

EVALUASI DAN PENGEMBANGAN MODEL SKOR KEPATUHAN CUKAI PENGUSAHA MINUMAN BERALKOHOL DENGAN STRATEGI ENCODING INOVATIF

BOY RIANSYAH



**PROGRAM STUDI MAGISTER STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR**

**BOGOR
2026**



- 

PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Evaluasi dan Pengembangan Model Skor Kepatuhan Cukai Pengusaha Minuman Beralkohol dengan Strategi Encoding Inovatif” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir tesis ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2026

Boy Riansyah
M0501241012

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

RINGKASAN

BOY RIANSYAH. Evaluasi dan Pengembangan Model Skor Kepatuhan Cukai Pengusaha Minuman Beralkohol dengan Strategi Encoding Inovatif. Dibimbing oleh BAGUS SARTONO dan FARIT MOCHAMAD AFENDI.

Direktorat Jenderal Bea dan Cukai (DJBC) saat ini menggunakan model skor kepatuhan sebagai dasar penentuan prioritas dan strategi pengawasan terhadap pengusaha minuman beralkohol. Model tersebut disusun dalam bentuk *scorecard* manual menggunakan 91 peubah prediktor kategorik dengan struktur kategori dan pembobotan yang ditetapkan melalui pendekatan normatif dan justifikasi ahli. Pendekatan ini belum mempertimbangkan hubungan statistik empiris antara karakteristik pengusaha dan kejadian pelanggaran sehingga diperlukan evaluasi dan pengembangan model yang lebih berbasis data. Berdasarkan latar belakang tersebut, tujuan penelitian ini adalah (1) mengevaluasi kinerja model skor kepatuhan yang saat ini digunakan oleh DJBC; (2) mengembangkan model skor kepatuhan baru dengan membandingkan kinerja algoritma pemodelan regresi logistik dengan *encoding Weight of Evidence* (WoE), *Ordered Target* (OT) dan algoritma CatBoost; serta (3) mengidentifikasi peubah prediktor yang memiliki tingkat kepentingan paling tinggi berdasarkan model terbaik dari hasil pengembangan model.

Model skor kepatuhan yang digunakan oleh DJBC saat ini terdiri atas 309 level kategori yang berasal dari 91 peubah prediktor. Distribusi pengusaha pada setiap level kategori tidak merata sehingga masih terdapat level kategori yang hanya berisikan kurang dari 5% pengusaha minuman beralkohol di Indonesia (*sparse categories*). Kondisi tersebut berpotensi mempengaruhi kestabilan proses *encoding* dan kinerja model prediktif yang dihasilkan. Berdasarkan kondisi tersebut, penelitian ini menerapkan proses restrukturisasi level kategori melalui penggabungan beberapa level kategori dengan frekuensi yang rendah untuk memperoleh representasi kategori yang lebih stabil. Hasil restrukturisasi level kategori ini berhasil menurunkan jumlah level kategori menjadi 242 buah.

Proses evaluasi terhadap model skor kepatuhan yang saat ini digunakan oleh DJBC menunjukkan bahwa model memiliki kemampuan diskriminasi yang terbatas dengan nilai statistik Kolmogorov-Smirnov (KS) sebesar 0,27. Pola distribusi skor kepatuhan yang dihasilkan juga belum mencerminkan tingkat risiko secara memadai karena kelompok dengan skor kepatuhan tinggi didominasi oleh pengusaha yang melakukan pelanggaran. Pengembangan model skor kepatuhan menggunakan regresi logistik dan CatBoost berhasil meningkatkan kinerja klasifikasi dibandingkan model skor kepatuhan yang digunakan saat ini. Penerapan *encoding* OT pada model regresi logistik secara konsisten menghasilkan kinerja yang lebih baik dibandingkan *encoding* WoE pada seluruh skenario pemodelan. Model regresi logistik terbaik menghasilkan nilai KS sebesar 0,42 dan AUC sebesar 0,68. Adapun kinerja terbaik dari seluruh skenario pemodelan dihasilkan dari model CatBoost dengan nilai KS sebesar 0,47 dan AUC sebesar 0,72. Penelitian ini juga menggunakan skema *repeated cross validation* (5 *fold* dan 20 pengulangan) untuk memvalidasi kinerja model sehingga menghasilkan pengukuran kinerja yang lebih andal.



Analisis tingkat kepentingan peubah prediktor menunjukkan bahwa X_{21} merupakan peubah prediktor yang memiliki tingkat kepentingan tertinggi dari model CatBoost terbaik. Sebagian besar peubah prediktor dengan tingkat kepentingan tertinggi berasal dari dimensi risiko inheren dan operasional sedangkan kontribusi peubah dari dimensi risiko rekam jejak relatif lebih rendah. Hasil penelitian secara keseluruhan menunjukkan bahwa pendekatan berbasis *machine learning* dan strategi *encoding* yang tepat mampu meningkatkan kemampuan diskriminasi model dibandingkan *scorecard* manual yang saat ini digunakan oleh DJBC. Temuan ini dapat menjadi dasar empiris bagi pengembangan sistem skor kepatuhan yang lebih akurat, objektif dan berbasis data dalam mendukung pengawasan pengusaha minuman beralkohol di Indonesia.

Kata kunci: cukai, catboost, ordered target, regresi logistik, weight of evidence

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

SUMMARY

BOY RIANSYAH. Evaluation and Development of an Excise Compliance Scoring Model for Alcoholic Beverage Entrepreneurs with an Innovative Encoding Strategy. Supervised by BAGUS SARTONO and FARIT M AFENDI.

The Directorate General of Customs and Excise (DGCE) currently uses a compliance scoring model to determine supervisory priorities and strategies for alcoholic beverage entrepreneurs. The model is implemented as a manual scorecard comprising 91 categorical predictor variables. Its category structure and variable weights are determined through a normative approach and expert judgment. This approach does not consider the empirical statistical relationship between entrepreneurs' characteristics and the occurrence of violations. As a result, the existing model requires further evaluation and the development of a more data-driven approach. This study therefore aims to (1) evaluate the performance of the compliance scoring model currently used by the DGCE, (2) develop a new compliance scoring model by comparing the performance of logistic regression with Weight of Evidence (WoE) encoding, logistic regression with Ordered Target (OT) encoding, and the CatBoost algorithm, and (3) identify the most influential predictor variables based on the best-performing model.

The compliance scoring model currently used by the DGCE consists of 309 category levels derived from 91 predictor variables. The distribution of entrepreneurs across these category levels is uneven. Several category levels contain fewer than 5% of all alcoholic beverage entrepreneurs in Indonesia and are therefore classified as sparse categories. This condition may reduce the stability of the encoding process and the predictive performance of the resulting models. To address this issue, this study restructured the category levels by merging low-frequency categories to obtain more stable category representations. This restructuring process reduced the number of category levels from 309 to 242.

The evaluation of the compliance scoring model currently used by the DGCE showed that the model had limited discriminative ability, with a Kolmogorov–Smirnov (KS) statistic of 0.27. The distribution of compliance scores also failed to adequately reflect risk levels because entrepreneurs with high compliance scores were predominantly found among those who committed violations. The newly developed logistic regression and CatBoost models achieved better classification performance than the existing compliance scoring model. Logistic regression with Ordered Target (OT) encoding consistently outperformed logistic regression with Weight of Evidence (WoE) encoding across all modeling scenarios. The best logistic regression model achieved a KS statistic of 0.42 and an area under the receiver operating characteristic curve (AUC) of 0.68. The best overall performance was achieved by the CatBoost model, which attained a KS statistic of 0.47 and an AUC of 0.72. Model performance was validated using repeated five-fold cross-validation with 20 repetitions to provide more reliable performance estimates.

The predictor importance analysis identified X_{21} as the most influential predictor in the best-performing CatBoost model. Most of the highly influential predictors belonged to the inherent risk and operational risk dimensions. In contrast, predictor variables from the historical risk dimension made relatively smaller contributions to the model. Overall, the findings demonstrate that machine learning

approaches combined with appropriate encoding strategies improve the discriminative ability of the compliance scoring model compared with the manual scorecard currently used by the DGCE. These findings provide an empirical basis for developing a more accurate, objective, and data-driven compliance scoring system to support the supervision of alcoholic beverage entrepreneurs in Indonesia.

Keywords: excise, catboost, logistic regression, ordered target, weight of evidence

@Hak Cipta milik IPB University

IPB University



- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

EVALUASI DAN PENGEMBANGAN MODEL SKOR KEPATUHAN CUKAI PENGUSAHA MINUMAN BERALKOHOL DENGAN STRATEGI ENCODING INOVATIF

BOY RIANSYAH

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI MAGISTER STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

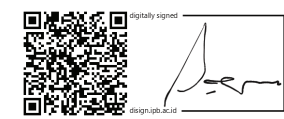


Judul Tesis : Evaluasi dan Pengembangan Model Skor Kepatuhan Cukai
Pengusaha Minuman Beralkohol dengan Strategi Encoding
Inovatif

Nama : Boy Riansyah
NIM : M0501241012

Disetujui oleh

Pembimbing 1:
Dr. Bagus Sartono, S.Si., M.Si.



Pembimbing 2:
Dr. Farit M Afendi, S.Si., M.Si.



Diketahui oleh

Ketua Program Studi:
Dr. Agus Mohamad Soleh, S.Si, MT.
NIP 197503151999031004



Dekan Sekolah Sains Data, Matematika, dan Informatika:
Prof. Dr. Ir. Agus Bueno, M.Si., M.Kom.
NIP 19660702 199302 1 001



Tanggal Ujian:
1 Juli 2026

Tanggal Lulus:



PRAKATA

Puji dan syukur penulis panjatkan ke hadirat Allah subhanahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini dapat diselesaikan dengan baik. Karya Ilmiah ini disusun sebagai bagian dari pemenuhan tugas akhir dengan judul “Evaluasi dan Pengembangan Model Skor Kepatuhan Cukai Pengusaha Minuman Beralkohol dengan Strategi Encoding Inovatif”.

Dalam proses penyusunan karya ilmiah ini, penulis memperoleh banyak bantuan, bimbingan dan dukungan dari berbagai pihak. Atas segala bantuan tersebut, penulis menyampaikan terima kasih dan penghargaan setinggi-tingginya kepada:

- Bapak Dr. Bagus Sartono, S.Si., M.Si. dan Dr. Farit M Afendi, S.Si., M.Si. selaku dosen pembimbing yang telah memberikan bimbingan, arahan, serta dukungan motivasi yang sangat berharga selama proses penelitian tesis.
- Ibu Dr. Ir. Indahwati, M.Si. selaku moderator pada kolokium yang telah memberikan masukan dan saran demi penyempurnaan proposal penelitian.
- Ibu Dr. Ir. Erfiani, M.Si. selaku penguji luar komisi yang telah memberikan masukan dan saran penyempurnaan penelitian.
- Ibu Sachnaz Desta Oktarina, S.Stat., M.Agr.Sc., Ph.D. dan Bapak Ali Winoto, S.Sos., MM. yang telah memberikan rekomendasi kepada penulis untuk dapat melanjutkan pendidikan magister di IPB University.
- Orang tua dan seluruh keluarga, khususnya almarhumah ibu Ati Sunarti dan adik Lis Diana Putri atas dukungan, doa, dan kasih sayang yang senantiasa menyertai penulis selama proses akademik dan penyusunan tesis.
- Direktorat Jenderal Bea dan Cukai, khususnya Direktorat Teknis dan Fasilitas Cukai dan Direktorat Penindakan dan Penyidikan yang telah memberikan kemudahan akses data penelitian.
- Lembaga Pengelola Dana Pendidikan (LPDP) Kementerian Keuangan Republik Indonesia atas segala dukungan pendanaan serta fasilitas yang diberikan kepada penulis untuk menyelesaikan pendidikan dan penelitian tesis ini dengan lancar.
- Kementerian Pendidikan Sains Tinggi, Sains dan Teknologi atas dukungan pendanaan melalui program riset BIMA yang diberikan kepada Dr. Bagus Sartono, S.Si., M.Si., sehingga turut mendukung kelancaran pelaksanaan penelitian tesis ini dengan baik.
- Teman-teman program studi Statistika dan Sains Data angkatan 2024 yang telah memberikan dukungan, semangat, serta kebersamaan selama proses pembelajaran dan penyelesaian penelitian tesis.

Akhir kata, semoga karya ilmiah ini dapat memberikan manfaat bagi semua pihak yang membutuhkan dan berkontribusi nyata bagi pengembangan ilmu pengetahuan dan perumusan kebijakan cukai berbasis data di Indonesia.

Bogor, Juli 2026

Boy Riansyah



- 

DAFTAR ISI

DAFTAR ISI	vii
DAFTAR TABEL	viii
DAFTAR GAMBAR	viii
DAFTAR LAMPIRAN	viii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
II TINJAUAN PUSTAKA	5
2.1 Cukai dan Kepatuhan Pengusaha Barang Kena Cukai	5
2.2 Model Skor Kepatuhan Cukai	5
2.3 Data Kategorik dan Proses <i>Encoding</i>	7
2.4 Regresi Logistik	7
2.5 <i>Weight of Evidence</i> (WoE)	8
2.6 Model CatBoost dan <i>Encoding Ordered Target</i> (OT)	9
2.7 Statistik Kolmogorov-Smirnov	11
2.8 Seleksi Peubah Prediktor	12
2.9 Perbandingan Kinerja Pengembangan Model	12
2.10 Identifikasi Tingkat Kepentingan Peubah Prediktor	13
III METODE	14
3.1 Pendekatan Penelitian	14
3.2 Data	14
3.3 Prosedur Analisis Data	15
IV HASIL ANALISIS DATA	19
4.1 Statistika Deskriptif Peubah Respons	19
4.2 Statistika Deskriptif Peubah Prediktor	19
4.3 Restrukturisasi Kategori Peubah Prediktor	22
4.4 Evaluasi Model Skor Kepatuhan yang Saat Ini Berlaku	24
4.5 Hasil Regresi Logistik dengan <i>Encoding</i> WoE	25
4.6 Hasil Regresi Logistik dengan <i>Encoding</i> OT	28
4.7 Hasil CatBoost	31
4.8 Analisis Tingkat Kepentingan Peubah Prediktor	32
V KESIMPULAN DAN SARAN	34
5.1 Kesimpulan	34
5.2 Saran	34
DAFTAR PUSTAKA	36
LAMPIRAN	40
RIWAYAT HIDUP	53



DAFTAR TABEL

1	Scorecard manual kepatuhan cukai	6
2	Dimensi peubah prediktor	14
3	Ringkasan struktur level kategori model scorecard INS- 3/BC/2024	20
4	Ringkasan struktur level kategori yang ditempati oleh pengusaha	20
5	Level kategori yang masuk ke dalam sparse categories	22
6	Ringkasan struktur level kategori setelah proses restrukturisasi	23
7	Hasil regresi logistik dengan encoding WoE	27
8	Hasil regresi logistik dengan encoding OT	29
9	Hasil algoritma CatBoost	31

DAFTAR GAMBAR

1	Distribusi pengusaha berdasarkan status kepatuhan	19
2	Sebaran jumlah pengusaha per level kategori dari tiap dimensi risiko	21
3	Nilai absolut korelasi point biserial dari peubah prediktor dengan peubah respons sebelum proses encoding	21
4	Boxplot sebaran skor kepatuhan berdasarkan INS-3/BC/2024	24
5	Plot KS (a) dan boxplot skor kepatuhan dari setiap kelas status kepatuhan pengusaha menggunakan model skor berdasarkan INS-3/BC/2024 (b)	24
6	Boxplot nilai WoE dari seluruh kategori peubah prediktor	26
7	Perbandingan nilai absolut korelasi point biserial peubah prediktor antara sebelum proses encoding dan setelah proses encoding WoE	26
8	Perbandingan nilai absolut korelasi point biserial peubah prediktor antara sebelum proses encoding dan setelah proses encoding OT	29
9	Perbandingan metrik accuracy, KS dan AUC dari regresi logistik encoding OT dan WoE di seluruh skenario pemodelan	30
10	Plot KS dari model CatBoost terbaik	32
11	Nilai feature importance dari peubah prediktor	33

DAFTAR LAMPIRAN

1	Lampiran 1 Peubah respons (Y) dan peubah prediktor (X)	41
2	Lampiran 2 Jumlah kategori per peubah prediktor hasil restrukturisasi	43
3	Lampiran 3 Skor kepatuhan dan status pelanggaran dari pengusaha	44
4	Lampiran 4 Nilai WoE dari peubah prediktor	45
5	Lampiran 5 Nilai Information Value dari peubah prediktor	48
6	Lampiran 6 Nilai OT dari peubah prediktor	49
7	Lampiran 7 Nilai AUC Univariat dari peubah prediktor	51
8	Lampiran 8 Nilai feature importance dari peubah prediktor	52