

Analisis *Customer churn* Dengan Implementasi Data Mining Berbasis Algoritma Random Forest (Studi Kasus: PT Prioritas Purwokerto Group)

1st Azka Nabalah Putri Armeira
Direktorat Universitas Telkom

Purwokerto

Universitas Telkom Purwokerto

Purwokerto, Indonesia

azkaarmeira@student.telkomuniversity.ac.id

2nd Sarah Astiti

Direktorat Universitas Telkom

Purwokerto

Universitas Telkom Purwokerto

Purwokerto, Indonesia

sarahas@telkomuniversity.ac.id

Abstrak — PT Prioritas Purwokerto Group merupakan pelaku bisnis ritel yang bergerak di sektor furnitur dan elektronik. Perusahaan ritel dituntut untuk beradaptasi dengan cepat terhadap perubahan lanskap pasar. Transformasi digital menjadi kunci utama dalam menghadapi tantangan ini, terutama dengan adanya pergeseran pola konsumsi pasca-pandemi dan pertumbuhan pesat *e-commerce*. Salah satu isu krusial yang perlu diatasi adalah *customer churn*. Studi ini berfokus pada identifikasi pelanggan yang berpotensi *churn* dan faktor-faktor yang mempengaruhinya. Penelitian ini memanfaatkan platform data RapidMiner dengan menerapkan algoritma *Random Forest* menghasilkan nilai sebesar 97,28% yang menunjukkan bahwa model ini memiliki kemampuan yang baik dalam mengklasifikasikan data dengan benar dan segmentasi *Recency, Frequency, Monetary* (RFM) untuk menganalisis data pelanggan yang diperoleh dari database Perusahaan sebanyak 6961 data. Setelah dilakukan segmentasi RFM dihasilkan data sebanyak 340 data pelanggan yang diklasifikasikan ke dalam empat kategori loyalitas: *Platinum, Gold, Diamond*, dan *Silver*. Hasil analisis menunjukkan bahwa 54% pelanggan PT Prioritas Purwokerto Group termasuk dalam kategori *Gold*, diikuti oleh 29% *Silver*, 14% *Platinum*, dan 3% *Diamond*. Dominasi kategori *Gold* mengindikasikan tingkat *churn* pelanggan yang cukup tinggi sehingga perlu dilakukannya strategi retensi yang efektif untuk meningkatkan loyalitas *customer*. Hal ini berkontribusi pada peningkatan loyalitas pelanggan, optimalisasi profitabilitas, dan pencapaian keunggulan kompetitif yang berkelanjutan.

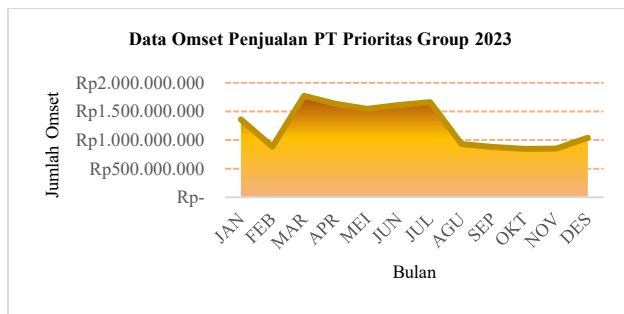
Kata kunci— *Bisnis Ritel, Customer churn, Random Forest, RFM, RapidMiner*.

I. PENDAHULUAN

Industri ritel merupakan salah satu sektor yang memiliki tingkat persaingan ketat dalam era perkembangan bisnis saat ini, penggunaan teknologi informasi menjadi salah satu pendorong serta pengaruh pada transformasi berjalannya bisnis. Ritel merupakan serangkaian kegiatan bisnis yang bertujuan untuk meningkatkan nilai dari barang atau jasa yang di jual kepada *customer* baik yang dikonsumsi secara pribadi atau rumah tangga [1]. PT Prioritas Purwokerto

Group merupakan salah satu bisnis yang bergerak di bidang ritel cash dan credit khususnya pada penjualan produk elektronik serta furnitur sejak tahun 2009. Dalam menjalankan bisnisnya yang semakin ketat menjaga pangsa pasar dan menjaga keberlanjutan merupakan hal yang perlu dilakukan PT Prioritas Purwokerto Group. Menjaga pangsa pasar adalah salah satu hal yang dapat dilakukan dalam usaha mempertahankan keberadaan dari loyal customer melalui prediksi akan kemungkinan terjadinya terjadinya *customer churn* atau kehilangan pelanggan pada suatu bisnis [2]. Mengingat dampak buruknya terhadap profitabilitas bisnis, analisis *churn customer* telah menjadi area fokus penting bagi perusahaan di berbagai industri termasuk industri ritel [3]. Tujuan dari analisis *churn customer* adalah untuk mengidentifikasi pola dan faktor yang berkontribusi terhadap *churn customer*, guna mengembangkan strategi dan intervensi untuk mempertahankan customer dan memitigasi dampak negatif *churn* [4]. Salah satu metode yang dapat digunakan dalam melakukan analisis *customer churn* yakni algoritma *Random Forest*, yang memanfaatkan pemodelan *data mining* untuk memprediksi dan mengidentifikasi customer dengan kemungkinan *churn* yang tinggi. Algoritma *Random Forest* merupakan bentuk pengembangan model dari *decision tree* yang melakukan validasi secara berulang kali pada proses pembandingan banyak *tree* yang ada serta mengambil keputusan berdasarkan hasil mayoritas *tree* [5]. Selain itu, algoritma ini efektif dalam mengidentifikasi pelanggan yang berisiko *churn*, memungkinkan perusahaan menerapkan strategi retensi yang lebih tepat sasaran [6].

Menurut Bapak Adif yang merupakan salah satu karyawan di PT Prioritas Purwokerto Group saat ini perusahaan tersebut masih mengandalkan penjualan di toko-toko fisik yang dimiliki di mana pada setiap cabang yang ada mereka memiliki fokus area pemasarannya masing-masing. Namun, dalam kurun waktu beberapa bulan PT Prioritas Purwokerto Group telah mengalami penurunan omset yang signifikan sebagaimana yang tertera pada data gambar 1.1 di bawah ini:



GAMBAR 1
(DATA OMSET PENJUALAN)

Berdasarkan pemaparan latar belakang diatas, maka peneliti mengusung penelitian mengenai analisis *churn* customer pada PT Prioritas Purwokerto Group berbasis algoritma Random Forest dengan menggunakan pemodelan data mining yang diharapkan dapat menjadi langkah pencegahan terjadinya *customer churn* di kemudian hari.

II. KAJIAN TEORI

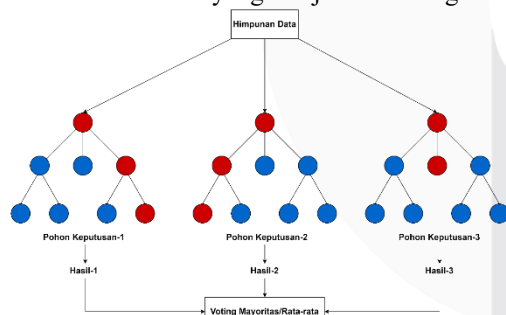
Menyajikan dan menjelaskan teori-teori yang berkaitan dengan variabel-variabel penelitian. Poin subjudul ditulis dalam abjad.

A. Customer churn

Customer churn merupakan persentase banyaknya customer yang berhenti membeli produk atau menggunakan atau menggunakan layanan dari suatu Perusahaan [7].

B. Algoritma Random Forest

Random Forest merupakan salah satu prosedur machine learning yang populer yang digunakan dalam mengembangkan model prediksi, model ini terdiri dari kumpulan pohon klasifikasi serta regresi. Model prediksi ini diperkenalkan pertama kali oleh Breiman pada tahun 2001 [8], [9]. Berikut merupakan tahapan pemodelan dalam algoritma Random Forest yang disajikan dalam gambar 2.1 :



GAMBAR 2
(PROSES KERJA PADA RANDOM FOREST)

Pada penelitian ini untuk membentuk model algoritma random forest maka dibuat sebuah pohon keputusan dengan menghitung nilai Entropy, Split, dan Gain dari masing masing variabel bebas pada data latih. Rumus untuk menghitung nilai Entropy, Split dan Gain dapat dilihat dari persamaan berikut:

$$Entropy(S) = \sum_{i=1}^c p_i \log_2 p_i \quad (1)$$

$$Entropy_{split} = \sum_{i=1}^c \frac{n_i}{n} Entropy(c) \quad (2)$$

$$Gain(A) = Entropy(S) - Entropy_{split} \\ = Entropy(S) - \sum_{i=1}^c \frac{|S_i|}{|S|} \times Entropy(S) \quad (3)$$

C. Synthetic Minority Oversampling Technique (SMOTE)

Metode SMOTE diperkenalkan Nithes V. Chawla pada tahun 2002. Metode ini digunakan untuk mengatasi ketidakseimbangan data dengan membuat data sintetis untuk kelas minoritas, berbeda dengan metode oversampling acak yang dipergunakan sebelum adanya metode SMOTE [10].

D. K- Fold Cross Validation

K-Fold Cross Validation merupakan teknik evaluasi model *machine learning* yang membagi dataset menjadi K bagian acak. Model dilatih pada sebagian data dan diuji pada sisanya, diulang K kali dengan setiap data point digunakan sebagai data uji [11]. Teknik ini efektif dalam meminimalkan bias dan overfitting, sehingga menghasilkan estimasi kinerja model yang lebih stabil dan akurat [12].

E. Recency, Frequency, dan Monetary (RFM) Segmentasi

Model RFM (Recency, Frequency, dan Monetary) merupakan model populer dalam bidang pemasaran yang membantu pengambilan keputusan mengidentifikasi pelanggan bernilai. Dengan memelihara informasi tentang kapan pelanggan terakhir membeli (recency), seberapa sering mereka membeli (frequency), dan berapa banyak uang yang mereka habiskan (monetary), RFM memfasilitasi segmentasi pasar dan pengembangan strategi pemasaran yang lebih efektif [13].

F. Confusion Matrix

Confusion Matix merupakan matriks yang digunakan untuk mempresentasikan hubungan antara atribut kelas sebenarnya dan kelas prediksi. Pada klasifikasi biner terdapat beberapa nilai evaluasi yang sering digunakan.

G. Data Mining

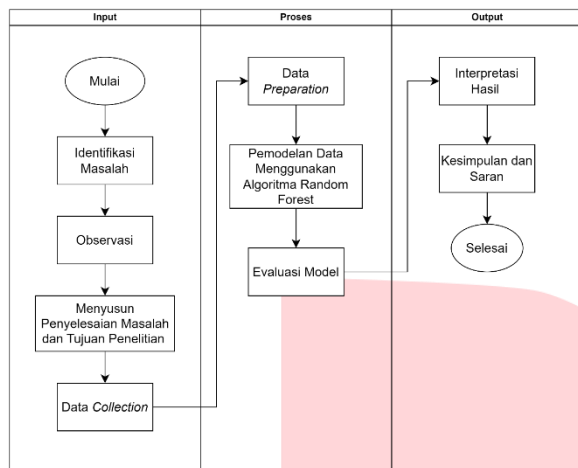
Data mining adalah proses menemukan pola, model serta jenis pengetahuan lain yang menarik pada sekumpulan data yang besar. Pengembangan Teknik data mining dilakukan dengan fokus pada analisa sekumpulan data yang besar [14].

H. RapidMiner

RapidMiner adalah perangkat lunak analisis data penambangan data mandiri ikembangkan oleh Dr. Markus Hofmann dan Ralf Klinkenberg, perangkat lunak ini dapat diintegrasikan ke dalam berbagai bahasa pemrograman. RapidMiner ditulis dalam bahasa pemrograman Java, sehingga dapat dijalankan di beberapa sistem operasi. RapidMiner memungkinkan pengguna untuk merancang alur analisis melalui antarmuka pengguna (UI). Desain ini kemudian disimpan dalam file XML yang mendefinisikan langkah-langkah analisis data yang ingin diterapkan, dan RapidMiner akan menjalankan proses tersebut secara otomatis berdasarkan file XML [15].

III. METODE

Penelitian ini dilakukan dalam beberapa proses dan prosedur yang berurutan seperti yang tertera pada Gambar 3 dibawah ini:



GAMBAR 3
(ALUR PENELITIAN)

A. Identifikasi Masalah

Penelitian ini diawali dengan identifikasi masalah, yaitu penurunan penjualan signifikan di PT Prioritas Group Purwokerto yang mengindikasikan *customer churn*. Analisis mendalam dilakukan untuk merumuskan masalah dan membatasi ruang lingkup penelitian pada isu *customer churn* di perusahaan tersebut.

B. Observasi,

Observasi merupakan langkah krusial untuk memahami *customer churn* di PT Prioritas Group Purwokerto. Peneliti menggunakan pendekatan beragam, meliputi kunjungan lapangan untuk mengamati proses bisnis dan interaksi pelanggan secara langsung, serta pengumpulan data deskriptif seperti jumlah pelanggan, tren penjualan, dan demografi pelanggan.

C. Menyusun Penyelesaian Masalah dan Tujuan Penelitian

Setelah identifikasi masalah dan observasi, penelitian ini secara sistematis merumuskan tujuan dan langkah-langkah untuk mengatasinya. Tujuannya adalah mengidentifikasi pelanggan yang berpotensi *churn* di PT Prioritas Group Purwokerto dan faktor-faktor penyebabnya, yang akan dicapai melalui pengumpulan data, segmentasi RFM, prediksi dengan Random Forest, dan analisis faktor *churn*.

D. Data Collection

Analisis *customer churn* di PT Prioritas Group Purwokerto sangat bergantung pada pengumpulan data yang sistematis dari *database* internal perusahaan. Data transaksi pelanggan menjadi fokus utama dan divalidasi untuk memastikan kualitas dan akurasi sebelum diproses.

E. Data Preparation

Data preparation adalah tahap penting dalam analisis *customer churn* menggunakan Random Forest. Data mentah dari PT Prioritas Group Purwokerto dibersihkan (*data cleaning*), diubah formatnya (*data transformation*), dipilih fiturnya (*feature selection*), dan direduksi dimensinya (*data*

reduction) agar siap dianalisis dan menghasilkan model yang akurat.

F. Pemodelan Data Menggunakan Algoritma Random Forest

Pada tahap ini, peneliti menerapkan Algoritma Random Forest untuk memodelkan perilaku *customer churn*. Langkah-langkah pemodelan melibatkan pelatihan model dengan data latih, pengaturan parameter algoritma, dan validasi model menggunakan data uji.

G. Evaluasi Model

Tujuan evaluasi model adalah untuk mengukur kinerja dan keandalan model Random Forest dalam memprediksi *customer churn*. Hasil evaluasi ini (berdasarkan *accuracy*, *precision*, dan *recall*) akan menentukan apakah model tersebut dapat diimplementasikan di PT Prioritas Group Purwokerto.

H. Interpretasi Hasil

Interpretasi hasil analisis Random Forest memberikan insight tentang *customer churn* di PT Prioritas Group Purwokerto dengan mengidentifikasi faktor-faktor penting, menganalisis segmentasi RFM, dan pola *churn* pada berbagai kelompok pelanggan

I. Kesimpulan dan Saran

Penelitian ini menyimpulkan temuan terkait *customer churn* di PT Prioritas Group Purwokerto dan memberikan saran untuk mengurangi *churn* melalui strategi retensi yang lebih efektif, peningkatan pelayanan, optimasi pemasaran, dan pemanfaatan teknologi digital. Saran ini didasarkan pada analisis data dan kondisi perusahaan.

IV. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan serta memproses data dengan implementasi penggunaan algoritma Random Forest untuk mengetahui fenomena *customer churn* di PT Prioritas Purwokerto Group. Data yang digunakan sebesar 6.961 data pada tahun 2021-2023 yang kemudian setelah di lakukan pemilihan data sehingga tidak terdapat data yang memiliki missing value menghasilkan data sebanyak 6949 data dan setelah melalui proses Recency, Frequency, Monetary Values (RFM) menghasilkan data sebanyak 340 data yang dapat di analisis.

A. Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data transaksi pembelian produk. Berikut Atribut data yang telah dikumpulkan pada tabel 1 berikut:

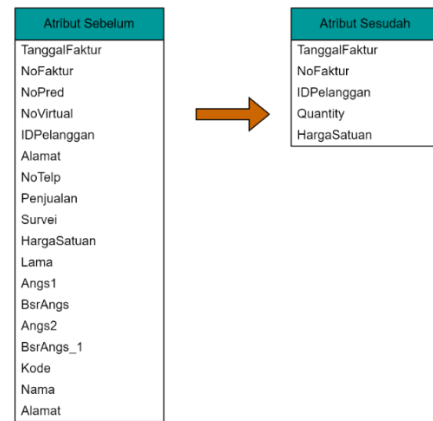
TABEL 1
(Atribut Data)

No	Atribut	Keterangan
1	TanggalFaktur	Tanggal invoice dikeluarkan kepada <i>customer</i> .
2	NoFaktur	Nomor unik yang diberikan pada faktur, berfungsi sebagai identifikasi spesifik untuk setiap transaksi penjualan.
3	NoPred	Nomor prediksi atau kode tertentu yang digunakan untuk keperluan internal atau manajemen.

4	NoVirtual	Nomor akun virtual yang sering digunakan untuk transaksi online atau pembayaran elektronik.
5	IDCustomer	Nomor identifikasi unik yang diberikan kepada setiap <i>customer</i> untuk mengidentifikasi mereka dalam sistem.
6	Alamat	Alamat lengkap <i>customer</i> yang berkaitan dengan faktur tersebut.
7	NoTelp	Nomor telepon <i>customer</i> yang dapat dihubungi.
8	Penjualan	Jumlah penjualan atau identifikasi dari penjual yang bertanggung jawab atas transaksi tersebut.
9	Survei	Data hasil survei
10	HargaSatuan	Harga per unit dari barang atau jasa yang dijual.
11	Lama	Durasi atau periode waktu, biasanya menunjukkan lamanya waktu pembayaran atau durasi kontrak.
12	DP	Down Payment (Uang Muka) yang dibayarkan oleh <i>customer</i> pada saat pembelian.
13	Angs1	Angsuran pertama yang dibayarkan oleh <i>customer</i> sebagai bagian dari pembayaran berkala.
14	BsrAngs	Besaran atau jumlah uang dari angsuran pertama yang harus dibayarkan.
15	Angs2	Angsuran kedua yang dibayarkan oleh <i>customer</i> sebagai bagian dari pembayaran berkala.
16	BsrAngs_1	Besaran atau jumlah uang dari angsuran kedua yang harus dibayarkan.
17	Kode	Kode identifikasi yang dapat merujuk pada produk, transaksi, atau referensi internal lainnya.
18	Nama	Nama produk yang dijual atau nama <i>customer</i> yang melakukan transaksi.
19	Keterangan	Catatan tambahan yang memberikan informasi lebih lanjut

B. Preprocessing

Tahapan *preprocessing* yaitu tahapan mengolah dataset mentah yang telah diperoleh menjadi data yang siap untuk diolah atau dimodelkan Berikut hasil *preprocessing* pada gambar 4:



GAMBAR 4

(HASIL *PREPROCESSING*)

Pada gambar 4 menunjukkan ilustrasi dari proses pemilihan atribut yaitu mereduksi atribut yang tidak akan digunakan. Hal ini dikarenakan pada analisis RFM (Recency, Frequency, Monetary) hanya menggunakan atribut TanggalFaktur, NoFaktur, IDCustomer, Quantity, dan HargaSatuan karena fokusnya pada tiga metrik utama yaitu Recency mengukur kapan terakhir kali customer melakukan pembelian yang diperoleh dari TanggalFaktur. Lalu Frequency mengukur seberapa sering customer melakukan pembelian yang dihitung dari jumlah NoFaktur per IDCustomer dan Monetary mengukur total pengeluaran customer, yang dihitung dari Quantity dikalikan dengan HargaSatuan. Adapun atribut lainnya seperti Alamat, NoTelp, Survei, Lama, DP, Angs1, Kode, Nama dan Keterangan tidak diperlukan.

C. Proses Recency, , Frequency, dan Monetary Value (RFM)

Dalam membuat RFM model untuk analisis customer churn dengan implementasi data mining berbasis algoritma Random Forest dimulai dari menghitung RFM Values dengan hasil pada tabel berikut:

TABEL 2
(Hasil RFM Values)

Customer	Frequency	Recency	Monetary
1	1	367	3.900.000
2	1	35	5.888.000
3	2	610	4.440.000
4	1	203	2.940.000
5	1	194	3.475.000
6	1	96	4.460.000
7	1	467	3.000.000
8	1	622	3.555.000
9	1	579	1.810.000
...	...	...	...
340	1	547	5.990.000

Setelah di dapatkan RFM Value dilakukan perhitungan RFM Score, Adapun untuk memberikan rentang nilai pada setiap skor R, F, dan M maka berikut adalah tabel yang menunjukkan bagaimana setiap nilai diklasifikasikan ke dalam skor:

TABEL 3
(KATEGORI RFM SCORE)

Skor	Recency	Frequency	Monetary
1	0 - 60	0 - 1	0 - 2940000
2	61 - 128	2	2940001 - 3900000
3	129 - 221	3	3900001 - 5888000
4	222+	3+	>5888000

Berdasarkan tabel di atas di dapatkan perhitungan RFM Score sebagai berikut:

TABEL 4
(Hasil RFM Score)

CustomerID	Frequency	Recency	Monetary Value	R	F	M
1	1	367	3900000	1	1	2
2	1	35	5888000	1	1	3
3	2	610	4440000	2	3	3
4	1	203	2940000	1	1	1
5	1	194	3475000	1	1	2
6	1	96	4460000	1	1	3
7	1	467	3000000	1	1	2
8	1	622	3555000	2	1	2
9	1	579	1810000	2	1	1
...	...	...	...	...	...	...
340	1	547	5.990.000	2	1	4

Untuk mengetahui RFM Score secara keseluruhan maka nilai dari skor Recency, Frequency, dan Monetary (R, F, M) dijumlahkan kemudian diberi label atau kelas menggunakan aturan tertentu. Berikut adalah tabel aturan pelabelan pada tabel 5 dan hasil pelabelan RFM pada tabel 6:

TABEL 5
(KATEGORI RFM LABEL)

RFM Score	Label
≤ 2	Bronze
≤ 4	Silver
≤ 6	Gold
≤ 8	Platinum
> 10	Diamond

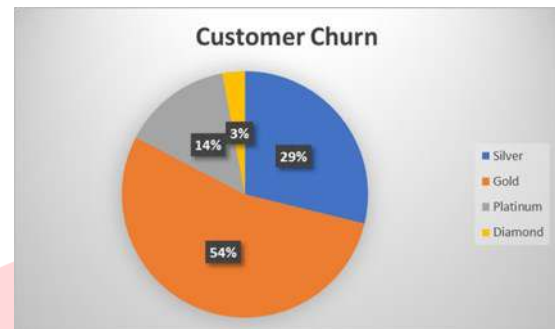
TABEL 6
(HASIL PELABELAN RFM)

CustomerID	R	F	M	RFM score	label
1	1	1	2	4	Silver
2	1	1	3	5	Gold
3	2	3	3	8	Platinum
4	1	1	1	3	Silver
5	1	1	2	4	Silver
6	1	1	3	5	Gold
7	1	1	2	4	Silver
8	2	1	2	5	Gold
9	2	1	1	4	Silver
...	...	...	...	...	...

CustomerID	R	F	M	RFM score	label
340	2	1	4	7	Platinum

D. Hasil Analisis Customer Churn

Berikut adalah hasil dari hasil RFM yang dapat digunakan untuk menganalisa customer:



GAMBAR 5
(Hasil RFM)

Berdasarkan analisis RFM maka didapatkan informasi terkait tingkat churn customer yang berbeda. Dalam analisis ini peneliti mengklasifikasikan customer menjadi empat kategori berdasarkan nilai RFM dan tingkat kesetiaan customer terhadap perusahaan. Terdapat level Gold yaitu sebagai kategori dengan nilai RFM “tinggi” walaupun customer dalam kategori ini masih dapat dikatakan setia namun hasil presentase menunjukkan tingkat loyalitas customer yang cukup rendah dan cenderung memiliki tingkat churn yang memiliki resiko cukup tinggi dengan loyalitas sebesar 54%.

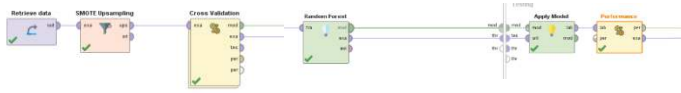
Selanjutnya terdapat level Silver yang memiliki nilai RFM yang sedikit lebih rendah dari Gold tetapi masih menunjukkan keterlibatan yang baik dengan perusahaan dengan persentase 29%. Adapun customer dalam kategori ini kurang loyal dengan perusahaan dan oleh karena itu lebih besar kemungkinan beralih ke produk perusahaan lain.

Lalu terdapat level Platinum dengan nilai RFM yang lebih namun tingkat churn yang sangat rendah dengan persentase customer pada level ini yaitu 14% dari keseluruhan customer. Customer dalam kategori ini menunjukkan tingkat loyalitas yang tinggi sehingga cenderung memiliki tingkat churn yang rendah.

Terakhir level Diamond yang menunjukkan keterlibatan customer yang rendah dengan perusahaan di mana nilai RFM yang rendah dengan nilai persentase customer sebesar 4%. Adapun jumlah customer dalam kategori ini paling sedikit tetapi loyalitas customer ini sangat kuat sehingga membuat mereka cenderung tetap setia terhadap perusahaan.

E. Hasil Pemodelan

Pemodelan data dalam penelitian ini merujuk pada penggunaan algoritma Random Forest untuk analisis customer churn di PT Prioritas Purwokerto Group. Tujuan utamanya adalah membangun model yang dapat melakukan prediksi customer churn PT Prioritas Purwokerto Group berdasarkan informasi dan nilai-nilai yang ada pada data yang sudah memiliki label RFM. Berikut adalah prosesnya menggunakan RapidMiner:



GAMBAR 5
(HASIL PEMODELAN RAPIDMINER)

F. Evaluasi

Penelitian ini menggunakan customer churn dengan akurasi, precision dan recall untuk melakukan evaluasi model berikut dengan menggunakan aplikasi Rapid Miner 11.0.0 hasilnya yakni:

	true. Platinum	true.Silver	true.Diamond	true.Gold	class precision
pred. Platinum	98	1	0	0	98.99%
pred.Silver	0	181	1	0	99.45%
pred. Diamond	0	1	38	1	95.00%
pred. Gold	0	0	10	182	94.79%
class recall	100.00%	98.91%	77.55%	99.45%	

TABEL 7
(Hasil Confusion Matrix)

Berdasarkan hasil evaluasi model klasifikasi dapat disimpulkan bahwa model yang digunakan menunjukkan kinerja yang sangat baik dalam memprediksi customer churn untuk PT Prioritas Purwokerto Group. Adapun dengan model klasifikasi yang dibuat memiliki nilai akurasi sebesar 97.28% yang menunjukkan bahwa model ini memiliki kemampuan yang baik dalam mengklasifikasikan data dengan benar. Hal ini berarti sebagian besar prediksi yang dilakukan oleh model adalah tepat.

Lalu pada nilai precision yang mengukur seberapa banyak dari prediksi positif model yang benar menunjukkan nilai 97% yang menunjukkan bahwa model memiliki nilai precision yang baik untuk hampir semua kelas. Hal ini menunjukkan bahwa model yang diperoleh berhasil memprediksi customer sebagai anggota kelas tertentu dan jarang sekali terjadi kesalahan.

Sementara itu hasil nilai recall yang mengukur seberapa banyak dari kelas positif yang diidentifikasi secara tepat oleh model menunjukkan nilai 94%. Hal ini berarti model hampir selalu berhasil mengidentifikasi customer yang sebenarnya sebagai anggota kelas tersebut.

G. Interpretasi Hasil

Hasil penelitian yang telah dilakukan customer dengan level Platinum terdapat sebesar 52% yang menandakan bahwa loyalitas yang dimiliki pelanggan PT Prioritas Purwokerto tinggi. Namun, sebagai pelaku bisnis tentu perlu adanya strategi retensi tingkat loyalitas customer agar tidak menurun. Berikut strategi retensi customer yang dapat dilakukan:

1. Segmentasi Pelanggan Berdasarkan RFM

- Platinum: Berikan program eksklusif untuk menjaga loyalitas.
- Gold & Diamond: Tawarkan insentif seperti promo spesial dan bundel produk.
- Silver: Fokus pada peningkatan layanan, program loyalitas, dan penanganan keluhan cepat.

- Peningkatan Pengalaman Pelanggan
 - Identifikasi sumber ketidakpuasan melalui survei.
 - Tingkatkan kualitas layanan dengan pelatihan karyawan.
 - Personalisasi layanan dan rekomendasi produk
- Pemanfaatan Algoritma Random Forest
 - Prediksi dini pelanggan berisiko churn untuk intervensi cepat.
 - Optimalkan promosi berdasarkan pola preferensi pelanggan.
- Program Loyalitas Berkelanjutan
 - Penghargaan untuk kesetiaan pelanggan Platinum dan kategori lainnya.
 - Edukasi produk untuk meningkatkan keterlibatan pelanggan Silver.

V. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan maka dapat diambil beberapa kesimpulan dari analisis customer churn dengan implementasi data mining berbasis algoritma Random Forest yaitu sebagai berikut:

- Berdasarkan hasil penelitian yang telah dilakukan faktor-faktor utama yang menyebabkan churn customer di PT Prioritas Purwokerto Group meliputi loyalitas customer yang rendah, nilai RFM (Recency, Frequency, Monetary) yang rendah, serta pengalaman customer yang tidak memuaskan. Customer yang memiliki keterlibatan rendah dan nilai RFM yang rendah menunjukkan kesetiaan yang rendah dan lebih cenderung untuk churn. Selain itu pengalaman negatif atau ketidakpuasan terhadap layanan atau produk juga menjadi penyebab utama customer berpindah ke produk perusahaan lain.
- Algoritma Random Forest pada penelitian ini berhasil digunakan untuk menganalisis dan memprediksi customer churn. Proses ini diawali dari pengumpulan data, preprocessing, pemodelan dan evaluasi. Adapun hasil model klasifikasi yang diperoleh memiliki model performance yang baik menunjukkan bahwa algoritma Random Forest dapat memprediksi kelas data customer yang berpotensi churn.
- Penelitian ini berhasil memperoleh hasil analisis customer churn yang dapat menjadi strategi retensi customer dengan melakukan segmentasi customer berdasarkan nilai RFM. Adapun pada penelitian ini hasil churn customer dikategorikan dalam level Platinum, Gold, Diamond dan Silver di mana customer dengan resiko churn tinggi yaitu level Silver dapat diberikan strategi peningkatan layanan untuk dapat menaikkan level loyalitas. Adapun berdasarkan hasil penelitian menunjukkan bahwa 54% customer di PT Prioritas Purwokerto Group memiliki kategori Gold yaitu tingkat loyalitas customer yang berpotensi mengalami churn yang tinggi.
- Hasil penelitian menunjukkan model prediksi customer churn menggunakan algoritma Random Forest menunjukkan akurasi, precision dan recall masing-masing sebesar 97,28%, 99,45%, 99,45% yang menunjukkan bahwa model sangat baik dalam mengklasifikasikan customer kedalam 4 kelas yaitu Platinum, Gold, Diamond dan Silver. Simpulan harus diuraikan dalam bentuk paragraf yang berisi poin utama

pembahasan hasil penelitian, berupa uraian dan tidak boleh menggunakan pointer.

REFERENSI

- [1] R. Kurniawan, "Transformasi Bisnis Melalui Penerapan Teknologi Sistem Informasi: Studi Kasus Pada Industri Retail."
- [2] Firmansyah and A. Yulianto, "Prediksi Customer Churn Pada Bisnis Retail Menggunakan Algoritma Naive Bayes," *Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 6, no. 1, 2021.
- [3] P. Jeyapraakhaash and R. K. Sashi, "Accuracy Measure of Customer Churn Prediction in Telecom Industry using Adaboost over Random Forest Algorithm," *J Pharm Negat Results*, vol. 13, Jan. 2022.
- [4] B. Zhang, "Customer Churn in Subscription Business Model—Predictive Analytics on Customer Churn," *BCP Business & Management*, vol. 44, pp. 870–876, Apr. 2023.
- [5] P. Lalwani, M. K. Mishra, J. S. Chadha, and P. Sethi, "Customer churn prediction system: a machine learning approach," *Computing Springer Nature*, vol. 104, no. 2, pp. 271–294, Feb. 2022.
- [6] Andy H, J. Nila R, Zia T, T. Faizal Y, Aji S, and H. S. Muhammad R, "Membangun Model Prediksi Churn Pelanggan yang Akurat," *Merkurius : Jurnal Riset Sistem Informasi dan Teknik Informatika*, vol. 2, no. 6, pp. 67–81, Oct. 2024.
- [7] M. Nadhira Faladiba and R. Haryastuti, "Churn Analisis Pada Data Pelanggan Telekomunikasi Menggunakan Ensemble Learning," *STATMAT (Jurnal Statistika dan Matematika)*, vol. 5, no. 1, pp. 45–54, 2023.
- [8] L. Breiman, "Random Forest," *Mach Learn*, vol. 45, no. 1, pp. 5–32, 2001.
- [9] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification And Regression Trees*. Routledge, 2017.
- [10] S. Sofyan and A. Prasetyo, "Penerapan Synthetic Minority Oversampling Technique (SMOTE) Terhadap Data Tidak Seimbang Pada Tingkat Pendapatan Pekerja Informal Di Provinsi D.I. Yogyakarta Tahun 2019," *Seminar Nasional Official Statistics*, pp. 686–687, 2021.
- [11] J. J. Purnama, Nina Kurnia Hikmawati, and Sri Rahayu, "Analisis Algoritma Klasifikasi Untuk Mengidentifikasi Potensi Risiko Kesehatan Ibu Hamil," *Journal of Applied Computer Science and Technology*, vol. 5, no. 1, pp. 120–127, Jun. 2024, doi: 10.52158/jacost.v5i1.809.
- [12] W. Wijianto, A. I. Pradana, S. Sopingi, and V. Atina, "Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa," *Jurnal Algoritma*, vol. 21, no. 1, May 2024.
- [13] V. Kumar, *Managing Customers for Profit : Strategies To Increase Profits And Build Loyalty*. New Jersey: Pearson Education, 2008.
- [14] H. Jiawei, P. Jian, and T. Hanghang, *Data Mining Concepts and Techniques*, Fourt Edition. Cambrige: Katey Birtcher, 2023.
- [15] Miryam C and Arum, "Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Naive Bayes dan ID3," *Jurnal Ilmiah Sistem Informasi, Teknologi Informasi dan Sistem Komputer*, vol. 17, no. 1, pp. 9–18, May 2022.