

KLASIFIKASI PELANGGAN *CHURN* PADA KEGIATAN SERVIS DAN PENJUALAN SPAREPART DENGAN METODE *RANDOM FOREST*

Sisca Dea Yuliani^{1*}
Puspita Kartikasari²

¹Program Studi Statistika, Fakultas Sains dan Teknologi, Universitas Terbuka, Tangerang Selatan, Indonesia

²Program Studi Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro, Semarang, Indonesia

*Penulis korespondensi-mail: siscadeayuliani01@gmail.com

ABSTRAK

Churn pelanggan adalah kondisi ketika pelanggan berhenti menggunakan produk atau layanan suatu perusahaan dalam periode tertentu. Tujuan dari penelitian ini untuk mengklasifikasikan pelanggan *churn* di PT. XYZ, perusahaan ritel otomotif yang menyediakan jasa servis dan penjualan sparepart. Metode *random forest* berbasis fitur RFM (*Recency, Frequency, Monetary*). Data mencakup riwayat transaksi servis dan penjualan sparepart periode Januari 2024 – Desember 2024. Metode *random forest* dipilih karena kemampuannya menangani data tidak seimbang dan mampu mengidentifikasi variabel kunci untuk pengambilan keputusan. Penelitian melibatkan *preprocessing* data dan pembagian dataset menjadi data pelatihan (80%) dan data pengujian (20%). Model dievaluasi menggunakan *confusion matrix* dengan akurasi, presisi, *recall*, F1-Score dan nilai AUC. Evaluasi model menunjukkan performa yang sangat baik dengan akurasi mencapai (67,71%), presisi (72,93%), *recall* (85,16%), F1-Score (78,6%), dan nilai AUC (0,6798) membuktikan efektivitas metode ini dalam mengidentifikasi pelanggan berisiko *churn*. Hasil dari penelitian ini diharapkan dapat membantu perusahaan mengurangi pelanggan *churn* dan meningkatkan profit secara berkelanjutan.

Kata Kunci: Pelanggan *Churn*, *Random Forest*, RFM, Ritel Otomotif.

1 PENDAHULUAN

Dunia bisnis yang penuh persaingan membuat retensi pelanggan yang sudah ada saat ini secara jelas lebih menguntungkan daripada mengejar pelanggan baru. PT. XYZ merupakan salah satu perusahaan yang fokus pada sektor ritel otomotif yang mengkhuskan penyedia jasa servis dan penjualan sparepart kendaraan, menghadapi hambatan besar dalam mempertahankan loyalitas pelanggan. Fenomena pelanggan yang berhenti menggunakan layanan perusahaan, atau dikenal dengan istilah pelanggan *churn*, menjadi perhatian serius karena dapat berdampak signifikan terhadap pendapatan dan keberlanjutan bisnis. Pelanggan yang melakukan *churn* berarti berhenti menggunakan produk atau layanan (Mufida et al., 2025). Tingkat *churn* pelanggan yang cukup besar menjadi indikator tingkat angka kegagalan bisnis, sehingga perlu dilakukan upaya pengurangan tingkat *churn* pelanggan. (Arina & Ulfah, 2022). *Churn* berfungsi sebagai metrik penurunan bisnis yang dapat berdampak buruk pada profitabilitas Perusahaan (Arina & Ulfah, 2022). Menurut Husein & Harahap (2021), *churn* adalah perhitungan statistik proporsi konsumen yang menghentikan aktivitas transaksi, sementara Pondel et al. (2021) menyatakan bahwa biaya memperoleh pelanggan lebih mahal dibandingkan biaya mempertahankan. Industri otomotif

sebagai sektor prioritas dalam revolusi industri 4.0, menghadapi tantangan unik dimana ketersediaan sparepart dan kualitas servis menjadi faktor penentu loyalitas pelanggan (Makin et al., 2023) (Usman & Syam, 2024).

Permasalahan utama yang dihadapi PT. XYZ meliputi ketidaktahuan terhadap variabel-variabel yang mempengaruhi keputusan *churn* pelanggan, belum optimalnya pemanfaatan data historis untuk memprediksi kecenderungan *churn*, serta kurangnya strategi berbasis data historis untuk mempertahankan pelanggan yang berisiko *churn*. Data historis pembelian sparepart dan servis menyimpan pola berharga yang dapat diprediksi menggunakan Teknik *machine learning* (Hidayat et al., 2025). *Machine learning* merupakan salah satu cabang kecerdasan buatan yang memberikan kemampuan kepada komputer untuk belajar tentang data baru tanpa memerlukan pemrograman manual secara rinci (Id, 2021). Prioritas utamanya terletak pada perancangan program yang mampu melakukan pembelajaran data secara mandiri, kemudian menghasilkan lalu sebuah model yang siap digunakan untuk memecahkan kasus tertentu (Id, 2021). Metode klasifikasi yang dapat digunakan untuk memprediksi *churn*, diantaranya *Decision Tree*, *Logistic Regression*, *Support Vector Machine*, *Neural Network*, *Naïve Bayes*, *K-Nearest Neighbors* (KNN) dan *Random Forest*. *Random forest* termasuk algoritma yang paling efektif dalam klasifikasi dan prediksi *churn* pelanggan. Metode ini unggul dalam menangani data yang kompleks dengan banyak variabel, ketahanan terhadap *overfitting*, tingkat akurasi yang tinggi, dan kemampuan untuk menentukan variabel-variabel yang paling berdampak pada proses pengambilan keputusan (Sulehu et al., 2025).

Penelitian mengenai pelanggan *churn* digunakan pada berbagai kasus, terutama dalam industri yang berbasis langganan atau layanan berulang (*recurring service*), seperti industri telekomunikasi (Abdulsalam et al., 2022), Perbankan dan Keuangan (Husein & Harahap, 2021), hasil panen padi (Nur et al., 2023), serta *Retail* berbasis Keanggotaan (Yulianto, 2021). Penelitian sebelumnya telah menunjukkan efektivitas metode *Random Forest* dalam memprediksi *churn* pelanggan di berbagai industri. (Mufida et al., 2025) menggunakan *Random Forest* untuk memprediksi *churn* pelanggan di sektor perbankan di Indonesia pada 1 Januari 2025 dan mencapai akurasi 83,12%, dengan *precision* 84%, *recall* 96%, dan *F1-Score* 90%, menjadikannya lebih efektif dalam mengidentifikasi pelanggan yang berpotensi *churn*. (Britanthia et al., 2020) menggunakan *random forest* untuk memprediksi klasifikasi fitur mode audio spotify dari data fitur audio Spotify yang rilis dari rentang waktu antara 2009 – 2019 dan mencapai akurasi 71,96% menandakan bahwa *Random forest* memberikan hasil yang konsisten dalam mengidentifikasi fitur-fitur penting seperti *speechiness*, *danceability*, dan *valence*, yang sesuai dengan teori musik.

Penelitian tersebut telah membuktikan keefektifan *random forest* dalam berbagai konteks, belum ada studi yang secara spesifik menganalisis faktor-faktor penyebab *churn* pada bisnis servis dan penjualan sparepart dengan karakteristik pasar Indonesia yang unik, khususnya di PT. XYZ. Persaingan yang terus meningkat dalam industri membuat perusahaan untuk serius mengendalikan customer *churn* (Geiler et al., 2022). Penelitian ini berfokus pada penerapan metode *random forest* untuk mengklasifikasikan pelanggan yang berpotensi *churn* dalam konteks bisnis servis dan penjualan sparepart di PT. XYZ. PT. XYZ dapat mengembangkan strategi implementasi berbasis hasil klasifikasi untuk meningkatkan strategi pemasaran, layanan purna jual, dan profitabilitas bisnis yang lebih efektif dengan mengidentifikasi variabel-variabel yang signifikan mempengaruhi *churn* dan mengukur tingkat akurasi model klasifikasi. Dampak positif yang diharapkan dari penelitian ini adalah mampu menyumbang pada peningkatan model prediksi *churn* di industri otomotif yang dapat membantu PT. XYZ dalam menyusun program loyalitas, meningkatkan

kualitas servis serta mengoptimalkan strategi pemasaran berbasis data (Apriana & Yuliansyah, 2024).

2 METODE

Penelitian ini fokus kepada analisis pelanggan churn di PT. XYZ yaitu sebuah Perusahaan yang bergerak di bidang servis dan penjualan sparepart kendaraan mobil. Latar belakang utama pada penelitian ini adalah tingginya tingkat churn pelanggan. Mengurangi tingkat pelanggan churn melalui pendekatan data science dengan memanfaatkan algoritma random forest, sehingga Perusahaan PT. XYZ dapat mengambil tindakan berdasarkan prediksi yang akurat. peneliti melakukan studi literatur yang bersumber dari buku dan jurnal pada terbitan 10 tahun terakhir yang relevan dengan permasalahan yang ada.

Penelitian ini menggunakan data sekunder berbasis deret waktu, adapun data ini merupakan data riwayat penggunaan jasa servis dan pembelian sparepart di PT. XYZ pada Januari 2024 – Desember 2024. Data berupa file excel dengan jumlah baris data sebanyak 2388 baris dan 15 kolom. Variabel pada data tersebut meliputi; Nama, Tanggal Masuk, No Polisi, Mobil, KM, Servis, No Wo, No Faktur, Jasa, Part, Chemical, Oli, Total, Jumlah, Keterangan. Data yang digunakan untuk *training* set pada periode Januari 2024 – November 2024 dan *test* set pada Desember 2024.

Tabel 1. Statistik Deskriptif

Deskripsi	Nilai
count	2388
Mean	865482
Standar deviasi	3327726
Min	0
25%	330000
50%	569372
75%	1036593
Max	147812200
Total servis (jasa)	580042460
Total pembelian spearpart	537677017
Jumlah pelanggan unik	1123

Tabel 1 menunjukkan statistik deskriptif dari dataset transaksi servis dan penjualan sparepart selama periode Januari 2024 sampai Desember 2024. Dataset ini terdapat 2388 transaksi valid dan jumlah pelanggan unik ada 1123 pelanggan dengan nilai rata-rata transaksi sebesar Rp 865.482. Variasi nilai cukup besar ditunjukkan oleh standar deviasi sebesar Rp 3.327.726 yang menunjukkan bahwa transaksi berkisar dari Rp 0 hingga Rp 14.781.220 . Total pendapatan dari layanan servis mencapai Rp 58.004.260 sedangkan total pendapatan dari penjualan sparepart mencapai Rp 537.677.017

Penelitian ini memerlukan identifikasi dan pengukuran variabel-variabel yang relevan untuk dapat memprediksi perilaku *churn* pelanggan secara akurat. Penelitian ini memanfaatkan sejumlah variabel dependen dan variabel independent

1. Variable dependen

Churn status adalah status yang menandakan apakah seorang pelanggan telah berhenti menggunakan layanan Perusahaan atau masih aktif sebagai pelanggan. Skala pengukurannya nominal/kategorikal (*Binary*). Nilai 1 = *chun* dan 0 = *non churn*.

2. Variable independen

Variabel independen dalam penelitian ini menggunakan pendekatan RFM (Recency, Frequency, Monetary), yang merupakan metode segmentasi pelanggan berdasarkan perilaku transaksi meliputi nama, tanggal masuk, servis, part serta total.

a) Recency

Mengukur seberapa baru pelanggan terakhir kali melakukan transaksi, dihitung dalam satuan hari sejak transaksi terakhir hingga tanggal analisis. Semakin tinggi nilai recency, semakin lama pelanggan tidak bertransaksi, yang menandakan risiko churn yang lebih besar.

b) Frequency

mencerminkan seberapa sering pelanggan melakukan transaksi dalam periode tertentu, seperti jumlah kunjungan servis atau pembelian sparepart dalam satu tahun. Pelanggan dengan frequency rendah cenderung kurang loyal dibandingkan pelanggan dengan frequency tinggi.

c) Monetary

Merepresentasikan total nilai transaksi pelanggan dalam periode yang sama, menunjukkan seberapa besar kontribusi finansial pelanggan terhadap bisnis. Pelanggan dengan Monetary tinggi biasanya lebih dipertahankan karena nilai ekonomisnya, sedangkan pelanggan dengan Monetary rendah lebih rentan churn.

Tahapan metode pada penelitian ini terdiri dari mengidentifikasi masalah dan tujuan, pengumpulan data, data *preprocessing*, data *splitting*, klasifikasi *churn* pelanggan, evaluasi model, *feature importance*, strategi retensi, kesimpulan.

1. Mengidentifikasi masalah dan tujuan. Penulis merumuskan masalah dengan mengembangkan model prediksi *churn* pelanggan menggunakan metode *Random Forest* berbasis fitur RFM (*Recency, Frequency, Monetary*) pada bisnis servis dan penjualan sparepart PT. XYZ. Fokus penelitian adalah untuk: (1) mengidentifikasi variabel paling berpengaruh terhadap *churn*, (2) membangun model klasifikasi dengan akurasi tinggi, dan (3) menyusun strategi retensi pelanggan berdasarkan hasil prediksi model.
2. Pengumpulan data. Sumber data utama berasal dari catatan transaksi servis dan penjualan sparepart di PT. XYZ pada periode Januari 2024 – Desember 2024.
3. Data *preprocessing*. *Preprocessing* data dilakukan sebelum pembangunan model data digunakan untuk memastikan kualitas data dan meningkatkan akurasi hasil model. Langkah langkah yang dilakukan dalam *preprocessing* data meliputi: pembersihan data dan transformasi data serta pelabelan data.
4. Data *splitting*. Membagi data set menjadi data *training (train set)* dan data uji (*test set*), dengan proporsi 80%:20%, setelah mendapatkan hasil dari *ratio training testing (%)* maka dilakukan penyeimbang data.
5. Klasifikasi *churn* pelanggan. Penelitian ini menggunakan *random forest* untuk memprediksi *churn* pelanggan PT. XYZ. *Random forest* merupakan algoritma dalam machine learning, diterapkan untuk tugas klasifikasi maupun regresi. (Nur et al., 2023). Teknik ini menggabungkan berbagai decision tree yang dilatih secara terpisah untuk mendapatkan hasil prediksi yang lebih tepat dan konsisten. Proses membangun *random forest* seperti bootstrap sampling, pembangunan *decision tree* serta penggabungan agregasi prediksi.
6. Evaluasi model. Penulis menggunakan *Accuracy, Precision* dan *Recall*, dan *F1-Score* dan kurva ROC-AUC untuk mengevaluasi model, kemudian menganalisis hasil penelitian berdasarkan nilai evaluasi tersebut.

$$Accuracy = \frac{TP+TN}{TP} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

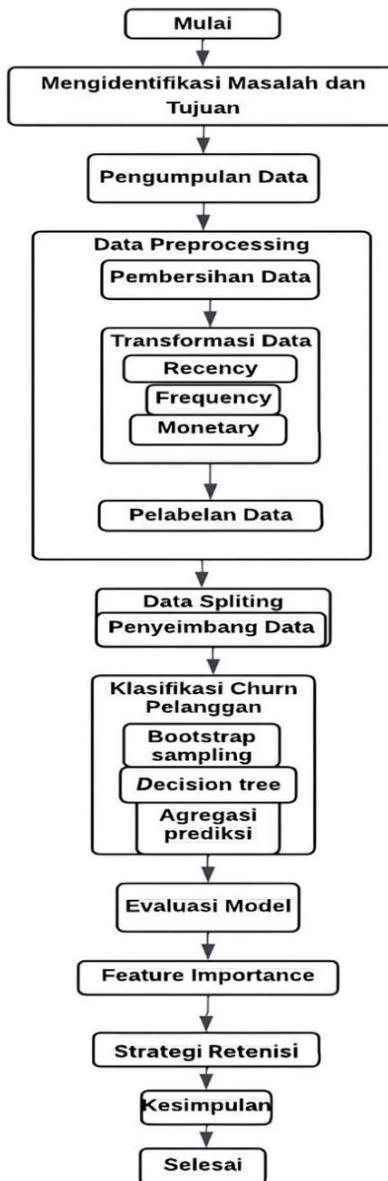
$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

$$True Positive Rate = Recall = \frac{TP}{TP+FN} \quad (5)$$

$$False Positive Rate = \frac{FP}{FP+TN} \quad (6)$$

7. *Feature importance*. Penulis memanfaatkan teknik *feature importance* guna menentukan variabel paling berpengaruh terhadap *churn*.
8. Strategi retensi, dilakukan untuk memberikan sebuah peluang untuk meningkatkan profitabilitas yang berkelanjutan pada PT. XYZ
9. Kesimpulan. Tahap terakhir adalah menyimpulkan hasil penelitian untuk menjawab permasalahan penelitian dan memberikan rekomendasi sehingga dapat meningkatkan profitabilitas bisnis.



Gambar 1. Diagram alir penelitian

3 HASIL DAN PEMBAHASAN

3.1 Data Preprocessing

Data *preprocessing* dilakukan sebagai persiapan sebelum penggunaan dataset pada model random forest, meliputi menghilangkan kolom yang tidak relevan sehingga hanya tersisa 4 kolom (Nama, Tanggal masuk, Servis, Part, total) dan memutasi *missing value*. Proses transformasi data dilakukan untuk membentuk dataset yang sesuai untuk klasifikasi *churn* menggunakan *random forest* dengan variabel RFM dan menentukan label *churn* berdasarkan waktu tertentu, yaitu dengan aturan:

1. Jika *Recency* > 90 hari, maka pelanggan dikategorikan sebagai *churn* (1)
2. Jika *Recency* ≤ 90 hari, maka pelanggan dikategorikan sebagai *non-churn* (0)

Tabel 2. Tampilan 10 baris pertama dari *transformasi data*

Nama	<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>	<i>Churn</i>
AA ASEP JAENUDIN	180	2	1819549	1
AAF SOPANDI	7	3	2237443	0
AAH AHMAD SYAHID	137	2	1221733	1
AAM AMIN HASAN	82	4	5325131	0
AAN SUPRIATMAN	328	2	2150629	1
ABDUL AZIZ SETIAPUTRA	121	1	300300	1
ABDUL KODIR	175	2	1250700	1
ABDUL ROHMAN	141	4	2888749	1
ACE DJ LATMA, IR	292	1	282433	1
ACEF SYAEFUL R	138	1	1027277	1

3.2 Data splitting

Dataset pada penelitian ini termasuk ke dalam *supervised learning*, dan merupakan kasus klasifikasi. Pembangunan model *machine learning*, dataset dibagi menjadi *training set* dan *test set* dengan proporsi 80%:20%. Pembagian ini bertujuan untuk melatih model klasifikasi menggunakan sebagian besar data.

Tabel 3. Hasil data splitting

<i>Ratio training testing (%)</i>	Jumlah data total	Jumlah data		Jumlah <i>Churn</i>	
		<i>Train</i>	<i>Test</i>	(1)	(0)
80:20	1112	889	223	0.690583	0.309417

Hasil jumlah data *splitting* tidak seimbang antar kelas *churn* dan *non-churn*, maka dilakukan penyeimbangan data dengan menggunakan Teknik *random oversampling* melalui algoritma SMOTE (*Synthetic Minority Over-sampling Technique*). Proses ini bertujuan guna meningkatkan kemampuan model dalam mendeteksi pelanggan yang berisiko *churn*. Sementara itu, *test set* dibiarkan utuh untuk memberikan hasil evaluasi yang objektif.

3.3 Klasifikasi *Churn Pelanggan*

Klasifikasi merupakan suatu teknik untuk menemukan karakteristik pembeda antar kelompok data guna memprediksi kategori objek baru (Dinata et al., 2023). Algoritma *random forest* termasuk kedalam *ensemble* yang bekerja dengan mengkombinasikan berbagai model untuk mencapai performa prediksi yang lebih unggul. (Id, 2021). *Machine learning* telah mentransformasi banyak aspek pengembangan teknologi, dengan penerapan utama dalam sistem pemrosesan data analisis, data prediktif termasuk klasifikasi *churn pelanggan* (Dinata et al., 2023). Teknik ini menggabungkan berbagai *decision tree* yang dilatih secara terpisah untuk mendapatkan hasil prediksi yang lebih tepat dan konsisten. Proses membangun *random forest* sebagai berikut

1. Bootstrap sampling, yaitu membagi data *training* acak menjadi beberapa bagian. Bagian data digunakan untuk melatih sebuah *decision tree* yang berbeda.
2. Pembangunan *decision tree*. Simpul dalam pembentukan *decision tree*, hanya sebagian fitur saja yang dipilih secara acak untuk proses membelah. Strategi ini menciptakan keragaman antar *decision tree* yang berbeda.

3. Penggabungan agregasi prediksi. Hasil dari semua *decision tree* dikombinasikan menggunakan pemungutan suara mayoritas.

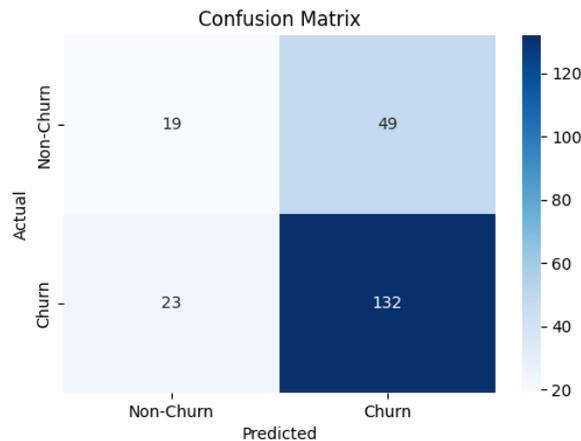
Klasifikasi *churn* pelanggan dilakukan dengan melatih model *random forest* menggunakan *training* set yang telah diseimbangkan. Pembangunan klasifikasi *churn* model *random forest*, menerapkan dua parameter utama: $n_estimators=100$ dan $max_depth=10$. Parameter $n_estimators$ mengatur jumlah *decision tree* dalam model *ensemble*, di mana jumlah yang lebih besar akan meningkatkan stabilitas prediksi tetapi memerlukan lebih banyak sumber daya komputasi. Parameter max_depth membatasi tingkat kedalaman setiap *decision tree* yang bertujuan untuk menghindari kecenderungan model yang terlalu kompleks terhadap data pelatihan.

3.4 Evaluasi Model

Penerapan metode *random forest* untuk prediksi *churn* pelanggan menghasilkan model yang dievaluasi menggunakan 20% data uji dari total dataset. Evaluasi memanfaatkan *confusion matrix* sebagai dasar perhitungan *accuracy*, *precision*, *recall*, dan *F1-score*. *Accuracy* mengukur ketepatan prediksi secara keseluruhan, sementara *precision* menunjukkan proporsi prediksi *churn* yang benar, *recall* mengindikasikan kemampuan model menemukan kasus sebenarnya, dan *F1-score* menyeimbangkan antara ketelitian dan kepekaan. Evaluasi model menggunakan $20\% \times 1112 = 223$ data dari total keseluruhan 2388 data.

Table 4. Hasil *confusion matrix*

		Actual Values	
		Positive	Negative
Predicted values	Positive	19	49
	Negative	23	132



Gambar 2. Plot *confusion matrix*

Hasil dari *confusion matrix* kemudian menghitung nilai evaluasi dengan rumus berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{132+19}{132+19+49+23} = \frac{151}{223} = 0,6771 = 67,71\%$$

$$Precision = \frac{TP}{TP+FP} = \frac{132}{132+44} = \frac{132}{181} = 0,7293 = 72,93\%$$

$$Recall = \frac{TP}{TP+FN} = \frac{132}{132+23} = \frac{132}{155} = 0,8516 = 85,16\%$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} = 2 \times \frac{0,7293 \times 0,8516}{0,7293+0,8516} = 0,786 = 78,6\%$$

Hasil evaluasi menunjukkan performa model yang baik dengan nilai *accuracy* (67,71%), *precision* (72,93%), *recall* (85,16%), dan *F1-score* (78,6%). Model ini berhasil menangkap pola penting dari fitur RFM (*Recency, Frequency, Monetary*) yang diekstrak dari data transaksi.

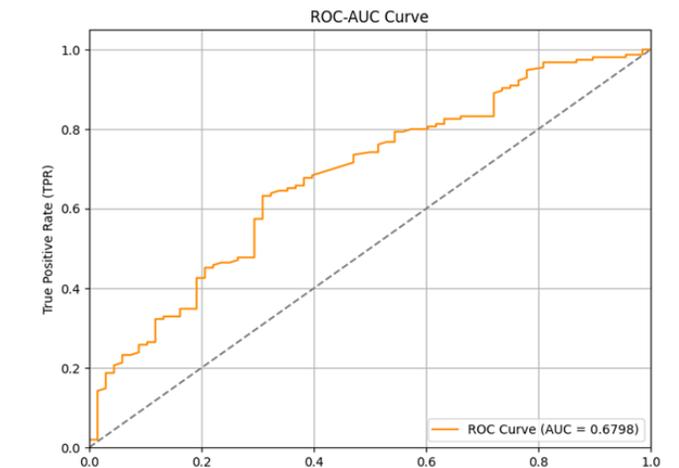
Kurva ROC-AUC juga digunakan untuk menilai performa model dalam membedakan pelanggan *churn* dan tidak *churn*, dimana $AUC > 0.9$ model dianggap baik dalam membedakan *churn* dan *non churn*.

$$True\ Positive\ Rate = Recall = \frac{TP}{TP+FN} = \frac{132}{132+23} = \frac{132}{155} = 0,8516 = 85,16\%$$

Jumlah pelanggan yang benar-benar terdeteksi dibandingkan dengan total pelanggan *churn*

$$False\ Positive\ Rate = \frac{FP}{FP+TN} = \frac{49}{49+19} = 0,7205$$

Jumlah pelanggan *churn* yang benar-benar terdeteksi dibandingkan dengan total pelanggan.

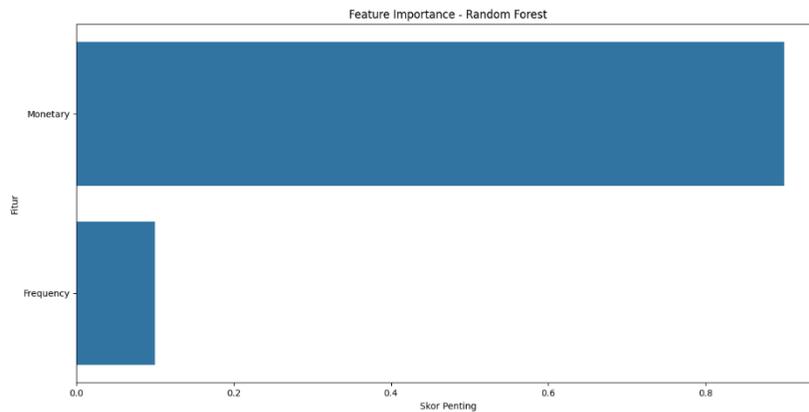


Gambar 3. Plot kurva ROC-AUC

Hasil Gambar 3 menunjukkan bahwa model *random forest* yang telah dibangun memiliki performa moderat hingga cukup baik dalam memprediksi *churn* pelanggan. Nilai AUC 0.6798, model ini dapat membantu PT. XYZ dalam mengidentifikasi pelanggan yang rentan *churn*.

3.5 Feature Importance

Teknik evaluasi *Feature Importance* berfungsi sebagai alat ukur sejauh mana pengaruh setiap variabel pada akurasi prediksi model (AISagri & Ykhlef, 2020). Evaluasi model dilanjutkan dengan analisis *feature importance* guna mengidentifikasi variabel-variabel yang paling menentukan dalam memprediksi apakah seorang pelanggan akan *churn* atau *nonchurn*, berdasarkan data servis dan penjualan sparepart.

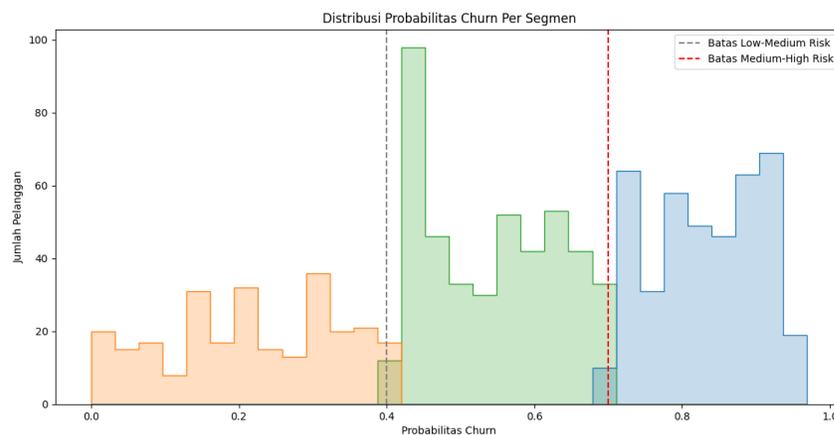


Gambar 4. Plot *feature importance*

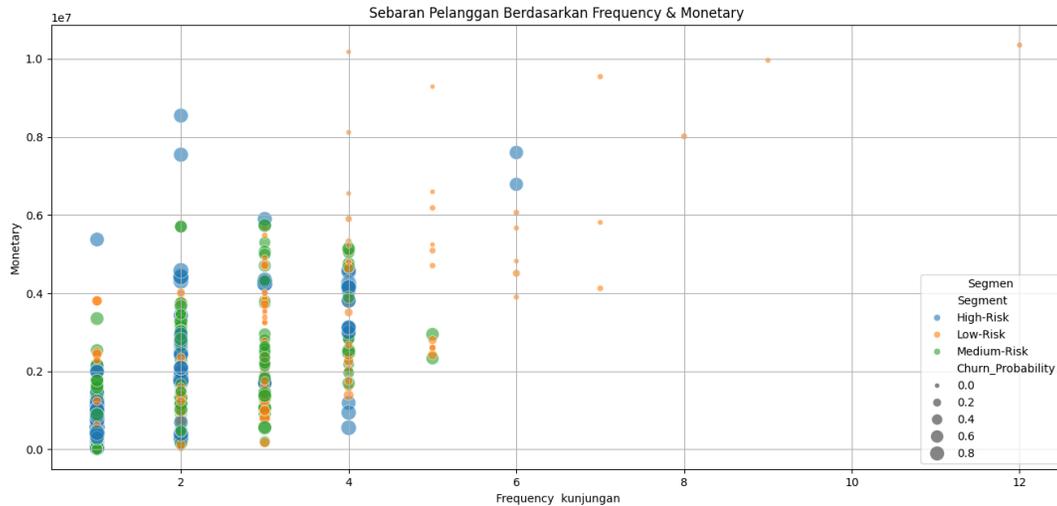
Hasil Gambar 4 menyatakan bahwa fitur *monetary* memiliki kontribusi yang jauh lebih besar dibandingkan *frequency* dalam memprediksi perilaku *churn*. Nilai skor penting menunjukkan bahwa *Monetary* mendominasi dengan nilai mendekati 0.9, sedangkan *Frequency* hanya sekitar 0.1 artinya besarnya pengeluaran pelanggan menjadi indikator utama apakah pelanggan akan berhenti atau tetap aktif. Frekuensi kunjungan memang berpengaruh, namun dampaknya relatif kecil dalam model ini. Temuan tersebut memberikan wawasan strategis bahwa program retensi pelanggan sebaiknya tidak hanya fokus pada kunjungan, tetapi juga pada upaya meningkatkan nilai transaksi pelanggan melalui promosi produk atau layanan premium.

3.6 Strategi Retensi

PT. XYZ mengembangkan strategi retensi berbasis *feature importance* untuk memberikan peluang meningkatkan profitabilitas yang berkelanjutan. Hasil analisis prediktif yang dikelompokkan menjadi: *high-risk* sebanyak 409 pelanggan, *medium-risk* sebanyak 441 pelanggan, dan *low-risk* sebanyak 262 pelanggan.



Gambar 5. Plot distribusi probabilitas churn per segmen



Gambar 6. Scatter plot pelanggan berdasarkan *frequency & monetary*

Hasil analisis segmentasi churn pada Gambar 5 memperlihatkan tiga kategori pelanggan PT. XYZ dengan karakteristik berbeda. Pelanggan *low-risk* (probabilitas churn $\leq 0,4$) yang berjumlah 262 pelanggan menunjukkan profil yang stabil dengan frekuensi transaksi tinggi dan pengeluaran besar, sehingga hanya memerlukan 5% alokasi sumber daya untuk program pemeliharaan seperti penawaran produk eksklusif atau penghargaan khusus. Sebanyak 441 pelanggan *medium-risk* (probabilitas churn 0.4 sampai 0.8) membutuhkan 15% upaya retensi melalui program *loyalitas* seperti sistem poin dan rekomendasi produk yang disesuaikan, mengingat pola transaksi mereka menunjukkan frekuensi transaksi menengah (4 sampai 8 kunjungan) dengan nilai transaksi seperti terlihat dalam Gambar 6 yang paling kritis adalah 409 pelanggan *high-risk* (probabilitas churn $> 0,8$) yang memerlukan 80% prioritas sumber daya melalui tindakan khusus seperti diskon 30% dan layanan prioritas, karena Gambar 6 menunjukkan aktivitas transaksi pelanggan yang sangat (2 sampai 4 kunjungan) dengan transaksi minimal. Kombinasi kedua visualisasi ini memungkinkan strategi retensi yang tepat sasaran, dimana Gambar 5 memberikan gambaran menyeluruh tentang distribusi risiko dan Gambar 6 memberikan konteks perilaku mendalam untuk masing-masing segmen, sehingga intervensi dapat disesuaikan secara lebih profesional, seperti memberikan promo khusus untuk high-risk. Pendekatan ini tidak hanya meningkatkan efektivitas retensi tetapi juga mengoptimalkan pemanfaatan sumber daya perusahaan secara proporsional berdasarkan tingkat prioritas masing-masing segmen pelanggan.

4 Kesimpulan

Model *random forest* berbasis fitur RFM (*Recency, Frequency, Monetary*) ini berhasil mengklasifikasikan pelanggan churn pada PT. XYZ, sebuah perusahaan ritel otomotif. Model yang dibangun menunjukkan kinerja yang cukup baik dengan *accuracy* (67,71%), *precision* (72,93%), *recall* (85,16%), *F1-Score* (78,6%), nilai AUC (0,6798) yang mengindikasikan kemampuan dalam membedakan pelanggan *churn* dan *non-churn*. Analisis *feature importance* mengungkapkan bahwa *monetary* merupakan variabel paling berpengaruh dalam prediksi *churn*, diikuti oleh *frequency* dan *recency*. Berdasarkan probabilitas *churn*, pelanggan terbagi menjadi tiga segmen: *high-risk*, *medium-risk*, dan *low-risk*, yang memerlukan strategi retensi berbeda, seperti program loyalitas atau diskon eksklusif bagi pelanggan high-risk. Hasil penelitian ini memberikan landasan bagi PT. XYZ untuk mengoptimalkan strategi retensi pelanggan dengan fokus pada peningkatan

nilai transaksi (*monetary*), sekaligus mengubah ancaman *churn* menjadi peluang peningkatan profitabilitas berkelanjutan. Rekomendasi untuk penelitian selanjutnya, dapat dipertimbangkan untuk menambahkan variabel lain seperti kepuasan pelanggan atau faktor eksternal untuk meningkatkan akurasi model.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Ibu Puspita Kartikasari, S.Si., M.Si. selaku dosen pembimbing atas bimbingan dan arahnya selama proses penelitian ini. Ucapan terima kasih juga disampaikan kepada PT. XYZ atas bantuan dalam bentuk perizinan serta pemberian data yang sangat mendukung kelancaran penelitian. Penghargaan juga disampaikan kepada keluarga dan teman-teman yang telah memberikan dukungan dan semangat dari awal hingga penelitian ini terselesaikan.

DAFTAR PUSTAKA

- Abdulsalam, S. O., Arowolo, M. O., Saheed, Y. K., & Afolayan, J. O. (2022). Customer Churn Prediction in Telecommunication Industry Using Classification and Regression Trees and Artificial Neural Network Algorithms. *Indonesian Journal of Electrical Engineering and Informatics*, 10(2), 431–440. <https://doi.org/10.52549/ijeei.v10i2.2985>
- AlSagri, H., & Ykhlef, M. (2020). Quantifying feature importance for detecting depression using random forest. *International Journal of Advanced Computer Science and Applications*, 11(5), 628–635. <https://dx.doi.org/10.14569/IJACSA.2020.0110577>
- Apriana, D., & Yuliansyah, C. (2024). Mengoptimalkan Penjualan Online Melalui Teknik Data Mining (Studi Kasus E-Commerce). *AL-MIKRAJ Jurnal Studi Islam Dan Humaniora (E-ISSN 2745-4584)*, 4(02), 514–527. <https://doi.org/10.37680/almikraj.v4i02.4774>
- Arina, F., & Ulfah, M. (2022). Analisa survival untuk mengurangi customer churn pada perusahaan telekomunikasi. *Journal Industrial Servicess*, 8(1), 59. <https://doi.org/10.36055/jiss.v8i1.14313>
- Britanithia, L., Tanujaya, C., Susanto, B., & Saragih, A. (2020). Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Fitur Mode Audio Spotify. *Indonesian Journal of Data and Science (IJODAS)*, 1(3), 68–78.
- Dinata, N. A. S., Abdurrahman, G., & Fitriyah, N. Q. (2023). Perbandingan Optimasi Algoritma Random Forest Menggunakan Teknik Boosting Terhadap Kasus Klasifikasi Churn Pelanggan Di Industri Telekomunikasi. *Jurnal Aplikasi Sistem Informasi Dan Elektronika*, 5(1), 28–37.
- Geiler, L., Affeldt, S., & Nadif, M. (n.d.). A survey on machine learning methods for churn prediction A survey on machine learning methods for churn prediction A survey on machine learning methods for churn prediction. *International Journal of Data Science and Analytics*, 2022(3), 217–242. <https://doi.org/10.1007/s41060-022-00312-5>
- Hidayat, R., Tri Saputra, H., Husnah, M., Nabila, N., Hidayatullah, M. B., Naufal Nazhmi, M., Azra, J., & Rana, A. (2025). Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produksi di Supermarket. *SIMKOM*, 10(1), 101–109. <https://doi.org/10.51717/simkom.v10i1.703>
- Husein, A. M., & Harahap, M. (2021). Pendekatan Data Science untuk Menemukan Churn Pelanggan pada Sector Perbankan dengan Machine Learning. *Data Sciences Indonesia (DSI)*, 1(1), 8–13. <https://doi.org/10.47709/dsi.v1i1.1169>
- Id, I. D. (2021). *Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python* (Vol. 1). Unri Press.

- Makin, S., Supriana, N., & Kurniawan, M. A. (2023). Model Peramalan Jumlah Penjualan Sparepart dengan Algoritma Forecasting Time Series: Studi Kasus di PT. XYZ. *Jurnal Pendidikan Dan Teknologi Indonesia*, 3(6), 265–271. <https://doi.org/10.52436/1.jpti.306>
- Mufida, E., Andriansyah, D., & Hertiana, H. (2025). Customer Churn Prediction Pada Sektor Perbankan Dengan Model Logistic Regression dan Random Forest. *Computer Science (CO-SCIENCE)*, 5(1), 58-66. <https://doi.org/10.31294/coscience.v5i1.7576>
- Nur, N., Wajidi, F., Sulfayanti, S., & Wildayani, W. (2023). Implementasi Algoritma Random Forest Regression untuk Memprediksi Hasil Panen Padi di Desa Minanga. *Jurnal Komputer Terapan*, 9(1), 58–64. <https://doi.org/10.35143/jkt.v9i1.5917>
- Pondel, M., Wuczyński, M., Gryniewicz, W., Łysik, Ł., Hernes, M., Rot, A., & Kozina, A. (2021). Deep learning for customer churn prediction in e-commerce decision support. *Business Information Systems*, 1, 3–12. <https://doi.org/10.52825/bis.v1i.42>
- Sulehu, M., Wisda, W., Wanita, F., & Markani, M. (2025). Optimasi Prediksi Kelulusan Mahasiswa Menggunakan Random Forest untuk Meningkatkan Tingkat Retensi. *Jurnal Minfo Polgan*, 13(2), 2364–2374. <https://doi.org/10.33395/jmp.v13i2.14472>
- Usman, S., & Syam, R. F. (2024). Predictive Sparepart Maintenance Menggunakan Algoritma Machine Learning Extreme Gradient Boosting Regressor. *Journal of System and Computer Engineering*, 5(2), 249–258.
- Yulianto, A. (2021). Prediksi Customer Churn Pada Bisnis Retail Menggunakan Algoritma Naïve Bayes. *REMIK: Riset Dan E-Jurnal Manajemen Informatika Komputer*, 6(1), 41–47.