Djtechno : Jurnal Teknologi Informasi
Vol. 6, No. 2, Agustus 2025
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

HYBRID DEEP LEARNING RANDOM FOREST OPTIMASI PEMILIHAN FITUR UNTUK PREDIKSI *CHURN* INDUSTRI TELEKOMUNIKASI

Atika Mutiarachim[1], Dhendra Marutho[2], Nur Atika Yuniarti[3], Ryan Arya Pramudya[4], Jaluanto Sunu Punjul Tyoso[5]
1,5) Bisnis Digital, Fakultas Ekonomika dan Bisnis, Universitas 17 Agustus 1945 Semarang
2) Informatika, Fakultas Ilmu Komputer, Universitas Muhammadiyah Semarang
3,4) Manajemen, Fakultas Ekonomika dan Bisnis, Universitas 17 Agustus 1945 Semarang

Article Info

Article history:

Received: 01 Juli 20205
Revised: 16 Juli 20205
Accepted: 22 Juli 20205

ABSTRACT

Abstrak

Customer churn merupakan tantangan kritis dalam industri telekomunikasi yang berdampak signifikan terhadap profitabilitas perusahaan. Penelitian ini mengusulkan pendekatan *hybrid machine learning* untuk memprediksi *customer churn* dengan mengintegrasikan *deep learning* dan *random forest* serta mengoptimalkan performa melalui seleksi fitur *chi square* dan *information gain*. Dataset IBM Telco *Customer churn* yang terdiri dari 7.043 sampel dengan 31 atribut digunakan dalam penelitian ini. Metodologi penelitian meliputi *preprocessing* data, implementasi 10-*fold cross validation*, aplikasi metode seleksi fitur, dan evaluasi performa menggunakan *confusion matrix* serta metrik klasifikasi biner. Hasil penelitian menunjukkan bahwa implementasi seleksi fitur secara signifikan meningkatkan akurasi prediksi, di mana akurasi tanpa seleksi fitur mencapai 97.00% (*Deep Learning*) dan 98.68% (*Random Forest*), sedangkan dengan seleksi fitur *chi square* meningkat menjadi 97.97% (*Deep Learning*) dan 98.72% (*Random Forest*). Performa terbaik dicapai oleh kombinasi *Random Forest* dengan seleksi fitur *information gain* yang menghasilkan akurasi 98.75%, *precision* 98.37%, *recall* 99.96%, dan F-*measure* 99.16%. Temuan ini membuktikan efektivitas kombinasi algoritma *ensemble* dengan teknik seleksi fitur dalam mengoptimalkan prediksi *customer churn* untuk mendukung strategi retensi pelanggan yang lebih tepat sasaran

Kata Kunci: Telco Customer churn, Deep Learning, Random Forest, Chi Square, Information Gain

Abstract

Customer churn represents a critical challenge in the telecommunications industry with significant impact on corporate profitability. This research proposes a hybrid machine learning approach for customer churn prediction by integrating deep learning and random forest algorithms, optimized through chi square and information gain feature selection techniques. The IBM Telco Customer churn dataset comprising 7,043 samples with 31 attributes was utilized in this study. The research methodology encompasses data preprocessing, 10-fold cross validation implementation, feature selection methods application, and performance evaluation using confusion matrix and binary classification metrics. Results demonstrate that feature selection implementation significantly enhances prediction accuracy, where accuracy without feature selection achieved 97.00% (Deep Learning) and 98.68% (Random Forest), while chi square feature selection improved performance to 97.97% (Deep Learning) and 98.72% (Random Forest). The optimal performance was achieved by Random Forest combined with information gain feature selection, yielding 98.75% accuracy, 98.37% precision, 99.96% recall, and 99.16% F-measure. These findings validate the effectiveness of ensemble algorithm combinations with feature selection techniques in optimizing customer churn prediction to support more targeted customer retention strategies.

Keywords: Telco Customer churn, Deep Learning, Random Forest, Chi Square, Information Gain

Djtechno : Jurnal Teknologi Informasi
Vol. 6, No. 2, Agustus 2025
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Corresponding Author:
E-mail: amutiarachim@gmail.com

1. PENDAHULUAN

Era digitalisasi mengubah *landscape* industri telekomunikasi secara fundamental [1], menciptakan persaingan yang semakin ketat di antara penyedia layanan telekomunikasi. Revolusi teknologi informasi dan komunikasi membuka peluang bagi munculnya berbagai operator baru dengan penawaran layanan yang beragam dan kompetitif. Indonesia mengalami pertumbuhan cepat pada jumlah perusahaan penyedia layanan komunikasi, tahun 2018 berjumlah 331 dan pada tahun 2023 mencapai 1011 [2][3]. Hal tersebut menciptakan kondisi pasar yang jenuh, persaingan intensif, yang sekaligus menimbulkan dampak *customer churn* atau perpindahan pelanggan. Hilangnya pelanggan menjadi ancaman serius bagi keberlangsungan bisnis perusahaan.

Customer churn merupakan fenomena dimana pelanggan memutuskan untuk menghentikan penggunaan layanan dan beralih ke competitor, yang berdampak langsung terhadap penurunan *revenue* dan *market share* Perusahaan [4]. Biaya untuk memperoleh pelanggan baru dapat mencapai 5-25 kali lebih mahal dibandingkan mempertahankan pelanggan yang sudah ada. Pelayanan yang buruk kepada pelanggan dapat menyebabkan pelanggan yang tidak memiliki niat untuk berhenti berlangganan, menjadi berhenti berlangganan dan beralih ke produk competitor [5]. Riset dari sebuah Perusahaan penyedia layanan telekomunikasi digital memberikan fakta bahwa pelanggan churn yang tidak direncanakan dapat menyebabkan kerugian hingga $35.5 sampai dengan $168 milliar per tahun, bahkan jumlahnya dapat melebihi pelanggan yang berencana untuk melakukan churn, yaitu mencapai 43.3 juta pelanggan [6]. Peningkatan customer *retention rate* sebesar 5% data meningkatkan profit Perusahaan hingga 25-95%.

Pada konteks industri telekomunikasi Indonesia, tingkat *churn rate* yang tinggi mengindikasikan urgensi penerapan strategi retensi pelanggan yang efektif [7]. Angka *churn* tinggi dapat mengindikasikan adanya permasalahan fundamental pada produk

Djtechno : Jurnal Teknologi Informasi
Vol. 6, No. 2, Agustus 2025
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

atau layanan yang ditawarkan [8]. Kemampuan untuk memprediksi *customer churn* secara akurat menjadi *key point* dalam mengembangkan program retensi yang tepat sasaran dan *cost effective*. Prediksi yang akurat memungkinkan perusahaan mengidentifikasi pelanggan yang berisiko tinggi melakukan *churn*, sehingga dapat dilakukan intervensi proaktif sebelum pelanggan tersebut benar-benar berhenti menggunakan produk atau layanan perusahaan [9].

Perkembangan teknologi *machine learning* dan *artificial intelligence* membuka peluang baru dalam pengembangan model prediksi *customer churn* yang lebih akurat dan *robust* [10]. Deep learning dan Random Forest sebagai subset *machine learning* umumnya memberikan performa yang baik dalam menangani data dengan kompleksitas tinggi dan pola non-linear, namun performa klasifikasi tetap bergantung pada kualitas dan relevansi fitur yang digunakan sebagai input.

Seleksi fitur merupakan tahap p*reprocessing* yang krusial dalam pengembangan model *machine learning*, khususnya untuk dataset dengan dimensi tinggi. Metode seleksi fitur yang tepat dapat meningkatkan akurasi model, mengurangi *computational cost* dan mencegah *overfitting*. Dua metode seleksi fitur yang telah terbukti efektif yaitu *chi square* dan *information gain* [11] [12]*. Chi square* mengukur tingkat ketergantungan antara fitur kategorik dengan variable target [13], [14], [15]. *Information gain* mengukur pengurangan *entropy* yang dihasilkan suatu fitur [16], [17], [18]. Penelitian terdahulu melakukan pengujian klasifikasi data *churn customer* untuk menemukan akurasi terbaik.

Table 1.Penelitian Terdahulu

No	Jurnal	Dataset	Selection	Classifiers	Pembagian Data	Akurasi Terbaik
1	[19]	IBM Telco & Cell2cell	Lasso Regularization & XGBoost	Deep-BP-ANN	Simple Holdout (SH) & 10-fold cross validation	SH, IBM-88.12%, Cell2cell-79.38%
2	[20]	telco_dataset.csv	-	kNN	Split 80:20	81%
3	[21]	Perusahaan Fashion Campus	Ada, namun tidak dijelaskan	RF	Cross Validation	100%
4	[22]	Kaggle bank-customers	Anova & Chi square	RF, RT, Linear SVC, LR, kNN	-	RF- 86%
5	[23]	IBM Telco *customer churn*	CFS & RFE	CNN & RF	Split 60:40, 70:30, 80:20	CNN&CFS 80:20 -98% pada data test, 99% pada data validation

Djtechno : Jurnal Teknologi Informasi
Vol. 6, No. 2, Agustus 2025
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

6	[24]	Kaggle blastchar/telco-customerchurn	-	DNN, RF, kNN, DT	Split 90:10	DNN 3 hidden layer-83.09%
7	[25]	Kaggle blastchar/telco-customerchurn	SFS, SBS, SFFS, SBFS	Naïve Bayes	10-fold cross validation	77.74% pada seluruh selection namun hanya dengan 2 fitur
8	[26]	IBM Telecom's Kaggle Dataset	Listing features (seleksi fitur berdasarkan kepentingan)	GradientBoost, AdaBoost, XGBoost, ANN, LR, RF	cross validation	XGBoost-82.20%
9	[27]	IBM Watson Telco-Customer-Churn Kaggle	Pearson Correlation	DT, kNN, RF	Tidak disebutkan secara spesifik	99%-RF
10	[28]	Telecom industry *customer churn*	Gravitational Search Algorithm	LR, DT, Adaboost, Knn, RF, Naïve Bayes, SCM, XGBoost, CatBoost	Split 80:20 & k-fold cross validation untuk train set	81.71%-Adaboost & XGBoost
11	[29]	IBM Telecom's Kaggle Dataset	PSO	DT C.4.5	Split 80:20	91.55%

Penelitian terdahulu menunjukkan penerapan *machine learning* dalam prediksi *customer churn* dengan variasi performa yang cukup signifikan, mulai dari 77.74% hingga 100%. Mayoritas penelitian membuktikan bahwa penerapan seleksi fitur mampu meningkatkan akurasi klasifikasi. Beberapa penelitian telah mengekplorasi metode seleksi fitur dengan algoritma klasifikasi, namun eksplorasi yang *menggabungkan deep learning* dan *random forest* dengan metode seleksi fitur *chi square* dan *information gain* masih terbatas.

Penelitian sebelumnya juga menunjukkan bahwa tidak ada algoritma tunggal yang konsisten superior, dimana CNN mencapai 98-99%, RF bervariasi dari 86%-100%, XGBoost mencapai 82.20% dan DT dengan optimasi mencapai 91.55%. Hal tersebut menjadi dasar *research gap* dalam penelitian ini yaitu eksplorasi spesifik kombinasi *deep learning* dan *random forest* dengan seleksi fitur *chi square* dan *information gain* untuk mengoptimalkan performa klasifikasi, yang dapat menjadi *knowledge* untuk mengembangkan strategi retensi pelanggan yang lebih efektif.

Djtechno : Jurnal Teknologi Informasi
Vol. 6, No. 2, Agustus 2025
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

2. METODE PENELITIAN

Penelitian ini menggunakan dataset IBM Telco *customer churn* diperoleh dari websitehttps://github.com/Pranjali/Telco_Customer_Churn_Analysis/tree/main/Data%20Source dengan total 33 atribut. Atribut Churn Label merupakan *class*/label, berisi Yes (pelanggan berhenti menggunakan produk) dan No (pelanggan tidak berhenti menggunakan produk). Atribut Churn Value di *exclude* karena merupakan *encoding* dari atribut label, sehingga jika digunakan menjadi atribut *dummy* dan tidak memberikan dampak signifikan. Total atribut yang digunakan sebanyak 31 atribut dan 1 label.

Table 2. *Atribut*

No	Nama Atribut	Tipe Data/ Role	Keterangan
1	Customer ID	P/id	ID pelanggan
2	Count	I	Jumlah pelanggan
3	Country	P	Negara
4	State	P	Negara bagian
5	City	P	Kota asal
6	Zip Code	P	Kode pos
7	Lat Long	P	Garis Bujur Lintang
8	Latitude	R	Garis Lintang
9	Longitude	R	Garis Bujur
10	Gender	B	Jenis kelamin
11	Senior Citizen	B	Penghuni Lama
12	Partner	B	Pasangan yang dimiliki pelanggan
13	Dependents	B	Tanggungan yang dimiliki pelanggan
14	Tenure Months	I	Tenor sewa dalam bulan
15	Phone Service	B	Layanan telepon
16	Multiple Lines	P	Layanan multi-line
17	Internet Service	P	Layanan internet
18	Online Security	P	Pengamanan jaringan
19	Online Backup	P	Layanan cadangan dalam jaringan
20	Device Protection	P	Paket perlindungan perangkat
21	Tech Support	P	Pelayanan teknikal support
22	Streaming TV	P	Layanan streaming TV
23	Streaming Movie	P	Layanan streaming film
24	Contract	P	Jenis kontrak pelanggan
25	Paperless Billing	B	Pembayaran tanpa surat tagihan
26	Payment Method	P	Metode pembayaran oleh pelanggan
27	Monthly Charges	R	Tagihan bulanan
28	Total Charges	R	Total tagihan
29	Churn Score	I	Skor Churn
30	CLTV	I	Customer Lifetime Value
31	Churn Reason	P	Alasan berhenti berlangganan
32	Churn Label (Yes dan No)	B/ label	Pelanggan berhenti berlangganan

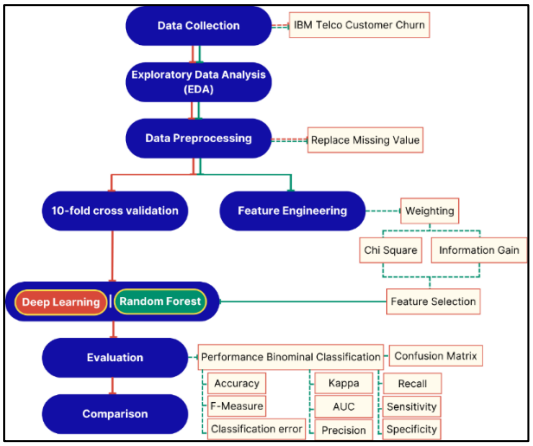
I: Integer
B: Binominal
P: Polynominal

Penelitian dilakukan menggunakan *tool* Altair AI Studio. Proses *data collection* dilakukan dengan mengunduh dataset, kemudian dilakukan EDA untuk memahami pola data, menemukan *outlier,* dan memahami hubungan antar atribut dan apa pengaruhnya pada Churn Label. Hasil EDA ditemukan *imbalance* jumlah data dan terdapat *missing value.* Label terdiri dari data Yes sebanyak 5174 dan No 1869. Perbedaan jumlah data

Djtechno : Jurnal Teknologi Informasi
Vol. 6, No. 2, Agustus 2025
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

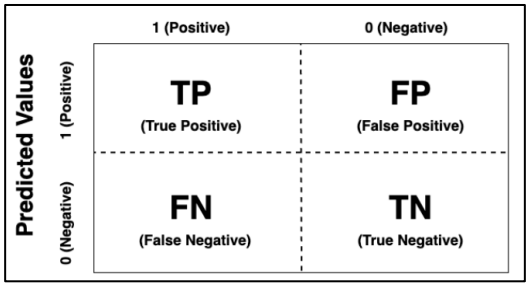
Yes dan No dengan nilai *imbalance ratio* sebesar 26.5%, termasuk dalam kategori *mild imbalance* pada klasifikasi biner, sehingga tidak perlu dilakukan *balancing data* [30], [31]. Tahap *preprocessing* dilakukan untuk memastikan data konsisten. Atribut Total Charges memiliki 11 data dengan *missing value*. Operator Replace Missing Value *with average* digunakan untuk mengisi data yang hilang dengan nilai rata-rata kolom.

Pembagian data menggunakan metode 10-*fold cross validation.* Pengujian pertama dilakukan tanpa seleksi fitur. Pengujian kedua dengan seleksi fitur, dilakukan dengan mengukur bobot atribut terlebih dahulu. Hasil pembobotan digunakan untuk menyeleksi atribut apa saja yang digunakan dalam algoritma klasifikasi DL dan RF.



Gambar 1 Alur Penelitian

Operator Performance Binominal Classification dipilih karena sesuai dengan label dataset yang bertipe *binominal*. Hasil *performance* terdiri dari *confusion matrix, accuracy,* F-*measure, classification error,* Kappa, AUC, *precision, recall, sensitivity* dan *specifity.* Evaluasi dilakukan dengan membandingkan *performance* terbaik.



Gambar 2 Confusion Matrix

Confusion matrix berisikan jumlah nilai prediksi dan nilai actual, untuk mengetahui kemampuan algoritma dalam memprediksi data dengan benar, nilai TP, FP, FN dan TN digunakan untuk menghitung perolehan *accuracy,* F-*measure, precision, recall, sensitivity* dan *specifity* [30].

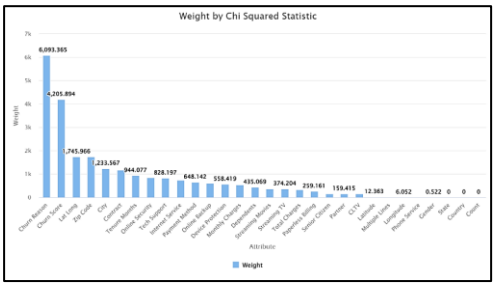
Djtechno : Jurnal Teknologi Informasi
Vol. 6, No. 2, Agustus 2025
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

3. HASIL DAN PEMBAHASAN

Perhitungan bobot dihitung untuk 30 atribut, karena atribut Customer ID di *set role* sebagai id dan atribut Customer Label merupakan label. Grafik bobot atribut terdapat pada gambar 3 dan 4. Pemilihan atribut dilakukan dengan mereduksi atribut dengan bobot terkecil, bertujuan untuk mengurangi pengaruh yang kurang relevan pada dataset, sehingga kinerja klasifikasi lebih optimal dan efisien [11]. Reduksi dapat dilakukan dengan menghilangkan atribut bernilai 0, atau berdasarkan batas nilai tertentu. Penelitian ini melakukan uji klasifikasi dengan menghapus atribut bernilai 0, namun *performance* yang dihasilkan kurang maksimal. Uji coba selanjutnya dilakukan perhitungan median untuk menentukan nilai tengah dari masing-masing bobot, kemudian dilakukan reduksi pada atribut dengan bobot kurang dari median. Uji klasifikasi dengan bobot atribut ≥ median menghasilkan *performance* yang lebih baik.

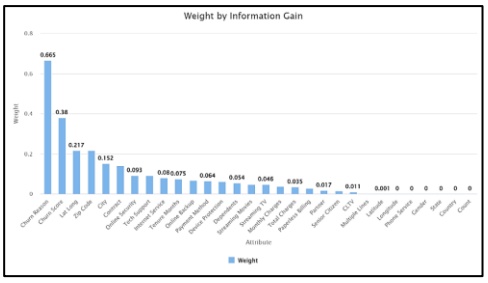
Hasil perhitungan bobot *chi square* pada 30 atribut adalah 0 sampai 6093.365. Atribut pada urutan 15 dan 16 menjadi median, Dependents dan Streaming Movie, dengan bobot 435.069 dan 375.661, perhitungan median sebagai berikut:

Median = $\frac{(435.069+375.661)}{2}$ = 405.365, sehingga dipilih atribut dengan bobot ≥ 405.365.



Gambar 3 Bobot Seleksi Fitur Chi Square

Pada perhitungan bobot *information gain*, dihasilkan bobot 0 sampai 0.665 untuk 30 atribut. Median dari perhitungan bobot *information gain* adalah 0.046, berasal dari median bobot atribut Streaming Movie dan Streaming TV sebagai urutan ke 15 dan 16, yang keduanya bernilai 0.046, sehingga diambil atribut dengan bobot ≥ 0.046.



Gambar 4 Bobot Seleksi Fitur Information Gain

Djtechno : Jurnal Teknologi Informasi
Vol. 6, No. 2, Agustus 2025
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

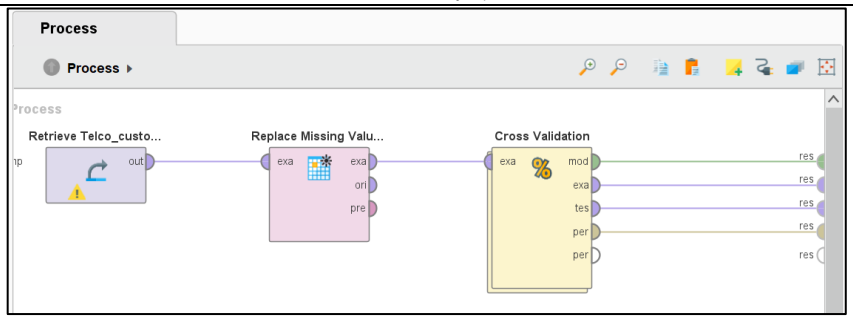
Tabel 3 menjelaskan atribut-atribut yang digunakan pada setiap pengujian. Tanda x menunjukkan atribut yang di reduksi setelah pembobotan, sehingga tidak digunakan dalam klasifikasi. Pada pengujian pertama, seluruh atribut digunakan, pada pengujian dengan seleksi chi squae, dan information gain, terdapat 15 atribut di reduksi. *Chi square* tidak menggunakan atribut Straming TV, *information gain* tidak menggunakan Monthly Charges, 14 atribut yang tidak digunakan pada keduanya adalah sama.

Table 3. *Atribut* pada Uji Klasifikasi

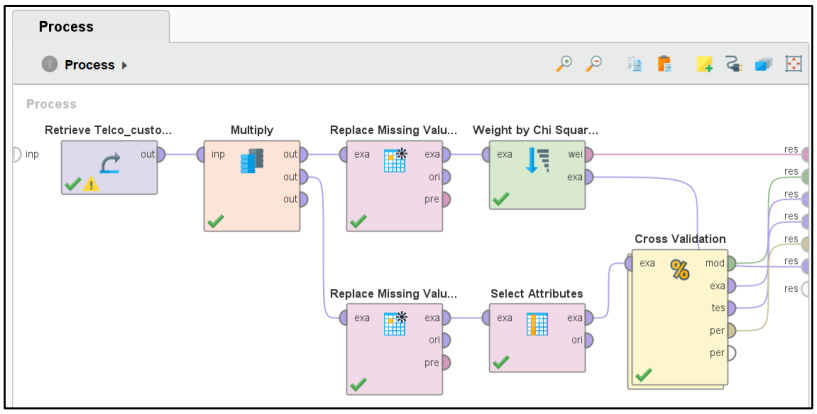
No	Nama Atribut	Tanpa Seleksi Fitur	Seleksi Fitur	
			Chi Square	Information Gain
1	Customer ID	v	x	X
2	Count	v	x	X
3	Country	v	x	X
4	State	v	x	X
5	City	v	v	V
6	Zip Code	v	v	V
7	Lat Long	v	v	V
8	Latitude	v	x	X
9	Longitude	v	x	X
10	Gender	v	x	X
11	Senior Citizen	v	x	X
12	Partner	v	x	X
13	Dependents	v	v	V
14	Tenure Months	v	v	v
15	Phone Service	v	x	x
16	Multiple Lines	v	x	x
17	Internet Service	v	v	v
18	Online Security	v	v	v
19	Online Backup	v	v	v
20	Device Protection	v	v	v
21	Tech Support	v	v	v
22	Streaming TV	v	x	v
23	Streaming Movie	v	v	v
24	Contract	v	v	v
25	Paperless Billing	v	x	x
26	Payment Method	v	v	v
27	Monthly Charges	v	v	X
28	Total Charges	v	x	X
29	Churn Score	v	v	v
30	CLTV	v	x	x
31	Churn Reason	v	v	v
32	Churn Label	v	v	v

Pengujian klasifikasi tanpa seleksi fitur terdapat pada gambar 5. Data yang sudah *preprocessing* langsung di klasifikasi pada algoritma DL dan RF, kemudian dihasilkan nilai *performance.* Pada pengujian dengan seleksi fitur *chi square*, data yang telah dipilih sesuai pembobotan *chi square* diklasifikasi pada algoritma DL dan RF sesuai gambar 6, dilanjutkan pada data yang telah dipilih sesuai pembobotan *information gain* sesuai gambar 7. Pemilihan atribut dilakukan dengan operator Select Atribut. 10-*fold cross validation* terhubung dengan algoritma DL dan RF, dengan operator Performance Binominal Classification pada gambar 8 dan 9.

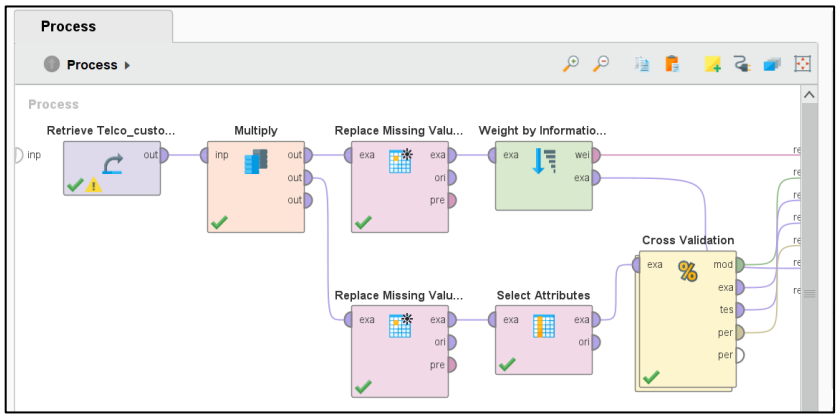
Djtechno : Jurnal Teknologi Informasi
Vol. 6, No. 2, Agustus 2025
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno



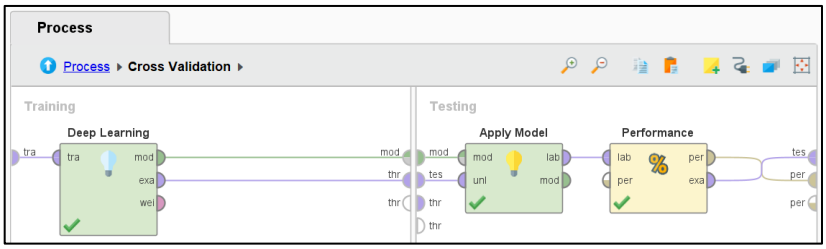
Gambar 5 Tanpa Weighting Seleksi Fitur



Gambar 6 Chi Square

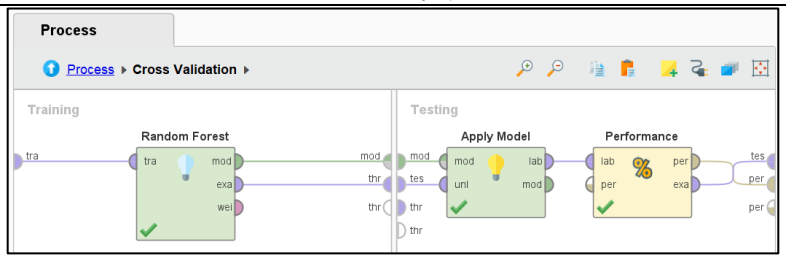


Gambar 7 Information Gain



Gambar 8 Deep Learning

Djtechno : Jurnal Teknologi Informasi
Vol. 6, No. 2, Agustus 2025
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno



Gambar 9 Random Forest

Hasil *confusion matrix* terdapat pada tabel 4 sampai 9, menunjukkan bahwa hasil TP dan TN memiliki jumlah yang besar, artinya hanya sedikit kesalahan prediksi yang terjadi.

Table 4. Confusion Matrix Deep Learning

	True Yes	True No	Class Precision
Pred Yes	1869	211	89.86%
Pred No	0	4963	100%
Class Recall	100%	95.92%	

Table 5. Confusion Matrix Deep Learning Chi Square

	True Yes	True No	Class Precision
Pred Yes	1772	46	97.47%
Pred No	97	5128	98.14%
Class Recall	94.81%	99.11%	

Table 6. Confusion Matrix Deep Learning Information Gain

	True Yes	True No	Class Precision
Pred Yes	1782	57	96.90%
Pred No	87	5117	98.33%
Class Recall	95.35%	98.90%	

Table 7. Confusion Matrix Random Forest

	True Yes	True No	Class Precision
Pred Yes	1777	1	99.94%
Pred No	92	5173	98.25%
Class Recall	95.08%	99.98%	

Table 8. Confusion Matrix Random Forest Chi Square

	True Yes	True No	Class Precision
Pred Yes	1781	2	99.89%
Pred No	88	5172	98.33%
Class Recall	95.29%	99.96%	

Table 9. Confusion Matrix Random Forest Information Gain

	True Yes	True No	Class Precision
Pred Yes	1783	2	99.89%
Pred No	86	5172	98.36%
Class Recall	95.40%	99.96%	

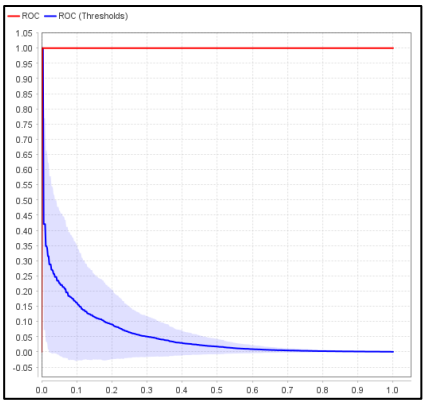
Nilai *performance* lengkap terdapat pada tabel 10. Hasil menunjukkan bahwa seleksi fitur terbukti meningkatkan akurasi klasifikasi dibandingkan tanpa menggunakan seleksi fitur. Hasil *performance* terbaik diperoleh algoritma RF dengan seleksi fitur *information gain,* yaitu akurasi 98.75%.

Djtechno : Jurnal Teknologi Informasi
Vol. 6, No. 2, Agustus 2025
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

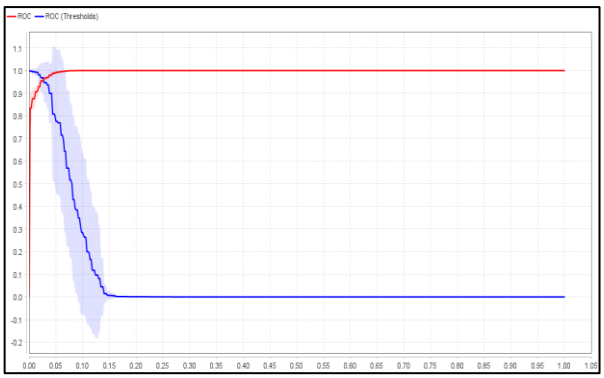
Table 10. *Performance*

Deep Learning 10-*fold Cross Validation*	Tanpa Seleksi Fitur		Seleksi Fitur			
			Chi Square		Information Gain	
	DL	RF	DL	RF	DL	RF
Accuracy	97.00%	98.68%	97.97%	98.72%	97.96%	98.75%
Classification error	3.00%	1.32%	2.03%	1.28%	2.04%	1.25%
Kappa	0.926	0.966	0.947	0.967	0.947	0.967
AUC	1	0.998	0.997	0.998	0.997	0.998
Precision	100%	98.25%	98.15%	98.33%	98.33%	98.37%
Recall	95.92%	99.98%	99.11%	99.96%	98.90%	99.96%
F-Measure	97.92%	99.11%	98.62%	99.14%	98.61%	99.16%
Sensitivity	95.92%	99.98%	99.11%	99.96%	98.90%	99.96%
Specificity	100%	95.08%	94.81%	95.29%	95.35%	95.40%

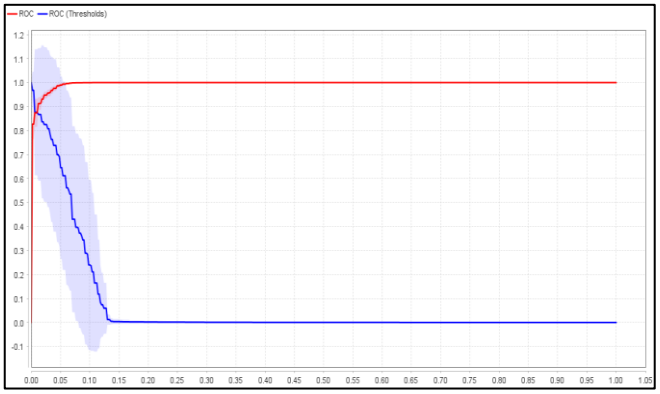
Akurasi merupakan rasio prediksi yang benar (baik positif maupun negatif) dibandingkan dengan total sampel. AUC Mengukur kemampuan model membedakan antara kelas positif dan negatif. Nilai AUC antara 0 sampai 1, semakin mendekati 1 maka model memiliki kemampuan yang baik dalam membedakan kelas. Kurva AUC yang baik adalah mendekati pojok kanan atas, pada angka 1. Gambar 10 sampai 15 menunjukkan seluruh kurva yang dihasilkan mendekati angka 1, sesuai dengan nilai AUC yang tertera pada tabel 10.



Gambar 10 Grafik AUC Deep Learning

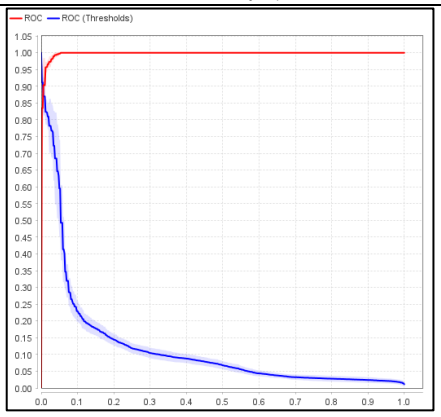


Gambar 11 Grafik AUC Deep Learning Chi Square

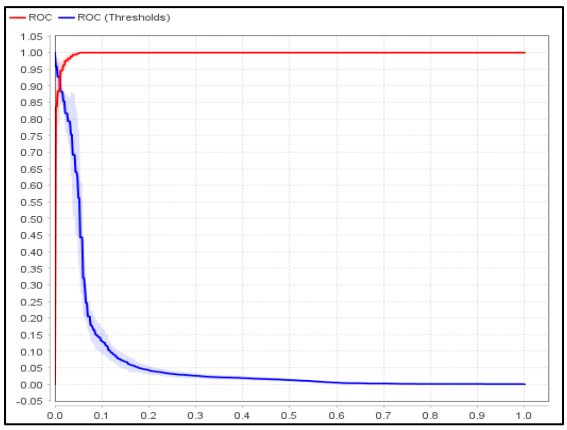


Gambar 12 Grafik AUC Deep Learning Information Gain

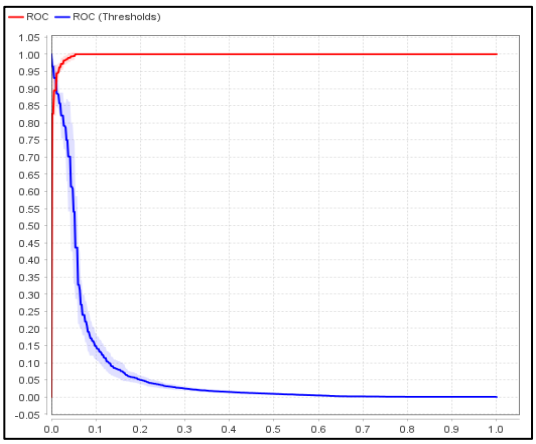
Djtechno : Jurnal Teknologi Informasi
Vol. 6, No. 2, Agustus 2025
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno



Gambar 13 Grafik AUC Random Forest



Gambar 14 Grafik AUC Random Forest Chi Square



Gambar 15 Grafik AUC Random Forest Information Gain

4. SIMPULAN

Penerapan teknik seleksi fitur, khususnya *chi square* dan *information gain*, secara signifikan meningkatkan performa prediksi *customer churn* pada industri telekomunikasi, dengan kombinasi Random Forest dan seleksi fitur *information gain* mencapai performa optimal (akurasi 98.75%, *precision* 98.37%, *recall* 99.96%, F-*measure* 99.16%, dan AUC 0.998), menunjukkan superioritas algoritma ensemble Random Forest dibandingkan *deep learning* pada dataset yang digunakan. Kontribusi utama penelitian meliputi validasi empiris efektivitas seleksi fitur, perbandingan komprehensif performa algoritma Random Forest dan *deep learning*, serta identifikasi kombinasi optimal untuk implementasi praktis, dengan implikasi manajerial bahwa perusahaan telekomunikasi dapat mengimplementasikan model prediksi Random Forest ini untuk mengidentifikasi pelanggan berisiko *churn* secara proaktif, memungkinkan alokasi sumber daya retensi yang lebih efisien dan *cost-effective*.

Djtechno : Jurnal Teknologi Informasi
Vol. 6, No. 2, Agustus 2025
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

REFERENCES

[1] S. H. Sahir *et al.*, *Ekonomi Global Tantangan dan Peluang di Era Digital*. Medan: Yayasan Kita Menulis, 2024.
[2] C. M. Annur, "Jumlah Perusahaan Internet Service Provider di Indonesia (2017-2022)," https://databoks.katadata.co.id/.
[3] B. Reynaldy, "Jumlah Penyedia Internet Mencapai 1.011 Perusahaan di 2023," https://data.goodstats.id/.
[4] E. Erwin, S. Tinggi, I. Ekonomi, C. Makassar, L. Judijanto, and R. Musprihadi, "Manajemen Pemasaran (Teori dan Strategi)," 2024. [Online]. Available: https://www.researchgate.net/publication/379927743
[5] A. P, "57 Customer Retention Statistics You Should Know in 2025," https://www.qrcode-tiger.com/.
[6] "US-Churn-Index-Summary-min," 2020.
[7] N. Salma and A. Aprianingsih, Ph.D, "*Customer churn* Analysis: Analyzing *Customer churn* Determinants on an ISP Company in Indonesia," *Buletin Pos dan Telekomunikasi*, pp. 29–40, Sep. 2021, doi: 10.17933/bpostel.2021.190103.
[8] "What is an average churn rate? Here's how to figure it out," https://stripe.com/.
[9] Y. Bharambe, P. Deshmukh, P. Karanjawane, D. Chaudhari, and N. M. Ranjan, "Churn Prediction in Telecommunication Industry," in *2023 International Conference for Advancement in Technology, ICONAT 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICONAT57137.2023.10080425.
[10] A. Mutiarachim, "Integration of Artificial Intelligence and Big Data Analytics in Customer-Centric Organizations," in *INTELLIGENT TRANSFORMATION: AI's Role in Business, Governance, Learning, and Spiritual Growth*, Semarang: Untag Press, 2024, ch. XII, pp. 180–195.
[11] A. K. B. Ginting, M. S. Lydia, and E. M. Zamzami, "Reduksi Atribut Menggunakan Chi Square untuk Optimasi Kinerja Metode Decision Tree C.4.5," *Jurnal Edukasi dan Penelitian Informatika*, vol. 9, pp. 44–49, Apr. 2023.
[12] A. Mutiarachim, F. K. Fikriah, B. Ansor, and A. P. Ramdani, "Boosting Performance Klasifikasi kNN Customer Loyalty dengan Chi square dan Information Gain," *Jurnal Transformatika*, vol. 22, no. 2, pp. 81–89, Mar. 2025, doi: 10.26623/6wgy1097.
[13] M. Onesime, Z. Yang, and Q. Dai, "Genomic Island Prediction via Chi square Test and Random Forest Algorithm," *Comput Math Methods Med*, vol. 2021, 2021, doi: 10.1155/2021/9969751.
[14] H. Bhoria, A. Dhankhar, and K. Solanki, "Chi square Feature Selection Technique for Student's performance prediction," *Indian J Sci Technol*, vol. 16, no. 38, pp. 3250–3257, 2023, doi: 10.17485/IJST/v16i38.921.
[15] A. S. Jaddoa and Z. T. M. Al-Ta'i, "Diagnosis of Diabetes Mellitus using (chi square-information gain) selectors and (SVM and KNN) Classifiers," in *AIP Conference Proceedings*, American Institute of Physics Inc., Mar. 2023. doi: 10.1063/5.0102761.
[16] N. Devian *et al.*, "Prediksi Penyakit Diabetes dengan Metode K-Nearest Neighbor (kNN) dan Seleksi Fitur Information Gain," 2024.
[17] N. L. Putri, R. A. Nugroho, R. Herteno, and P. Korespondensi, "Instrusion Detection System Berbasis Seleksi Fitur dengan Kombinasi Filter Information Gain Ratio dan Correlation Intrusion Detection System Based on Feature Selection with Filter Combination of Information Gain Ratio and Correlation," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 8, no. 3, pp. 457–464, Jun. 2021, doi: 10.25126/jtiik.202183154.
[18] F. Nabiel Syahreza, P. Nurul Sabrina, E. Ramadhan Teknik Informatika, U. Jendral Achmad Yani Jl Terusan Jend Sudirman, J. Barat, and K. Cimahi, "Prediksi Penyakit Stroke Menggunakan Metode K-Nearest Neigbors dan Information Gain," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 6, Dec. 2024.
[19] S. W. Fujo, S. Subramanian, and M. A. Khder, "*Customer churn* prediction in telecommunication industry using deep learning," *Information Sciences Letters*, vol. 11, no. 1, pp. 185–198, Jan. 2022, doi: 10.18576/isl/110120.
[20] M. Amirulhaq Iskandar, U. Latifa, U. Singaperbangsa Karawang, and J. H. Ronggo Waluyo, "Website Prediksi *Customer churn* untuk Mempertahankan Pelanggan pada Perusahaan Telekomunikasi," *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 2, pp. 1308–1316, Apr. 2023.
[21] T. N. Muthmainnah and A. Voutama, "Pendekatan Data Science untuk Menemukan *Customer churn* pada Perusahaan Fashion dengan Metode Machine Learning," *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD*, vol. 6, pp. 463–471, Jul. 2023, [Online]. Available: https://ojs.trigunadharma.ac.id/index.php/jsk/index
[22] A. M. Husein and M. Harahap, "Pendekatan Data Science untuk Menemukan Churn Pelanggan pada Sector Perbankan dengan Machine Learning," *Data Sciences Indonesia (DSI)*, vol. 1, no. 1, pp. 8–13, Nov. 2021, doi: 10.47709/dsi.v1i1.1169.

Djtechno : Jurnal Teknologi Informasi
Vol. 6, No. 2, Agustus 2025
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

[23] D. Adji Kusuma, A. Ratna Dewi, and A. Rony Wijaya, "Perbandingan Random Forest dan Convolutional Neural Network dalam Memprediksi Peralihan Pelanggan," MEI, 2025.

[24] H. Nalatissifa and H. F. Pardede, "Customer Decision Prediction Using Deep Neural Network on Telco *Customer churn* Data," *Jurnal Elektronika dan Telekomunikasi*, vol. 21, no. 2, p. 122, Dec. 2021, doi: 10.14203/jet.v21.122-127.

[25] Y. Yulianti and A. Saifudin, "Sequential Feature Selection in *Customer churn* Prediction Based on Naive Bayes," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing Ltd, Aug. 2020. doi: 10.1088/1757-899X/879/1/012090.

[26] L. Hota, "Computational Intelligence and Machine Learning Prediction of *Customer churn* in Telecom Industry: A Machine Learning Perspective".

[27] S. K. Wagh, A. A. Andhale, K. S. Wagh, J. R. Pansare, S. P. Ambadekar, and S. H. Gawande, "*Customer churn* prediction in telecom sector using machine learning techniques," *Results in Control and Optimization*, vol. 14, Mar. 2024, doi: 10.1016/j.rico.2023.100342.

[28] P. Lalwani, M. K. Mishra, J. S. Chadha, and P. Sethi, "*Customer churn* prediction system: a machine learning approach," *Computing*, vol. 104, no. 2, pp. 271–294, Feb. 2022, doi: 10.1007/s00607-021-00908-y.

[29] M. Rizki Kurniawan *et al.*, "Prediksi *Customer churn* Pada Perusahaan Telekomunikasi Menggunakan Algoritma C4.5 Berbasis Particle Swarm," 2023.

[30] A. Mutiarachim and J. S. P. Tyoso, "Optimasi Prediksi Pemasaran Nasabah Deposito Bank dengan Metode Klasifikasi Logistic Regression," *Jurnal Cakrawala Informasi*, vol. 4, no. 1, pp. 20–28, Jun. 2024, doi: 10.54066/jci.v2i1.176.

[31] V. Kumar *et al.*, "Addressing Binary Classification over Class Imbalanced Clinical Datasets Using Computationally Intelligent Techniques," *Healthcare (Switzerland)*, vol. 10, no. 7, Jul. 2022, doi: 10.3390/healthcare10071293.