Optimalisasi Hybrid Sampling pada SVM dan Ensemble Learning untuk Prediksi Churn

Hartono Wijaya, Hairani Hairani, Viviana Herlita Vidiasari, Farda Milanda Amin

Universitas Bumigora, Mataram, Indonesia

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

Abstrak

Churn merupakan kondisi ketika pelanggan menghentikan penggunaan produk atau layanan suatu perusahaan, yang secara langsung berdampak terhadap penurunan pendapatan dan peningkatan biaya akuisisi pelanggan baru. Ketidakmampuan model pembelajaran mesin dalam mengenali pelanggan yang berisiko churn akibat ketidakseimbangan data menjadi tantangan utama dalam sistem prediksi churn. Penelitian ini bertujuan untuk mengembangkan model prediksi churn yang lebih akurat dan sensitif dengan mengombinasikan algoritma Support Vector Machine (SVM) dan metode ensemble (Bagging dan Stacking), disertai penerapan teknik hybrid sampling seperti SMOTE. Metode penelitian meliputi pengumpulan data, preprocessing data, pembagian data, pelatihan model dan evaluasi . Hasil penelitian menunjukkan bahwa model Stacking memberikan performa terbaik dengan akurasi dan F1-score mencapai 86%, serta nilai AUC sebesar 0,93. Fitur products number, age, dan active member teridentifikasi sebagai variabel paling berpengaruh terhadap churn. Batasan utama penelitian ini terletak pada keterbatasan sumber data dan belum dilakukannya tuning parameter secara mendalam. Penelitian ini juga memberikan kontribusi praktis bagi strategi retensi pelanggan dan dapat dikembangkan lebih lanjut dengan validasi pada dataset lintas industri dan tuning parameter yang lebih luas.

Kata kunci: *Churn, Support Vector Machine, Ensemble learning, Hybrid sampling, Data Tidak Seimbang*

Abstract

Churn is a condition when customers stop using a company's products or services, which directly impacts revenue decline and increased costs of acquiring new customers. The inability of machine learning models to recognize customers at risk of churn due to data imbalance is a major challenge in churn prediction systems. This research aims to develop a more accurate and sensitive churn prediction model by combining the Support Vector Machine (SVM) algorithm and ensemble methods (Bagging and Stacking), along with the application of hybrid sampling techniques such as SMOTE. The research methods include data collection, data preprocessing, data splitting, model training, and evaluation. The research results show that the Stacking model provides the best performance with an accuracy and F1-score of 86%, as well as an AUC value of 0.93. The features products number, age, and active member were identified as the most influential variables on churn. The main limitation of this research lies in the limited data sources and the lack of in depth parameter tuning. This research also provides practical contributions to customer retention strategies and can be further developed with validation on cross-industry datasets and broader parameter tuning.

Keywords*: Churn, Support Vector Machine, Ensemble learning, Hybrid sampling, Imbalanced Data*

1. Pendahuluan

Churn merupakan kondisi ketika pelanggan menghentikan penggunaan produk atau layanan suatu perusahaan, yang secara langsung berdampak terhadap penurunan pendapatan dan peningkatan biaya akuisisi pelanggan baru. Fenomena ini menjadi isu strategis di berbagai sektor industri seperti telekomunikasi, perbankan, dan e-commerce, terutama dalam menghadapi persaingan pasar yang semakin kompetitif. Oleh karena itu, pengembangan model prediksi churn yang akurat dan adaptif sangat diperlukan untuk mendukung strategi retensi pelanggan berbasis data [1]. Salah satu tantangan utama dalam membangun model prediksi churn adalah ketidakseimbangan kelas, di mana jumlah pelanggan yang churn jauh lebih sedikit dibandingkan pelanggan yang tetap. Ketimpangan ini menyebabkan algoritma pembelajaran mesin cenderung bias terhadap kelas mayoritas, sehingga mengabaikan pola-pola penting

dari pelanggan yang berisiko churn. Akibatnya, meskipun akurasi model terlihat tinggi, kemampuan dalam mendeteksi churn sering kali rendah, sebagaimana ditunjukkan oleh nilai recall yang kecil [2].

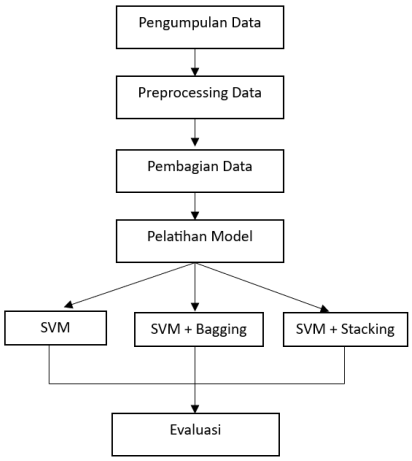
Berbagai penelitian sebelumnya telah mengusulkan pendekatan untuk mengatasi tantangan tersebut. Fannisa Salsabila dkk (2025). menerapkan teknik SMOTE dan Random Over-Sampling (ROS) dan menemukan bahwa Random Forest tanpa oversampling memberikan akurasi tertinggi sebesar 87,13%, namun recall churn masih rendah. Penerapan SMOTE dan ROS meningkatkan sensitivitas terhadap kelas minoritas, meskipun akurasi sedikit menurun menjadi 86,20% dan 81,47% secara berurutan [3]. Penelitian M. Ainur Syawaludin dkk (2024). juga menunjukkan dominasi Random Forest dibandingkan Naive Bayes dan Decision Tree, dengan akurasi 85% dan AUC 0,83. Namun, Naive Bayes hanya mencatat recall sebesar 0,22, yang berarti model gagal mengenali sebagian besar pelanggan yang benar-benar berhenti [4]. Selain pendekatan konvensional, deep learning juga mulai banyak diadopsi. Dewa Adji Kusuma et al. (2025) membuktikan bahwa model CNN dengan seleksi fitur CFS dan RFE serta tuning parameter mampu mencapai akurasi 89% dan F1-score 88%, menunjukkan keunggulan CNN dalam mengenali pola perilaku pelanggan yang kompleks [5]. Dalam konteks serupa, Reza et al. (2024) menggunakan gabungan RFE, SMOTE, dan AdaBoost untuk memprediksi attrition karyawan dan berhasil mencapai akurasi 90,7%, sekaligus mengidentifikasi fitur penting seperti Monthly Income dan Years Since Last Promotion yang juga dapat diaplikasikan dalam konteks prediksi churn [6].

Meskipun berbagai pendekatan dalam prediksi churn telah menunjukkan hasil yang menjanjikan, tantangan utama yang masih dihadapi adalah ketidakseimbangan data yang berdampak pada performa model, khususnya dalam hal recall terhadap kelas minoritas [7]. Permasalahan ini menuntut metode yang tidak hanya mampu mempertahankan akurasi keseluruhan, tetapi juga sensitif terhadap deteksi churn yang umumnya berjumlah lebih sedikit. Support Vector Machine (SVM) dikenal sebagai algoritma klasifikasi yang efektif dalam menangani data berdimensi tinggi melalui prinsip margin maksimum, namun memiliki kelemahan dalam menangani distribusi kelas yang tidak seimbang. Di sisi lain, metode ensemble learning memiliki kemampuan yang baik dalam meningkatkan stabilitas dan generalisasi model melalui kombinasi prediktor yang beragam. Oleh karena itu, integrasi antara SVM dan pendekatan ensemble membuka peluang untuk membangun model prediktif yang lebih adaptif dan akurat dalam konteks churn [8].

Dengan begitu Penelitian ini bertujuan untuk menguji efektivitas metode hybrid sampling pada SVM dan ensemble learning dalam meningkatkan akurasi dan recall prediksi churn pelanggan bank. Selain itu kontribusi utama dari penelitian ini terletak pada penyusunan pendekatan sistematis dalam pemilihan dan penerapan metode hybrid sampling yang optimal, serta pembuktian empiris terhadap peningkatan performa SVM dan ensemble dalam menangani data tidak seimbang pada kasus prediksi churn [9]. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi baik secara teoritis maupun praktis dalam pengembangan sistem prediktif berbasis data, serta mendukung pengambilan keputusan strategis dalam pengelolaan pelanggan di lingkungan bisnis yang dinamis dan kompetitif [10].

2. Metode Penelitian

Pada penelitian ini menerapkan algoritma machine learning yang terdiri dari Support Vector Machine (SVM), Bagging, dan Stacking untuk melakukan prediksi churn pelanggan bank. Pada gambar 1. tahapan metode penelitian terdiri dari lima tahap utama, yaitu pengumpulan dataset, preprocessing data, pemodelan, prediksi, dan evaluasi. Pada penelitian ini digunakan lima tahapan karena seluruh proses dilakukan berbasis data publik dan pendekatan eksperimental [11].



Gambar 1. Tahapan Penelitian

Tahapan pertama dimulai dari pengumpulan data, di mana data yang digunakan berasal dari dataset publik yang memuat informasi pelanggan perbankan [12]. Tahap selanjutnya adalah preprocessing data yang bertujuan untuk memastikan kualitas data sebelum proses pelatihan model dilakukan. Proses ini meliputi penghapusan kolom yang tidak memiliki kontribusi terhadap prediksi. Selanjutnya, dilakukan encoding terhadap atribut kategorikal menggunakan teknik Label Encoding dan One Hot Encoding. Untuk mengatasi masalah ketidakseimbangan data antar kelas, digunakan metode Synthetic Minority Oversampling Technique (SMOTE) yang bertujuan untuk memperbanyak data pada kelas minoritas secara sintetis. Setelah itu, dilakukan proses standarisasi data menggunakan Standard Scaler untuk menyamakan skala antar fitur numerik, mengingat algoritma Support Vector Machine (SVM) sangat sensitif terhadap perbedaan skala tersebut [13].

Data yang telah diproses kemudian dibagi menjadi dua bagian, yakni data latih dan data uji dengan rasio 80:20. Pembagian ini dilakukan secara acak untuk memastikan bahwa evaluasi model dilakukan pada data yang tidak pernah dilihat sebelumnya oleh model, sehingga dapat mencerminkan kemampuan generalisasi model yang sebenarnya [14]. Pada tahap pelatihan model, tiga pendekatan digunakan untuk membangun dan membandingkan performa model prediksi churn. Model pertama adalah Support Vector Machine (SVM) sebagai baseline. Model kedua menerapkan pendekatan ensemble menggunakan Bagging Classifier, di mana beberapa model SVM dibangun secara paralel dan hasil prediksinya digabungkan untuk meningkatkan kestabilan model. Model ketiga menggunakan teknik Stacking, yaitu gabungan dari dua algoritma (SVM dan Decision Tree) yang hasil prediksinya digunakan sebagai masukan bagi model meta-classifier berupa Logistic Regression. Seluruh model dilatih menggunakan data yang telah diseimbangkan dan distandarisasi [15].

Tahap terakhir dalam metode ini adalah evaluasi model, yang bertujuan untuk mengukur performa masing-masing pendekatan dalam melakukan klasifikasi churn pelanggan. Evaluasi dilakukan menggunakan beberapa metrik, yaitu akurasi, presisi, recall, F1-score. Metrik ini digunakan untuk memberikan gambaran menyeluruh mengenai kinerja model, khususnya dalam mengidentifikasi pelanggan yang berisiko churn [16].

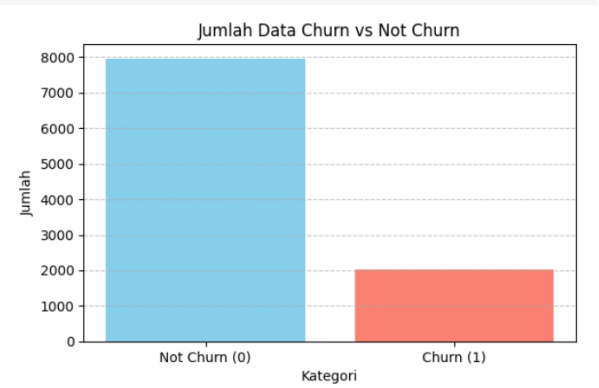
3. Hasil dan Pembahasan

Penelitian ini menggunakan data yang berasal dari dataset Kaggle berjudul *Bank Customer Churn Prediction*. Pada tabel 1. dataset terdiri dari 10.000 entri pelanggan dengan 10 atribut, yaitu: *customer id, credit score, country, gender, age, tenure, products number, credit card, active member, dan churn*.

Tabel 1. Dataset Bank Customer Churn

Customer id	Credit score	Country	Gender	Age	Tenure	Products Number	Credit Card	Active Number	Churn
15634602	619	France	Female	42	2	1	1	1	1
15647311	608	Spain	Female	41	1	1	0	1	0
15619304	502	France	Female	42	8	3	1	0	1
15737888	699	France	Female	39	1	2	0	0	0
15574012	850	Spain	Female	43	2	1	1	1	0

Selanjutnya dilakukan Preprocessing data untuk memastikan kualitas data sebelum pelatihan model. Kolom customer id dihapus karena tidak bersifat prediktif. Atribut kategorikal gender diencoding dengan Label Encoding, sementara country dikonversi menggunakan One Hot Encoding agar model tidak mengasumsikan urutan antar kategori.



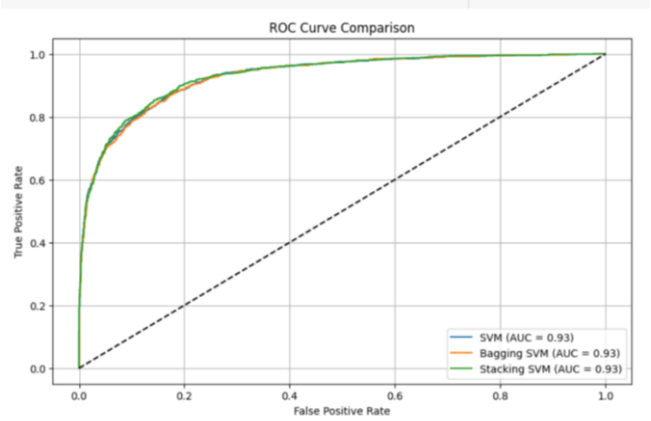
Gambar 2. Jumlah Data Pelanggan Churn dan Not Churn

Pada Gambar 2, ditunjukkan bahwa ketidakseimbangan kelas pada variabel target (churn vs. non-churn) diatasi menggunakan teknik SMOTE, yang berfungsi mensintesis data pada kelas minoritas sehingga distribusi antar kelas menjadi lebih seimbang. Langkah ini penting untuk menghindari bias model terhadap kelas mayoritas. Setelah proses oversampling, seluruh fitur numerik kemudian dinormalisasi menggunakan *Standard Scaler*, mengingat algoritma SVM sangat peka terhadap perbedaan skala antar fitur. Data yang telah diproses selanjutnya dibagi menjadi dua bagian, yaitu data latih dan data uji, dengan rasio 80:20. Pembagian dilakukan secara acak untuk memastikan evaluasi model berlangsung secara objektif dan tidak bias. Berdasarkan data tersebut, dikembangkan tiga pendekatan model klasifikasi, yaitu SVM dasar, SVM dengan teknik Bagging, dan SVM dengan pendekatan Stacking, yang kemudian dievaluasi untuk melihat pengaruh masing-masing terhadap akurasi dan sensitivitas dalam mendeteksi churn.

Tabel 2. Tabel Evaluasi Kinerja Model

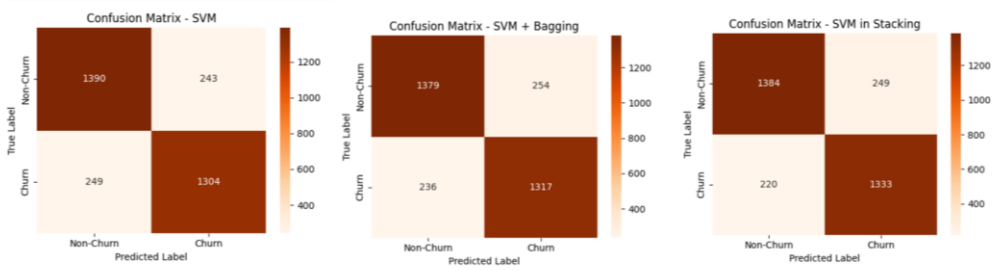
Model	Akurasi	Presisi	Recall	F1-Score
SVM	85%	85%	85%	85%
SVM + Bagging	85%	84%	85%	84%
SVM + Stacking	86%	86%	85%	86%

Hasil evaluasi performa model ditampilkan dalam Tabel 2. menggunakan metrik akurasi, presisi, recall, dan F1-score. Model dasar SVM menunjukkan performa seimbang dengan nilai 85% untuk semua metrik. Bagging tidak memberikan peningkatan signifikan, bahkan menunjukkan sedikit penurunan F1-score menjadi 84%, meskipun recall tetap stabil. Sebaliknya, pendekatan Stacking menghasilkan performa terbaik dengan F1-score dan akurasi sebesar 86%. Hal ini mengindikasikan bahwa Stacking mampu mengombinasikan kekuatan dua algoritma berbeda secara sinergis. Namun demikian, peningkatan 1% pada F1-score perlu dianalisis dari sisi signifikansi praktis. Meskipun secara statistik tidak dibuktikan dalam penelitian ini, perbedaan tersebut dapat berdampak penting dalam konteks bisnis, karena model lebih sensitif terhadap pelanggan yang benar-benar churn.



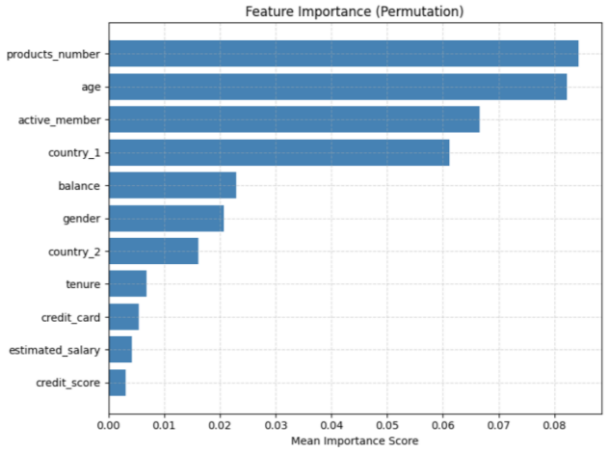
Gambar 3. Kurva ROC

Pada Gambar 3. menyajikan perbandingan kurva ROC dari tiga model klasifikasi, yaitu Support Vector Machine (SVM), Bagging SVM, dan Stacking SVM dalam prediksi churn. Berdasarkan hasil visualisasi, ketiga model menunjukkan performa yang serupa dengan nilai Area Under the Curve (AUC) sebesar 0,93. Nilai AUC tersebut mencerminkan tingkat akurasi yang tinggi dalam membedakan antara kelas churn dan non-churn, di mana seluruh kurva berada jauh di atas garis diagonal yang merepresentasikan klasifikasi acak (AUC = 0,5). Penerapan teknik ensemble seperti bagging dan stacking tidak menunjukkan peningkatan signifikan terhadap kinerja model dibandingkan dengan SVM tunggal, sebagaimana ditunjukkan oleh kemiripan bentuk kurva dan kesamaan nilai AUC. Temuan ini mengindikasikan bahwa model SVM telah memiliki kemampuan diskriminatif yang optimal dalam konteks prediksi churn, sehingga penambahan kompleksitas melalui metode ensemble tidak secara substansial meningkatkan performa pada metrik ini.



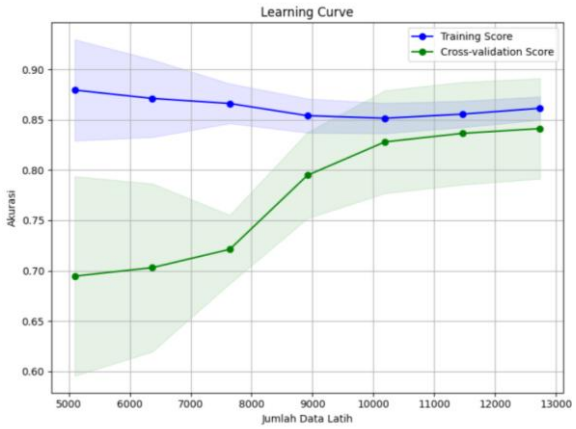
Gambar 4. Confusion Matrix

Pada Gambar 4. menampilkan hasil perbandingan confusion matrix dari tiga model klasifikasi, yaitu SVM, SVM dengan teknik bagging, dan SVM yang diterapkan dalam pendekatan stacking untuk prediksi churn. Model SVM dasar mencatat 1.304 prediksi benar untuk pelanggan churn (true positive) dan 249 kesalahan prediksi churn sebagai non-churn (false negative). Sementara itu, penerapan bagging pada SVM meningkatkan jumlah prediksi benar menjadi 1.317 dan mengurangi kesalahan menjadi 236. Model dengan performa tertinggi ditunjukkan oleh SVM dalam skema stacking, yang berhasil mengidentifikasi 1.333 kasus churn secara tepat dan hanya menghasilkan 220 kesalahan klasifikasi. Walaupun terjadi sedikit peningkatan false positive pada kedua model ensemble (254 untuk bagging dan 249 untuk stacking), hasil ini diimbangi oleh peningkatan kemampuan model dalam mengenali pelanggan yang benar-benar churn . Dengan kata lain, metode ensemble terutama stacking menunjukkan keunggulan dalam meningkatkan sensitivitas model terhadap kelas churn tanpa menyebabkan penurunan signifikan dalam akurasi prediksi keseluruhan. Oleh karena itu, pendekatan stacking dapat dianggap sebagai pilihan yang lebih efektif dalam konteks prediksi churn meskipun terjadi sedikit peningkatan false positive, model stacking memberikan keseimbangan lebih baik antara sensitivitas dan spesifisitas.



Gambar 5. Feature Importance

Gambar 5 menampilkan hasil analisis *feature importance* menggunakan metode *permutation importance* untuk mengevaluasi kontribusi masing-masing variabel terhadap kinerja model dalam memprediksi churn. Berdasarkan hasil analisis, fitur products number tercatat sebagai variabel dengan pengaruh terbesar, diikuti oleh age, active member, dan country 1. Temuan ini menunjukkan bahwa jumlah produk yang dimiliki pelanggan, usia, status keanggotaan aktif, dan asal negara merupakan faktor utama yang memengaruhi kemungkinan pelanggan melakukan churn. Sebaliknya, fitur seperti credit score, estimated salary, dan kepemilikan credit card menunjukkan tingkat pengaruh yang rendah, sehingga dianggap memiliki kontribusi yang minimal dalam proses klasifikasi. Informasi ini penting untuk mendukung pengambilan keputusan yang lebih tepat, khususnya dalam merancang strategi retensi pelanggan. Secara khusus, pengaruh besar dari products number mengindikasikan bahwa pelanggan dengan lebih banyak produk cenderung memiliki loyalitas yang lebih tinggi.



Gambar 6. Learning Curve

Gambar 6 menggambarkan *learning curve* yang mengilustrasikan pengaruh jumlah data latih terhadap akurasi model, baik pada data pelatihan maupun validasi silang (*cross-validation*). Kurva biru menunjukkan skor akurasi pada data pelatihan, sedangkan kurva hijau menunjukkan akurasi pada proses validasi silang. Dari grafik terlihat bahwa peningkatan jumlah data latih secara bertahap meningkatkan akurasi validasi, dari sekitar 70% hingga mendekati 85%. Sementara itu, akurasi pelatihan relatif stabil di rentang 85% hingga 90%. Perbedaan antara kedua kurva semakin kecil ketika data latih mencapai sekitar 12.000, yang menunjukkan bahwa model semakin mampu melakukan generalisasi dan mengurangi *variance*. Wilayah bayangan pada masing-masing kurva menggambarkan tingkat variasi atau ketidakpastian, yang juga menurun seiring bertambahnya data latih.

4. Kesimpulan

Penelitian ini dilakukan untuk mengatasi tantangan dalam prediksi churn pelanggan, khususnya akibat ketidakseimbangan data yang sering menyebabkan penurunan akurasi dan recall pada kelas

minoritas. Untuk menjawab permasalahan tersebut, diterapkan pendekatan kombinasi antara teknik hybrid sampling dan algoritma machine learning, yaitu Support Vector Machine (SVM) serta metode ensemble seperti Bagging dan Stacking. Hasil yang diperoleh menunjukkan bahwa model SVM dengan pendekatan Stacking menghasilkan performa terbaik, dengan akurasi dan F1-score mencapai 86%. Hal ini menunjukkan efektivitas kombinasi hybrid sampling dan ensemble learning dalam meningkatkan kemampuan model mengenali pelanggan yang berisiko churn. Selain itu, fitur products number, age, dan active member terbukti menjadi variabel yang paling berkontribusi terhadap prediksi churn. Kurva pembelajaran juga menunjukkan peningkatan generalisasi model seiring bertambahnya data, sedangkan kurva ROC menampilkan kemampuan klasifikasi yang tinggi dengan AUC sebesar 0,93. Untuk penelitian selanjutnya disarankan untuk menguji pendekatan ini pada dataset dari institusi yang berbeda guna mengukur validitas eksternal model. Selain itu, eksplorasi terhadap tuning parameter dan teknik sampling alternatif perlu dilakukan untuk meningkatkan adaptivitas dan performa model dalam skenario data yang lebih kompleks dan bervariasi.

Daftar Pustaka

[1] M. Rizki Kurniawan, P. Nurul Sabrina, and R. Ilyas, "Prediksi Customer Churn Pada Perusahaan Telekomunikasi Menggunakan Algoritma C4.5 Berbasis Particle Swarm Optimization," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 5, pp. 3369–3375, 2024, doi: 10.36040/jati.v7i5.7476.

[2] M. Amirulhaq Iskandar and U. Latifa, "Website Prediksi Customer Churn Untuk Mempertahankan Pelanggan Pada Perusahaan Telekomunikasi," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1308–1316, 2023, doi: 10.36040/jati.v7i2.6639.

[3] F. S. Pratiwi *et al.*, "Implementasi Metode Smote dan Random OverSampling pada Algoritma Machine Learning," vol. 8, no. 1, pp. 87–98, 2025.

[4] M. A. Syawaludin and R. Hidayat, "Prediksi Churn Pelanggan Multinational Bank Menggunakan Algoritma Machine Learning," vol. 4, no. 2, pp. 89–97, 2024.

[5] D. A. Kusuma, A. R. Dewi, and A. R. Wijaya, "Perbandingan Random Forest dan Convolutional Neural Network dalam Memprediksi Peralihan Pelanggan," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 10, no. 1, pp. 186–194, 2025.

[6] M. T. Rfe and D. A. N. Adaboost, "Optimalisasi prediksi kehilangan karyawan menggunakan teknik rfe, smote, dan adaboost," vol. 9, no. 4, pp. 2131–2145, 2024.

[7] M. Manalu and L. Wulandari, "Customer Churn Classification Using Support Vector Machine (SVM) Algorithm," vol. 9, no. 3, pp. 74–79, 2024.

[8] B. Durkaya Kurtcan and T. Ozcan, "Predicting customer churn using grey wolf optimization-based support vector machine with principal component analysis," *J. Forecast.*, vol. 42, no. 6, pp. 1329–1340, 2023, doi: 10.1002/for.2960.

[9] H. Hairani and D. Priyanto, "A New Approach of Hybrid Sampling SMOTE and ENN to the Accuracy of Machine Learning Methods on Unbalanced Diabetes Disease Data," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 8, pp. 585–590, 2023, doi: 10.14569/IJACSA.2023.0140864.

[10] N. Y. Nhu, T. Van Ly, and D. V. Truong Son, "Churn prediction in telecommunication industry using kernel Support Vector Machines," *PLoS One*, vol. 17, no. 5 May, 2022, doi: 10.1371/journal.pone.0267935.

[11] F. C. Lucky Lhaura Van, M. K. Anam, S. Bukhori, A. K. Mahamad, S. Saon, and R. L. V. Nyoto, "The Development of Stacking Techniques in Machine Learning for Breast Cancer Detection," *J. Appl. Data Sci.*, vol. 6, no. 1, pp. 71–85, 2025, doi: 10.47738/jads.v6i1.416.

[12] R. Pratama, M. I. Herdiansyah, D. Syamsuar, and A. Syazili, "Prediksi Customer Retention Perusahaan Asuransi Menggunakan Machine Learning," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 1, pp. 96–104, 2023, doi: 10.32736/sisfokom.v12i1.1507.

[13] A. M. Husein and M. Harahap, "Pendekatan Data Science untuk Menemukan Churn Pelanggan pada Sector Perbankan dengan Machine Learning," *Data Sci. Indones.*, vol. 1, no. 1, pp. 8–13, 2021, doi: 10.47709/dsi.v1i1.1169.

[14] J. Faran and A. Triayudi, "Analysis of the Effectiveness of Polynomial Fit Smote Mesh on Imbalance Dataset for Bank Customer Churn Prediction With Xgboost and Bayesian Optimization," *J. Tek. Inform.*, vol. 5, no. 3, pp. 661–667, 2024, [Online]. Available: https://doi.org/10.52436/1.jutif.2024.5.3.1284.

[15] J. K. Sana, M. Z. Abedin, M. S. Rahman, and M. S. Rahman, "A novel customer churn prediction model for the telecommunication industry using data transformation methods and feature selection," *PLoS One*, vol. 17, no. 12 December, 2022, doi: 10.1371/journal.pone.0278095.

[16] A. R. B. Jamroni, W. Hadikristanto, and M. Fatchan, "Analisis Faktor dan Prediksi Atrisi untuk

Optimalisasi Retensi Karyawan Menggunakan Machine Learning," vol. 7, no. 3, pp. 1059–1067, 2025, doi: 10.32877/bt.v7i3.2301.