Penerapan *Ensemble Learning* dengan *Hard Voting* untuk Klasifikasi *Customer Churn*

Andhika Rama Putra Astawa, Galih Hendro Martono, Mayadi

Universitas Bumigora, Mataram, Indonesia

Correspondence : *e-mail*: andhikaastawac@gmail.com

Abstrak

Customer churn menjadi salah satu tantangan terbesar bagi perusahaan telekomunikasi karena berdampak langsung pada pendapatan dan keberlanjutan bisnis. Penelitian ini bertujuan untuk meningkatkan akurasi prediksi churn dengan mengembangkan model ensemble learning berbasis Hard Voting Classifier yang menggabungkan tiga algoritma berbeda, yaitu Naïve Bayes, Random Forest, dan Nearest Centroid. Dataset pelanggan yang digunakan mencakup informasi demografis, perilaku penggunaan layanan, dan status churn, yang kemudian diproses melalui tahapan pembersihan data, seleksi fitur, normalisasi, serta teknik resampling SMOTE-Tomek untuk menyeimbangkan distribusi kelas. Pemilihan fitur dilakukan dengan metode Information Gain dan analisis korelasi, sehingga hanya atribut yang relevan digunakan dalam pemodelan. Hasil pengujian menunjukkan bahwa Hard Voting Classifier mampu mencapai akurasi sebesar 90% dengan nilai recall untuk kelas churn sebesar 81%, lebih tinggi dibandingkan Random Forest (78%), meskipun akurasi Random Forest lebih tinggi (95%). Nilai precision untuk kelas non-churn juga meningkat hingga 97%, menandakan model ini efektif mengurangi kesalahan dalam memprediksi pelanggan tetap. Temuan ini membuktikan bahwa pendekatan ensemble learning dengan base learner heterogen dapat memadukan keunggulan masing-masing algoritma untuk meningkatkan deteksi churn. Meski demikian, performa Hard Voting masih bergantung pada kualitas masing-masing classifier, sehingga optimasi hyperparameter dan eksplorasi kombinasi model lain direkomendasikan untuk penelitian selanjutnya. Hasil penelitian ini diharapkan dapat membantu perusahaan merumuskan strategi retensi pelanggan yang lebih tepat sasaran dan berkelanjutan.

Kata kunci: *customer churn, ensemble learning, hard voting classifier, random forest, SMOTE-Tomek.*

Abstract

Customer churn remains one of the most critical challenges for telecommunication companies as it directly affects revenue and business sustainability. This study aims to enhance churn prediction accuracy by developing an ensemble learning model based on a Hard Voting Classifier that combines three distinct algorithms: Naïve Bayes, Random Forest, and Nearest Centroid. The customer dataset used includes demographic information, service usage behavior, and churn status, which were processed through data cleaning, feature selection, normalization, and a SMOTE-Tomek hybrid resampling technique to address class imbalance. Feature selection was performed using the Information Gain method and correlation analysis to ensure that only relevant attributes were included in the modeling process. Experimental results show that the Hard Voting Classifier achieved an accuracy of 90% with a recall rate of 81% for the churn class, outperforming Random Forest's recall of 78%, although Random Forest maintained a higher overall accuracy (95%). The precision for the non-churn class also reached 97%, indicating that the model effectively minimizes errors in identifying loyal customers. These findings demonstrate that an ensemble learning approach with heterogeneous base learners can successfully leverage the strengths of each algorithm to improve churn detection performance. Nevertheless, the model's performance still depends on the quality of its base classifiers, suggesting that future research should focus on hyperparameter optimization and exploring alternative classifier combinations. The outcomes of this study are expected to support companies in designing more targeted and sustainable customer retention strategies.

Keywords *customer churn, ensemble learning, hard voting classifier, random forest, SMOTE-Tomek.*

1. Pendahuluan

Perkembangan pesat di sektor telekomunikasi menjadikan industri ini salah satu pilar penting dalam mendukung konektivitas masyarakat modern. Namun, di balik pertumbuhan tersebut, tantangan terkait loyalitas pelanggan kian kompleks. Fenomena *customer churn*, yakni keputusan pelanggan untuk berhenti berlangganan dan beralih ke penyedia layanan lain, menjadi persoalan yang tak terelakkan. Studi terbaru [1] menunjukkan bahwa industri telekomunikasi global menghadapi rata-rata *churn rate* sebesar 1,9% per bulan untuk layanan pascabayar dan mencapai 67% per tahun untuk layanan prabayar. Bagi perusahaan telekomunikasi, kehilangan pelanggan lama berarti harus menanggung biaya akuisisi pelanggan baru yang relatif lebih tinggi dibandingkan biaya mempertahankan pelanggan yang sudah ada. Fenomena ini mendorong perlunya pengembangan sistem prediksi *churn* yang lebih akurat dan handal [2]. Dalam upaya menekan angka *churn*, perusahaan dituntut untuk mampu mendeteksi kemungkinan pelanggan akan berhenti sedini mungkin. Di sinilah teknologi analitik prediktif, khususnya teknik klasifikasi, memegang peranan strategis. Berbagai pendekatan telah dikembangkan untuk memprediksi *churn*, mulai dari metode statistik konvensional hingga algoritma *machine learning* yang lebih adaptif. Meskipun demikian, penggunaan model tunggal seringkali belum cukup mampu menangkap pola perilaku pelanggan yang semakin dinamis.

Salah satu pendekatan yang belakangan banyak diterapkan untuk meningkatkan akurasi prediksi adalah *ensemble learning*. Teknik ini memanfaatkan keunggulan beberapa model dasar yang digabungkan, sehingga kelemahan satu model dapat ditutupi oleh keunggulan model lainnya. Beragam penelitian terdahulu telah merumuskan dan menerapkan berbagai metode *machine learning* untuk memetakan pola perilaku pelanggan dan memprediksi potensi *customer churn*. Berbagai algoritma, mulai dari model klasifikasi sederhana hingga pendekatan komputasi yang lebih kompleks, telah diuji untuk memperoleh prediksi yang semakin akurat dan dapat diandalkan seperti pada penelitian [3] meggunakan Naïve Bayes dengan pendekatan pembobotan fitur menggunakan Algoritma Genetika (subkelas Algoritma Evolusioner) dievaluasi pada dataset publik (BigML Telco churn, IBM Telco, dan Cell2Cell). Hasil menunjukkan rata-rata presisi masing-masing 0,97, 0,97, dan 0,98; recall 0,84, 0,94, dan 0,97; F1-score 0,89, 0,96, dan 0,97; MCC 0,89, 0,96, dan 0,97; serta akurasi 0,95, 0,97, dan 0,98. Penggunaan model *ensemble learning*, salah satunya teknik *voting classifier,* telah banyak dibuktikan mampu memberikan performa prediksi yang lebih konsisten dan akurat dibandingkan pendekatan yang hanya mengandalkan satu model saja. Dengan menggabungkan beberapa algoritma pembelajaran, *voting classifier* dapat menyeimbangkan kelemahan masing-masing model dasar sehingga hasil prediksi menjadi lebih andal [4]. Contohnya, penelitian, [5] berhasil merancang kombinasi antara *Random Forest, AdaBoost*, dan SVM dengan capaian akurasi 88,7%. Namun demikian, metode tersebut masih memiliki keterbatasan dalam mengolah data dengan distribusi yang timpang tanpa dukungan SMOTE. Di sisi lain [6] mengimplementasikan algoritma C4.5 dengan optimasi *hyperparameter* sehingga mampu memperoleh akurasi hingga 80,05%, tetapi proses komputasinya terbilang rumit. Pendekatan *hybrid* yang memadukan SMOTE dan *Ensemble* dilakukan pada penelitian [7] dengan meningkatkan nilai recall hingga 93%, meskipun masih terbatas pada pemanfaatan model *homogen* seperti *Random Forest* dan *Adaboost*. Sementara itu, studi [8] menunjukkan bahwa penerapan *Soft Voting Ensemble* mampu mengungguli performa model individual. Penelitian lain yang dilakukan oleh [9], [10] juga memperlihatkan bahwa *Naïve Bayes* dapat memperoleh tingkat akurasi mencapai 98,6%, namun sejauh ini belum dipadukan dengan pendekatan berbasis *centroid* seperti *Nearest Centroid*. Di sisi lain, penelitian [11], [12] menegaskan pentingnya penggunaan teknik *resampling* seperti SMOTE untuk menangani data di sektor perbankan, sedangkan [13] juga menggarisbawahi kelemahan model tunggal yang masih rentan terhadap gangguan *noise*. Penelitian [14] membuktikan bahwa Voting Classifier dengan basis model heterogen mampu mencapai akurasi sekitar 85%, meskipun kombinasi spesifik antara *Naïve Bayes, Random Forest*, dan *Nearest Centroid* belum pernah diuji sebelumnya. Sementara itu, studi terbaru [15], [16] berhasil mengidentifikasi celah riset terkait optimasi kombinasi classifier untuk data pelanggan telekomunikasi, sedangkan [17], [18] mendorong perlunya pengkajian lebih lanjut terhadap integrasi model berbasis deret waktu ke dalam skema ensemble.

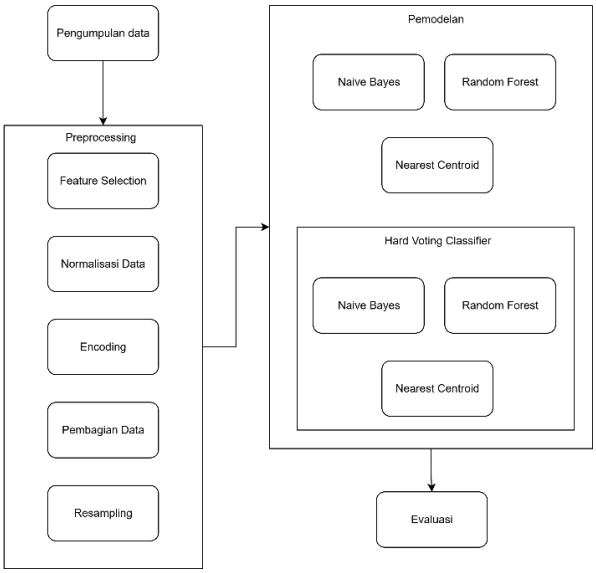
Berdasarkan berbagai temuan tersebut, dapat disimpulkan bahwa tren penelitian terkait prediksi *customer churn* semakin diarahkan pada pemanfaatan teknik *ensemble learning* serta strategi penanganan data yang tidak seimbang. Kendati demikian, variasi kombinasi model dasar yang diterapkan masih didominasi oleh algoritma yang seragam, padahal karakteristik data *churn* kerap kali bersifat tidak seimbang dan rawan terpengaruh *noise* sehingga belum sepenuhnya terakomodasi secara optimal. Sebagian besar studi juga cenderung hanya menitikberatkan pada pengujian *classifier* yang umum digunakan, sehingga potensi pemanfaatan model yang kurang populer seperti *Nearest Centroid* masih jarang diangkat. Beberapa penelitian sebelumnya, seperti yang dilakukan oleh [6], [9] pun belum menguji penggunaan *Hard*

Voting Classifier yang secara simultan memadukan *Naïve Bayes, Random Forest*, dan *Nearest Centroid*. Di samping itu, teknik *resampling* pada [7] masih terbatas pada *oversampling* konvensional tanpa integrasi pendekatan lain seperti Tomek Links, yang sebenarnya dapat membantu memperbaiki kualitas data. Selain itu, evaluasi performa model pada[14] dinilai belum komprehensif dalam membandingkan *trade-off* antara akurasi dan *recall churn*, terutama dalam konteks data telekomunikasi. Oleh karena itu, penelitian ini menawarkan kontribusi baru dengan merancang skema *ensemble learning* berbasis *Hard Voting* yang menggabungkan tiga algoritma berbeda, sehingga diharapkan dapat memaksimalkan kelebihan masing-masing model, meningkatkan akurasi prediksi *churn*, serta mendukung perumusan strategi retensi pelanggan yang lebih tepat sasaran dan berkelanjutan di industri telekomunikasi.

Untuk menjawab berbagai keterbatasan tersebut, penelitian ini merumuskan sebuah *framework* prediksi churn yang dirancang secara inovatif dengan tiga kontribusi utama. Pertama, penelitian ini menghadirkan penerapan perdana *Hard Voting Classifier* yang secara simultan memadukan *Naïve Bayes* (untuk menjaga efisiensi komputasi), *Random Forest* (untuk memperoleh akurasi yang lebih tinggi), dan *Nearest Centroid* (untuk mengenali pola *non-linear*) sebagaimana diilustrasikan dalam temuan [13], [19]. Kedua, penelitian ini menerapkan SMOTE-Tomek sebagai strategi *hybrid* guna menanggulangi permasalahan ketidakseimbangan kelas, sekaligus meminimalkan *noise* pada *decision boundary*, dengan memperluas gagasan yang telah dibahas pada [7], [13]. Ketiga, dilakukan optimasi seleksi fitur menggunakan pendekatan *Information Gain* dan korelasi, merujuk pada rekomendasi [6], [8] untuk memastikan hanya fitur yang paling relevan dan informatif yang digunakan dalam proses klasifikasi. Pendekatan terintegrasi ini diharapkan mampu menjaga keseimbangan antara akurasi prediksi dan tingkat *recall churn* secara lebih optimal.

2. Metode Penelitian

Metode penelitian ini diawali dengan pengumpulan dataset pelanggan dari perusahaan telekomunikasi yang diperoleh dari Kaggle. Data yang diperoleh kemudian melalui tahap *preprocessing* yang mencakup tahapan pertama yaitu *feature selection* dengan analisis korelasi yang bertujuan mempertahankan atribut paling relevan bagi proses prediksi. Kedua, normalisasi data bertujuan untuk menyamakan skala data numerik. *Encoding* untuk mengubah fitur dengan data kategori menjadi data numerik. Lalu pembagian data *training* dan *testing* dengan rasio 80:20. Serta *resampling hybrid* SMOTE-Tomek untuk mengatasi ketidakseimbangan distribusi kelas sekaligus mengurangi *noise* pada area *decision boundary*. Setelah data siap, model dikembangkan dengan membangun tiga algoritma pembelajaran dasar, yakni Naïve Bayes, Random Forest, dan Nearest Centroid, yang kemudian digabungkan melalui skema *Hard Voting Classifier* untuk memperoleh prediksi akhir berdasarkan suara mayoritas. Evaluasi kinerja dilakukan dengan membandingkan metrik akurasi, *precision, recall, f1-score*, serta analisis *confusion matrix*, sehingga validitas hasil dapat terukur secara menyeluruh dan dapat dijadikan dasar rekomendasi strategi retensi pelanggan yang lebih efektif. Tahapan enelitian yang digunakan dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Metodologi Penelitian

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari platform Kaggle dengan judul "*Telechurn*" yang dipublikasikan oleh CrowdAnalytix [20]. Dataset ini terdiri dari 3.333 sampel pelanggan dan 20 fitur, termasuk informasi seperti riwayat transaksi, penggunaan layanan, dan interaksi pelanggan. Fitur target adalah status *churn* (berhenti berlangganan atau tidak). Data dikumpulkan dalam format CSV dan diimpor ke lingkungan Python menggunakan *library* Pandas untuk memudahkan eksplorasi dan analisis lebih lanjut. Pemilihan dataset ini didasarkan pada kelengkapan atribut dan relevansinya dengan masalah *customer churn* di industri telekomunikasi. Penjelasan fitur-fitur dalam dataset yang digunakan dapat dilihat pada Tabel 1.

Tabel 1. Penjelasan Fitur dalam Dataset

Nama Variabel	Tipe Data	Keterangan
State	String	Lokasi geografis pelanggan
Account length	Integer	Lama waktu dalam hari pelanggan telah berlangganan
Area code	Integer	Kode area telepon pelanggan
International plan	String	Indikator apakah pelanggan memiliki paket layanan internasional
Voice mail plan	String	Indikator apakah pelanggan menggunakan layanan voice mail
Number vmail messages	Integer	Jumlah pesan voice mail yang diterima pelanggan
Total day minutes	Double	Total menit panggilan yang dilakukan pelanggan pada jam pagi/siang hari
Total day calls	Integer	Jumlah panggilan yang dilakukan pelanggan pada jam pagi/siang hari
Total day charge	Double	Biaya total panggilan pada jam pagi/siang hari
Total eve minutes	Double	Total menit panggilan pada jam sore/malam hari
Total eve calls	Integer	Jumlah panggilan pada jam sore/malam hari
Total eve charge	Double	Biaya total panggilan pada jam sore/malam har
Total night minutes	Double	Total menit panggilan pada jam malam hari
Total night calls	Integer	Jumlah panggilan pada jam malam hari
Total night charge	Double	Biaya total panggilan pada jam malam hari
Total intl minutes	Double	Total menit panggilan internasional yang dilakukan pelanggan
Total intl calls	Integer	Jumlah panggilan internasional yang dilakukan pelanggan
Total intl charge	Double	Biaya total panggilan internasional
Customer service calls	Integer	Jumlah panggilan yang dilakukan pelanggan ke layanan pelanggan, yang bisa menjadi 209ndicator masalah atau ketidakpuasan
Churn	Boolean	Variabel biner yang menunjukkan apakah pelanggan berhenti menggunakan layanan (1 = churned/berhenti, 0 = tidak)

2.2. Preprocessing

Tahap *preprocessing* merupakan tahapan persiapan data sebelum pemodelan untuk meningkatkan kualitas dan kinerja algoritma *machine learning* [21], berikut merupakan tahapan *preprocessing* yang digunakan dalam penelitian ini.

1. *Feature Selection* adalah pemilihan subset fitur dari *dataset* paling relevan untuk pelatihan model. Salah satu cara *Feature Selection* adalah penggunaan *p-value* untuk mengukur relevansi dari fitur terhadap model selain itu pula fitur dapat dipilih berdasarkan redundansinya terhadap fitur lainnya [22].
2. Normalisasi Data merupakan teknik standar dalam *preprocessing* data yang bertujuan untuk menyamakan skala nilai pada fitur-fitur. Proses ini penting karena banyak algoritma *machine learning* yang bergantung pada skala fitur-fitur seragam, sehingga fitur dengan nilai lebih besar secara signifikan dapat mendominasi perhitungan [23]. Metode normalisasi data yang digunakan dalam penelitian ini adalah *Robust Scaler*, yang melakukan transformasi data berdasarkan statistik kuartil alih-alih *mean* dan standar deviasi sehingga dapat memitigasi dampak negatif dari *outlier* dalam dataset, berikut merupakan rumus dari *robust scaler* yang digunakan pada persamaan 1 [24].

$$X_{norm} = \frac{X - Q_1(X)}{Q_3(X) - Q_1(X)} \quad (1)$$

Persamaan 1 digunakan untuk menghitung hasil normalisasi dengan X_{norm} sebagai hasil normalisasi x sebagai data sebelum normalisasi, $Q_1(X)$ dan $Q_3(X)$ berturut-turut sebagai quartil pertama dan quartil ketiga dari fitur yang akan dinormalisasikan.

3. *Encoding* merupakan proses transformasi fitur menjadi bentuk numerik agar dapat diproses dalam metode *machine learning* yang digunakan [25]. Contoh dari metode *encoding* yang umum digunakan merupakan *one-hot-encoding* dimana data dipisah menjadi kolom terpisah yang merepresentasikan tiap kategori dengan nilai biner, selain itu pula terdapat label *encoding* yang mengganti tiap nilai kategori pada fitur menjadi nilai numerik unik [26].
4. Pembagian Data adalah proses di mana *dataset* dibagi menjadi dua subset utama yakni *data training* digunakan untuk melatih model, dan juga *data testing* yang berfungsi sebagai simulasi data baru untuk mengevaluasi performa model [27]. Umumnya *dataset* dipisah dengan rasio data *training* berbanding data *testing* 70:30 atau 80:20 [28].
5. Resampling merupakan teknik *preprocessing* yang digunakan untuk menangani masalah ketidakseimbangan kelas [29]. Salah satu metode *resampling* yang efektif adalah kombinasi *SMOTE* sebagai *oversampling* dan *Tomek Links* sebagai *undersampling*, yang bekerja secara sinergis untuk memperbaiki distribusi kelas minoritas sekaligus membersihkan *noise* di perbatasan kelas [29], [30].

2.3. Pemodelan

Pada tahapan ini *data training* akan digunakan untuk melatih model *Hard voting classifier* yang terdiri dari Naïve Bayes, *Random Forest*, dan *Nearest Centroid*. Hasil prediksi akhir dari *Hard voting classifier* merupakan prediksi mayoritas dari ketiga model tersebut [31]. Pendekatan ini didasarkan pada prinsip "kebijaksanaan kerumunan" (*wisdom of the crowd*), di mana agregasi dari beberapa pendapat independen cenderung lebih baik daripada pendapat individu [32].

2.4. Evaluasi

Evaluasi model dilakukan menggunakan metrik-metrik berikut yakni, akurasi sebagai pengukur persentase prediksi benar secara keseluruhan, presisi menunjukkan ketepatan prediksi kelas positif, *recall* mengukur kemampuan model mendeteksi kelas positif, *f1-score* menampilkan rata-rata harmonik presisi dan *recall*. *Confusion matrix* digunakan untuk memvisualisasikan kinerja model. Evaluasi dilakukan pada data uji yang tidak terlibat dalam pelatihan untuk memastikan generalisasi model.

3. Hasil dan Pembahasan

Bagian ini menyajikan temuan utama penelitian beserta analisis mendalam terhadap hasil yang diperoleh. Data yang telah melalui tahap *preprocessing* dan pemodelan dianalisis menggunakan metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score* untuk mengukur kinerja model.

3.1. Data Penelitian

Dataset yang digunakan dilakukan analisis statistik deskriptif dan visualisasi distribusi fitur target sehingga menunjukkan ketidakseimbangan kelas (*class imbalance*) dengan 14,5% pelanggan *churn* (483 sampel) dan 85,5% *non-churn* (2.850 sampel). Pemeriksaan *missing value* mengonfirmasi tidak ada data yang hilang, sehingga tidak diperlukan imputasi.

3.2. Hasil Preprocessing

Proses *preprocessing* data dilakukan guna memastikan mutu data tetap terjaga sebelum tahap pemodelan dijalankan. Pemilihan fitur diawali dengan mengevaluasi korelasi antarvariabel serta signifikansi statistiknya terhadap target *churn*. Beberapa atribut, seperti *Total day charge* dan *Total eve charge*, dihapus karena menunjukkan korelasi sempurna (1,0) dengan variabel lain, yaitu *Total day minutes* dan *Total eve minutes*. Sementara itu, fitur *Account length* dan *Total night calls* dieliminasi lantaran nilai p-value di atas 0,5, yang menandakan pengaruhnya terhadap *churn* tidak signifikan. Tahap seleksi ini berhasil mereduksi jumlah fitur dari 20 menjadi 13 atribut utama. Setelah seleksi fitur, normalisasi variabel numerik diterapkan dengan *Robust Scaler* untuk meminimalkan dampak keberadaan outlier. Untuk data kategorikal, dua metode pengkodean digunakan: *label encoding* pada fitur biner seperti *International plan* (No=0, Yes=1) serta *one-hot encoding* pada fitur multinominal seperti *Area code* agar tidak muncul bias urutan. Usai transformasi, dataset dibagi menjadi data latih (80%) dan data uji (20%) dengan teknik *stratify* guna menjaga proporsi kelas tetap seimbang. Mengingat distribusi kelas yang tidak seimbang (483 data *churn* berbanding 2.850 data *non-churn*), teknik resampling *SMOTE-Tomek* diterapkan khusus pada data latih. Di tahap ini, SMOTE menghasilkan data sintetis untuk kelas minoritas, sedangkan *Tomek Links* membantu membersihkan data di area perbatasan kelas yang rawan

ambigu. Langkah ini membuat distribusi menjadi lebih proporsional, yakni sekitar 2.830 sampel non-churn dan 1.237 sampel *churn*, tanpa kehilangan informasi penting. Rangkaian *preprocessing* ini secara keseluruhan mendukung kesiapan data untuk dikembangkan lebih lanjut melalui model *ensemble learning*.

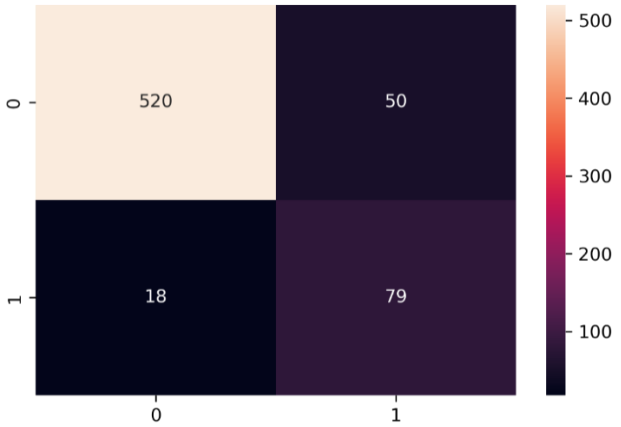
3.3. Implementasi dan Evaluasi

Setelah dataset melewati seluruh tahap *preprocessing*, proses pelatihan dilanjutkan dengan menguji model yang telah di-fit menggunakan data latih, kemudian dievaluasi menggunakan data uji. Kinerja model diukur melalui sejumlah metrik, seperti akurasi, *precision, recall, f1-score*, serta analisis *confusion matrix*. Ringkasan hasil evaluasi berdasarkan *classification report* disajikan pada Tabel 2.

Tabel 2. *Classification Report* dari *Hard Voting Classifier*

Kelas	Precision	Recall	F1-score	Support
0	0,97	0,91	0,94	570
1	0,63	0,81	0,71	97
Accuracy	-	-	0,90	667
Macro Avg	0,79	0,87	0,82	667
Weighted Avg	0,92	0,90	0,90	667

Hard Voting Classifier menunjukkan kinerja yang cukup seimbang dalam hal akurasi keseluruhan maupun kemampuan mendeteksi kelas minoritas, yaitu pelanggan *churn*. Berdasarkan *classification report,* model ini berhasil meraih akurasi sebesar 90%, dengan *precision* mencapai 97% untuk kelas *non-churn* (0) dan *recall* sebesar 81% pada kelas *churn* (1). Nilai *F1-score* masing-masing tercatat 94% untuk kelas 0 dan 71% untuk kelas 1, yang menggambarkan adanya keseimbangan antara *precision* dan *recall*. Rincian distribusi prediksi dapat dilihat melalui *confusion matrix* yang ditampilkan pada Gambar 2..



Gambar 2. *Confusion Matrix* dari *Hard Voting Classifier*

Confusion matrix dari *Hard Voting Classifier* memperlihatkan bahwa model ini mampu mengidentifikasi sebanyak 520 pelanggan *non-churn* secara akurat (*true negative*), serta berhasil memprediksi 79 pelanggan *churn* dengan benar (*true positive*). Meski demikian, masih terdapat 50 kasus *false positive*, yaitu pelanggan *non-churn* yang keliru terdeteksi sebagai *churn*, serta 18 kasus *false negative* di mana pelanggan *churn* tidak terdeteksi oleh model. Temuan ini mengindikasikan bahwa model cenderung lebih waspada dalam menangkap potensi churn meskipun berisiko salah mengklasifikasikan sebagian pelanggan setia. Sebagai pembanding, dilakukan pula pengujian masing-masing *base classifier — Naïve Bayes, Random Forest*, dan *Nearest Centroid —* yang dilatih secara terpisah menggunakan dataset yang telah melewati tahap *preprocessing* serupa. Hasil perbandingan performa antara *Hard Voting Classifier* dan ketiga model dasar tersebut dirangkum pada Tabel 3.

Tabel 3. Perbandingan Performa dari tiap Model

Metrik	Kelas	Naïve Bayes	Random Forest	Nearest Centroid	Hard Voting Classifier
Precision	0	0,95	0,96	0,94	0,97
Recall	0	0,88	0,98	0,72	0,91
F1-score	0	0,91	0,97	0,82	0,94

Precision	1	0,50	0,87	0,31	0,61
Recall	1	0,70	0,78	0,74	0,81
F1-score	1	0,58	0,83	0,44	0,70
Accuracy	-	0,85	0,95	0,73	0,90

Hard Voting Classifier memiliki sejumlah keunggulan jika dibandingkan dengan masing-masing model dasarnya, yaitu *Naïve Bayes, Random Forest*, dan *Nearest Centroid*. Salah satu nilai lebihnya terlihat dari peningkatan *recall* pada kelas minoritas (*churn*) yang mampu mencapai 81%, lebih tinggi dibandingkan *recall* yang dihasilkan oleh *Random Forest* (78%), *Naïve Bayes* (70%), maupun *Nearest Centroid* (74%). Pencapaian ini menegaskan bahwa pendekatan *ensemble* efektif memadukan kelebihan tiap algoritma, khususnya dalam mengidentifikasi pelanggan yang benar-benar berhenti menggunakan layanan. Selain itu, nilai *precision* pada kelas *non-churn* juga sangat baik, yaitu 97%, yang berarti model jarang salah memprediksi pelanggan tetap. Capaian ini penting karena membantu perusahaan menekan pemborosan biaya dalam menjalankan program retensi. Namun demikian, meskipun *recall* meningkat, nilai *f1-score* untuk churn (71%) pada *Hard Voting Classifier* masih lebih rendah dibandingkan *f1-score Random Forest* yang mencapai 83%. Salah satu penyebabnya adalah kontribusi *Nearest Centroid* yang *f1-scorenya* hanya 44% sehingga menurunkan rata-rata performa pada kelas minoritas. Di sisi lain, akurasi keseluruhan *Hard Voting Classifier* berada di angka 90%, sedikit lebih rendah daripada akurasi *Random Forest* yang mampu mencapai 95%, menunjukkan bahwa strategi *ensemble* ini memang sedikit mengorbankan akurasi total demi mendongkrak kemampuan mendeteksi pelanggan *churn.*

4. Kesimpulan

Penelitian ini berhasil merancang model prediksi *customer churn* berbasis *Hard Voting Classifier* yang menggabungkan tiga algoritma, yakni *Naïve Bayes, Random Forest*, dan Nearest Centroid. Berdasarkan hasil pengujian, model *ensemble* ini mampu mencapai tingkat akurasi sebesar 90% dengan nilai *recall* untuk *churn* sebesar 81%. Salah satu temuan utama dari studi ini menunjukkan bahwa *Random Forest* tampil lebih baik dalam hal akurasi keseluruhan (95%) dan precision (87%), namun *Hard Voting Classifier* terbukti lebih andal dalam mendeteksi pelanggan *churn* dengan *recall* 81% dibandingkan *Random Forest* yang hanya 78%, serta memiliki *precision* kelas *non-churn* yang sedikit lebih tinggi (97% berbanding 96%). Meski demikian, riset ini masih memiliki beberapa keterbatasan yang dapat dijadikan titik tolak untuk pengembangan di masa mendatang. Kinerja *Hard Voting Classifier* yang diusulkan masih bergantung pada performa masing-masing base classifier yang digunakan. Oleh sebab itu, untuk mendapatkan hasil prediksi yang lebih presisi dan stabil, disarankan dilakukan penyetelan *hyperparameter* pada setiap *base classifier* serta eksplorasi kombinasi model dasar lainnya yang lebih sesuai. Dengan pendekatan ini, diharapkan kualitas klasifikasi dapat meningkat dan tetap handal meskipun ada ketidakpastian pada salah satu model pembentuknya.

Daftar Pustaka

[1] S. Saleh and S. Saha, "Customer retention and churn prediction in the telecommunication industry: a case study on a Danish university," *SN Appl Sci*, vol. 5, no. 7, p. 173, 2023.

[2] O. Çelik and U. O. Osmanoglu, "Comparing to techniques used in customer churn analysis," *Journal of Multidisciplinary Developments*, vol. 4, no. 1, pp. 30–38, 2019.

[3] A. Amin, A. Adnan, and S. Anwar, "An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and Naïve Bayes," *Appl Soft Comput*, vol. 137, p. 110103, 2023, doi: https://doi.org/10.1016/j.asoc.2023.110103.

[4] M. K. Awang, M. Makhtar, N. Udin, and N. F. Mansor, "Improving customer churn classification with ensemble stacking method," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 11, 2021.

[5] A. Muneer, R. F. Ali, A. Alghamdi, S. M. Taib, A. Almaghthawi, and E. A. A. Ghaleb, "Predicting customers churning in banking industry: A machine learning approach," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 1, pp. 539–549, 2022.

[6] S. Antoh, R. Herteno, I. Budiman, D. Kartini, and M. I. Mazdadi, "Prediksi Churn Pelanggan Telekomunikasi dengan Optimalisasi Seleksi Fitur dan Tuning Hyperparameter pada Algoritma Klasifikasi C4. 5," *Jurnal Sistem Informasi Bisnis*, vol. 15, no. 1, pp. 60–67, 2025.

[7] Y. Zhou, W. Chen, X. Sun, and D. Yang, "Early warning of telecom enterprise customer churn based on ensemble learning," *PLoS One*, vol. 18, no. 10, p. e0292466, 2023.

[8] E. Manro, D. Malhotra, and D. Kamthania, "Customer Churn Prediction using Machine Learning," *Journal of Innovations in Computer Science and Trends in IT*, vol. 2, no. 1, pp. 3048–4707, 2025.

[9] B. R. Agasti and S. Satpathy, "Predicting customer churn in telecommunication sector using Naïve Bayes algorithm," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 35, no. 3, pp. 1610–1617, 2024.

[10] J. Gerald Manju, A. Dharini, B. Kiruthika, and A. Malini, "Online Food Delivery Customer Churn Prediction: A Quantitative Analysis on the Performance of Machine Learning Classifiers," in *International Conference on Data Analytics & Management*, Springer, 2023, pp. 95–104.

[11] J. Latheef and S. Vineetha, "Predicting customer loyalty in banking sector with mixed ensemble model and hybrid model," in *Smart Computing Techniques and Applications: Proceedings of the Fourth International Conference on Smart Computing and Informatics, Volume 2*, Springer, 2021, pp. 363–371.

[12] R. Bhuria *et al.*, "Ensemble-based customer churn prediction in banking: a voting classifier approach for improved client retention using demographic and behavioral data," *Discover Sustainability*, vol. 6, no. 1, p. 28, 2025.

[13] R. Bhuria *et al.*, "Ensemble-based customer churn prediction in banking: a voting classifier approach for improved client retention using demographic and behavioral data," *Discover Sustainability*, vol. 6, no. 1, p. 28, 2025.

[14] A. Manzoor, M. A. Qureshi, E. Kidney, and L. Longo, "A review on machine learning methods for customer churn prediction and recommendations for business practitioners," *IEEE access*, vol. 12, pp. 70434–70463, 2024.

[15] M. Z. Alotaibi and M. A. Haq, "Customer churn prediction for telecommunication companies using machine learning and ensemble methods," *Engineering, Technology & Applied Science Research*, vol. 14, no. 3, pp. 14572–14578, 2024.

[16] S. Wu, W.-C. Yau, T.-S. Ong, and S.-C. Chong, "Integrated churn prediction and customer segmentation framework for telco business," *Ieee Access*, vol. 9, pp. 62118–62136, 2021.

[17] M. Vasudevan, R. S. Narayanan, S. F. Nakeeb, and A. Abhishek, "Customer churn analysis using XGBoosted decision trees," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 25, no. 1, pp. 488–495, 2022.

[18] Reyad Hussien, Mohamed Mahgoub, Shahenda Youssef, Ashraqat Torky, and Nermin K. Negied, "A novel artificial intelligent-based approach for real time prediction of telecom customer's coming interaction," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 33, no. 1, pp. 540–556, Jan. 2024.

[19] J. Gerald Manju, A. Dharini, B. Kiruthika, and A. Malini, "Online food delivery customer churn prediction: a quantitative analysis on the performance of machine learning classifiers," in *International Conference on Data Analytics & Management*, Springer, 2023, pp. 95–104.

[20] Sudip Chatterjee, CrowdANALYTIX, and Kaggle, "Telechurn," Kaggle. Accessed: Jun. 05, 2025. [Online]. Available: https://www.kaggle.com/datasets/remonic97/telechurn/data

[21] R. F. Putra *et al.*, *Data Mining: Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia, 2023.

[22] S. Lonang and D. Normawati, "Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination," *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 49, 2022.

[23] P. H. Artanti, "Penerapan Neural Network dengan optimasi Ant Colony Optimization dan Backpropagation untuk membangun model prediksi diabetes tahap awal," Universitas Islam Negeri Maulana Malik Ibrahim, 2023.

[24] K. Elsa Virantika and J. Ipmawati, "Evaluasi Hasil Pengujian Tingkat Clusterisasi Penerapan Metode K-Means Dalam Menentukan Tingkat Penyebaran Covid-19 di Indonesia," 2022.

[25] B. Aribowo and S. Fairuz, *Panduan Praktis Machine Learning Klasifikasi Menggunakan Python: Diandra Kreatif*. Diandra Kreatif, 2024.

[26] S. E. Caria Ningsih, M. P. Sukemi, M. S. Andi Reni Syamsuddin, and S. P. M. Fitra Gustiar, "Statistik: panduan praktis untuk analisis data," 2024, *PT. Media Penerbit Indonesia*.

[27] P. W. Rahayu *et al.*, *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia, 2024.

[28] I. Maulana, N. Khairunisa, and R. Mufidah, "Deteksi bentuk wajah menggunakan convolutional neural network (CNN)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, pp. 3348–3355, 2023.

[29] E. F. Swana, W. Doorsamy, and P. Bokoro, "Tomek link and SMOTE approaches for machine fault classification with an imbalanced dataset," *Sensors*, vol. 22, no. 9, p. 3246, 2022.

[30] Z. Xu, D. Shen, T. Nie, Y. Kou, N. Yin, and X. Han, "A cluster-based oversampling algorithm combining SMOTE and k-means for imbalanced medical data," *Inf Sci (N Y)*, vol. 572, pp. 574–589, 2021.

[31] A. Taha, "Intelligent ensemble learning approach for phishing website detection based on weighted soft voting," *Mathematics*, vol. 9, no. 21, p. 2799, 2021.

[32] H. B. Truong and V. C. Tran, "A framework for fake news detection based on the wisdom of crowds and the ensemble learning model," *Computer Science and Information Systems*, vol. 20, no. 4, pp. 1439–1457, 2023.