghasilkan performa lebih baik dibandingkan *encoding* WoE pada seluruh skenario model regresi logistik dengan peningkatan KS dari 0,27 menjadi 0,34 untuk WoE dan

*Penulis korespondensi

0,42 untuk OT. Model CatBoost memberikan kinerja yang lebih tinggi dibandingkan dengan *encoding* regresi logistik (AUC = 0,715; KS = 0,471) yang menandakan kemampuan diskriminasi yang superior dalam mendeteksi pelanggaran kepatuhan. Temuan ini membuktikan bahwa integrasi *encoding* berbasis target serta algoritma *boosting* dapat menghasilkan model skor yang lebih adaptif dan akurat untuk mendukung sistem profil risiko cukai di Indonesia.

Kata Kunci: Catboost, Ordered Target Encoding, Weight of Evidence

Abstract. *The transformation of categorical variables as part of the feature engineering process is a critical stage in classification modeling because it determines the quality of the information representation used throughout model development. This issue is particularly relevant to the development of a compliance scoring model for excise taxpayers producing or importing alcoholic beverages in Indonesia, where all predictor variables are categorical. The compliance scoring model currently implemented by the Directorate General of Customs and Excise (DGCE) is based on a manually constructed scorecard in which the categorization of predictor variables is determined subjectively. Moreover, the existing model has not been empirically validated to represent the probability of compliance violations. This study aims to evaluate the current excise compliance scoring model while developing data-driven alternatives using logistic regression and the Categorical Boosting (CatBoost) algorithm with innovative categorical encoding strategies. Two encoding methods are compared: Weight of Evidence (WoE), representing a conventional statistical approach, and Ordered Target (OT), which is adapted from the CatBoost encoding mechanism. All models are validated using repeated k-fold cross-validation and evaluated based on the Area Under the Receiver Operating Characteristic Curve (AUC), the Kolmogorov–Smirnov (KS) statistic, and several additional performance metrics. The results demonstrate that OT encoding consistently outperforms WoE across all logistic regression scenarios, increasing the KS statistic from 0.27 for the existing compliance scoring model to 0.34 with WoE and 0.42 with OT. Furthermore, the CatBoost model achieves superior predictive performance compared with both logistic regression models (AUC = 0.715; KS = 0.471), indicating stronger discriminative ability in identifying compliance violations. These findings demonstrate that integrating target-based encoding with boosting algorithms produces a more adaptive and accurate compliance scoring model, providing a robust analytical framework to support risk-based excise supervision in Indonesia.*

Keywords: Catboost, Ordered Target Encoding, Weight of Evidence

1. Pendahuluan

Cukai merupakan instrumen fiskal yang tidak hanya berfungsi sebagai sumber penerimaan negara, melainkan juga sebagai instrumen pengendalian konsumsi barang yang menimbulkan eksternalitas negatif [1]. Salah satu komoditas yang dikenakan cukai adalah Minuman Mengandung Etil Alkohol (MMEA). Efektivitas pengawasan terhadap pengusaha MMEA sangat bergantung pada kemampuan pemerintah dalam mengidentifikasi tingkat kepatuhan pengusaha secara objektif sehingga sumber daya pengawasan dapat dialokasikan secara lebih tepat.

Saat ini Direktorat Jenderal Bea dan Cukai (DJBC) mengimplementasikan suatu model penilaian skor kepatuhan melalui instrumen Instruksi Direktur Jenderal Bea dan Cukai Nomor INS-3/BC/2024. Model ini berfungsi untuk mengukur tingkat kepatuhan pengusaha MMEA terhadap ketentuan cukai. Kerangka penilaian yang berlaku dalam model ini masih menggunakan pendekatan *scorecard* manual yang berbentuk fungsi linier dengan penetapan bobot relatif serta urutan kategori dari variabel penjelas didasarkan pada pertimbangan praktisi dan

pengalaman pengawasan di lapangan. Hal ini menunjukkan bahwa skor kepatuhan yang dihasilkan belum memiliki validitas empiris terhadap peluang terjadinya pelanggaran.

Pengembangan model kepatuhan cukai berbasis data untuk pengusaha MMEA menghadapi tantangan metodologis akibat ukuran sampel yang relatif terbatas dibandingkan dengan jumlah variabel prediktor. Populasi pengusaha MMEA yang terdaftar secara legal dan aktif di Indonesia hanya terdiri atas 109 pengusaha sedangkan model kepatuhan menggunakan 91 variabel prediktor yang seluruhnya berbentuk kategorik. Karakteristik tersebut menyebabkan model berada pada kondisi yang rentan dengan risiko *overfitting* dan penurunan kemampuan generalisasi model [3]. Kondisi ini menuntut strategi *feature engineering*, seleksi variabel serta algoritma pembelajaran mesin yang mampu menangani variabel kategorik secara efisien dan tidak hanya meningkatan akurasi prediksi tetapi juga menjaga kemampuan generalisasi model terhadap data baru. Sebagai respons atas tantangan tersebut, penelitian ini mengintegrasikan beberapa mekanisme pengendalian kompleksitas model seperti validasi silang berulang, seleksi variabel serta metode pembelajaran mesin yang dirancang untuk mengurangi risiko *overfitting* pada data berdimensi tinggi dengan ukuran sampel terbatas.

Pendekatan yang telah banyak digunakan dalam pengembangan *scorecard* berbasis data adalah regresi logistik dengan *encoding* WoE. *Encoding* ini mengubah setiap kategori menjadi representasi numerik berdasarkan hubungan empirisnya dengan variabel respons sehingga dapat meningkatkan keterbacaan model sekaligus mempertahankan interpretabilitas koefisien regresi logistik. WoE juga memiliki keterkaitan dengan Information Value (InfVal) yang memungkinkan proses seleksi variabel dilakukan secara sistematis sehingga pendekatan ini banyak diterapkan pada pengembangan model klasifikasi berbasis data kategorik [4].

Perkembangan metode pembelajaran mesin menghadirkan algoritma *Categorical Boosting* (CatBoost) yang dirancang secara khusus untuk menangani dominasi data kategorik. Berbeda dengan pendekatan regresi logistik yang memerlukan proses *encoding* secara terpisah, CatBoost mengintegrasikan mekanisme *encoding* OT, *ordered boosting* serta regularisasi dan *shrinkage* selama proses pembelajaran untuk mengurangi risiko *overfitting* dan meningkatkan kemampuan generalisasi model. Karakteristik tersebut menjadikan CatBoost sebagai salah satu pendekatan yang menjanjikan untuk diterapkan pada dataset berdimensi tinggi dengan jumlah observasi yang relatif terbatas [5].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk melakukan evaluasi model skor kepatuhan yang saat ini berlaku serta melakukan pengembangan model alternatif dengan mengkomparasi efektivitas kinerja antara regresi logistik berbasis *encoding* variabel prediktor dan algoritma CatBoost. Kontribusi kebaruan penelitian ini terletak pada penerapan *encoding* OT yang secara inheren terdapat pada algoritma CatBoost sebagai strategi *feature engineering* ke dalam kerangka regresi logistik. Temuan dari penelitian ini diharapkan dapat memberikan kontribusi nyata bagi upaya reformasi sistem *profiling* risiko di DJBC sekaligus menawarkan kerangka metodologis yang relevan bagi pengembangan model klasifikasi yang andal pada kondisi data terbatas dan kompleks.

2. Landasan Teori

2.1. *Regresi Logistik dengan Encoding WoE*

Regresi logistik merupakan metode klasifikasi untuk peubah respons biner yang banyak digunakan dalam pengembangan model *scorecard* karena menghasilkan probabilitas kejadian yang mudah diinterpretasikan. Probabilitas kejadian untuk setiap observasi dengan vektor prediktor x_i dinyatakan sebagai berikut.

$$\Pr(y_i = 1 \mid x_i) = \frac{1}{1 + \exp(-\eta(x_i; \beta))}, \tag{2.1}$$

dengan $\eta(x_i; \beta) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$ adalah *index function* atau kombinasi linear dari variabel prediktor. Model ini memiliki keunggulan utama berupa interpretabilitas yang tinggi karena setiap koefisien β_{j} dapat ditafsirkan sebagai efek marginal terhadap log odds dari kejadian respons [5].

WoE merupakan teknik *encoding* variabel yang berasal dari teori informasi dan banyak digunakan dalam pengembangan model *scorecard* berbasis regresi logistik. WoE mengukur kekuatan hubungan antara kategori dalam variabel prediktor dan status dari variabel respon biner. WoE untuk suatu bin kategori ke-i secara matematis didefinisikan menggunakan persamaan berikut:

$$\mathrm{WoE}_i = \ln\left(\frac{\text{Proporsi good (non-event)}_i}{\text{Proporsi bad (event)}_i}\right). \tag{2.2}$$

Proses transformasi variabel kategorik menjadi bentuk WoE menghasilkan nilai numerik pada setiap kategori yang secara langsung mempresentasikan *log odds* terhadap kejadian target. Hal ini memudahkan analis atau pembuat kebijakan untuk memahami arah dan besarnya pengaruh suatu variabel prediktor terhadap probabilitas respons [6].

WoE juga digunakan untuk menghitung InfVal sebagai ukuran kekuatan prediksi suatu variabel prediktor terhadap respons dan dasar dalam proses seleksi variabel melalui persamaan berikut:

$$IV = \sum_{i=1}^{k} (\text{Proporsi Non-Event}_i - \text{Proporsi Event}_i) \times \mathrm{WoE}_i, \tag{2.3}$$

dengan k adalah jumlah kategori dalam suatu peubah prediktor. Kombinasi WoE dan IV memberikan hasil superior dalam pemilihan variabel prediktor pada dataset berdimensi tinggi karena dapat mengeliminasi variabel prediktor yang bersifat tidak informatif sejak tahap awal pemodelan [7].

2.2. *Catboost dan Transformasi OT*

Dalam beberapa tahun terakhir, pendekatan *gradient boosting* menjadi metode utama dalam berbagai aplikasi *machine learning*, khususnya pada kondisi yang didominasi data kategorik. Salah satu inovasi terbaru dari pendekatan ini adalah algoritma Catboost yang secara spesifik dirancang untuk menangani data kategorik secara efisien tanpa memerlukan transformasi eksplisit seperti *one-hot encoding* maupun *label encoding*. Struktur pohon keputusan simetris yang digunakan

CatBoost membuat model dapat menangkap pola dan hubungan non linear antar variabel prediktor [8].

CatBoost dirancang untuk mengurangi risiko *overfitting* dan meningkatkan kemampuan generalisasi model, khususnya ketika diterapkan pada dataset berdimensi tinggi dengan jumlah observasi terbatas [9]. Kemampuan tersebut didukung oleh beberapa mekanisme bawaan CatBoost seperti *ordered boosting* untuk mengurangi *prediction shift*, *shrinkage* melalui *learning rate* yang mengendalikan besarnya pembaruan model pada setiap iterasi serta regularisasi yang membatasi kompleksitas pohon keputusan sehingga model menjadi lebih stabil terhadap variasi data. Penanganan variabel kategorik pada CatBoost dilakukan secara otomatis melalui teknik OT *encoding* dengan menghitung rata-rata label berdasarkan urutan acak data pelatihan dan menerapkan *prior smoothing* untuk mengurangi pengaruh kategori dengan frekuensi rendah sekaligus mencegah target leakage [10]. Proses *encoding* tersebut dapat diformulasikan sebagai berikut.

$$\mathrm{Enc}(x_{\sigma_i}) = \frac{\sum_{j=1}^{i-1} \mathbf{1}\left(x_{\sigma_{j,k}} = x_{\sigma_{p,k}}\right) \cdot y_{\sigma_j} \; + \; a \cdot P}{\sum_{j=1}^{i-1} \mathbf{1}\left(x_{\sigma_{j,k}} = x_{\sigma_{p,k}}\right) \; + \; a}, \tag{2.4}$$

dengan:

a : Parameter *Smoothing*,
p : Observasi yang sedang ditransformasi,
P : Nilai prior global, biasanya μ,
$x_{\sigma_{j,k}}$: Nilai fitur ke-k pada observasi ke-j dalam permutasi σ,
$x_{\sigma_{p,k}}$: Nilai fitur ke-k pada observasi ke-p yang sedang ditransformasi,
y_{σ_j} : Target dari observasi ke-j dalam permutasi σ,
$\left[x_{\sigma_{j,k}} = x_{\sigma_{p,k}}\right]$: Indikator, 1 jika bernilai sama, 0 jika tidak.

Melalui pendekatan ini, proses *encoding* bersifat *data-driven*, kontekstual terhadap urutan data dan tidak memerlukan *preprocessing* tambahan sehingga cocok untuk diaplikasikan pada dataset yang memiliki informasi kategori sangat dominan seperti pada data pengusaha cukai MMEA. Dari sisi optimasi, CatBoost membangun model prediksi $Ft(x)$ secara iteratif melalui fungsi boosting sebagai berikut.

$$F_t(x) = F_{t-1}(x) + \eta \cdot h_t(x), \tag{2.5}$$

dimana η sebagai *learning rate* dan $h_t(x)$ sebagai pohon keputusan berdasarkan residual atau gradien dari fungsi kerugian logistik (*log loss*).

3. Hasil dan Pembahasan

3.1. *Data*

Penelitian ini menggunakan data profil risiko pengusaha cukai MMEA tahun 2023, data Surat Tagihan Cukai (STCK) tahun 2024, serta data penindakan pelanggaran tahun 2024 yang bersumber dari sistem pengawasan internal DJBC. Data STCK dan penindakan pelanggaran digunakan untuk mengkonstruksi variabel respons biner (melanggar dan tidak melanggar) sedangkan data profil risiko sebagaimana

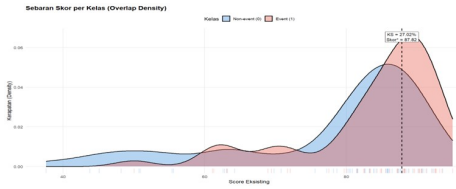
diatur dalam INS-3/BC/2024 digunakan sebagai variabel prediktor yang terdiri atas 91 variabel kategorik. Penelitian ini melibatkan 109 pengusaha cukai MMEA yang terdiri atas 80 pabrik dan 29 importir. Berdasarkan hasil konstruksi variabel respons sebanyak 52 pengusaha (47,7%) termasuk kategori melanggar dan 57 pengusaha (52,3%) termasuk kategori tidak melanggar sehingga distribusi kelas pada penelitian ini relatif seimbang dan tidak memerlukan penanganan khusus terhadap ketidakseimbangan kelas.

3.2. *Evaluasi Model Skor Kepatuhan yang Berlaku*

Evaluasi model skor kepatuhan yang saat ini berlaku dilakukan dengan menggunakan statistik Kolmogorov-Smirnov (KS) untuk mengukur kemampuan model dalam membedakan antara kelompok pengusaha yang melanggar dan tidak melanggar. Nilai KS dihitung berdasarkan selisih maksimum antara fungsi distribusi kumulatif dari dua kelompok yang dapat diformulasikan sebagai berikut.

$$KS = \max_{x} \left| F_1(x) - F_0(x) \right|, \tag{3.1}$$

dengan $F_1(x)$ merupakan fungsi distribusi kumulatif dari kelompok pengusaha yang melakukan pelanggaran dan $F_0(x)$ merupakan fungsi distribusi kumulatif dari kelompok pengusaha yang tidak melakukan pelanggaran. Nilai KS berkisar antara 0 hingga 1 dengan nilai yang lebih besar menunjukkan kemampuan diskriminasi yang lebih baik.



Gambar 1. Sebaran Skor per Kelas dari Model Eksisting

Dari Gambar 1 terlihat skor kepatuhan pada kelompok pengusaha yang melakukan pelanggaran cenderung lebih tinggi dibandingkan kelompok pengusaha yang tidak melanggar dengan median di sekitar 86 – 87. Sebaran skor antara kelompok *event* (1) dan *non-event* (0) masih memperlihatkan area tumpang tindih yang cukup besar sehingga pemisahan antar kelompok belum optimal. Perbedaan distribusi mulai terlihat lebih jelas pada rentang skor tinggi sebagaimana ditunjukkan oleh titik maksimum statistik KS yang tercapai pada skor sekitar $87,82$ dengan nilai KS sebesar 27,02%.

Kondisi tersebut menunjukkan bahwa model yang saat ini berlaku tidak cukup baik dalam merepresentasikan hubungan antara skor dan perilaku kepatuhan yang sebenarnya. Dalam praktik *credit scoring*, ambang$KS \geq 40\%$ lazim digunakan sebagai indikator kemampuan diskriminasi yang baik [11]. Berdasarkan capaian tersebut, dapat disimpulkan bahwa terdapat kesempatan yang cukup luas untuk dilakukan

pengembangan terhadap model skor kepatuhan pengusaha cukai MMEA agar menghasilkan model yang lebih optimal.

3.3. *Validasi Silang*

Seluruh pemodelan dalam penelitian ini divalidasi menggunakan *repeated cross validation* sebanyak 20 kali untuk memperoleh estimasi performa yang lebih stabil pada dataset dengan jumlah observasi yang relatif terbatas. Penelitian ini berorientasi pada kemampuan prediksi sehingga evaluasi performa dilakukan secara eksklusif pada data pengujian yang tidak digunakan selama proses pelatihan meskipun rasio jumlah variabel prediktor terhadap ukuran sampel relatif tinggi. Pendekatan ini memastikan bahwa setiap observasi secara bergantian berperan sebagai data pelatihan dan data pengujian. Seluruh metrik evaluasi yang dilaporkan pada penelitian ini merupakan rata-rata hasil evaluasi 100 *testing fold* sehingga mencerminkan kemampuan model dalam melakukan generalisasi terhadap data yang tidak digunakan selama proses pelatihan. Prosedur ini juga dapat mengurangi ketergantungan hasil evaluasi terhadap satu kali pembagian data sehingga memberikan dasar yang lebih andal dalam membandingkan performa antar pendekatan pemodelan pada kondisi data berdimensi tinggi dengan ukuran sampel yang terbatas [12].

3.4. *Analisis Model Regresi Logistik dengan Encoding WoE*

Tahapan *pre-processing* sebelum proses pemodelan mencakup *cleansing data*, identifikasi *sparse categories*, *merging sparse categories* ke kategori yang berdekatan pada peubah prediktor yang sama serta penerapan *Laplace additive smoothing* pada kategori yang hanya memiliki satu kelas respons. Prosedur tersebut bertujuan menghasilkan nilai WoE yang terdefinisi dengan baik [13].

Penelitian ini membangun empat spesifikasi model regresi logistik dengan memanfaatkan prediktor hasil *encoding* WoE untuk dibandingkan yaitu: (a) model penuh dengan 91 variabel prediktor; (b) model seleksi InfVal yang hanya menggunakan variabel prediktor dengan nilai InfVal di atas 0,1 [14]; (c) model *forward selection* yang melakukan pemilihan maju dari himpunan 91 variabel prediktor; dan (d) model *forward selection* pasca InfVal dengan melakukan pemilihan maju yang dibatasi pada variabel prediktor yang telah disaring berdasarkan InfVal. Keseluruhan rangkaian spesifikasi ini dirancang untuk mengevaluasi kompromi antara kompleksitas model dan relevansi prediktor dengan hasil sebagai berikut.

Tabel 1. Hasil Analisis Regresi Logistik dengan Encoding WoE

Model	Jumlah Variabel	AUC ± SD	KS ± SD	Accuracy ± SD	F1-Score ± SD	Bal Accuracy ± SD	Sensitivity ± SD
(a)	91	0.581 ± 0.088	0.209 ± 0.138	0.540 ± 0.105	0.514 ± 0.128	0.539 ± 0.106	0.525 ± 0.170
(b)	53	0.622 ± 0.097	0.272 ± 0.166	0.585 ± 0.110	0.570 ± 0.125	0.585 ± 0.111	0.587 ± 0.156
(c)	18	0.598 ± 0.087	0.210 ± 0.147	0.575 ± 0.110	0.556 ± 0.127	0.576 ± 0.110	0.574 ± 0.175
(d)	16	0.630 ± 0.103	0.336 ± 0.156	0.621 ± 0.090	0.606 ± 0.097	0.621 ± 0.090	0.620 ± 0.137

Tabel 1 menunjukkan bahwa model dengan performa terbaik tidak dihasilkan oleh model yang menggunakan 91 variabel prediktor. Penerapan strategi seleksi

variabel pada spesifikasi selain model (a) menghasilkan peningkatan pada hampir seluruh metrik evaluasi dibandingkan model penuh. Temuan ini mengindikasikan bahwa reduksi kompleksitas melalui pemilihan prediktor yang paling informatif mampu meningkatkan performa validasi. Hasil tersebut menegaskan bahwa proses seleksi variabel merupakan tahapan penting dalam pengembangan model regresi logistik berbasis *encoding* WoE pada kondisi data berdimensi tinggi dengan jumlah observasi yang relatif terbatas.

3.5. *Analisis Model Regresi Logistik dengan Encoding OT*

Tahapan *preprocessing* sebelum pemodelan regresi logistik dengan *encoding* OT mengikuti prosedur yang sama seperti regresi logistik berbasis *encoding* WoE. Perbedaannya terletak pada proses optimasi *hyperparameter* jumlah permutasi dan nilai *smoothing*. Jumlah permutasi yang diuji adalah 5, 10, 15, dan 20 sedangkan nilai *smoothing* yang diuji adalah 1, 5, 10, 15, dan 20. Seluruh kombinasi *hyperparameter* kemudian dievaluasi berdasarkan kestabilan hasil transformasi OT. Hasil evaluasi menunjukkan bahwa jumlah permutasi sebesar 5 dan nilai *smoothing* sebesar 20 memberikan hasil transformasi yang paling stabil sehingga dipilih sebagai *hyperparameter* optimal pada penelitian ini.

Penelitian ini membangun empat spesifikasi model regresi logistik dengan memanfaatkan prediktor hasil *encoding* OT untuk dibandingkan yaitu: (a) model penuh dengan 91 variabel prediktor; (b) model seleksi AUC Univariat yang hanya menggunakan variabel dengan nilai AUC Univariat lebih dari 0,55 [15]; (c) model *forward selection* yang melakukan pemilihan maju dari himpunan 91 variabel prediktor; dan (d) model *forward selection* pasca seleksi AUC Univariat dengan melakukan pemilihan maju yang dibatasi pada variabel prediktor yang telah disaring berdasarkan AUC Univariat. Rangkaian spesifikasi ini dirancang untuk mengevaluasi kompromi antara kompleksitas model dan relevansi prediktor dengan hasil sebagai berikut.

Tabel 2. Hasil Analisis Regresi Logistik dengan Encoding OT

Model	Jumlah Variabel	AUC ± SD	KS ± SD	Accuracy ± SD	F1-Score ± SD	Bal Accuracy ± SD	Sensitivity ± SD
(a)	91	0.607 ± 0.080	0.253 ± 0.137	0.566 ± 0.104	0.540 ± 0.128	0.566 ± 0.103	0.550 ± 0.165
(b)	35	0.608 ± 0.096	0.292 ± 0.143	0.559 ± 0.108	0.542 ± 0.120	0.560 ± 0.108	0.560 ± 0.160
(c)	16	0.615 ± 0.107	0.243 ± 0.180	0.607 ± 0.106	0.593 ± 0.115	0.607 ± 0.106	0.610 ± 0.151
(d)	8	0.683 ± 0.106	0.419 ± 0.148	0.615 ± 0.115	0.623 ± 0.119	0.617 ± 0.116	0.678 ± 0.164

Tabel 2 menunjukkan bahwa performa terbaik dihasilkan oleh spesifikasi yang menerapkan seleksi variabel sebelum proses pemodelan. Seluruh spesifikasi selain model penuh menghasilkan peningkatan pada hampir seluruh metrik evaluasi. Temuan ini menunjukkan bahwa manfaat seleksi variabel tetap konsisten pada *encoding* OT. Hasil tersebut menegaskan bahwa kombinasi transformasi OT dan seleksi variabel mampu menghasilkan model yang lebih sederhana dengan performa validasi yang lebih baik pada data berdimensi tinggi dan jumlah observasi yang relatif terbatas.

3.6. *Analisis Model Catboost*

Penelitian ini menerapkan proses optimasi *hyperparameter* untuk memperoleh konfigurasi model CatBoost yang paling optimal. *Hyperparameter* yang dioptimasi meliputi *learning rate*, *l2_leaf_reg*, *bagging temperature*, *random strength*, *one hot max size*, *leaf estimation iterations* dan *depth*. Proses optimasi dilakukan secara sistematis menggunakan skema *repeated cross validation* untuk memperoleh kombinasi parameter yang menghasilkan performa terbaik pada data validasi. Pemilihan parameter ini mengacu pada panduan dan prinsip yang dijabarkan dalam publikasi CatBoost [9].

Penelitian ini membangun tiga spesifikasi model CatBoost untuk dibandingkan yaitu: (a) CatBoost yang menggunakan 91 variabel prediktor; (b) CatBoost WoE *Selected* yang menggunakan himpunan variabel hasil seleksi yang sama dengan model regresi logistik berbasis *encoding* WoE; dan (c) CatBoost OT *Selected* yang menggunakan himpunan variabel hasil seleksi yang sama dengan model regresi logistik berbasis *encoding* OT dengan hasil sebagai berikut.

Tabel 3. Hasil Analisis Model CatBoost

Model	Jumlah Variabel	AUC ± SD	KS ± SD	Accuracy ± SD	F1-Score ± SD	Bal Accuracy ± SD	Sensitivity ± SD
(a)	91	0.715 ± 0.107	0.471 ± 0.105	0.725 ± 0.106	0.722 ± 0.161	0.726 ± 0.107	0.750 ± 0.167
(b)	53	0.733 ± 0.110	0.426 ± 0.122	0.697 ± 0.091	0.673 ± 0.154	0.695 ± 0.095	0.654 ± 0.184
(c)	35	0.736 ± 0.091	0.429 ± 0.096	0.633 ± 0.087	0.565 ± 0.122	0.627 ± 0.087	0.500 ± 0.132

Kombinasi *hyperparameter* optimal berbeda pada setiap spesifikasi model CatBoost karena masing-masing spesifikasi menggunakan jumlah dan himpunan variabel prediktor yang berbeda. Konfigurasi *hyperparameter* optimal untuk setiap spesifikasi disajikan secara terpisah pada Tabel 4 sebagai berikut.

Tabel 4. Kombinasi Hyperparameter Optimal Model CatBoost

Model	Learning Rate	Depth	L2 Leaf Reg	Bagging Temp.	Random Strength	One Hot Max Size	Leaf Est. Iter.
(a) CatBoost Full	0.04	7	20	0.25	0.5	2	1
(b) CatBoost WoE Selected	0.04	7	10	1	1	2	1
(c) CatBoost OT Selected	0.04	6	10	1	0.5	3	3

4. Kesimpulan

Hasil evaluasi menunjukkan bahwa model skor kepatuhan yang saat ini digunakan oleh DJBC memiliki kemampuan diskriminasi yang relatif terbatas dengan skor KS sebesar 0,27. Pengembangan model berbasis data yang dilakukan pada penelitian ini mampu meningkatkan performa klasifikasi secara konsisten sehingga menunjukkan bahwa pendekatan empiris lebih efektif dalam merepresentasikan risiko pelanggaran dibandingkan *scorecard* manual yang saat ini digunakan.

Hasil pengembangan model secara keseluruhan mengindikasikan bahwa model CatBoost menunjukkan kinerja yang secara signifikan lebih unggul dibandingkan regresi logistik dengan *encoding* WoE maupun OT dalam mengklasifikasikan status

kepatuhan pengusaha cukai MMEA. Keunggulan tersebut tercermin pada hampir seluruh metrik evaluasi. Hal ini mengonfirmasi kemampuan CatBoost dalam menangani data kategorik berdimensi tinggi secara efisien melalui mekanisme *ordered boosting* yang mengurangi bias akibat *target leakage* sekaligus meningkatkan daya diskriminasi model terhadap perilaku kepatuhan. Model CatBoost dengan seluruh variabel direkomendasikan untuk digunakan apabila prioritas utama adalah pencapaian akurasi dan daya diskriminasi tertinggi. Alternatif yang lebih efisien ditunjukkan oleh model CatBoost dengan 35 variabel hasil seleksi AUC Univariat yang mampu mempertahankan performa kompetitif dengan kompleksitas model yang lebih rendah.

Penerapan *encoding* OT dalam regresi logistik terbukti memberikan hasil yang lebih baik dibandingkan *encoding* WoE pada seluruh spesifikasi model yang dilakukan. Kondisi ini memberikan alternatif pilihan kepada DJBC selaku *data owner* ketika menginginkan model yang bersifat *non blackbox* dan mudah diinterpretasikan. Regresi logistik dengan *encoding* OT dapat menjadi opsi metodologis yang lebih transparan sekaligus memberikan dasar empiris kuat untuk mendukung kebijakan pengawasan kepatuhan cukai.

5. Ucapan Terima kasih

Penulis menyampaikan terima kasih kepada Direktorat Jenderal Bea dan Cukai atas izin yang diberikan dalam pemanfaatan data internal yang menjadi dasar utama analisis dalam penelitian ini. Ucapan terima kasih juga disampaikan kepada Basis Informasi Penelitian dan Pengabdian Masyarakat (BIMA) dari Kementerian Pendidikan Tinggi, Sains, dan Teknologi atas dukungan dan fasilitasi yang diberikan sehingga penelitian ini dapat terlaksana dengan baik dan menghasilkan luaran sebagaimana yang diharapkan.

Bibliography

[1] Williem, C.A., Julitawaty, W., 2024, Peran kebijakan ekstensifikasi cukai dalam melindungi kesehatan masyarakat, *Jurnal Akuntansi, Keuangan, Perpajakan dan Tata Kelola Perusahaan* Vol. **1**(3): 178 – 184

[2] Direktorat Jenderal Bea dan Cukai, 2024, *Instruksi Nomor INS-3/BC/2024 tentang Profil Risiko Pengusaha Barang Kena Cukai dalam Rangka Pengawasan dan Pelayanan di Bidang Cukai*, Kementerian Keuangan Republik Indonesia.

[3] Margeloiu, A., 2025, *Tabular Machine Learning on Small-Size and High-Dimensional Data*, University of Cambridge Repository. https://doi.org/10.17863/CAM.117151

[4] Sartono, B., Dharmawan, H., 2023, *Pemodelan Prediktif Berbasis Pohon Klasifikasi* Vol. 1, IPB Press.

[5] Siddiqi, N., 2017, *Intelligent Credit Scoring: Building and Implementing Better Credit Risk Scorecards*, 2nd ed., Hoboken, NJ: John Wiley & Sons, Inc.

[6] Duong, B. T., Sariyev, O., Zeller, M., 2025, Can Weight of Evidence Method Improve Poverty Targeting?, *Social Indicators Research* Vol. 178: 305 – 333, https://doi.org/10.1007/s11205-025-03576-z

[7] Pomazanov, M., 2023, Continuous transformation of nonmonotonic factors of

the scoring model increasing its discriminating power, *Procedia Computer Science* Vol. **221**: 237 – 244

[8] Hancock, J. T., Khoshgoftaar, T. M., 2020, CatBoost for Big Data: An Interdisciplinary Review, *Journal of Big Data* Vol. **7**(1): 1 – 45, https://doi.org/10.1186/s40537-020-00369-8

[9] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., Gulin, A., 2018, CatBoost: Unbiased boosting with categorical features, *Advances in Neural Information Processing Systems* Vol. **31**. https://arxiv.org/abs/1706.09516

[10] Biarnes, A., 2021, How CatBoost encodes categorical variables? *Medium*, https://medium.com/data-science/how-catboost-encodes-categorical-variables-3866fb2ae640

[11] Chen, H., Xiang, Y., 2017, The study of credit scoring model based on group lasso, *Procedia Computer Science* Vol. **122**: 677 – 684 https://doi.org/10.1016/j.procs.2017.11.423

[12] James, G., Witten, D., Hastie, T., Tibshirani, R., 2023, *An Introduction to Statistical Learning: With Applications in R and Python* (3rd ed.). Springer.

[13] Persson, R., 2021, Weight of evidence transformation in credit scoring models: How does it affect the discriminatory power?, *Master's thesis*, Lund University, Department of Statistics, LUP Student Papers, http://lup.lub.lu.se/student-papers/record/9066332

[14] Kritzinger, N., Van Vuuren, G. W., 2018, An optimised credit scorecard to enhance cut-off score determination, *South African Journal of Economic and Management Sciences* Vol. **21**(1). https://doi.org/10.4102/sajems.v21i1.1571

[15] Abellana, D.P.M., Lao, D.M.M., 2023, A new univariate feature selection algorithm based on the best-worst multi-attribute decision-making method, *Decision Analytics Journal* Vol. **7**: 100240. https://doi.org/10.1016/j.dajour.2023.100240