

Penanganan Data *Churn* Tidak Seimbang Menggunakan Pembobotan pada Model Supervised Machine Learning

Handling Imbalanced Churn Data Using Weighting in Supervised Machine Learning Models

Sitti Nurhaliza^{a,1,*}, Andi Harismahyanti^{a,2}, Morina A. Fathan^{b,3}, Muhammad Edy Rizal^{a,4},
Muh. Zarkawi Yahya^{b,5}, Asfar^{a,6}

^aProgram Studi Sains Data, Univeristas Tadulako, Palu, Indonesia

^bProgram Studi Statistika, Univeristas Tadulako, Palu, Indonesia

¹nurhaliza@untad.ac.id; ²andiharismahyanti@untad.ac.id; ³morinafathan@untad.ac.id; ⁴muhedyrizal@untad.ac.id;

⁵zarkawi@untad.ac.id; ⁶asfar@untad.ac.id

*Corresponding author

Informasi Artikel	ABSTRAK
<p>Diserahkan : 17 September 2025 Diterima : 28 November 2025 Direvisi : 23 Februari 2026 Diterbitkan : 25 Februari 2026</p> <p>Kata Kunci: Churn Imbalance Data Weighted Supervised Machine Learning</p>	<p>Customer churn merupakan tantangan strategis dalam industri digital karena berdampak langsung pada pendapatan dan biaya akuisisi pelanggan baru. Salah satu kendala utama dalam membangun model prediksi churn adalah ketidakseimbangan kelas, dimana proporsi pelanggan churn hanya 11,4% dibandingkan 88,6% non-churn, dengan imbalance ratio hampir 8:1. Ketidakseimbangan ini berpotensi menurunkan sensitivitas model terhadap kelas minoritas. Meskipun berbagai teknik penanganan imbalance telah banyak diteliti, studi yang secara sistematis mengevaluasi efektivitas class weighting pada model-model klasifikasi dasar dalam konteks churn dengan ketidakseimbangan ekstrem masih terbatas. Penelitian ini bertujuan mengevaluasi efektivitas teknik pembobotan kelas (class weighting) dalam meningkatkan kinerja model klasifikasi churn pada data telekomunikasi JABODETABEK tahun 2019. Pendekatan supervised machine learning digunakan dengan lima algoritma utama: regresi logistik, K-Nearest Neighbors (KNN), decision tree, naive Bayes, dan random forest. Evaluasi dilakukan menggunakan stratified 5-fold cross-validation dan metrik yang relevan untuk data tidak seimbang, yaitu recall, specificity, F1-score, dan AUC-ROC. Hasil penelitian menunjukkan bahwa penerapan class weighting memberikan peningkatan signifikan pada nilai recall, khususnya pada model decision tree, KNN, dan naive Bayes. Model Naive Bayes Balanced memberikan performa terbaik dengan recall di atas 0,75, meskipun terjadi sedikit penurunan specificity sebagai trade-off. Secara umum, strategi pembobotan kelas terbukti mengurangi bias terhadap kelas mayoritas dan menghasilkan keseimbangan metrik yang lebih baik. Temuan ini menegaskan bahwa teknik penyeimbangan kelas, meskipun sederhana, tetap krusial untuk meningkatkan akurasi identifikasi pelanggan berisiko churn dan dapat dijadikan referensi praktis dalam pengembangan sistem retensi pelanggan di sektor industri telekomunikasi.</p>
<p>Keywords: Churn Imbalance Data Class Weighting Supervised Machine Learning</p>	<p>ABSTRACT</p> <p><i>Customer churn is a strategic challenge in the digital industry as it directly impacts revenue and the cost of acquiring new customers. One of the main obstacles in building churn prediction models is class imbalance, where the proportion of churned customers is only 11.4% compared to 88.6% non-churn, with an imbalance ratio of nearly 8:1. This imbalance can potentially reduce the model's sensitivity to the minority class. Although various imbalance-handling techniques have been widely studied, systematic evaluations of the effectiveness of class weighting on basic classification models in churn prediction with extreme class imbalance remain limited. This study aims to evaluate the effectiveness of class weighting techniques in improving the performance of churn classification models using 2019 telecommunication data from the JABODETABEK region. A supervised machine learning approach was employed with five main algorithms: logistic regression, K-Nearest Neighbors (KNN), decision tree, naive Bayes, and random forest. The evaluation was conducted using stratified 5-fold cross-validation and metrics relevant for imbalanced data, namely recall, specificity, F1-score, and AUC-ROC. The results show that the application of class weighting</i></p>

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



significantly improved recall values, particularly in decision tree, KNN, and naive Bayes models. The Balanced Naive Bayes model achieved the best performance with a recall above 0.75, although accompanied by a slight decrease in specificity as a trade-off. Overall, class weighting strategies were proven to reduce bias toward the majority class and yield better-balanced metrics. These findings highlight that class balancing techniques, although simple, remain crucial for enhancing the accuracy of identifying customers at risk of churn and can serve as a practical reference in developing customer retention systems in the telecommunications industry.

I. Pendahuluan

Dalam era transformasi digital, customer churn yakni keputusan pelanggan untuk berhenti menggunakan layanan menjadi persoalan strategis karena tidak hanya mengurangi pendapatan, tetapi juga meningkatkan biaya akuisisi pelanggan baru [1]. Kemampuan mengidentifikasi pelanggan yang berisiko churn sejak dini merupakan fondasi penting bagi strategi retensi serta pengambilan keputusan bisnis yang efektif. Selain berdampak pada aspek finansial, churn juga memengaruhi metrik kunci seperti *customer lifetime value* dan efektivitas kampanye pemasaran. Oleh karena itu, peningkatan kemampuan prediksi churn menjadi kebutuhan operasional sekaligus strategis bagi banyak organisasi [2].

Namun, membangun model prediksi churn bukanlah hal yang sederhana. Salah satu tantangan teknis utama adalah ketidakseimbangan kelas, di mana jumlah kasus churn (kelas minoritas) jauh lebih sedikit dibandingkan dengan kasus tidak churn (kelas mayoritas) [3]. Pada studi ini, data aktual menunjukkan proporsi churn hanya sebesar 11,4%, sehingga terjadi ketimpangan rasio hampir 8:1 antara kelas mayoritas (non-churn) dan kelas minoritas (churn). Ketidakseimbangan tersebut berpotensi menimbulkan bias pada model prediksi sehingga cenderung mengabaikan pelanggan yang benar-benar berisiko churn.

Berbagai pendekatan telah dikembangkan untuk mengatasi masalah ini, baik pada level data maupun algoritma. Pendekatan berbasis data seperti SMOTE dan variannya misalnya Borderline-SMOTE serta ADASYN digunakan secara luas untuk menyeimbangkan distribusi kelas dengan menciptakan data sintetis [4]. Beberapa studi melaporkan bahwa SMOTE yang dikombinasikan dengan model ensemble seperti AdaBoost mampu meningkatkan performa prediksi churn hingga mencapai nilai F1-score sebesar 87,6% [5]. Pendekatan *hybrid oversampling* berbasis GAN yang dipadukan dengan *cost-sensitive learning* (misalnya CostLearnGAN) juga terbukti memberikan peningkatan signifikan pada AUC dan F1-score dibandingkan dengan pendekatan tradisional [6].

Walaupun metode canggih seperti *GAN-hybrid* dan *ensemble cost-sensitive* menunjukkan kinerja tinggi, class weighting tetap relevan karena bersifat praktis, mudah diterapkan, dan tidak memerlukan perubahan distribusi data [7],[8]. Pada sisi lain, pendekatan berbasis algoritma seperti *class weighting* dan *cost-sensitive learning* menawarkan solusi efisien dengan cara menyesuaikan bobot kesalahan selama proses pelatihan model [9]. Sejumlah studi juga mencatat bahwa metode resampling seperti SMOTE dan ADASYN tidak selalu stabil pada data dengan noise atau ketidakseimbangan tinggi, sehingga class weighting menjadi alternatif yang penting dipertimbangkan. Namun, kajian yang secara khusus membandingkan performa model klasifikasi dasar dengan dan tanpa pembobotan kelas pada konteks prediksi churn masih relatif terbatas. Kondisi ini menunjukkan adanya gap yang jelas dalam literatur, yaitu belum adanya bukti empiris yang sistematis mengenai efektivitas class weighting pada data churn yang sangat tidak seimbang. Selain itu, pendekatan ensemble modern seperti *Balanced Random Forest* dan *Easy Ensemble* juga dilaporkan memberikan hasil yang menjanjikan dalam klasifikasi data tidak seimbang [10], [11].

Berdasarkan latar belakang tersebut, penelitian ini berfokus pada evaluasi kinerja sejumlah model klasifikasi yaitu regresi logistik, *k-nearest neighbors* (KNN), decision tree, random forest, dan naive Bayes baik dengan maupun tanpa penerapan pembobotan kelas. Evaluasi dilakukan dengan menggunakan metrik yang relevan untuk data tidak seimbang, yakni recall, F1-score, specificity, dan AUC-ROC. Hasil penelitian ini diharapkan dapat memberikan bukti empiris terkait efektivitas pembobotan kelas dalam menangani data churn yang tidak seimbang, serta menjadi referensi praktis dalam pengembangan sistem prediksi churn di sektor industri digital.

II. Metode

A. Pendekatan Penelitian

Penelitian ini menerapkan pendekatan *supervised machine learning* untuk memprediksi *customer churn* pada dataset pelanggan industri yang memiliki ketidakseimbangan kelas. Dataset bersumber dari industri

telekomunikasi di wilayah JABODETABEK tahun 2019. Dataset yang digunakan memiliki fitur-fitur yaitu usia pelanggan, kecepatan internet, jumlah tagihan, penggunaan internet per bulan, dan durasi menonton TV. Seluruh proses analisis dilakukan menggunakan bahasa pemrograman R, dengan dukungan paket *tidyverse* untuk manajemen dan eksplorasi data [12]. Pipeline pemodelan, *resampling*, dan evaluasi dibangun melalui *mlr3verse*, khususnya *mlr3pipelines*, yang memungkinkan perancangan alur kerja dalam bentuk *directed acyclic graph* (DAG) yang fleksibel serta dapat dibenchmark secara sistematis, sehingga memudahkan replikasi dan eksperimen terkontrol [13].

B. Deskripsi Data

Dataset penelitian berisi informasi pelanggan yang mencakup data demografi, perilaku penggunaan layanan, dan status *churn* dengan nilai biner (1 = *churn*, 0 = *non-churn*). Distribusi awal data menunjukkan proporsi pelanggan *churn* sebesar 11,4% dan *non-churn* sebesar 88,6%, menghasilkan *imbalance ratio* (IR) sebesar 8:1. *Imbalance ratio* dihitung menggunakan formula (1).

$$IR = \frac{\max(n+,n-)}{\min(n+,n-)} \quad (1)$$

$$IR \approx 88.6 / 11.4 \approx 8:1$$

Berdasarkan literatur *imbalanced learning*, ketidakseimbangan dengan rasio ini dapat menyebabkan bias klasifikasi terhadap kelas mayoritas dan menurunkan sensitivitas terhadap kelas minoritas, sehingga memerlukan strategi penanganan khusus [14].

C. Persiapan Data dan Eksplorasi

Tahap ini mencakup data *cleaning* dan transformasi (tipe faktor untuk kategorikal, numerik untuk kontinu). Praktik dan resep teknis mengacu pada panduan data *cleaning* di R, termasuk pembuatan *dummy variables* (*one-hot encoding*) untuk kategorikal bila dibutuhkan. Standarisasi z-score juga dilakukan untuk menyamakan skala data dan menghindari bias dengan formula $X' = \frac{(X - \mu)}{\sigma}$, dengan μ rata-rata dan σ simpangan baku [15].

D. Analisis *churn Rate*

Churn rate merupakan metrik kunci yang mengukur persentase pelanggan yang berhenti menggunakan layanan dalam periode tertentu. Dalam konteks bisnis telekomunikasi, *churn rate* yang tinggi mengindikasikan adanya masalah dalam retensi pelanggan yang dapat berdampak signifikan terhadap revenue dan profitabilitas perusahaan.

Churn rate dihitung menggunakan formula (2):

$$churn\ Rate = \left(\frac{n_{Churn}}{n_{Total}} \right) \times 100\% \quad (2)$$

dimana n_{Churn} adalah jumlah pelanggan yang melakukan *churn* dan n_{Total} adalah total pelanggan dalam dataset.

Analisis *churn rate* dilakukan secara menyeluruh dengan mempertimbangkan berbagai segmentasi pelanggan berdasarkan karakteristik demografi, perilaku penggunaan layanan, dan durasi berlangganan. Hal ini penting untuk memahami pola *churn* yang spesifik pada setiap segmen pelanggan, yang dapat memberikan insight untuk strategi retensi yang lebih targeted.

Tingkat ketidakseimbangan dikuantifikasi menggunakan *imbalance ratio* sebagai dasar pemilihan strategi penanganan. Interpretasi *imbalance ratio* mengikuti kriteria:

- Rasio 1:1 = seimbang
- Rasio > 3:1 = tidak seimbang sedang
- Rasio > 10:1 = sangat tidak seimbang

Selain itu, nilai IR dipakai untuk mengkuantifikasi tingkat ketidakseimbangan dan menjadi dasar pemilihan strategi penanganan. Konsekuensi ketidakseimbangan terhadap pembelajaran mesin dibahas luas oleh He & Garcia [16].

E. Pembagian Data dan Validasi

Evaluasi model dilakukan menggunakan *Stratified K-Fold Cross-Validation* dengan $K=5$ untuk mempertahankan proporsi kelas di setiap fold. Pemilihan 5-fold *cross-validation* memberikan keseimbangan antara bias dan varians estimasi dengan biaya komputasi yang *reasonable* [17].

F. Algoritma Pembelajaran Mesin

1. Regresi Logistik

Regresi logistik dipilih sebagai baseline model karena kesederhanaan interpretasi dan kemampuan memberikan estimasi probabilitas yang dapat digunakan untuk analisis risiko. Model ini memodelkan log-odds probabilitas churn sebagai fungsi linier dari variabel prediktor.

$$P(Y = 1|X) = \frac{1}{1 + \exp\{-(\beta^0 + \beta^1 X^1 + \dots + \beta_p X_p)\}} \quad (3)$$

Parameter diestimasi menggunakan *maximum likelihood estimation*. Model ini dipilih karena transparansi interpretasi koefisien dan kemampuan memberikan estimasi probabilitas

2. K-Nearest Neighbors (KNN)

KNN merupakan algoritma non-parametrik yang mengklasifikasikan observasi berdasarkan similaritas dengan $K=10$ tetangga terdekat dalam ruang fitur. Algoritma ini menggunakan jarak Euclidean setelah standardisasi data:

$$\hat{y}(x) = \operatorname{argmax}_c \sum_{\{x_j \in N^{10}(x)\}} 1_{\{y_j = c\}} \quad (4)$$

Dengan *kernel rectangular*, setiap tetangga memiliki bobot yang sama dalam proses voting. Pemilihan $K=10$ didasarkan pada keseimbangan antara bias (K terlalu kecil) dan variance (K terlalu besar). KNN memiliki keunggulan dalam menangkap pola lokal dan tidak membuat asumsi tentang distribusi data, namun sensitif terhadap *curse of dimensionality* dan memerlukan komputasi yang intensif pada fase prediksi [18].

3. Decision Tree

Decision Tree dibangun menggunakan algoritma CART dengan kriteria pemisahan Gini impurity [19].

$$\text{Gini}(t) = 1 - \sum_c p(c|t)^2 \quad (5)$$

Pemisahan optimal dipilih berdasarkan maksimalisasi *information gain*. *Pruning* dilakukan menggunakan *cost-complexity pruning* untuk mencegah *overfitting*.

4. Naive Bayes

Klasifikasi berdasarkan teorema Bayes dengan asumsi independensi kondisional antar fitur [20].

$$P(Y = c|X = x) \propto P(Y = c) \prod_{i=1}^p P(X_i = x_i|Y = c) \quad (6)$$

Untuk fitur numerik digunakan distribusi Gaussian, sedangkan fitur kategorikal menggunakan distribusi multinomial. *Laplace smoothing* ($\alpha=1$) diterapkan untuk mengatasi masalah *zero probability*.

5. Random Forest

Random Forest terdiri dari ensemble $B=500$ pohon keputusan yang dilatih pada sampel bootstrap. Pada setiap split, sejumlah $m_{try} = \lfloor \sqrt{p} \rfloor$ fitur dipilih secara acak sebagai kandidat pemisahan. Prediksi akhir diperoleh melalui *majority voting* dari seluruh pohon dalam *ensemble* [21].

G. Penanganan Ketidakseimbangan Kelas

Untuk mengatasi masalah ketidakseimbangan kelas, diterapkan strategi pembobotan kelas (*class weighting*) dengan rasio target 8:1 sesuai dengan *imbalance ratio* data. *Class weighting* diterapkan pada regresi logistik dan random forest, sedangkan *class balancing* diterapkan pada *decision tree*, KNN, dan *Naive Bayes* [16].

Bobot kelas dihitung menggunakan formula:

$$w_i = \frac{n}{k \times n_i} \quad (7)$$

dimana n adalah total sampel, k adalah jumlah kelas, dan n_i adalah ukuran kelas ke- i .

H. Evaluasi Model

Evaluasi model menggunakan metrik yang sesuai untuk data tidak seimbang [19]:

1. **Recall (Sensitivity)**: mengukur proporsi kasus *churn* yang berhasil diidentifikasi

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

2. **Specificity**: mengukur proporsi kasus *non-churn* yang berhasil diidentifikasi dengan benar

$$Specificity = \frac{TN}{TN + FP} \quad (9)$$

3. **F1-Score**: harmonic mean antara precision dan recall

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (10)$$

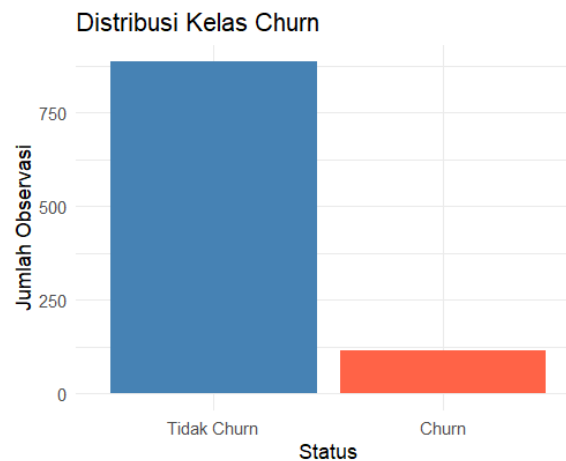
4. **AUC-ROC**: mengukur kemampuan model dalam membedakan kelas di berbagai *threshold*

I. Perbandingan Model

Performa model dibandingkan sebelum dan sesudah penanganan ketidakseimbangan kelas menggunakan seluruh metrik evaluasi. Analisis dilakukan pada data testing untuk memberikan estimasi performa yang tidak bias. Model terbaik dipilih berdasarkan keseimbangan antara recall, specificity, dan F1-score dengan mempertimbangkan konteks bisnis dimana identifikasi customer *churn* (recall) menjadi prioritas utama [22].

III. Hasil dan Pembahasan

Distribusi kelas sangat memengaruhi hasil pemodelan klasifikasi. Oleh karena itu, sebelum melakukan pemodelan, terlebih dahulu dilakukan eksplorasi data. *Column chart* pada Gambar 1 menunjukkan distribusi status *churn* pelanggan.



Gambar 1. Distribusi status pelanggan

Gambar 1 menunjukkan bahwa jumlah pelanggan yang tidak tetap menggunakan layanan jauh lebih banyak dibandingkan dengan pelanggan yang *berhenti* menggunakan layanan. Hal ini mengindikasikan adanya ketidakseimbangan data yang parah sebab hanya sekitar 11,4% pelanggan yang berstatus *churn*.

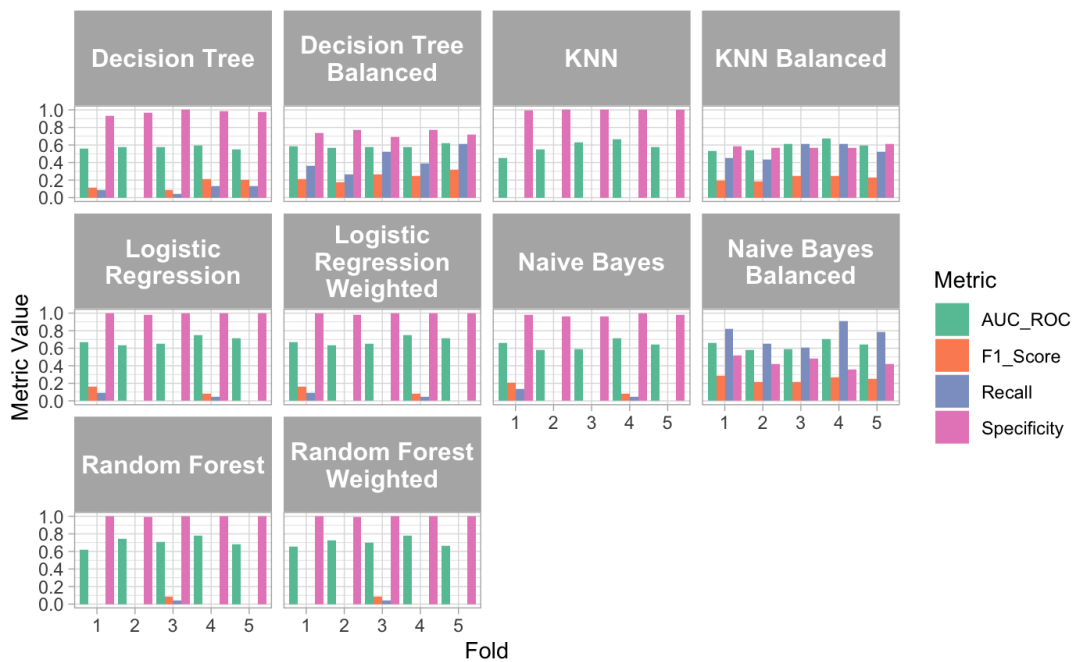
Untuk mengatasi ketidakseimbangan data, kami menggunakan strategi pembobotan kelas (*class weighting*), yaitu dengan memberikan bobot yang lebih tinggi pada kelas minoritas (*churn*) agar model lebih memperhatikan dan belajar mengenali pola-pola yang terkait dengan pelanggan yang berpotensi berhenti menggunakan layanan. Bobot yang digunakan adalah 8:1 berdasarkan perbandingan rasio yang telah dijabarkan pada bagian 3.2. Penggunaan bobot ini diharapkan dapat menghasilkan model klasifikasi yang lebih seimbang, baik dari segi sensitivitas terhadap kelas minoritas maupun keseimbangan antar metrik evaluasi seperti *recall* dan F1-score.

Adapun proses evaluasi dilakukan dengan menggunakan *stratified 5-fold cross validation*, sebagaimana dijelaskan pada bagian metode penelitian. Hal ini untuk memastikan kestabilan model yang digunakan. Tabel berikut ini menunjukkan jumlah responden pada tiap *fold*.

Tabel 1. Jumlah responden per fold

Fold	Status	Train Data	Test Data
1	Tidak <i>churn</i>	709	177
	<i>Churn</i>	92	22
2	Tidak <i>churn</i>	708	178
	<i>Churn</i>	91	23
3	Tidak <i>churn</i>	709	177
	<i>Churn</i>	91	23
4	Tidak <i>churn</i>	709	177
	<i>Churn</i>	91	23
5	Tidak <i>churn</i>	709	177
	<i>Churn</i>	91	23

Pemodelan kemudian dilakukan pada kelima data set pada tabel tersebut menggunakan lima model *machine learning*, masing-masing dengan dua varian: terbobot dan tidak terbobot, sehingga diperoleh sepuluh model *machine learning*. Model optimal kemudian diuji pada data *test*. Gambar berikut menunjukkan performa kesepuluh model.



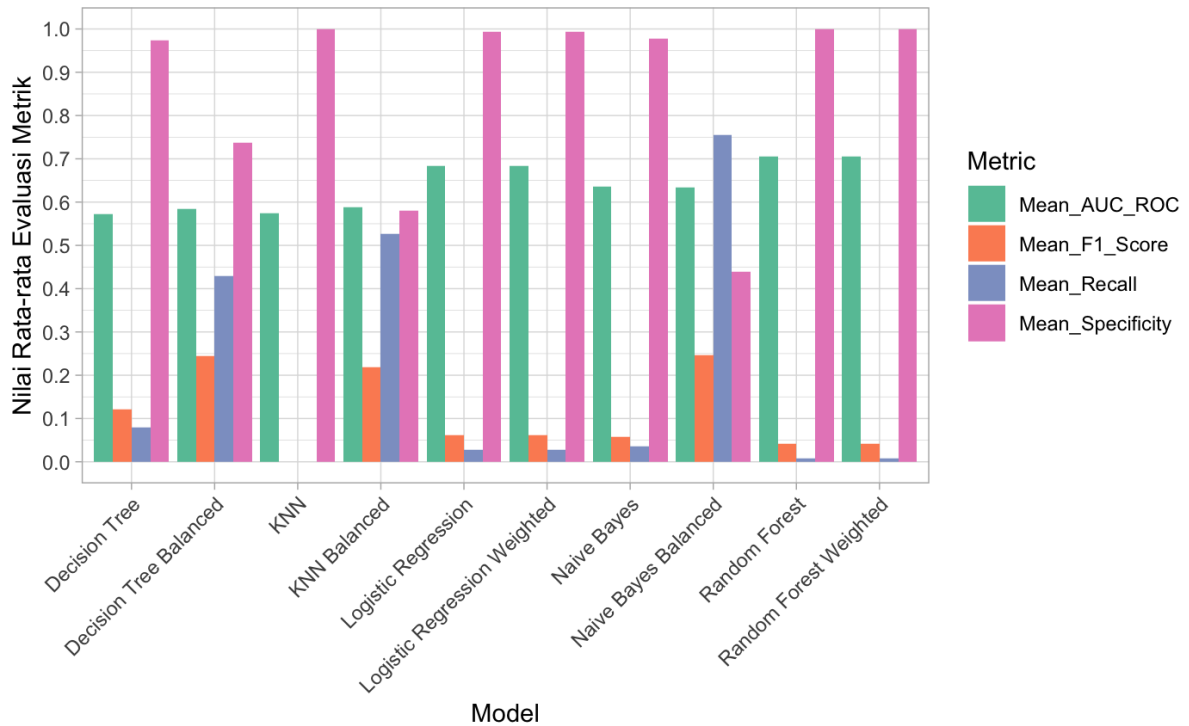
Gambar 2. Evaluasi metrik setiap model (fold=5)

Gambar 2 menampilkan empat metrik yang mengilustrasikan performa kesepuluh model. Salah satu metrik yang paling penting adalah *recall*, dimana metrik ini menunjukkan proporsi kelas positif yang dideteksi dengan benar. Dalam hal ini, nilai *recall* yang tinggi menunjukkan bahwa model yang dievaluasi mampu mendeteksi pelanggan yang *churn* dengan akurat. Gambar 2 menunjukkan peningkatan *recall* setelah penggunaan pembobotan, khususnya pada model *decision tree*, KNN, dan *naive bayes*. Adapun model *logistic regression* dan *random forest* tidak menunjukkan peningkatan *recall* meskipun telah menggunakan pembobotan.

Selain *recall*, metrik *specificity* digunakan untuk mengevaluasi klasifikasi kelas non-*churn*. Metrik ini menunjukkan nilai yang relatif tinggi sebab proporsi kelas non-*churn* merupakan kelas mayoritas. Adapun *F1-score* dihitung dari nilai *recall* dan *specificity*, sehingga peningkatan nilai *recall* pada model *decision tree balanced*, KNN *balanced*, dan *naive bayes balanced* menyebabkan peningkatan *F1-score* pada model-model tersebut. Adapun metrik ROC-AUC mengindikasikan hasil yang kontradiktif dengan pola yang cenderung tinggi dan tidak berubah diantara kesepuluh model. Meskipun begitu, perlu dipertimbangkan bahwa hasil ROC-

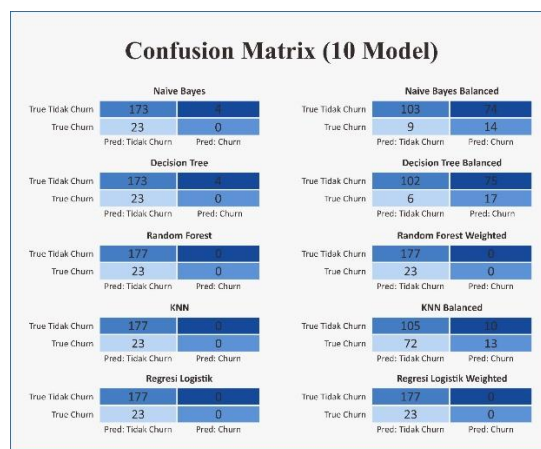
AUC memang cenderung tinggi pada kasus dengan data tak seimbang, sehingga skor ROC-AUC tidak dapat diandalkan sepenuhnya dalam kondisi ini.

Secara umum, penerapan teknik penyeimbangan data terbukti memberikan dampak positif terhadap performa model secara keseluruhan. *Specificity* memang sedikit menurun, namun ini adalah *trade-off* yang umum terjadi ketika model mulai lebih sensitif terhadap kelas yang sebelumnya jarang terdeteksi. Keseimbangan antara metrik menjadi lebih baik, mengindikasikan bahwa model tidak lagi bias terhadap satu kelas saja.



Gambar 3. Barplot Rata-rata 5-fold evaluasi metrik setiap model berdasarkan

Gambar 3 menunjukkan rata-rata keempat metrik dari sepuluh model yang diuji, mengilustrasikan ringkasan hasil dari Gambar 2. Berdasarkan Gambar 3, di antara semua model yang diuji, *Decision Tree Balanced* dan *Naive Bayes Balanced* menunjukkan hasil terbaik diantara kelima *fold*. Dalam kasus ini, model *Naive Bayes Balanced* lebih direkomendasikan karena memiliki nilai *recall* yang lebih baik, dengan skor *recall* lebih dari 0,75. Nilai *recall* yang lebih baik memungkinkan model *Naive Bayes Balanced* untuk mendeteksi pelanggan *churn* dengan lebih baik.



Gambar 4. Confusion matrix dari setiap model

Gambar 4 menunjukkan hasil confusion matrix dari semua model. Berdasarkan gambar tersebut menunjukkan efek balancing atau class weighting berbeda-beda antar model. Naive Bayes memperoleh peningkatan signifikan pada kemampuan mendeteksi pelanggan churn karena model ini sangat bergantung pada proporsi kelas, sedangkan Decision Tree hanya mengalami perubahan kecil dan cenderung menambah false positive tanpa meningkatkan recall secara berarti. Hasil dari KNN memburuk setelah balancing karena metode ini sangat sensitif terhadap distribusi data.

Random Forest dan Regresi Logistik tidak menunjukkan perbaikan recall karena sejak awal sudah mampu menangkap seluruh kasus churn. Pada logistic regression, keterbatasan model yang hanya membentuk decision boundary linear membuat weighting tidak mampu mengatasi kompleksitas pemisahan kelas pada data non-linier. Bobot tambahan hanya menggeser boundary secara marginal sehingga tidak cukup untuk meningkatkan sensitivitas terhadap kelas churn. Sementara itu, random forest secara alami cukup robust terhadap ketidakseimbangan kelas karena mekanisme bagging dan ensemble-nya. Struktur pohon keputusan pada masing-masing tree tetap didominasi oleh distribusi kelas mayoritas dalam bootstrap sample, sehingga perubahan impurity akibat weighting tidak banyak mempengaruhi pola split. Akibatnya, penyesuaian bobot tidak menghasilkan peningkatan recall yang bermakna pada model ini.

Secara keseluruhan, balancing paling efektif pada model berbasis probabilistik sederhana. Dengan demikian, hasil ini menggarisbawahi pentingnya penanganan ketidakseimbangan data dalam proses pelatihan model klasifikasi. Tanpa penyeimbangan, model akan bias terhadap kelas mayoritas dan mengabaikan kelas minoritas, yang dalam konteks dunia nyata bisa berarti melewatkan kasus-kasus penting. Oleh karena itu, penggunaan teknik penyeimbangan seperti *weighting* atau *resampling* sangat direkomendasikan dalam situasi serupa untuk menghasilkan model yang lebih stabil.

IV. Kesimpulan dan saran

Penelitian ini menegaskan bahwa ketidakseimbangan kelas (rasio 8:1) sangat memengaruhi performa model prediksi churn. Penerapan class weighting terbukti meningkatkan sensitivitas model terhadap kelas minoritas, khususnya pada Decision Tree Balanced dan Naive Bayes Balanced. Model Naive Bayes Balanced menunjukkan performa terbaik dengan recall di atas 0,75, meskipun specificity sedikit menurun sebagai trade-off yang masih dapat diterima. Secara manajerial, hasil ini memberikan dasar bagi perusahaan telekomunikasi untuk lebih proaktif dalam strategi retensi, misalnya dengan memberikan insentif, personalisasi layanan, atau kampanye targeted kepada pelanggan yang berisiko churn. Dengan pendekatan ini, perusahaan dapat mengurangi kehilangan pelanggan, menekan biaya akuisisi baru, serta meningkatkan profitabilitas jangka panjang. Meskipun demikian, penelitian ini memiliki keterbatasan karena hanya menggunakan satu teknik penanganan imbalance, yaitu class weighting, serta memanfaatkan satu dataset telekomunikasi tahun 2019 dengan fitur yang terbatas sehingga generalisasi hasil masih perlu diuji. Oleh karena itu, penelitian selanjutnya perlu membandingkan weighting dengan metode imbalance lainnya seperti SMOTE, ADASYN, atau teknik hybrid berbasis GAN, serta memperluas cakupan data dengan menambah fitur perilaku pelanggan dan melakukan validasi temporal maupun lintas wilayah untuk mendapatkan model prediksi churn yang lebih robust dan aplikatif.

Daftar Pustaka

- [1] Rofik, J. Unjung, dan B. Prasetyo, "Enhancing customer churn prediction with stacking ensemble and stratified k-fold," *Bull. Electr. Eng. Informatics*, vol. 14, no. 1, hal. 398–408, 2025, doi: <https://doi.org/10.11591/eei.v14i1.8112>.
- [2] P. Boozary, S. Sheykhan, H. GhorbanTanhaei, dan C. Magazzino, "Enhancing customer retention with machine learning: A comparative analysis of ensemble models for accurate churn prediction," *Int. J. Inf. Manag. Data Insights*, vol. 5, no. 1, hal. 100331, 2025, doi: <https://doi.org/10.1016/j.jjime.2025.100331>.
- [3] I. N. M. Adiputra, P. Wanchai, dan P. C. Lin, "Optimized customer churn prediction using tabular generative adversarial network (GAN)-based hybrid sampling method and cost-sensitive learning," *PeerJ Comput. Sci.*, vol. 11, hal. 1–29, 2025, doi: <https://doi.org/10.7717/peerj-cs.2949>.
- [4] H. T. T. Binh dan Y.-K. Kwon, "An Effective SMOTE-Based oversampling technique for class imbalance in software defect prediction," *IEEE Access*, vol. 2, no. 8, hal. 24379–24392, 2020, doi: <https://doi.org/10.1109/ACCESS.2020.2970401>.
- [5] R. Suguna, J. Suriya Prakash, H. Aditya Pai, T. R. Mahesh, V. Vinoth Kumar, dan T. E. Yimer, "Mitigating class imbalance in churn prediction with ensemble methods and SMOTE," *Sci. Rep.*, vol. 15, no. 1, hal. 1–21, 2025, doi: [10.1038/s41598-025-01031-0](https://doi.org/10.1038/s41598-025-01031-0).

- [6] I. N. M. Adiputra dan P. Wanchai, "CTGAN-ENN: a tabular GAN-based hybrid sampling method for imbalanced and overlapped data in customer churn prediction," *J. Big Data*, vol. 11, no. 1, 2024, doi: <https://doi.org/10.1186/s40537-024-00982-x>.
- [7] A. C. Bahnsen, D. Aouada, dan B. Ottersten, "A novel cost-sensitive framework for customer churn predictive modeling," *Decis. Anal.*, vol. 2, no. 1, 2015, doi: <https://doi.org/10.1186/s40165-015-0014-6>.
- [8] D. J. Benkendorf, S. D. Schwartz, D. R. Cutler, dan C. P. Hawkins, "Correcting for the effects of class imbalance improves the performance of machine-learning based species distribution models," *Ecol. Modell.*, vol. 483, hal. 110414, 2023, doi: <https://doi.org/10.1016/j.ecolmodel.2023.110414>.
- [9] X. Liu, L. Guo, dan Y. Guo, "Cost-sensitive learning for imbalanced classification.," *Pattern Recognit.*, vol. 105, hal. 107298, 2019, doi: <https://doi.org/10.1016/j.patcog.2020.107298>.
- [10] J. Wang dan H. Yao, "A Comparative study of ensemble learning techniques for imbalanced classification," *IEEE Access*, vol. 8, hal. 111763–111780, 2020, doi: <https://doi.org/10.1109/ACCESS.2020.3003177>.
- [11] A. Sharma dan A. Bhardwaj, "A Review on ensemble models for class imbalance in churn prediction," *J. King Saud Univ. – Comput. Inf. Sci.*, vol. 35, no. 10, hal. 101806, 2023, doi: <https://doi.org/10.1016/j.jksuci.2023.101806>.
- [12] H. Wickham *et al.*, "Welcome to the tidyverse," *J. open source Softw.*, vol. 4, no. 43, hal. 1686, 2019, doi: [10.21105/joss.01686](https://doi.org/10.21105/joss.01686).
- [13] M. Binder, F. Pfisterer, M. Lang, L. Schneider, L. Kotthof, dan B. Bischl, "Mlr3Pipelines - Flexible machine learning Pipelines in R," *J. Mach. Learn. Res.*, vol. 22, hal. 1–7, 2021, [Daring]. Tersedia pada: <http://jmlr.org/papers/v22/21-0281.html>
- [14] M. Steininger, K. Kobs, P. Davidson, A. Krause, dan A. Hotho, "Density-based weighting for imbalanced regression," *Mach. Learn.*, vol. 110, no. 8, hal. 2187–2211, 2021, doi: <https://doi.org/10.1007/s10994-021-06023-5>.
- [15] A. Harismahyanti, Indahwati, A. Fitrianto, dan Erfiani, "Outlier detection on high dimensional data using minimum vector variance (Mvv)," *Barekeng*, vol. 16, no. 3, hal. 797–804, 2022, doi: <https://doi.org/10.30598/barekengvol16iss3pp797-804>.
- [16] H. He dan E. A. Garcia, "Learning from Imbalanced Data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, hal. 1263–1284, 2009, doi: <https://doi.org/10.1109/TKDE.2008.239>.
- [17] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *IJCAI'95: Proceedings of the 14th international joint conference on Artificial intelligence*, 1995, hal. 1137–1143. doi: <https://dl.acm.org/doi/10.5555/1643031.1643047>.
- [18] D. W. Hosmer, S. Lemeshow, dan R. X. Sturdivant, *Applied Logistic Regression*, Third Edit. Canada: John Wiley & Sons, Inc., 2013. doi: <https://doi.org/10.1002/9781118548387>.
- [19] L. Breiman, J. Friedman, R. A. Olshen, dan C. J. Stone, *Classification and Regression Trees*, 1st Editio. New York: Chapman and Hall/CRC, 2017. doi: <https://doi.org/10.1201/9781315139470>.
- [20] H. Zhang, "The Optimality of Naive Bayes," in *Proceedings of the seventeenth international florida artificial intelligence research society conference (FLAIRS 2004)*, California: The AAAI Press, 2004, hal. 1–6.
- [21] L. Breiman, *Random Forests*. Netherlands: Kluwer Academic Publishers, 2001. doi: <https://doi.org/10.1023/A:1010933404324>.
- [22] M. Sokolova dan G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, hal. 427–437, 2009, doi: <https://doi.org/10.1016/j.ipm.2009.03.002>.