e-ISSN : 2964-5115
p-ISSN : 2964-4364

Evaluasi Regresi Logistik dengan Pembobotan Kelas dan Optimasi Ambang Keputusan untuk Prediksi Pembelian pada E-Commerce

Putri Mayang[*1], Casto Uripto[2], M. Rifa'i Susanto[3]
[*1]*Logistik bisnis, Politeknik Bisnis Digital Indonesia, Bogor*
[2,3]*Bisnis Digital, Politeknik Bisnis Digital Indonesia, Bogor*
e-mail: [*1]mayangputri85@gmail.com, [2]castouripto@gmail.com, [3]rifai.temprina@gmail.com

Abstrak

Penelitian ini bertujuan mengevaluasi pengaruh pembobotan kelas dan optimasi ambang keputusan terhadap kinerja regresi logistik dalam memprediksi pembelian pada e-commerce. Penelitian menggunakan Online Shoppers Purchasing Intention Dataset yang terdiri atas 12.330 sesi kunjungan dengan distribusi kelas tidak seimbang. Empat skenario dievaluasi, yaitu regresi logistik standar dan berbobot kelas, masing-masing menggunakan ambang keputusan 0,5 dan ambang optimal. Data dibagi secara stratifikasi menjadi data pelatihan dan pengujian dengan proporsi 80:20. Ambang optimal ditentukan berdasarkan F1-score dari prediksi out-of-fold pada data pelatihan. Hasil menunjukkan bahwa regresi logistik berbobot kelas dengan ambang optimal menghasilkan F1-score tertinggi sebesar 0,6306, dengan akurasi 0,8755, precision 0,5835, recall 0,6859, dan ROC-AUC 0,8962. Sebaliknya, regresi logistik standar dengan ambang 0,5 menghasilkan akurasi tertinggi sebesar 0,8812, tetapi recall hanya 0,3560. Analisis sensitivitas menunjukkan bahwa penghapusan variabel PageValues menurunkan F1-score model terbaik menjadi 0,3902. Temuan ini menunjukkan bahwa pembobotan kelas dan optimasi ambang dapat meningkatkan keseimbangan precision dan recall, sedangkan penggunaan PageValues perlu mempertimbangkan ketersediaan informasi pada waktu prediksi.

Article Info

Kata Kunci:
regresi logistik;
pembobotan kelas;
optimasi ambang keputusan;
prediksi pembelian;
e-commerce

Abstract

This study evaluates the effects of class weighting and decision threshold optimization on logistic regression performance for e-commerce purchase prediction. The Online Shoppers Purchasing Intention Dataset, comprising 12,330 browsing sessions with an imbalanced class distribution, was used. Four scenarios were evaluated: standard and class-weighted logistic regression, each with a default threshold of 0.5 and an optimized threshold. The data were stratified into training and test sets using an 80:20 split. The optimal threshold was selected by maximizing the F1-score based on out-of-fold predictions from the training data. The results show that class-weighted logistic regression with an optimized threshold achieved the highest F1-score of 0.6306, with an accuracy of 0.8755, precision of 0.5835, recall of 0.6859, and ROC-AUC of 0.8962. In contrast, standard logistic regression with a threshold of 0.5 achieved the highest accuracy of 0.8812, but its recall was only 0.3560. Sensitivity analysis revealed that removing the PageValues feature reduced the best model's F1-score to 0.3902. These findings indicate that class weighting and threshold optimization can improve the balance between precision and recall, while the use of PageValues requires consideration of its availability at prediction time

Keywords:
logistic regression;
class weighting;
decision threshold optimization;
purchase prediction;
e-commerce

I. PENDAHULUAN

E-commerce menjadi salah satu sarana perdagangan yang memungkinkan konsumen mencari produk, membandingkan pilihan, dan melakukan pembelian melalui platform digital. Bagi pengelola bisnis, data aktivitas pengunjung dapat digunakan untuk memahami pola kunjungan yang

e-ISSN : 2964-5115
p-ISSN : 2964-4364

berkaitan dengan transaksi. Informasi seperti jumlah halaman yang dikunjungi, durasi penelusuran, dan pola navigasi dapat dianalisis untuk memperkirakan kemungkinan terjadinya pembelian. Dalam konteks interaksi manusia dan komputer (IMK), data tersebut memberikan gambaran tentang perilaku pengguna ketika berinteraksi dengan sistem, meskipun tidak secara langsung mengukur persepsi atau kepuasan pengguna. Oleh karena itu, analisis perilaku interaksi dapat menjadi salah satu pendekatan untuk memahami aktivitas pengunjung dalam lingkungan bisnis digital.

Penelitian mengenai prediksi pembelian pada e-commerce telah dikembangkan melalui berbagai metode machine learning. Satu dan Islam (2023), misalnya, menggunakan Online Shoppers Purchasing Intention Dataset untuk mengevaluasi sejumlah teknik transformasi fitur, penyeimbangan data, seleksi fitur, dan algoritma klasifikasi. Penelitian tersebut melibatkan regresi logistik bersama sejumlah algoritma lainnya dan menemukan bahwa Random Forest memberikan hasil yang stabil dalam eksperimen mereka. Wen et al. (2023) juga mengembangkan pendekatan prediksi niat pembelian berdasarkan data clickstream, sedangkan Trivedi et al. (2024) mengkaji prediksi pembelian dengan menggabungkan karakteristik pelanggan dan pola penelusuran. Penelitian-penelitian tersebut menunjukkan bahwa aktivitas pengguna dan karakteristik kunjungan dapat dimanfaatkan untuk membangun model prediksi pembelian. Namun, kinerja model tetap perlu dinilai berdasarkan karakteristik data dan tujuan penggunaan hasil prediksi.

Salah satu metode yang dapat digunakan untuk klasifikasi pembelian adalah regresi logistik. Metode ini menghasilkan probabilitas suatu sesi berakhir dengan pembelian berdasarkan variabel prediktor yang digunakan. Selain menghasilkan klasifikasi, regresi logistik memungkinkan interpretasi koefisien dan odds ratio untuk menjelaskan hubungan statistik antara variabel prediktor dan peluang pembelian. Akan tetapi, hasil klasifikasi juga dipengaruhi oleh distribusi kelas dan ambang keputusan (decision threshold) yang digunakan untuk mengubah probabilitas menjadi label kelas. Jika ambang standar 0,5 digunakan tanpa mempertimbangkan tujuan analisis, model mungkin kurang optimal dalam mengenali sesi yang benar-benar menghasilkan pembelian. Pembobotan kelas dan penyesuaian ambang keputusan dapat menjadi alternatif yang perlu diuji untuk mengetahui pengaruhnya terhadap kinerja model.

Penelitian ini menggunakan Online Shoppers Purchasing Intention Dataset yang tersedia di Kaggle dan berasal dari UCI Machine Learning Repository (Sakar & Kastro, 2018). Dataset tersebut berisi 12.330 sesi kunjungan dengan 18 atribut, termasuk variabel target Revenue yang menunjukkan apakah suatu sesi berakhir dengan pembelian. Sebanyak 1.908 sesi atau sekitar 15,5% menghasilkan pembelian, sedangkan 10.422 sesi atau sekitar 84,5% tidak menghasilkan pembelian (Sakar & Kastro, 2018). Perbedaan proporsi tersebut menunjukkan adanya ketidakseimbangan kelas yang perlu diperhatikan dalam pemodelan. Kondisi ini juga menjadi alasan untuk tidak menilai model hanya berdasarkan akurasi, sebab model yang lebih sering memprediksi kelas mayoritas dapat menghasilkan akurasi yang relatif tinggi tanpa mampu mengenali sebagian besar sesi pembelian dengan baik.

Meskipun Satu dan Islam (2023) telah menguji regresi logistik pada dataset yang sama, penelitian mereka berfokus pada perbandingan berbagai teknik pengolahan fitur dan algoritma klasifikasi. Penelitian ini mengambil fokus yang lebih terarah, yaitu membandingkan regresi logistik standar dengan regresi logistik berbobot kelas serta menguji pengaruh optimasi ambang keputusan terhadap hasil prediksi. Perbandingan tersebut dilakukan untuk mengetahui apakah kedua pendekatan dapat memperbaiki kemampuan model dalam mengenali sesi pembelian dan bagaimana perubahan ambang keputusan memengaruhi keseimbangan antara presisi dan recall. Selain itu, koefisien dan odds ratio akan dianalisis untuk mengidentifikasi karakteristik interaksi pengguna yang berasosiasi dengan peluang pembelian. Dengan demikian, kesenjangan penelitian yang dituju terletak pada evaluasi terfokus terhadap strategi pembobotan kelas, pemilihan ambang keputusan, dan interpretasi model. Kebaruan ini tetap perlu dinilai berdasarkan perbandingan dengan penelitian terdahulu yang menggunakan rancangan serupa.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengevaluasi regresi logistik dengan

e-ISSN : 2964-5115
p-ISSN : 2964-4364

pembobotan kelas dan optimasi ambang keputusan untuk memprediksi pembelian pada e-commerce. Kinerja model akan diukur menggunakan confusion matrix, akurasi, presisi, recall, F1-score, dan ROC-AUC. Proses penentuan ambang keputusan akan menggunakan data validasi agar informasi dari data pengujian tidak memengaruhi pemilihan model. Hasil evaluasi selanjutnya dibandingkan dengan model regresi logistik standar untuk mengetahui perubahan kinerja yang diperoleh. Penelitian ini diharapkan dapat memberikan gambaran mengenai penggunaan regresi logistik pada data pembelian yang tidak seimbang sekaligus menjelaskan karakteristik interaksi pengguna yang berkaitan dengan hasil transaksi. Temuan yang diperoleh dapat menjadi bahan pertimbangan bagi pengelola e-commerce dalam memahami perilaku kunjungan, dengan tetap membatasi interpretasi pada hubungan statistik yang ditemukan dalam dataset.

II. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komputasional untuk mengevaluasi regresi logistik dalam memprediksi pembelian pada e-commerce. Eksperimen dilakukan dengan membandingkan regresi logistik standar dan regresi logistik berbobot kelas (class-weighted logistic regression), serta menguji pengaruh optimasi ambang keputusan (decision threshold optimization) terhadap hasil klasifikasi. Dataset yang digunakan adalah Online Shoppers Purchasing Intention Dataset yang tersedia melalui Kaggle dan didokumentasikan oleh UCI Machine Learning Repository (Sakar & Kastro, 2018). Unit analisis penelitian adalah sesi kunjungan pengguna, sedangkan variabel target Revenue menunjukkan apakah suatu sesi berakhir dengan pembelian.

Dataset dan Variabel Penelitian

Dataset berisi 12.330 sesi kunjungan dan 18 atribut yang mencakup variabel prediktor serta variabel target. Variabel prediktor menggambarkan aktivitas penelusuran, karakteristik sesi, dan karakteristik pengunjung. Tabel 1 menyajikan pengelompokan variabel berdasarkan fungsi analisisnya.

Tabel 1. Pengelompokan variabel penelitian

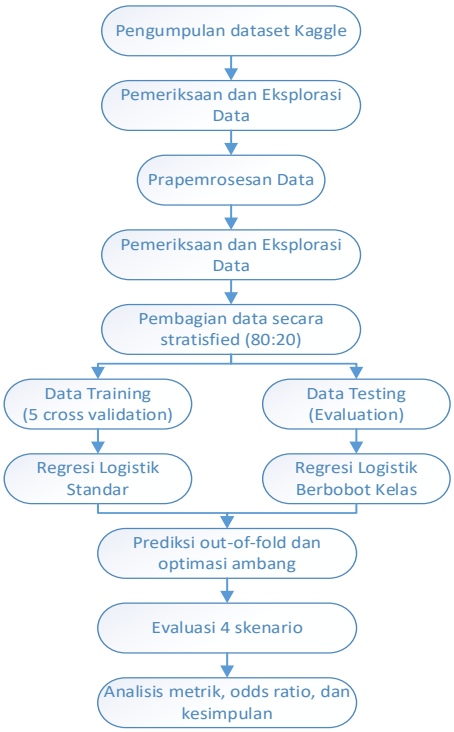
Kelompok	Variabel	Keterangan
Aktivitas navigasi	`Administrative, Informational, ProductRelated`	Jumlah halaman yang dikunjungi pada kategori terkait
Durasi penelusuran	`Administrative_Duration,Informational_Duration,ProductRelated_Duratio`	Durasi kunjungan pada kategori halaman terkait
Karakteristik sesi	`BounceRates, ExitRates, PageValues, SpecialDay, Weekend, Month`	Indikator perilaku kunjungan dan kondisi sesi
Karakteristik teknis dan pengunjung	`OperatingSystems, Browser, Region, TrafficType, VisitorType`	Informasi perangkat lunak, wilayah, sumber lalu lintas, dan jenis pengunjung
Variabel target	`Revenue`	Status pembelian pada sesi kunjungan

Variabel Revenue dikodekan sebagai 1 apabila sesi menghasilkan pembelian dan 0 apabila tidak menghasilkan pembelian. Berdasarkan dokumentasi dataset, sebanyak 1.908 sesi menghasilkan pembelian dan 10.422 sesi tidak menghasilkan pembelian (Sakar & Kastro, 2018). Dengan demikian, proporsi kelas pembelian sekitar 15,5%, sedangkan kelas nonpembelian sekitar 84,5%. Ketidakseimbangan tersebut menjadi pertimbangan utama dalam pemilihan strategi pelatihan dan metrik evaluasi.

Tahapan Pengolahan Data

Pemeriksaan awal dilakukan terhadap tipe data, nilai kosong, data duplikat, rentang nilai, dan distribusi kelas target. Variabel kategorikal, seperti Month dan VisitorType, diubah menjadi representasi numerik menggunakan one-hot encoding. Variabel numerik diproses sesuai kebutuhan model, termasuk standardisasi apabila diperlukan. Seluruh parameter prapemrosesan dipelajari hanya dari data pelatihan dan diterapkan pada data lain melalui pipeline yang sama untuk mencegah kebocoran data.

e-ISSN : 2964-5115
p-ISSN : 2964-4364



Gambar 1. Alur Penelitian

Dataset kemudian dibagi menjadi 80% data pelatihan dan 20% data pengujian dengan metode stratified split. Pembagian ini mempertahankan proporsi kelas pada kedua bagian data. Data pengujian disimpan untuk evaluasi akhir dan tidak digunakan dalam pemilihan parameter, model, atau ambang keputusan. Pada data pelatihan, digunakan validasi silang berstrata lima lipatan (stratified five-fold cross-validation) untuk menghasilkan prediksi out-of-fold yang diperlukan dalam evaluasi dan pemilihan ambang keputusan.

Pemodelan Regresi Logistik

Regresi logistik digunakan untuk memperkirakan probabilitas suatu sesi menghasilkan pembelian. Model ini dipilih karena sesuai untuk klasifikasi biner dan memungkinkan interpretasi hubungan statistik antara variabel prediktor dan peluang pembelian. Probabilitas pembelian dirumuskan sebagai berikut:

$$P(Y_i = 1 \mid X_i) = p_i = \frac{1}{1 + \exp(-z_i)}$$

Dengan

$$z_i = \beta_0 + \sum_{j=1}^{m} \beta_j x_{ij}$$

Model regresi logistik standar dilatih menggunakan fungsi kerugian logistik dengan regularisasi yang ditetapkan secara konsisten. Untuk mengatasi ketidakseimbangan kelas, penelitian juga menggunakan regresi logistik berbobot kelas. Bobot kelas dihitung berdasarkan jumlah sampel pada masing-masing kelas menggunakan persamaan:

$$w_c = \frac{N}{K, N_c}$$

dengan (w_c) sebagai bobot kelas ke-(c), (N) sebagai jumlah seluruh sampel pelatihan, (K) sebagai jumlah kelas, dan (N_c) sebagai jumlah sampel pada kelas ke-(c). Karena kelas pembelian lebih sedikit, pendekatan ini memberikan bobot relatif lebih besar kepada sampel pembelian dalam proses pelatihan.

Fungsi kerugian berbobot dapat dinyatakan sebagai:

$$L = -\sum_{i=1}^{N} w_{y_i}[y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)]$$
$$+ \frac{\lambda}{2} \sum_{j=1}^{m} \beta_j^2$$

dengan (w_{y_i}) sebagai bobot kelas dari sampel ke-(i) dan (λ) sebagai parameter regularisasi. Regularisasi diterapkan pada koefisien fitur, tidak termasuk intersep. Dalam implementasi, parameter model dan regularisasi harus ditetapkan secara konsisten agar perubahan hasil dapat dikaitkan dengan pembobotan kelas dan ambang keputusan, bukan perbedaan pengaturan lain.

Optimasi Ambang Keputusan

Model regresi logistik menghasilkan probabilitas, bukan label kelas secara langsung. Untuk mengubah probabilitas menjadi prediksi kelas, digunakan ambang keputusan (t), yang dirumuskan sebagai berikut:

$$\hat{y}_i = \begin{cases} 1, jika\ p_i \geq t; \\ 0, jika\ p_i \leq t; \end{cases}$$

Ambang standar yang digunakan sebagai pembanding adalah t=0.5. Penelitian ini juga mengevaluasi ambang alternatif yang dipilih berdasarkan F1-score tertinggi pada prediksi out-of-fold dari data pelatihan.

F1-score dihitung menggunakan persamaan:

$$F1 = 2\frac{P \times R}{P + R}$$

Rancangan Eksperimen

Tabel 2. Skenario eksperimen

Skenario	Model	Ambang keputusan	Tujuan
E1	Regresi logistik standar	0,5	Menjadi baseline
E2	Regresi logistik standar	Hasil optimasi	Mengukur dampak optimasi ambang
E3	Regresi logistik berbobot kelas	0,5	Mengukur dampak pembobotan kelas
E4	Regresi logistik berbobot kelas	Hasil optimasi	Mengukur kombinasi kedua pendekatan

Penelitian membandingkan empat skenario untuk memisahkan kontribusi pembobotan kelas dan optimasi ambang keputusan. Seluruh skenario menggunakan pembagian data dan prosedur prapemrosesan yang sama.

Interpretasi Model dan Reproduksibilitas

Koefisien regresi dan odds ratio digunakan untuk menginterpretasikan hubungan statistik antara fitur dan peluang pembelian. Odds ratio diperoleh melalui transformasi eksponensial koefisien:

$$\mathrm{OR_j} = \exp(\beta_j)$$

Nilai ($\mathrm{OR_j}$) menunjukkan bahwa kenaikan satu unit pada prediktor terkait dengan peningkatan odds pembelian, sedangkan nilai ($\mathrm{OR_j} < 1$) menunjukkan penurunan odds, dengan asumsi prediktor lain tetap. Untuk fitur hasil one-hot encoding, interpretasi dilakukan terhadap kategori referensi yang digunakan. Interpretasi tersebut tidak dianggap sebagai bukti hubungan sebab-akibat.

III. HASIL DAN PEMBAHASAN

Karakteristik Dataset

Dataset terdiri atas 12.330 sesi kunjungan pada situs e-commerce. Variabel target Revenue menandai apakah suatu sesi berakhir dengan pembelian, sedangkan prediktor mencakup aktivitas navigasi, durasi kunjungan, karakteristik waktu, tipe pengunjung, dan informasi teknis kunjungan. Data dibagi secara stratifikasi menjadi data pelatihan sebanyak 9.864 sesi dan data pengujian sebanyak 2.466 sesi. Ambang keputusan optimal ditentukan menggunakan F1-score dari prediksi out-of-fold pada data pelatihan melalui validasi silang lima lipatan. Seluruh metrik yang dilaporkan pada bagian ini dihitung pada data uji.

Distribusi kelas yang tidak seimbang menjadi alasan untuk tidak mengandalkan akurasi sebagai satu-satunya ukuran evaluasi. Pada kondisi tersebut, model dapat memperoleh akurasi tinggi dengan lebih sering memprediksi kelas mayoritas, tetapi gagal mendeteksi sebagian sesi pembelian. Karena itu, evaluasi juga mempertimbangkan precision, recall, F1-score, dan ROC-AUC.

Perbandingan Kinerja Empat Skenario

Tabel 3 memperlihatkan perbedaan kinerja antara regresi logistik standar dan regresi logistik berbobot kelas, masing-masing dengan ambang 0,5 dan ambang optimal. Regresi logistik standar pada ambang 0,5 menghasilkan akurasi tertinggi (0,8812) dan precision tertinggi (0,7432), tetapi recall hanya 0,3560 sehingga F1-score sebesar 0,4814. Artinya, prediksi positif dari model ini relatif tepat, tetapi banyak sesi pembelian aktual tidak teridentifikasi.

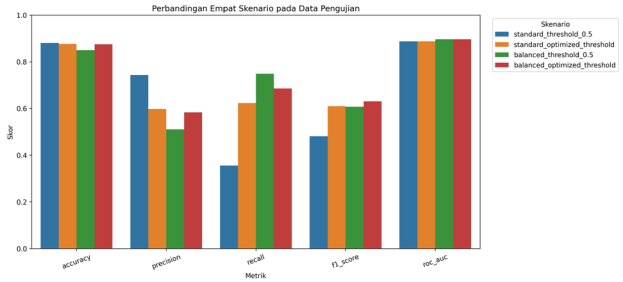
Pada regresi logistik standar, penurunan ambang menjadi 0,2264 meningkatkan recall menjadi 0,6230 dan F1-score menjadi 0,6103, sementara akurasi turun sedikit menjadi 0,8767. Model berbobot kelas dengan ambang 0,5 menghasilkan recall tertinggi (0,7487), tetapi precision turun menjadi 0,5107 dan akurasi menjadi 0,8500. Sementara itu, model berbobot kelas dengan ambang optimal 0,5623 menghasilkan F1-score tertinggi (0,6306), dengan precision 0,5835, recall 0,6859, dan akurasi 0,8755. Dibandingkan model standar pada ambang 0,5, F1-score meningkat sebesar 0,1492 atau 14,92 poin persentase.

Dengan demikian, kombinasi pembobotan kelas dan optimasi ambang memberikan keseimbangan precision dan recall terbaik di antara skenario yang diuji, tetapi bukan pilihan dengan akurasi atau precision tertinggi. Gambar 1 merangkum pola ini dan memperlihatkan bahwa pemilihan model perlu disesuaikan dengan tujuan operasional. Jika kesalahan berupa prediksi pembelian yang keliru memiliki biaya tinggi, precision mungkin lebih penting; jika kehilangan calon transaksi lebih merugikan, recall yang lebih tinggi dapat menjadi prioritas.

e-ISSN : 2964-5115
p-ISSN : 2964-4364

Tabel 3. Kinerja empat skenario pada data uji

	Standar	Standar+ ambang optimal	Berbobot kelas	Berbobot kelas + ambang optimal
Ambang	0.5000	0.2264	0.5000	0.5623
Akurasi	0.8812	0.8767	0.8500	0.8755
Precision	0.7432	0.5980	0.5107	0.5835
Recall	0.3560	0.6230	0.7487	0.6859
F1-Score	0.4814	0.6103	0.6072	0.6306
ROC-AUC	0.8876	0.8876	0.8962	0.8962

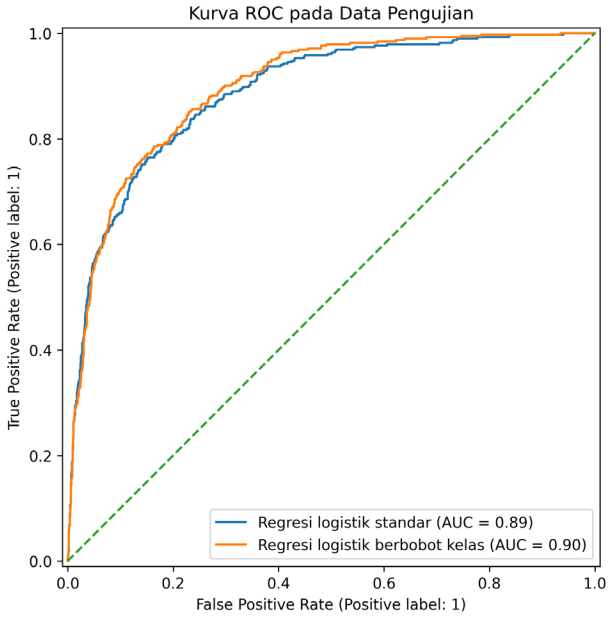


Gambar 2. Perbandingan metrik kinerja empat skenario pada data uji.

Pengaruh Ambang Keputusan dan Kemampuan Diskriminasi

Nilai ambang optimal yang diperoleh adalah 0,2264 untuk regresi logistik standar dan 0,5623 untuk model berbobot kelas. Perbedaan ini menunjukkan bahwa ambang yang menghasilkan F1-score terbaik bergantung pada distribusi probabilitas keluaran model. Ambang 0,5 tidak selalu menghasilkan keseimbangan precision dan recall terbaik, khususnya ketika distribusi kelas tidak seimbang atau model dilatih dengan pembobotan kelas.

ROC-AUC model standar sebesar 0,8876, sedangkan model berbobot kelas mencapai 0,8962. Nilai ROC-AUC tidak berubah saat ambang keputusan model yang sama dioptimalkan karena metrik tersebut menilai kemampuan pemeringkatan skor pada berbagai ambang, bukan keputusan pada satu ambang tertentu. Perbedaan ROC-AUC yang relatif kecil menunjukkan kemampuan diskriminasi sedikit lebih tinggi pada model berbobot kelas, sedangkan peningkatan F1-score pada skenario ambang optimal merefleksikan penyesuaian aturan klasifikasi.



Gambar 3. Kurva ROC regresi logistik standar dan berbobot kelas.

Interpretasi Koefisien Model

Interpretasi koefisien model standar menunjukkan bahwa PageValues memiliki koefisien positif terbesar di antara fitur yang ditampilkan (β = 1,5363; OR = 4,6475). ExitRates memiliki koefisien negatif (β = −0,8107; OR = 0,4445), sedangkan kategori Month_Feb memiliki koefisien −1,3037 (OR = 0,2715) dan Month_Nov memiliki koefisien 0,6426 (OR = 1,9013). Odds ratio untuk variabel kategorikal harus ditafsirkan relatif terhadap kategori referensi, sedangkan interpretasi prediktor numerik mengacu pada skala fitur setelah standardisasi.

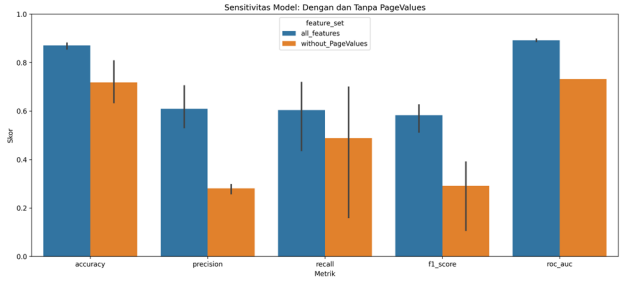
Koefisien tersebut menggambarkan asosiasi dalam model dan tidak membuktikan hubungan sebab-akibat. Karena pengujian signifikansi dan interval kepercayaan tidak disajikan, temuan ini sebaiknya diperlakukan sebagai interpretasi deskriptif atas model prediksi, bukan sebagai bukti pengaruh yang signifikan secara statistik.

Analisis Sensitivitas terhadap PageValues

Penghapusan PageValues menurunkan ROC-AUC dari 0,8876 menjadi 0,7316 pada model standar dan dari 0,8962 menjadi 0,7317 pada model berbobot kelas. Pada skenario berbobot kelas dengan ambang optimal, F1-score turun dari 0,6306 menjadi 0,3902. Hasil ini menunjukkan bahwa PageValues berkontribusi besar terhadap kinerja prediksi pada dataset yang digunakan.

e-ISSN : 2964-5115
p-ISSN : 2964-4364

Temuan tersebut tidak dengan sendirinya membuktikan adanya data leakage. Namun, definisi variabel dan waktu ketersediaannya perlu diperiksa sebelum model digunakan untuk prediksi prospektif. Jika PageValues dihitung menggunakan informasi yang baru tersedia setelah atau sangat dekat dengan transaksi, pemanfaatannya dapat membatasi kegunaan model untuk memperkirakan pembelian sebelum transaksi berlangsung. Gambar 4 menyajikan perbandingan kinerja model dengan dan tanpa fitur tersebut.



Gambar 4. Analisis sensitivitas kinerja model dengan dan tanpa PageValues.

Implikasi dan Keterbatasan

Secara empiris, hasil menunjukkan bahwa akurasi saja tidak memadai untuk menilai model pada data dengan kelas tidak seimbang. Pembobotan kelas meningkatkan perhatian terhadap kelas pembelian, sedangkan optimasi ambang menyesuaikan keseimbangan precision dan recall dengan kriteria F1-score. Bagi aplikasi e-commerce, pemilihan ambang perlu mempertimbangkan konsekuensi bisnis dari false positive dan false negative, bukan hanya nilai metrik statistik.

Penelitian ini terbatas pada satu dataset dan satu pembagian data uji. Kestabilan kinerja perlu diuji melalui pengulangan pembagian data atau validasi eksternal pada dataset lain. Selain itu, interpretasi koefisien belum disertai interval kepercayaan, dan peran PageValues perlu dievaluasi berdasarkan definisi serta waktu ketersediaan fitur. Batasan ini perlu dipertimbangkan sebelum hasil diterapkan pada sistem operasional.

IV. KESIMPULAN

Penelitian ini mengevaluasi kinerja regresi logistik standar dan regresi logistik berbobot kelas dengan optimasi ambang keputusan untuk memprediksi pembelian pada e-commerce. Hasil pengujian menunjukkan bahwa kombinasi pembobotan kelas dan optimasi ambang menghasilkan F1-score tertinggi sebesar 0,6306, dengan precision 0,5835, recall 0,6859, dan akurasi 0,8755. Sementara itu, regresi logistik standar dengan ambang 0,5 menghasilkan akurasi tertinggi sebesar 0,8812, tetapi recall hanya mencapai 0,3560. Temuan ini menegaskan bahwa akurasi saja belum memadai untuk mengevaluasi model pada data dengan distribusi kelas tidak seimbang. Pembobotan kelas dan optimasi ambang dapat meningkatkan keseimbangan antara precision dan recall, meskipun kinerja terbaik tetap bergantung pada tujuan prediksi yang ingin dicapai.

Analisis sensitivitas menunjukkan bahwa penghapusan variabel PageValues menurunkan F1-score model berbobot kelas dengan ambang optimal dari 0,6306 menjadi 0,3902. Hasil tersebut mengindikasikan bahwa PageValues berkontribusi besar terhadap kemampuan prediksi model, sehingga definisi variabel dan waktu ketersediaannya perlu diperiksa sebelum model diterapkan untuk prediksi pembelian secara prospektif. Secara keseluruhan, penelitian ini menunjukkan bahwa evaluasi model prediksi pembelian perlu mempertimbangkan strategi penanganan ketidakseimbangan kelas, pemilihan ambang keputusan, dan validitas fitur yang digunakan. Penelitian selanjutnya dapat menguji kestabilan kinerja melalui pengulangan pembagian data, mengevaluasi ambang berdasarkan biaya kesalahan bisnis, serta menguji generalisasi model pada dataset e-commerce lain..

V. REFERENSI

Sakar, C., & Kastro, Y. (2018). *Online Shoppers Purchasing Intention Dataset* [Data set]. UCI Machine Learning Repository. https://doi.org/10.24432/C5F88Q

Satu, M. S., & Islam, S. F. (2023). Modeling online customer purchase intention behavior applying different feature engineering and classification techniques. *Discover Artificial Intelligence*, 3, Article 36. https://doi.org/10.1007/s44163-023-00086-0

e-ISSN : 2964-5115
p-ISSN : 2964-4364

Tanvir, A.-A., Khandokar, I. A., Islam, A. K. M. M., Islam, S., & Shatabda, S. (2023). A gradient boosting classifier for purchase intention prediction of online shoppers. *Heliyon*, 9(4), Article e15163. https://doi.org/10.1016/j.heliyon.2023.e15163

Trivedi, S. K., Patra, P., Srivastava, P. R., Zhang, J. Z., & Zheng, L. J. (2024). What prompts consumers to purchase online? *A machine learning approach. Electronic Commerce Research,* 24(4), 2953–2989. https://doi.org/10.1007/s10660-022-09624-x

Wen, Z., Lin, W., & Liu, H. (2023). *Machine-learning-based approach for anonymous online customer purchase intentions using clickstream data. Systems,* 11(5), Article 255. https://doi.org/10.3390/systems11050255