

# Prediksi *Churn* Pelanggan pada Perusahaan Telekomunikasi Menggunakan Algoritma Pembelajaran Ensambel Berbasis *Bagging*

Hasna Dhiya Nafitra dan Arif Djunaidy

Departemen Sistem Informasi, Institut Teknologi Sepuluh Nopember (ITS)

*e-mail*: adjunaidy@its.ac.id

**Abstrak**—Perkembangan yang pesat dan munculnya banyak pesaing di industri telekomunikasi menuntut perusahaan untuk menerapkan strategi manajemen hubungan pelanggan yang efektif untuk mengatasi masalah *churn* yang dapat merugikan perusahaan. *Churn* adalah peristiwa berpindahannya pelanggan ke perusahaan lain karena alasan tertentu. Perusahaan harus mampu mengidentifikasi faktor-faktor yang memengaruhi pelanggan untuk melakukan *churn* karena biaya mempertahankan pelanggan lebih murah daripada memperoleh pelanggan baru. Salah satu upaya yang dapat dilakukan adalah membuat model prediksi *churn* bagi para pelanggannya. Dalam Tugas Akhir ini, model prediksi *Boosted Bagging* (BoBag) dipilih karena model ini menawarkan pendekatan ensambel berbasis *bagging* yang inovatif dengan *hyperparameter* yang dapat disesuaikan, seperti jumlah sampel, jumlah estimator, dan *learning rate*. Tugas Akhir ini menghasilkan sebuah model prediksi BoBag terbaik untuk kombinasi *hyperparameter* dengan jumlah estimator sebesar 200, jumlah sampel sebesar 0,5, dan *learning rate* sebesar 0,01. Kombinasi terbaik ini menghasilkan model dengan akurasi, F1, dan AUC berturut-turut sebesar 0,8589, 0,7112 dan 0,9176. Analisis terhadap model yang dihasilkan memberikan implikasi bahwa perusahaan harus memberikan perhatian khusus kepada para pelanggan bulanan dengan cara memberikan insentif atau diskon dan berusaha agar pelanggan tersebut bersedia untuk beralih ke kontrak yang lebih panjang. Selain itu, perusahaan harus tetap memperbaiki layanan internal agar dapat bersaing dengan para kompetitor di bidangnya.

**Kata Kunci**—*Bagging* Ensambel, *Boosted Bagging*, Prediksi *Churn*, Telekomunikasi.

## I. PENDAHULUAN

PERUSAHAAN telekomunikasi merupakan salah satu sektor yang mengalami perkembangan pesat di era digital saat ini, ditandain dengan munculnya banyak teknologi baru dan konektivitas yang semakin meningkat. Hal ini membuat perusahaan telekomunikasi harus tumbuh dengan cepat untuk memenuhi tuntutan pasar yang terus berkembang. Kehadiran banyak pesaing di sektor serupa menciptakan tekanan tambahan bagi perusahaan dalam menjaga hubungan dengan pelanggannya [1]. Strategi manajemen hubungan pelanggan yang efektif tidak hanya penting untuk mempertahankan pelanggan yang ada tetapi juga sebagai sarana untuk memenangkan persaingan di pasar yang kompetitif.

Setiap pelanggan memiliki kebutuhan dan ekspektasi berbeda yang dapat berubah seiring waktu dan pengalaman mereka dengan layanan yang diberikan. Beberapa pelanggan mungkin tidak puas dengan kualitas layanan, harga, atau fitur yang ditawarkan, sehingga mereka beralih ke pesaing yang menawarkan layanan lebih baik atau lebih murah [2]. *Customer churn*, atau pelanggan berhenti berlangganan, adalah masalah besar yang dihadapi perusahaan telekomunikasi karena dapat menurunkan pendapatan,

### Algoritma *Boosted Bagging*

**Input:** Training set  $D = \{(x_i, y_i)\}_{i=1}^m$ , where  $x_i \in \mathbb{R}^n$  and  $y_i \in \{0,1\}$ ,  $T$  as the number of bootstrap samples and  $k_1$  for the number of  $i$ th bootstrap samples from the first bootstrap sample  $i$  for  $i = 1, 2, \dots, T$

**Output:** A hybrid classifier  $B$

```

1 For  $i = 1$  to  $T$  do
2   Generate a sample dataset  $D_i$  by bootstrap sampling from  $D$ ;
3   for  $j = 1$  to  $k_1$  do
4     Generate a new sample dataset  $D_{ij}$  by bootstrap sampling from  $D_i$ ;
5   end
6   Apply Bagging algorithm  $b_i$  to samples  $D_{ij}$  for  $j = 1, 2, \dots, k_i$ ;
7 end

8 Return  $B(x) = \arg \max_{y \in \{0,1\}} \sum_{i=1}^T \mathbb{I}(b_i(x) = y)$ , where  $\mathbb{I}(\cdot)$  returns 1 if the statement is true, otherwise returns 0;
```

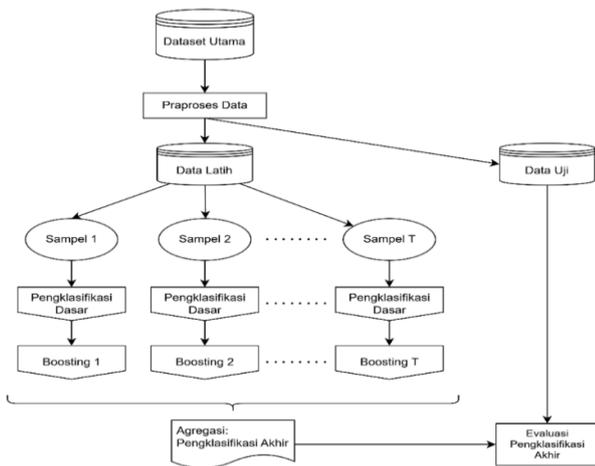
Gambar 1. Algoritma BoBag.

pangsa pasar, dan reputasi perusahaan [3]. Berdasarkan survei [4], industri telekomunikasi memiliki tingkat *churn* sebesar 20%—40% per tahun, dan biaya mempertahankan pelanggan yang ada 5—10 kali lebih murah daripada memperoleh pelanggan baru. Oleh karena itu, perusahaan telekomunikasi harus mengidentifikasi faktor-faktor yang memengaruhi *churn* dan mengambil langkah tepat untuk meminimalisirnya serta mempertahankan pelanggan loyal.

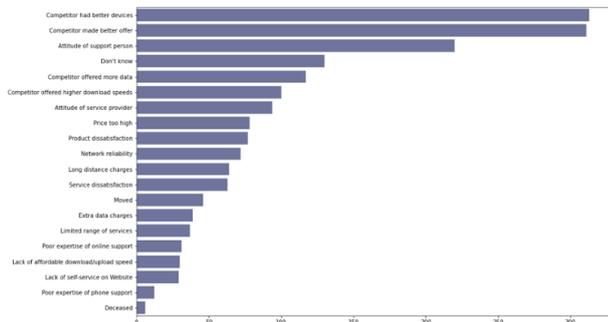
*Customer churn prediction* (CCP) menjadi pendekatan penting untuk meningkatkan retensi pelanggan. Melalui prediksi ini, perusahaan dapat menganalisis perilaku pelanggan untuk melihat kemungkinan mereka berhenti menggunakan layanan, sehingga keputusan yang tepat dapat diambil untuk mencegah *churn* [5]. CCP juga membantu perusahaan memahami pendapatan yang diharapkan di masa depan dan mengidentifikasi kelemahan layanan pelanggan [6].

Banyak penelitian menggunakan teknik *machine learning* dan *data mining* untuk melakukan CCP, terutama klasifikasi. Berbagai model prediksi *churn* diusulkan untuk mencari kinerja terbaik dari prediksi yang dilakukan. Peningkatan kinerja sekecil apa pun sangat berharga dalam tindakan yang diambil, sehingga peneliti terus menggunakan dan meningkatkan teknik yang ada dengan mengusulkan pendekatan hibrida dan ensambel [6]. Menurut Hsu (2017), pendekatan hibrida yang menggabungkan dua atau lebih algoritma klasifikasi berbeda dapat menciptakan model yang lebih kuat dengan memanfaatkan kekuatan masing-masing algoritma [7].

Penelitian Tavassoli dan Koosha (2002) mengusulkan kombinasi algoritma berbasis *bagging* untuk meningkatkan kinerja dari *ensemble base classifier* yang digunakan. Penelitian ini mengusulkan tiga algoritma hibrida, yaitu *Boosted Bagging* (BoBag), *Bagged Bagging* (BaBag), dan *Bagging of Neural Network with Learning Based on Genetic Algorithm* (BNNGA) [6]. Setiap algoritma menggunakan *Decision Tree* (DT) sebagai *bagging classifier* dan



Gambar 2. Diagram algoritma boosted bagging.



Gambar 3. Persebaran alasan pelanggan melakukan churn.

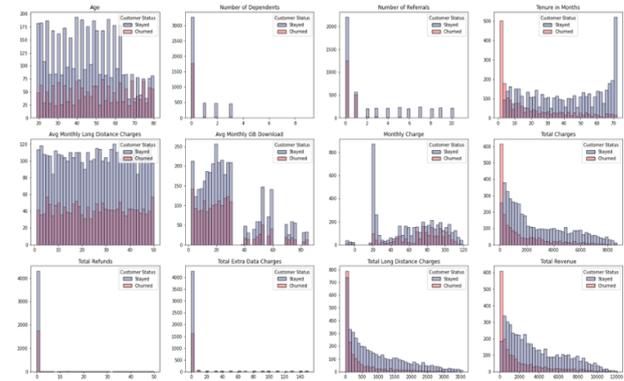
*RobustBoost* (RB) sebagai *boosted classifier* pada metode BoBag. Hasil model yang diusulkan memberikan kinerja lebih baik dibandingkan dengan pengklasifikasi dasar dan ensambel lainnya. Namun, hasil model ini tidak menyertakan analisis bisnis untuk perusahaan telekomunikasi dalam mengatasi *churn*.

Dengan demikian, penelitian ini akan melakukan percobaan menggunakan algoritma BoBag untuk memprediksi dan mencari karakteristik pelanggan yang *churn* di perusahaan telekomunikasi. Setelah model prediksi berhasil dibangun, rekomendasi strategi yang sesuai akan diberikan. Harapannya, perusahaan dapat menerapkan strategi retensi yang ditargetkan secara proaktif, sehingga mengurangi tingkat *churn* dan meningkatkan loyalitas pelanggan.

## II. DASAR TEORI

### A. Prediksi Churn Pelanggan

*Churn* adalah kondisi di mana pelanggan beralih dari satu perusahaan ke perusahaan lain atau kompetitor dan menjadikan pelanggan tersebut berada pada periode tidak aktif yang berkepanjangan [8]. Perusahaan telekomunikasi menghadapi masalah besar terkait dengan perputaran pelanggan, yang dapat memengaruhi pertumbuhan dan pendapatan. Pelanggan dapat *churn* karena berbagai alasan: kurangnya interaksi perusahaan, penanganan keluhan yang lambat, komentar negatif di media sosial, ketidakpuasan dengan layanan, ketidaksesuaian perangkat lunak, atau tawaran kompetitor yang lebih baik. Perusahaan sering tidak menyadari alasan di balik pelanggan yang pergi. Analisis data dapat membantu memahami dan memprediksi *churn* untuk menghindarinya.



Gambar 4. Persebaran nilai pada kolom numerik berdasarkan status pelanggan.

Tabel 1. *Confusion Matrix*

		PREDIKSI	
		Churn	Non-churn
AKTUAL	Churn	TP	FN
	Non-churn	FP	TN

Untuk membuat model prediksi *churn*, perusahaan membutuhkan data historis pelanggan. Model prediksi *churn* mendukung evaluasi data bisnis historis untuk mengidentifikasi pelanggan yang berisiko tinggi *churn*, memungkinkan perusahaan fokus pada kategori tertentu dengan teknik retensi yang dipersonalisasi. Dengan prediksi yang tepat, perusahaan dapat mengambil tindakan proaktif untuk mempertahankan pelanggan [9]. Banyak metode pemodelan digunakan untuk memprediksi *churn* di berbagai organisasi [10].

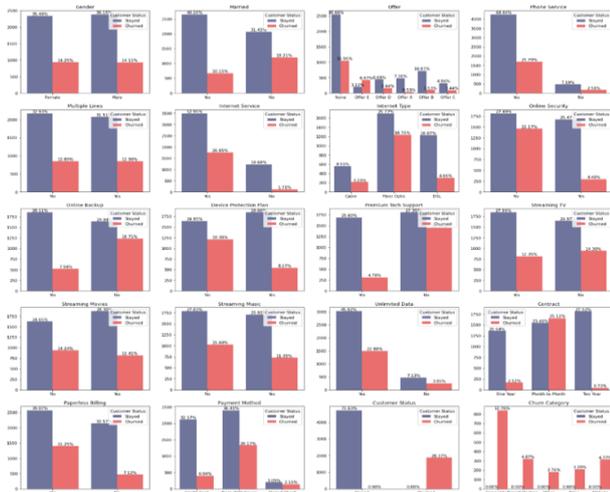
### B. Pembelajaran Ensambel

Metode ensambel merupakan kombinasi dari berbagai metode *machine learning* untuk meningkatkan kinerja prediksi. Dua teknik ensambel yang terkenal adalah *bagging* dan *boosting*. *Bagging* atau *bootstrap aggregating*, diperkenalkan oleh Breiman (2016) [11], menghasilkan beberapa set pelatihan baru melalui teknik *bootstrap sampling*, yaitu pengambilan sampel secara acak dengan pengembalian. Setiap set pelatihan baru digunakan untuk membangun model prediktif, dan hasil akhirnya ditentukan melalui pemungutan suara atau rata-rata. Metode ini efektif untuk algoritma pembelajaran yang tidak stabil, seperti *Decision Tree* [12].

*Boosting* bertujuan memperkuat pengklasifikasi lemah menjadi lebih kuat. Algoritma seperti *AdaBoost* dan *LogitBoost* digunakan untuk meningkatkan kinerja klasifikasi dengan memberikan bobot lebih besar pada data yang salah diklasifikasikan [6]. *RobustBoost* juga digunakan untuk mengatasi data yang tidak relevan atau bersifat *noise*. *AdaBoost* memiliki variasi seperti *AdaBoost.M1* dan *AdaBoost.M2* yang berbeda dalam prosedur pembobotannya [13]. Dengan menggunakan metode ensambel seperti *bagging* dan *boosting*, peneliti dapat meningkatkan kinerja prediksi model, mengurangi kesalahan klasifikasi, dan menangani data yang tidak stabil atau tidak seimbang secara lebih efektif [14].

### C. Boosted Bagging (BoBag)

Algoritma BoBag berdasarkan Gambar 1 akan melakukan metode *boosting* untuk meningkatkan kinerja prediksi pada



Gambar 6. Persebaran nilai pada kolom kategorial berdasarkan status pelanggan.

setiap sampel. *Boosting* dilakukan untuk meningkatkan pengklasifikasi dasar yang lemah untuk membuatnya lebih kuat karena melatih pengklasifikasi lemah lebih mudah daripada melatih pengklasifikasi yang kuat. Pada tahap ini, metode *boosting* yang digunakan dapat berbeda-beda seperti yang diilustrasikan pada Gambar 2. Agregasi akhir pada metode ini dapat berupa teknik yang sudah ada, seperti pemungutan suara [6].

D. Metrik Evaluasi Model

Metrik evaluasi adalah faktor kunci dalam penilaian kinerja klasifikasi dan panduan pemodelan pengklasifikasi [15]. Terdapat lima pengukuran yang biasanya digunakan untuk mengevaluasi model yang diimplementasi pada CCP, yaitu akurasi, presisi, *recall*, *F1-Score*, dan AUC. Kelima metrik ini dapat diformulasikan dari *confusion matrix* yang dapat dilihat pada Tabel 1 [16].

III. METODOLOGI

A. Pengumpulan Data

Dataset yang digunakan pada tugas akhir ini adalah data pelanggan dari perusahaan telekomunikasi di California, Maven Communication, yang diambil pada Q2 tahun 2022 dan diunduh melalui situs <https://www.mavenanalytics.io/data-playground>. Dataset ini terdiri dua tabel terpisah yang meliputi tabel yang berisi memerinci aktivitas pelanggan telekomunikasi dan tabel jumlah populasi pada tiap kode pos. Namun, pada tugas akhir ini hanya akan menggunakan tabel aktivitas dari 7.043 pelanggan yang setiap barisnya mewakili satu pelanggan beserta karakteristik masing-masing pelanggan yang dijabarkan dalam 38 kolom yang ada. Dari 38 kolom tersebut, terdapat 21 kolom kategorial, 14 kolom numerikal, dan 2 kolom dengan tipe karakter.

B. Eksplorasi Data

Sebelum melakukan eksplorasi lebih jauh, terlebih dulu melakukan penghapusan pelanggan dengan status *joined* karena pada tugas akhir ini hanya berfokus pada pelanggan yang *churned* dan *stayed* saja. Selain itu, kolom terkait “Customer ID” dan kolom terkait lokasi pelanggan, seperti “City”, “Zip Code”, “Latitude”, dan “Longitude” juga akan dihapus karena kolom-kolom tersebut tidak akan digunakan

Tabel 2. Hasil Pembagian Dataset

Jenis Data	Ukuran Data	Persebaran Kelas
Data Awal	6.589	(71,63 : 28,37)
Data Latih	4.216	(71,63 : 28,36)
Data Validasi	1.055	(71,65 : 28,34)
Data Uji	1.318	(71,62 : 28,37)



(a)

(b)

Gambar 5. Fitur yang dihapus (a) dan yang tersisa (b) setelah proses seleksi fitur.

secara langsung untuk melakukan prediksi. Hasil penghapusan pelanggan menyisakan 6.589 pelanggan dengan persebaran 71,6% pelanggan *stayed* dan 28,4% pelanggan yang *churned*.

Selanjutnya, dicari tahu apa alasan pelanggan yang menyebabkannya untuk *churn*. Gambar 3 menunjukkan beragam alasan pelanggan berhenti berlangganan dengan yang terbanyak berkaitan langsung dengan kompetitor. Dari alasan ini, bisa dilihat jika kompetitor memiliki keunggulan layanan yang ditawarkan dibandingkan layanan kita, terutama yang terkait dengan layanan internet.

Persebaran pelanggan *churned* dan *stayed* pada kolom numerik divisualisasikan menggunakan histogram pada Gambar 4. Kolom “Tenure in Months” menggambarkan bahwa karakteristik pelanggan yang *churned* tertinggi adalah pelanggan yang pergi beberapa bulan (5 bulan) pertama setelah berlangganan. Hal ini sejalan dengan “Total Charges” yang dibebankan kepada pelanggan dan “Total Revenue” yang diperoleh dari pelanggan yang berhenti berlangganan relatif rendah dibandingkan dengan pelanggan yang tetap berlangganan. Pelanggan dengan biaya bulanan yang lebih tinggi cenderung lebih banyak yang *churned*. Dari kolom “Number of Dependents”, terlihat bahwa pelanggan dengan jumlah tanggungan yang lebih tinggi cenderung lebih jarang untuk *churn*, kemungkinan karena lebih banyak anggota keluarga yang memanfaatkan layanan.

Selain distribusi untuk kolom numerik, distribusi pelanggan juga dilihat untuk kolom kategori pada Gambar 5. Berdasarkan gambar tersebut, dapat disimpulkan bahwa persebaran jenis kelamin pelanggan antara perempuan dan laki-laki memiliki persebaran yang seimbang. Status perkawinan pada pelanggan yang belum menikah cenderung memiliki tingkat *churn* yang lebih tinggi dibandingkan dengan pelanggan yang sudah menikah. Kemudian, jika dilihat dari berbagai penawaran yang ditawarkan kepada pelanggan, lebih banyak pelanggan yang tidak mengambil penawaran (*offer*) dan lebih banyak pelanggan yang *churned* setelah menerima *Offer E*. Hal ini mengindikasikan bahwa promosi dan penawaran khusus bisa menjadi alat efektif untuk mengurangi *churn*. Untuk langganan layanan internet dan telepon, tingkat *churned* yang lebih tinggi ada pada pelanggan yang berlangganan internet atau telepon. Namun,

Tabel 3.  
Perbandingan Hasil Model Klasifikasi Tunggal

	<i>Decision Tree</i>	<i>Bagging</i>	<i>AdaBoost</i>
Akurasi	0,8020	0,8369	0,8498
Presisi	0,6619	0,7751	0,7604
Recall	0,6176	0,5990	0,6872
F1-Score	0,6400	0,6757	0,7219

Tabel 4.  
Alternatif Ruang Hyperparameter

Hyperparameter	Ruang Pencarian
n_estimator	[50, 100, 150, 200, 250]
max_samples	[0.1, 0.2, 0.3, 0.4, 0.5]
learning_rate	[0.001, 0.01, 0.1, 0.2, 0.3]

untuk pelanggan yang menggunakan layanan tambahan internet, seperti “Online Security”, “Online Backup”, “Device Protection Plan”, dan lain-lain dapat mengurangi tingkat *churned* pelanggan. Jenis internet *fiber optic* menjadi yang paling banyak digunakan dibandingkan tipe yang lainnya. Akan tetapi, pengguna yang tipe ini juga yang paling banyak berhenti berlangganan. Jika dilihat dari jenis kontrak yang diambil, pelanggan dengan jenis kontrak bulanan paling banyak *churned* karena jenis kontrak ini memiliki jangka waktu yang lebih pendek daripada yang tahunan sehingga lebih mudah bagi pelanggan untuk berpindah ke penyedia layanan yang lain.

### C. Praproses Data

Tahap praproses meliputi pengecekan dan penanganan nilai kosong dan pencilan, pengkodean fitur, seleksi fitur, pembagian data latih dan uji, serta terakhir normalisasi data. Tahap ini dilakukan untuk mentransformasi data mentah menjadi data yang siap digunakan untuk membangun model prediksi. Langkah pertama adalah pengecekan nilai kosong dilakukan untuk kolom layanan telepon, layanan internet, dan status pelanggan, serta kolom-kolom turunannya. Proses ini mencakup identifikasi perhitungan jumlah nilai kosong pada setiap kolom terkait. Selanjutnya, data yang memenuhi kondisi tertentu disaring dan jumlah nilai kosongnya dihitung kembali. Hasil dari proses ini menunjukkan bahwa jumlah nilai kosong pada ketiga kolom dan yang memenuhi kondisi (jika tidak berlangganan dan kolom turunannya bernilai kosong) memiliki jumlah yang sama, mendukung hipotesis awal. Hasil imputasi untuk kolom numerikal menunjukkan banyak pelanggan yang tidak berlangganan layanan telepon atau internet memiliki tingkat *churn* yang tinggi. Imputasi juga dilakukan untuk kolom-kolom kategorial dengan mengisinya sesuai dengan deskripsi kolom, seperti “Offer” yang diisi dengan *None* jika pelanggan tidak menerima tawaran.

Selanjutnya, proses penanganan *outlier* (pencilan) dilakukan dengan menggunakan metode interkuartil, yang bertujuan untuk mengidentifikasi dan menangani nilai-nilai yang berada di luar rentang yang wajar dari distribusi data. Metode imputasi menggunakan nilai median yang dipilih karena metode ini dapat digunakan pada data *skewed* yang tidak bergantung pada rata-rata atau standar deviasi dan tidak membuat asumsi distribusi [17]. Kolom “Number of Referrals”, “Avg Monthly GB Download”, “Total Long Distance Charges”, dan “Total Revenue” dipilih untuk dilakukan imputasi nilai median karena memiliki pencilan yang signifikan. Kolom-kolom yang semua nilainya adalah

pencilan dan tidak memiliki kelompok distribusi tertentu, seperti kolom “Number of Dependents”, “Total Refunds”, dan “Total Extra Data Charges” tidak dilakukan penanganan. Hal ini karena jika dilakukan imputasi dengan nilai median, hasilnya akan terpusat pada satu nilai saja, yaitu mediannya yang dapat mengubah struktur data dan tidak merepresentasikan data aslinya.

Pengkodean fitur menggunakan *one-hot-encoding* (OHE) pada kolom kategorial selain “Customer Status” dengan mengubah setiap nilai unik menjadi kolom baru dengan nilai biner (0/1). *Label encoding* (LE) diterapkan pada kolom “Customer Status,” mengubah status *churned* menjadi 1 dan *stayed* menjadi 0. Penggunaan OHE pada proses ini digunakan untuk memastikan setiap kategori diberi bobot sama dan menghindari bias numerik, sehingga semua kategori diperlakukan setara dalam model [18].

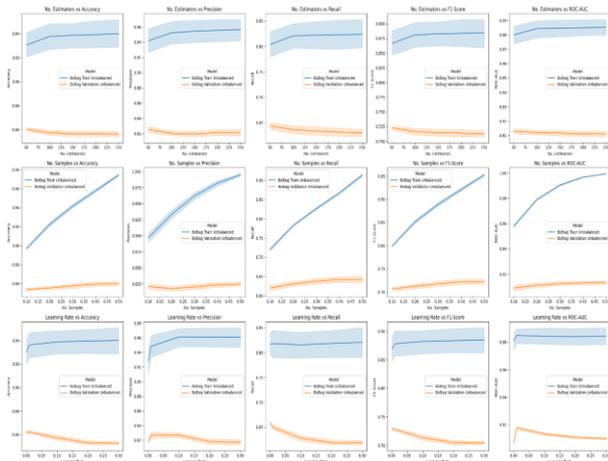
Setelah dilakukan pengkodean, fitur kembali diseleksi untuk memastikan bahwa ia memiliki korelasi dengan kelas *churned* berdasarkan kekuatan korelasi *Spearman*-nya dan korelasi antarfiturnya. Dari proses seleksi fitur ini, diketahui terdapat 19 fitur yang terindikasi redundan dan akan dilakukan penghapusan sehingga menyisakan 38 fitur. Adapun fitur-fiturnya yang dihapus maupun yang tersisa ditunjukkan pada Gambar 6 (a) dan (b).

Setelah memilih fitur-fitur yang akan digunakan, pembagian dataset awal dibagi menjadi data latih dan data uji dengan perbandingan 80:20. Metode *stratify* digunakan untuk memastikan bahwa proporsi kelas tetap sama seperti pada dataset asli. Selanjutnya, proses *hold out* dilakukan untuk membagi lagi data latih menjadi dua bagian: data latih dan data validasi, dengan perbandingan yang sama yaitu 80:20. Adapun hasil pembagian data ini seperti yang terlihat pada Tabel 2 menghasilkan persebaran kelas yang sama dengan data awal, yaitu sekitar 71,6% untuk kelas *stayed* dan 28,4% untuk kelas *churned*.

Tahap praproses terakhir sebelum membangun model adalah melakukan normalisasi data menggunakan *MinMaxScaler* untuk menstandarkan skala variabel-variabel yang memiliki varians tinggi menjadi nilai dengan rentang (0,1) [4], [19]. Nilai minimum dan maksimum yang sudah didapatkan dari inisialisasi *MinMaxScaler* kemudian diterapkan pada semua data X (variabel independen) untuk set data pelatihan, validasi, dan uji.

### D. Pembangunan Model Klasifikasi Tunggal

Pembangunan model klasifikasi tunggal dilakukan sebagai langkah awal sebelum implementasi model prediksi menggunakan *Boosted Bagging* (BoBag) untuk mendapatkan *baseline performance* dari model sederhana. Model yang digunakan meliputi *Decision Tree*, *Bagging*, dan *AdaBoost*. Ketiga model tersebut akan langsung dilakukan uji coba terhadap data tes menggunakan data latih tidak seimbang tanpa mengubah parameter *default*-nya. Hasil dari model klasifikasi tunggal ini dapat dilihat pada Tabel 3. Dari Tabel 3, terlihat bahwa nilai akurasi dan *F1-Score* mengalami peningkatan dari *Decision Tree*, *Bagging*, dan *AdaBoost*. Peningkatan ini dapat dikaitkan dengan metode ensambel yang digunakan dalam *Bagging* dan *AdaBoost*. *Bagging* mengurangi varians dengan merata-ratakan beberapa model (dalam hal ini *Decision Tree*), yang biasanya meningkatkan kinerja dibandingkan dengan *Decision Tree* tunggal.



Gambar 7. Efek perubahan *hyperparameter* terhadap metrik pada data pelatihan dan validasi.

AdaBoost, di sisi lain, berfokus pada pengurangan bias dan varians dengan meningkatkan model secara iteratif dengan memberikan bobot pada hasil klasifikasi yang salah, yang mengarah ke metrik kinerja keseluruhan yang lebih baik.

**E. Pembangunan Model Prediksi Berbasis Bagging**

Eksperimen pembangunan model ini diawali dengan proses *hyperparameter tuning* untuk menentukan kombinasi parameter terbaik yang akan digunakan dalam model. Dengan teknik *grid search*, setiap kemungkinan dalam ruang *hyperparameter* akan dicoba hingga diperoleh set *hyperparameter* yang paling optimal. *Hyperparameter* seperti jumlah estimator, ukuran sampel, dan *learning rate* ditentukan dan diatur dalam rentang nilai tertentu seperti pada Tabel 4.

Implementasi model dimulai dengan membangun dan melatih model menggunakan data latih dan validasi dengan berbagai kombinasi *hyperparameter*. Proses pelatihan dilakukan dengan melatih dan memprediksi data latih yang sama, sedangkan proses validasi membangun model menggunakan data latih untuk memprediksi data validasi. Visualisasi *lineplot* pada Gambar 7 menampilkan gambaran efek dari tiga *hyperparameter* terhadap berbagai metrik evaluasi, yaitu akurasi, presisi, *recall*, *F1-score*, dan ROC-AUC. Pada baris pertama, peningkatan jumlah estimator umumnya menghasilkan peningkatan dalam semua metrik evaluasi, menunjukkan bahwa lebih banyak estimator membantu model untuk belajar lebih baik dan menghasilkan prediksi yang lebih akurat, meskipun dengan beberapa fluktuasi yang lebih besar. Namun, pada hasil validasi peningkatan jumlah estimator membuat kinerja model stabil cenderung menurun. Hal ini mungkin disebabkan oleh model yang terlalu kompleks sehingga tidak dapat generalisasi dengan baik pada data yang tidak terlihat sebelumnya. Untuk peningkatan jumlah sampel juga berkontribusi positif pada semua metrik evaluasi dari hasil pelatihan maupun validasi. Hal ini bisa terjadi karena lebih banyak data yang digunakan dalam setiap iterasi model sehingga memperbaiki kemampuan model dalam membuat prediksi yang akurat. Baris ketiga memperlihatkan efek dari *learning rate*. Dari hasil pelatihan, peningkatan *learning rate* secara signifikan meningkatkan metrik evaluasi, tetapi *learning rate* yang terlalu tinggi dapat menyebabkan *overfitting* dan menurunkan kinerja model pada data validasi. Ini menunjukkan adanya *sweet spot* untuk *learning rate*. Selanjutnya pencarian

Tabel 5.

Hasil *Hyperparameter* Terbaik dari Tahap Pelatihan & Validasi

Tahap	Pelatihan	Validasi
Jumlah Estimator	250	200
Jumlah Sampel	0,5	0,5
<i>Learning Rate</i>	0,3	0,01
Akurasi	0,9834	0,8692
Presisi	0,9991	0,8397
<i>Recall</i>	0,9423	0,6656
<i>F1-Score</i>	0,9699	0,7425
ROC-AUC	0,9999	0,9203
Waktu Komputasi (s)	243,44	616,81

*hyperparameter* terbaik dilakukan berdasarkan nilai *F1* tertinggi dengan hasilnya ditampilkan pada Tabel 5.

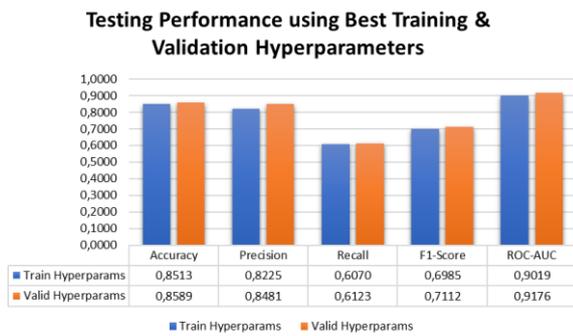
Dengan jumlah sampel yang sama, tahap pelatihan memerlukan lebih banyak estimator dan *learning rate* lebih tinggi, menunjukkan model yang lebih kompleks dan agresif dibandingkan dengan tahap validasi. Hal ini menunjukkan bahwa *learning rate* yang tinggi dapat menyebabkan *overfitting* pada data pelatihan, sementara *learning rate* lebih rendah pada tahap validasi menunjukkan generalisasi yang lebih baik tetapi dengan kinerja metrik yang lebih rendah. Model dengan banyak estimator mampu menangkap lebih banyak variasi dalam data, namun meningkatkan waktu komputasi. Hasil *hyperparameter* dari tahap validasi lebih baik digunakan untuk pengujian karena menunjukkan kemampuan yang lebih baik dalam menangani data baru, mengurangi risiko *overfitting*, dan memberikan prediksi yang lebih akurat dan dapat diandalkan.

**F. Uji Coba dan Evaluasi Model**

Pengujian ke data uji dilakukan menggunakan *hyperparameter* terbaik dari hasil pelatihan maupun validasi untuk membandingkan kinerja model dalam berbagai kondisi. Hasilnya ditunjukkan pada Gambar 8. Dari gambar tersebut, tetapi pengujian dengan *hyperparameter* dari tahap validasi sedikit lebih unggul di semua metrik evaluasi.

Model dengan *hyperparameter* validasi memiliki nilai akurasi yang sedikit lebih tinggi (0,8589) daripada model yang diuji menggunakan *hyperparameter* pelatihan (0,8513). Ini berarti model mampu memberikan prediksi yang akurat untuk setiap kelas yang ada karena akurasi mengindikasikan persentase prediksi yang benar dari total keseluruhan prediksi. Nilai *F1-Score* hasil pengujian *hyperparameter* validasi juga lebih tinggi (0,7112) dibandingkan dengan model pelatihan (0,6985) yang menunjukkan keseimbangan lebih baik antara presisi dan *recall*, sehingga model validasi lebih andal dalam prediksi *churn* yang seimbang dan akurat [20]. Jika dilihat dari nilai ROC-AUC, model validasi (0,9176) juga lebih tinggi dibandingkan model pelatihan (0,9019) yang berarti bahwa model validasi juga lebih baik dalam membedakan antara kelas positif (*churned*) dan negatif (*stayed*) secara jelas, mengurangi jumlah kesalahan yang terjadi karena kebingungan antara kedua kelas.

Jika dibandingkan dengan hasil evaluasi klasifikasi tunggal pada Tabel 3, terlihat bahwa model BoBag cukup efektif dalam memprediksi *churned* dan *stayed* secara keseluruhan dan cukup baik dalam mengurangi *false positive*. Hal ini ditandai dengan unggulnya nilai akurasi dan presisinya. Namun, jika dilihat dari nilai *recall* dan *F1-Score*, model



Gambar 9. Hasil evaluasi pengujian menggunakan *hyperparameter* pelatihan.

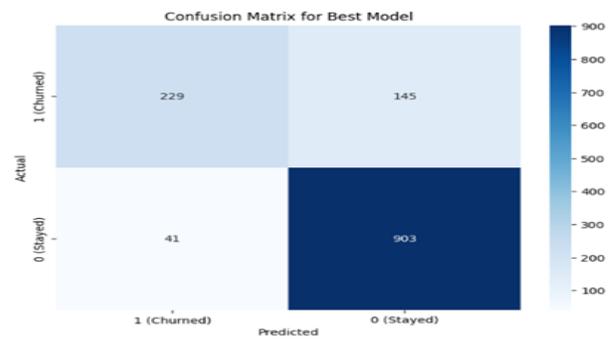
*AdaBoost* masih sedikit lebih tinggi (0,6872 untuk *recall* dan 0,7219 untuk *F1-Score*) dibandingkan hasil uji coba model *BoBag* menggunakan *hyperparameter* dari hasil pelatihan (0,6070 dan 0,6985) dan validasi (0,6123 dan 0,7112). Hal ini mungkin disebabkan oleh perbedaan dalam cara kedua model menangani data ketidakseimbangan. *AdaBoost* cenderung lebih efektif dalam meningkatkan kinerja model lemah dengan memberikan bobot lebih besar pada kesalahan, sehingga lebih mampu menangkap pola dari kelas minoritas (*churn*). Sebaliknya, *BoBag* yang menggabungkan *bagging* dan *boosting* mungkin masih terpengaruh oleh variasi dalam data latih, meskipun sudah dioptimalkan dengan *hyperparameter* terbaik. Hal ini menunjukkan bahwa meskipun *BoBag* dapat memberikan hasil yang seimbang antara akurasi dan presisi, tetapi kemampuan untuk mendeteksi pelanggan *churn* (*recall*) masih lebih unggul pada model *AdaBoos*.

Secara lebih detail, **Error! Reference source not found.** menampilkan *confusion matrix* dari model terbaik, yaitu model *Boosted Bagging* yang dilatih menggunakan *hyperparameter* hasil validasi. Model tersebut mampu memprediksi sekitar 229 pelanggan *churned* aktual (*true positive*) dan 903 pelanggan *stayed* aktual (*true negative*). Dengan kata lain, dari 374 pelanggan *churned*, model ini salah mengklasifikasikan 145 sebagai *stayed* (*false negative*), sedangkan dari 944 pelanggan *stayed*, model ini salah mengklasifikasikan 41 sebagai *churned* (*false positive*).

Selanjutnya, dilakukan juga *feature importance* dengan menghitung *permutation importance* untuk menilai pentingnya setiap fitur dalam mempengaruhi prediksi model. Hasilnya menunjukkan bahwa fitur “*Contract\_Month-to-Month*”, “*Total Charges*”, dan “*Monthly Charge*” memiliki nilai *mean importance* yang tinggi, yang berarti kontrak bulan dan biaya yang dibayarkan pelanggan memainkan peran penting dalam memprediksi *churn*.

#### IV. KESIMPULAN

Terdapat dua kesimpulan yang didapatkan dari hasil pengujian pembangunan model prediksi menggunakan algoritma *Boosted Bagging*, yaitu: (1) Model *Boosted Bagging* (*BoBag*) menunjukkan kinerja prediksi *churn* yang baik, khususnya ketika menggunakan *hyperparameter* model terbaik dari tahap validasi (jumlah estimator = 200, jumlah sampel = 0,5, dan *learning rate* = 0,01). Hasil pengujian menggunakan *hyperparameter* tersebut memberikan nilai akurasi, *F1-Score* dan AUC yang lebih tinggi daripada model yang dibangun menggunakan *hyperparameter* model yang



Gambar 8. *Confusion matrix* dari model terbaik.

dihasilkan dari tahap pelatihan. Adapun nilai akurasi, *F1-Score*, dan AUC berurutan-turut pada model terbaik adalah 0,8589; 0,7112; dan 0,9176. Sedangkan nilai akurasi, *F1-Score* dan AUC pada model yang menggunakan *hyperparameter* dari tahap pelatihan berturut-turut adalah 0,8513; 0,6985; 0,9016. Akurasi, *F1-Score*, dan AUC yang lebih tinggi menunjukkan bahwa model memiliki kemampuan lebih baik dalam memprediksi kasus *churned* dan *stayed* secara keseluruhan. Selain itu, model yang dihasilkan mengindikasikan keseimbangan yang lebih baik antara presisi dan *recall* dan memiliki kemampuan yang lebih baik dalam membedakan antara kelas *churned* dan *stayed* secara keseluruhan. (2) Hasil prediksi *churn* menggunakan model *BoBag* terbaik menunjukkan bahwa fitur “*Contract\_Month-to-Month*”, “*Total Charges*”, dan “*Monthly Charge*” memiliki pengaruh yang signifikan dalam memprediksi *churn*. Pelanggan dengan kontrak bulanan memiliki kemungkinan *churn* yang lebih tinggi dibandingkan dengan pelanggan dengan kontrak tahunan atau dua tahunan. Oleh karena itu, strategi retensi yang dapat disarankan adalah memberikan insentif atau diskon kepada pelanggan dengan kontrak bulanan untuk beralih ke kontrak yang lebih panjang, sambil tetap meningkatkan layanan untuk mempertahankan pelanggan dan mampu bersaing dengan kompetitor di bidangnya.

#### DAFTAR PUSTAKA

- [1] V. Haridasan and S. Venkatesh, “Crm implementation in Indian telecom industry – evaluating the effectiveness of mobile service providers using data envelopment analysis,” *Int. J. Bus. Res. Manag. (IJBRM), Comput. Sci. Journals (CSC Journals)*, vol. 2, no. 3, pp. 110–127, 2011.
- [2] S. M. Shrestha and A. Shakya, “A customer churn prediction model using XGBoost for the telecommunication industry in Nepal,” *Procedia Comput. Sci.*, vol. 215, pp. 652–661, 2022, doi: 10.1016/j.procs.2022.12.067.
- [3] A. K. Ahmad, A. Jafar, and K. Aljoumaa, “Customer churn prediction in telecom using machine learning in big data platform,” *J. Big Data*, vol. 6, no. 1, p. 28, Dec. 2019, doi: 10.1186/s40537-019-0191-6.
- [4] 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, Jul. 2023, doi: 10.1007/s42452-023-05389-6.
- [5] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. C. Chatzisavvas, “A comparison of machine learning techniques for customer churn prediction,” *Simul. Model. Pract. Theory*, vol. 55, pp. 1–9, Jun. 2015, doi: 10.1016/j.simpat.2015.03.003.
- [6] S. Tavassoli and H. Koosha, “Hybrid ensemble learning approaches to customer churn prediction,” *Kybernetes*, vol. 51, no. 3, pp. 1062–1088, Feb. 2022, doi: 10.1108/K-04-2020-0214.
- [7] K.-W. Hsu, “A theoretical analysis of why hybrid ensembles work,” *Comput. Intell. Neurosci.*, vol. 2017, pp. 1–12, 2017, doi: 10.1155/2017/1930702.
- [8] A. Perianez, A. Saas, A. Guitart, and C. Magne, “Churn prediction in mobile social games: towards a complete assessment using survival

- ensembles,” in *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, Oct. 2016, Quebec, Canada, pp. 564–573, doi: 10.1109/DSAA.2016.84.
- [9] P. K. Dalvi, S. K. Khandge, A. Deomore, A. Bankar, and V. A. Kanade, “Analysis of customer churn prediction in telecom industry using decision trees and logistic regression,” in *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*, Mar. 2016, Indore, India, pp. 1–4, doi: 10.1109/CDAN.2016.7570883.
- [10] S. W. Fujo, S. Subramanian, and M. A. Khder, “Customer churn prediction in telecommunication industry using deep learning,” *Inf. Sci. Lett.*, vol. 11, no. 1, pp. 185–198, Jan. 2022, doi: 10.18576/isl/110120.
- [11] L. Breiman, “Bagging predictors,” *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996, doi: 10.1007/BF00058655.
- [12] D. Opitz and R. Maclin, “Popular ensemble methods: an empirical study,” *J. Artif. Intell. Res.*, vol. 11, pp. 169–198, Aug. 1999, doi: 10.1613/jair.614.
- [13] P. Halimah and Y. Ekawati, “Penerapan metode Taguchi untuk meningkatkan kualitas bata ringan pada UD. XY Malang,” *Jiems (Journal Ind. Eng. Manag. Syst.)*, vol. 13, no. 1, Feb. 2020, doi: 10.30813/jiems.v13i1.1694.
- [14] P. Flach, *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*. United Kingdom: Cambridge University Press, ISBN: 9781107096394, 2012.
- [15] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, “A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches,” *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.)*, vol. 42, no. 4, pp. 463–484, Jul. 2012, doi: 10.1109/TSMCC.2011.2161285.
- [16] M. Fathian, Y. Hoseinpoor, and B. Minaei-Bidgoli, “Offering a hybrid approach of data mining to predict the customer churn based on bagging and boosting methods,” *Kybernetes*, vol. 45, no. 5, pp. 732–743, May 2016, doi: 10.1108/K-07-2015-0172.
- [17] B. Lglewicz and D. C. Hoaglin, *How to Detect and Handle Outliers*. USA: ASQC, ISBN: 0-87389-247-X, 1993.
- [18] M. Kuhn and K. Johnson, “An Introduction to Feature Selection,” in *Applied Predictive Modeling*. New York: Springer, ISBN: 978-1461468486, 2013, pp. 487–519.
- [19] M. Karanovic, M. Popovac, S. Sladojevic, M. Arsenovic, and D. Stefanovic, “Telecommunication services churn prediction - deep learning approach,” in *2018 26th Telecommunications Forum (TELFOR)*, Nov. 2018, pp. 420–425, doi: 10.1109/TELFOR.2018.8612067.
- [20] M. Kubat, R. C. Holte, and S. Matwin, “Machine learning for the detection of oil spills in satellite radar images,” *Mach. Learn.*, vol. 30, no. 2/3, pp. 195–215, 1998.