

PREDIKSI CUSTOMER CHURN DENGAN ALGORITMA DECISION TREE C4.5 BERDASARKAN SEGMENTASI PELANGGAN PADA PERUSAHAAN RETAIL

Ni Wayan Wardani¹, Gede Rasben Dantes², Gede Indrawan³

¹Program Studi Teknologi Informasi, STMIK STIKOM Indonesia
Denpasar, Indonesia

^{2,3}Program Studi Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha
Singaraja, Indonesia

e-mail: niwayan.wardani@stiki-indonesia.ac.id¹, rasben.dantes@pasca.undiksha.ac.id²,
gindrawan@undiksha.ac.id³

Received : February, 2018

Accepted : February, 2018

Published : April, 2018

Abstrak

Pelanggan merupakan aset yang sangat penting bagi perusahaan retail. Hal ini adalah alasan mengapa perusahaan retail harus merencanakan dan menggunakan strategi yang cukup jelas dalam memperlakukan pelanggan. Dengan banyaknya jumlah pelanggan yang dimiliki, maka masalah yang harus dihadapi adalah bagaimana mengidentifikasi karakteristik semua pelanggan dan mampu mempertahankan pelanggan yang telah ada agar tidak menghentikan pembelian dan berpindah ke perusahaan retail pesaing (churn). Dengan menerapkan konsep CRM (Customer Relationship Management), perusahaan dapat melakukan identifikasi pelanggan dengan melakukan segmentasi pelanggan sekaligus dapat melaksanakan program mempertahankan pelanggan dengan melakukan prediksi potensi churn pada setiap kelas pelanggan. Data yang digunakan berasal dari data pelanggan dan data penjualan pada UD.Mawar Sari. Proses segmentasi pelanggan menggunakan model RFM (Recency, Frequency, Monetary) sehingga mendapatkan kelas pelanggan UD. Mawar Sari yaitu Dormant, Everyday, Golden dan Superstar. Pembangunan model prediksi menggunakan algoritma Decision Tree C4.5. Penerapan model prediksi memperoleh hasil kinerja, yaitu: kelas Dormant : Recall 97.51%, Precision 75.18%, Accuracy 76.18%. Kelas Everyday : Recall 100%, Precision 99.04%, Accuracy 99.04%. Kelas Golden : Recall 100%, Precision 98.84%, Accuracy 98.84%. Kelas Superstar : Recall 96.15%, Precision 99.43%, Accuracy 95.63%. Dari hasil evaluasi dengan confusion matrix dapat disimpulkan bahwa kelas pelanggan dormant adalah kelas pelanggan yang berpotensi churn.

Kata Kunci: Churn, Retail, Segmentasi Pelanggan, RFM, C4.5

Abstract

Customer is a very important asset for retail companies. This is the reason why retail companies should plan and use a fairly clear strategy in treating customers. With the large number of customers, the problem that must be faced is how to identify the characteristics of all customers and able to retain existing customers in order not to stop buying and moving to a competitor retail company. By applying the concept of CRM, a company can identify customers by segmenting customers while also being able to implement customer retention programs by predicting potential churn on each customer class. The data used comes from UD.Mawar Sari. Customer segmentation process uses RFM model to get customer class. UD. Mawar Sari customer class is dormant, everyday, golden and superstar. The construction of prediction models using the Decision Tree C4.5. The application of the prediction model obtains performance results, that is: Dormant: Recall 97.51%, Precision 75.18%, Accuracy 76.18%. Everyday: Recall 100%, Precision 99.04%, Accuracy 99.04%. Golden: Recall 100%, Precision 98.84%, Accuracy

98.84%. Superstar: Recall 96.15%, Precision 99.43%, Accuracy 95.63%. Results of the evaluation with confusion matrix it can be concluded that the dormant customer class is a potentially churn customer class.

Keywords: Churn, Retail, Customer Segmentation, RFM, C4.5

1. PENDAHULUAN

Di tengah persaingan bisnis *retail* yang sangat ketat dewasa ini menjadikan perusahaan *retail* mulai mengalihkan perhatian mereka dari sekedar mengembangkan produk dan layanan yang unggul ke arah penciptaan pengalaman personal pelanggan[10].

Pelanggan bagi perusahaan adalah aset utama, karenanya berbagai cara ditempuh oleh perusahaan agar pelanggan tidak menghentikan pembelian. Apalagi di era digital saat ini mendapatkan pelanggan baru sangatlah sulit, terlebih lagi menjamurnya perusahaan *retail* sampai ke pelosok desa. Selain harus mampu mengenali pelanggan potensial, pihak manajemen perusahaan *retail* juga harus mampu mempertahankan pelanggan potensial (*customer retention*) sehingga dapat mencegah pelanggan menghentikan pembelian dan berpindah ke perusahaan pesaing (*churn*).

Dalam usaha *retail*, sejumlah biaya dihabiskan untuk menarik pelanggan baru, akan tetapi seiring waktu pelanggan tersebut dapat menghentikan pembelian mereka karena berbagai alasan. Hal tersebut tentunya tidak diinginkan oleh pihak perusahaan *retail* sehingga perlu mengetahui terlebih dahulu pelanggan yang memiliki kecenderungan untuk *churn* pada suatu waktu. Salah satu cara untuk mengetahui hal tersebut adalah dengan melakukan prediksi dan deteksi dini yaitu dengan melakukan *churn prediction*.

Menurut Emmet C. Murphy dan Mark A. Murphy dalam buku *Leading On The Edge of Chaos* bahwa memperoleh pelanggan baru menghabiskan biaya lima kali lipat dari biaya untuk memuaskan dan mempertahankan pelanggan lama, sementara itu peningkatan sebanyak 2% dalam mempertahankan pelanggan (*customer retention*) punya dampak terhadap laba seperti memangkas biaya sebesar 10%. Oleh karena itu, dengan menerapkan *churn prediction*, perusahaan dapat melakukan identifikasi pelanggan *churn* dan menerapkan strategi pemasaran yang tepat terhadap pelanggan – pelanggan lama dengan harapan dapat meningkatkan laba perusahaan.

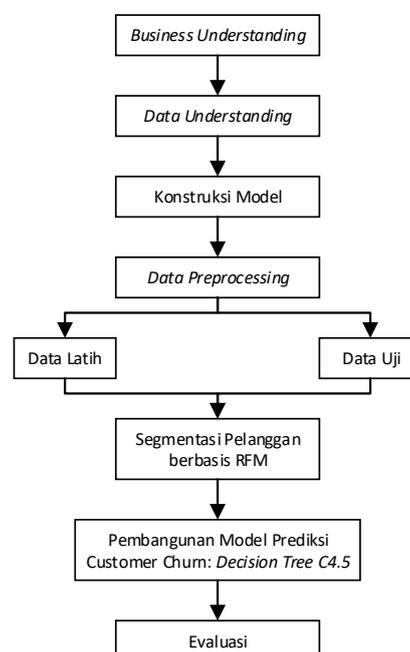
Pada penelitian ini akan digunakan model RFM (*recency, frequency, monetary*) untuk mengidentifikasi karakter pelanggan dan

membentuk segmentasi pelanggan sehingga menghasilkan beberapa kelas pelanggan. Data yang digunakan untuk mendapatkan fitur *recency, frequency, dan monetary* adalah berasal dari data pelanggan dan data transaksi pelanggan. Data setiap kelas pelanggan digunakan untuk pembangunan model prediksi *customer churn* menggunakan Algoritma *Decision Tree C4.5*. Algoritma *Decision Tree C4.5* merupakan salah satu teknik klasifikasi dalam *data mining*. Kefleksibelan membuat metode ini atraktif, khususnya karena memberikan keuntungan berupa visualisasi saran dalam bentuk pohon keputusan yang membuat prosedur prediksinya dapat diamati [5].

Dari paparan diatas, maka pada penelitian ini bertujuan mengetahui implementasi model RFM dalam menghasilkan kelas pelanggan melalui proses segmentasi pelanggan dan menerapkan algoritma *Decision Tree C4.5* dalam prediksi *customer churn* di setiap kelas pelanggan serta mengetahui tingkat akurasi yang dihasilkan setiap algoritma.

2. METODE PENELITIAN

Penelitian ini diselesaikan dengan melalui tahapan-tahapan yang ditunjukkan pada Gambar 1.



Gambar 1. Diagram Ikhtisar Penelitian

2.1 Business Understanding

Business understanding adalah pemahaman terhadap tujuan bisnis, penilaian situasi dan menerjemahkan tujuan bisnis ke dalam tujuan *data mining*. Dalam penelitian ini, salah satu tujuan dari data mining adalah untuk mendukung proses *customer retention* adalah mengidentifikasi karakter pelanggan dan dan membuat prediksi *customer churn*.

2.2 Data Understanding

Pada tahap ini dilakukan proses pengumpulan data, kemudian melakukan analisa data serta melakukan evaluasi kualitas data yang digunakan dalam penelitian ini. Dari proses perijinan data yang diajukan, diperoleh data transaksi pelanggan pada tahun 2015 sebanyak 30.765 *record* data, tahun 2016 sebanyak 64.008 *record* data dan pada tahun 2017 sebanyak 26.316 data.

2.3 Konstruksi model

Konstruksi model dalam penelitian ini adalah menentukan *data window* dan *forecasting window*. *Data window* adalah kerangka waktu untuk variabel input yang digunakan untuk mengkonstruksi model. *Forecasting window* adalah rentang waktu perilaku *churn* pelanggan yang akan diprediksi.

Dalam penelitian ini, kerangka waktu yang digunakan untuk konstruksi model data window di data latih adalah data transaksi pelanggan selama 9 bulan (Juli 2015 sd Maret 2016), sedangkan rentang waktu perilaku *churn* pelanggan sebagai *forecasting window* di data latih yang digunakan adalah data transaksi pelanggan selama 3 bulan (April 2016 sd Juni 2016).

Pada data uji, kerangka waktu yang digunakan untuk konstruksi model data window adalah data transaksi pelanggan selama 9 bulan (Mei 2016 sd Februari 2017), sedangkan rentang waktu perilaku *churn* pelanggan sebagai *forecasting window* yang digunakan adalah data transaksi pelanggan selama 3 bulan (Maret 2017 sd Mei 2017).

2.4 Data Preprocessing

Data preprocessing meliputi memilih data, pembersihan data, dan mengelompokkan atribut-atribut atau field yang telah terpilih menjadi 1 tabel (Tabel 1).

Tabel 1. Data Parsial dari Database UD.
Mawar Sari

Field	Keterangan
Id_Pelanggan	Kode Pelanggan
TglBeliAkhir	Menandakan <i>recency</i> , merupakan tanggal transaksi yang dilakukan oleh pelanggan
FrekuensiBeli	Menandakan <i>frequency</i> , merupakan jumlah transaksi selama periode yang ditentukan
TotalBeli	Menandakan <i>monetary</i> , merupakan jumlah uang selama periode yang ditentukan

2.5 Data Latih

Pada data latih, *data window* sebanyak 30.765 *record*, sedangkan *forecasting window* sebanyak 17.915 *record*.

2.6 Data Uji

Pada data uji, *data window* sebanyak 51.714 *record*, *forecasting window* sebanyak 14.511 *record*.

2.7 Segmentasi Pelanggan Berbasis RFM

Segmentasi adalah proses membagi pelanggan menjadi beberapa *cluster* dengan kategori loyalitas pelanggan untuk membangun strategi pemasaran. Segmentasi pelanggan dibagi menjadi 6 karakteristik berdasarkan nilai RFM [4] sebagai berikut:

1. Superstar

Pelanggan dengan *loyalty* yang tinggi dengan nilai *monetary*, frekuensi dan mempunyai transaksi paling tinggi.

2. Golden

Pelanggan yang mempunyai nilai *monetary* tertinggi kedua, frekuensi yang tinggi dan mempunyai rata-rata transaksi.

3. Typical

Pelanggan yang mempunyai rata-rata nilai *monetary* dan rata-rata transaksi.

4. Occasional

Pelanggan yang mempunyai nilai *monetary* terendah kedua setelah *dormant*, nilai *recency* paling rendah dan transaksi paling tinggi.

5. Everyday

Pelanggan yang memiliki peningkatan transaksi, transaksi yang rendah, dan mempunyai nilai *monetary* sedang sampai rendah.

6. Dormant

Pelanggan yang mempunyai frekuensi dan *monetary* yang paling rendah dan *recency* yang paling rendah.

Dalam pembentukan model RFM, diterapkan RFM skor satu sampai tiga (1- 3) pada domain nilai *recency*, *frequency*, dan *monetary* di setiap pelanggan. Skor 3 adalah nilai tertinggi dan skor 1 adalah nilai terendah. Skor akhir dihitung dari kombinasi setiap skor pada atribut. Pemberian skor sendiri akan memudahkan proses segmentasi pelanggan.

Nilai *recency*, *frequency*, *monetary* dibagi menjadi 3 bagian dengan nilai 3, 2 dan 1. Nilai *recency* dihitung berdasarkan tanggal transaksi terakhir atau interval waktu transaksi terakhir dengan saat ini. Nasabah dengan tanggal transaksi terbaru mempunyai nilai 3 sedangkan nasabah dengan tanggal transaksi terjauh di masa lalu mempunyai nilai 1. Begitu juga dengan nilai *frequency*, nasabah yang sering bertransaksi mempunyai nilai *frequency* yang tinggi, yaitu 3. Sedangkan nasabah yang jarang bertransaksi mempunyai nilai 1. Nasabah yang mempunyai total nilai transaksi terbesar mempunyai nilai *monetary* yang tinggi, dengan nilai 3. Sebaliknya nasabah yang mempunyai total nilai transaksi terkecil mempunyai nilai *monetary* yang rendah yaitu 1. Rentang skor dan domain nilai didapatkan dari hasil

wawancara dengan manajemen UD. Mawar Sari dan hasil bagi 3 dari nilai atribut terkecil ke nilai atribut tertinggi.

Skor dan domain nilai untuk atribut *recency*, *frequency*, *monetary* ditunjukkan dalam Tabel 2 berikut ini :

Tabel 2. Skor dan Domain Nilai RFM

Atribut	Skor	Domain nilai
Recency	1	$r > 150$ hari
	2	$61 \text{ hari} \leq r \leq 150$ hari
	3	$0 \leq r \leq 60$ hari
Frequency	1	$0 < f \leq 30$
	2	$31 \leq f \leq 60$
	3	$f > 60$
Monetary	1	$0 < m \leq \text{Rp. } 1.500.000$
	2	$\text{Rp. } 1.500.000 < m \leq \text{Rp. } 3.500.000$
	3	$m > \text{Rp. } 3.500.000$

2.8 Pembangunan Model Prediksi Customer Churn dengan Decision Tree C4.5

Model yang digunakan untuk membangun model prediksi *customer churn* dalam penelitian ini adalah algoritma *Decision Tree C4.5*. Data hasil pemodelan RFM masih dalam bentuk numerik, sedangkan pemodelan dengan algoritma *Decision Tree C4.5* membutuhkan data berbentuk kategorikal, sehingga data hasil pemodelan RFM harus di konversi ke dalam bentuk kategorikal. Tabel 3 adalah klasifikasi atribut untuk nilai data hasil konversi yang akan digunakan untuk memprediksi *customer churn*.

Tabel 3. Klasifikasi Atribut dengan Variabel Kategorikal

Atribut	Skor	Domain nilai	Variabel kategorikal
Recency	1	$r > 150$ hari	Lama
	2	$61 \text{ hari} \leq r \leq 150$ hari	Agak lama
	3	$0 \leq r \leq 60$ hari	Baru saja
Frequency	1	$0 < f \leq 30$	Jarang
	2	$31 \leq f \leq 60$	Agak sering
	3	$f > 60$	Sering
Monetary	1	$0 < m \leq \text{Rp. } 1.500.000$	Rendah
	2	$\text{Rp. } 1.500.000 < m \leq \text{Rp. } 3.500.000$	Sedang
	3	$m > \text{Rp. } 3.500.000$	Tinggi

Setelah ditentukan skor dan domain nilai pada *recency*, *frequency*, dan *monetary* maka akan didapatkan skor akhir RFM dan penentuan label pelanggan.

Terdapat 27 (3x3x3) kombinasi skor akhir RFM . Skor akhir RFM yang tertinggi adalah 333 dan

yang terendah adalah 111. Nasabah dengan skor akhir 333 merupakan nasabah yang dengan tingkat kelayakan yang tinggi sedangkan nasabah dengan skor akhir 111 merupakan nasabah dengan tingkat kelayakan yang rendah.

Tabel 4.Deskripsi Variabel Kategorikal dan Label Pelanggan

Kelas	Skor			Label pelanggan
	R	F	M	
K1	Baru saja	Jarang	Rendah	<i>Dormant D</i>
K2	Baru saja	Jarang	Sedang	<i>Dormant A</i>
K3	Baru saja	Jarang	Tinggi	<i>Occational A</i>
K4	Baru saja	Agak sering	Rendah	<i>Everyday D</i>
K5	Baru saja	Agak sering	Sedang	<i>Golden D</i>
K6	Baru saja	Agak sering	Tinggi	<i>Superstar D</i>
K7	Baru saja	Sering	Rendah	<i>Everyday A</i>
K8	Baru saja	Sering	Sedang	<i>Golden A</i>
K9	Baru saja	Sering	Tinggi	<i>Superstar A</i>
K10	Agak lama	Jarang	Rendah	<i>Dormant E</i>
K11	Agak lama	Jarang	Sedang	<i>Dormant B</i>
K12	Agak lama	Jarang	Tinggi	<i>Occational B</i>
K13	Agak lama	Agak sering	Rendah	<i>Everyday E</i>
K14	Agak lama	Agak sering	Sedang	<i>Golden E</i>
K15	Agak lama	Agak sering	Tinggi	<i>Superstar E</i>
K16	Agak lama	Sering	Rendah	<i>Everyday B</i>
K17	Agak lama	Sering	Sedang	<i>Golden B</i>
K18	Agak lama	Sering	Tinggi	<i>Superstar B</i>
K19	Lama	Jarang	Rendah	<i>Dormant F</i>
K20	Lama	Jarang	Sedang	<i>Dormant C</i>
K21	Lama	Jarang	Tinggi	<i>Occational C</i>
K22	Lama	Agak sering	Rendah	<i>Everyday F</i>
K23	Lama	Agak sering	Sedang	<i>Golden F</i>
K24	Lama	Agak sering	Tinggi	<i>Superstar F</i>
K25	Lama	Sering	Rendah	<i>Everyday C</i>
K26	Lama	Sering	Sedang	<i>Golden C</i>
K27	Lama	Sering	Tinggi	<i>Superstar C</i>

Proses *mining* dengan algoritma *Decision Tree* C4.5 dimulai dengan menghitung nilai *Entropy*, *Gain*, dan *Split* dari masing-masing atribut data training yang ada sehingga menghasilkan *Gain Ratio*. Rumus untuk menghitung *Entropy*, *Gain*, *Split* dan *Gain Ratio* bisa dilihat pada persamaan 1, 2, 3, 4 berikut :

$$Entropy(S) = \sum_{i=1}^k - p_i * \log_2 p_i \dots\dots\dots(1)$$

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots\dots(2)$$

$$SplitInfo\phi(s,j) = - \sum_{i=1}^k p(v_i | s) \log_2 p(v_i | s) \dots\dots\dots(3)$$

$$RasioGain(s,j) = \frac{Gain(s,j)}{SplitInfo\phi(s,j)} \dots\dots\dots(4)$$

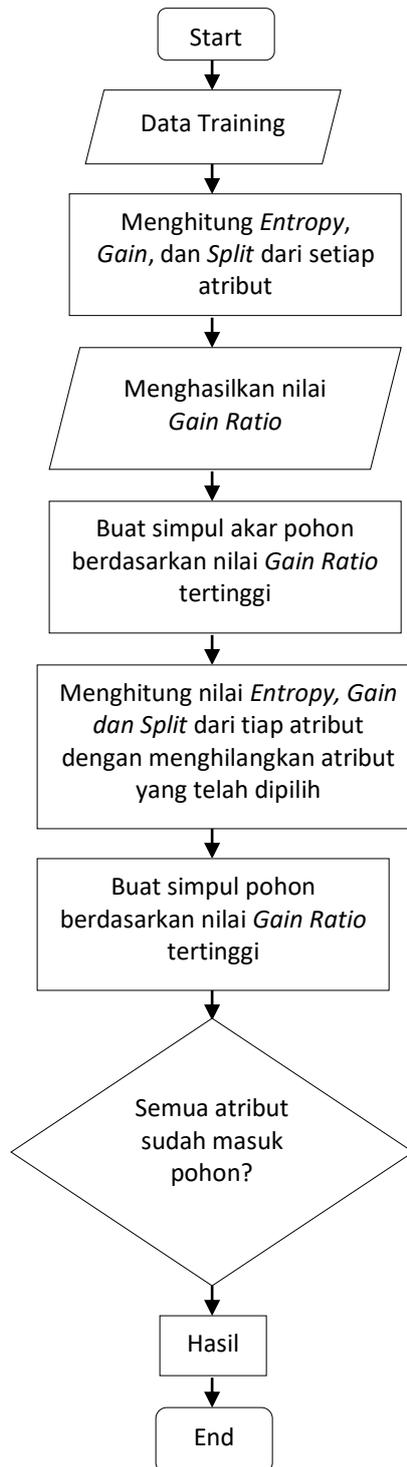
Dimana :

- S : himpunan kasus
- A : atribut
- n : jumlah partisi atribut A

|S_i| : jumlah kasus pada partisi ke-i

|S| : jumlah kasus dalam S

Atribut yang memiliki *Gain Ratio* terbesar dipilih untuk membuat simpul akar. Selanjutnya menghitung nilai *Entropy*, *Gain* dan *Split* dari masing – masing atribut dengan menghilangkan atribut yang telah dipilih sebelumnya. Atribut yang memiliki *Gain Ratio* terbesar dipilih untuk membuat simpul internal. Ulangi perhitungan tersebut hingga semua atribut memiliki kelas. Jika semua atribut/pohon sudah memiliki kelas, maka tampilkan pohon keputusan awal dan generate aturan keputusan awal. Tahapan alur kerja proses algoritma *Decision Tree* C4.5 diilustrasikan pada Gambar 2 berikut.



Gambar 2. Flowchart Decision Tree C4.5

2.9 Evaluasi

Confusion Matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep *data mining*. Berikut ini adalah tabel *confusion matrix* :

Tabel 5. *Confusion Matrix*

		Kelas Hasil Prediksi	
		Tidak (+)	Ya (-)
Kelas Asli	Tidak (+)	TP (<i>True Positive</i>)	FN (<i>False Negative</i>)
	Ya (-)	FP (<i>False Positive</i>)	TN (<i>True Negative</i>)

Evaluasi dengan *confusion matrix* menghasilkan nilai *accuracy*, *precision*, dan *recall*. *Accuracy* dalam klasifikasi adalah persentase ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi [9], sedangkan *precision* adalah proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. *Recall* adalah proporsi kasus positif yang sebenarnya yang diprediksi positif secara benar. Rumusnya sebagai berikut :

1. *Recall* = $TP / (TP + FN)$(5)
2. *Precision* = $TP / (TP + FP)$(6)
3. *Accuracy* = $(TP + TN) / (TP + FP + FN + TN)$ (7)

Keterangan :

1. TP (*True Positive*) adalah jumlah record positif yang diklasifikasikan sebagai positif
2. FP (*False Positive*) adalah jumlah record negatif yang diklasifikasikan sebagai positif
3. FN (*False Negative*) adalah jumlah record positif yang diklasifikasikan sebagai negative
4. TN (*True Negative*) adalah jumlah record negatif yang diklasifikasikan sebagai negatif

3. HASIL DAN PEMBAHASAN

Mengidentifikasi karakter pelanggan UD. Mawar Sari melalui proses segmentasi pelanggan berbasis RFM (*Recency*, *Frequency*, *Monetary*) menghasilkan kelas pelanggan *Dormant*, *Everyday*, *Golden* dan *Superstar*. Penerapan model prediksi *customer churn* diuji coba pada 4 data uji yaitu :

1. Data uji dengan data pelanggan kelas *Dormant* (2284 kasus)
2. Data uji dengan data pelanggan kelas *Everyday* (104 kasus).
3. Data uji dengan data pelanggan kelas *Golden* (259 kasus).
4. Data uji dengan data pelanggan kelas *Dormant* (183 kasus).

Berikut adalah hasil kinerja algoritma *Decision Tree* C4.5 pada setiap kelas pelanggan (Tabel 6).

Tabel 6. Hasil Kinerja Algoritma *Decision Tree* C4.5

Kelas Pelanggan	<i>Decision Tree C4.5</i>		
	<i>Recall</i>	<i>Precision</i>	<i>Accuracy</i>
<i>Dormant</i>	97.51%	75.18%	76.18%
<i>Everyday</i>	100%	99.04%	99.04%
<i>Golden</i>	100%	98.84%	98.84%
<i>Superstar</i>	96.15%	99.43%	95.63%

3.1 Hasil Pengujian pada Kelas Pelanggan *Dormant*

Hasil kinerja algoritma *Decision Tree* C4.5 pada data kelas pelanggan *Dormant* dengan total 2284 kasus adalah sebagai berikut:

Tabel 7. *Confusion Matrix* Kelas Pelanggan *Dormant*

		Kelas Hasil Prediksi	
		Tidak (+)	Ya (-)
Kelas Asli	Tidak (+) = 1569	1530	39
	Ya(-)= 715	505	210

1. Proporsi kasus positif yang sebenarnya diprediksi positif secara benar adalah sebanyak 1530 kasus dari 1569 kasus yang sebenarnya positif (*Recall* 97.51%).
2. Proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya sebanyak 1530 kasus dari 2035 kasus yang diprediksi positif (*Precision* 75.18%).
3. Jumlah kasus yang tepat diprediksi adalah sebanyak 1740 kasus terdiri dari 1530 kasus positif dan 210 kasus negatif yang diprediksi secara benar (*Accuracy* 76.18%).
4. Jumlah kasus yang diprediksi dengan tidak tepat sebesar 544 kasus.

3.2 Hasil Pengujian pada Kelas Pelanggan *Everyday*

Hasil kinerja algoritma *Decision Tree* C4.5 pada data kelas pelanggan *Everyday* dengan total 104 kasus adalah sebagai berikut :

Tabel 8. *Confusion Matrix* Kelas Pelanggan *Everyday*

		Kelas Hasil Prediksi	
		Tidak (+)	Ya (-)
Kelas Asli	Tidak (+) = 103	103	0
	Ya(-) = 1	1	0

1. Proporsi kasus positif yang sebenarnya diprediksi positif secara benar adalah sebanyak 103 kasus dari 103 kasus yang sebenarnya positif (*Recall* 100%).
2. Proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya sebanyak 103 kasus dari 104 kasus yang diprediksi positif (*Precision* 99.04%).
3. Jumlah kasus yang tepat diprediksi adalah sebanyak 103 kasus terdiri dari 103 kasus positif dan 0 kasus negatif yang diprediksi secara benar (*Accuracy* 99.04%).

4. Jumlah kasus yang diprediksi dengan tidak tepat sebesar 1 kasus.

3.3 Hasil Pengujian pada Kelas Pelanggan Golden

Hasil kinerja algoritma *Decision Tree C4.5* pada data kelas pelanggan *Golden* dengan total 259 kasus adalah sebagai berikut :

Tabel 9. *Confusion Matrix* Kelas Pelanggan *Golden*

		Kelas Hasil Prediksi	
		Tidak (+)	Ya (-)
Kelas Asli	Tidak (+) = 256	256	0
	Ya(-) = 3	3	0

1. Proporsi kasus positif yang sebenarnya diprediksi positif secara benar adalah sebanyak 256 kasus dari 256 kasus yang sebenarnya positif (*recall* 100%).
2. Proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya sebanyak 256 kasus dari 259 kasus yang diprediksi positif (*precision* 98.84%).
3. Jumlah kasus yang tepat diprediksi adalah sebanyak 256 kasus terdiri dari 256 kasus positif dan 0 kasus negatif yang diprediksi secara benar (*accuracy* 98.84%)
4. Jumlah kasus yang diprediksi dengan tidak tepat sebesar 3 kasus.

3.4 Hasil Pengujian pada Kelas Pelanggan Superstar

Hasil kinerja algoritma *Decision Tree C4.5* pada data kelas pelanggan *Superstar* dengan total 183 kasus adalah sebagai berikut :

Tabel 10. *Confusion Matrix* Kelas Pelanggan *Superstar*

		Kelas Hasil Prediksi	
		Tidak (+)	Ya (-)
Kelas Asli	Tidak (+) = 182	175	7
	Ya(-) = 1	1	0

1. Proporsi kasus positif yang sebenarnya diprediksi positif secara benar adalah sebanyak 175 kasus dari 182 kasus yang sebenarnya positif (*Recall* 96.15%).
2. Proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya sebanyak 175 kasus dari 176

kasus yang diprediksi positif (*Precision* 99.43%).

3. Jumlah kasus yang tepat diprediksi adalah sebanyak 175 kasus terdiri dari 175 kasus positif dan 0 kasus negatif yang diprediksi secara benar (*Accuracy* 95.63%).
4. Jumlah kasus yang diprediksi dengan tidak tepat sebesar 8 kasus.

4. KESIMPULAN

Berdasarkan hasil yang didapat dari penelitian ini maka dapat disimpulkan sebagai berikut :

1. Dari proses segmentasi pelanggan menggunakan pemodelan RFM (*Recency, Frequency, Monetary*) dapat diketahui karakteristik pelanggan UD. Mawar Sari. Dari proses segmentasi tersebut dapat diketahui kelas pelanggan yang dimiliki oleh UD. Mawar Sari yaitu kelas *Dormant, Everyday, Golden* dan *Superstar*.
2. Jumlah pelanggan terbesar yang dimiliki UD. Mawar Sari selama 2 tahun periode Juli 2015 – Mei 2017 berada di kelas *Dormant* yaitu pelanggan dengan karakteristik yang memiliki nilai *recency, frequency, dan monetary* yang rendah.
3. Jumlah pelanggan terkecil dimiliki oleh UD. Mawar Sari selama 2 tahun periode Juli 2015 – Mei 2017 berada di kelas *Everyday* yaitu pelanggan dengan karakteristik yang memiliki peningkatan transaksi dan kategori nilai *monetary* rendah atau sedang.
4. Hasil kinerja Algoritma *Decision Tree C4.5* pada kelas pelanggan *Dormant* yaitu *recall* 97.51%, *precision* 75.18%, dan *accuracy* 76.18%. Hasil kinerja pada kelas pelanggan *everyday* yaitu *recall* 100%, *precision* 99.04%, dan *accuracy* 99.04%. Hasil kinerja pada kelas pelanggan *golden* yaitu *recall* 100%, *precision* 98.84%, dan *accuracy* 98.84%. Hasil kinerja pada kelas pelanggan *superstar* yaitu *recall* 96.15%, *precision* 99.43%, dan *accuracy* 95.63%.
5. Dilihat dari hasil perhitungan *precision, recall* dan *accuracy* pada setiap kelas pelanggan maka dapat disimpulkan bahwa pelanggan UD. Mawar Sari yang berpotensi *churn* adalah kelas pelanggan *dormant*.
6. Divisi CRM UD. Mawar Sari perlu melakukan program mempertahankan pelanggan (*customer retention*) pada kelas pelanggan *dormant* mengingat jumlah terbesar pelanggan yang dimiliki UD. Mawar Sari adalah pelanggan dengan kelas *dormant*.

DAFTAR PUSTAKA

- [1] A A G. B. Ariana, "Customer Segmentation Dengan Metode Self Organizing Map (Studi Kasus : UD . Fenny)," *Neuron*, vol. 2, no. 1, pp. 1–7, 2011.
- [2] A. Chorianopoulos, *Effective CRM using Predictive Analytics*. Wiley, 2009.
- [3] E. C. Murphy and M. A. Murphy, *Leading On the Edge Of Chaos : The 10 Critical Elements for Success in Volatile Times*. USA: Prentice Hall Press, 2002.
- [4] E. Prasetyo, *Data Mining : Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Penerbit ANDI Yogyakarta, 2013.
- [5] E. Prasetyo, *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Penerbit ANDI Yogyakarta, 2014.
- [6] F. Gorunescu, *Data Mining Concept Model and Techniques*. Berlin: Springer, 2011.
- [7] G. Klepac, *Developing Churn Models Using Data mining Techniques and Social Network Analysis*. USA: IGI Global, 2015.
- [8] J. Burez and D. Van den Poel, "Handling class imbalance in customer churn prediction," *Expert Syst. Appl.*, vol. 36, no. 3 PART 1, pp. 4626–4636, 2009.
- [9] J. Han and M. Kamber, *Data Mining : Concepts and Techniques, 2nd ed.* Morgan Kaufmann Publishers, 2006.
- [10] K. W. Wong, "Data Mining Using Fuzzy Theory for Customer Relationship Management," vol. 4, no. Wawisr, pp. 188–200, 2001.
- [11] M. Listiana, Sudjalwo, and D. Gunawan, "Perbandingan Algoritma Decision Tree (C4.5) Dan Naïve Bayes Pada Data Mining Untuk Identifikasi Tumbuh Kembang Anak Balita (Studi Kasus Puskesmas Kartasura)," *Informatika*, vol. 1, no. 1, p. 18, 2015.
- [12] P. . Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Pearson Education, Inc, 2006.
- [13] P. S. Venatesan, *Data Mining and Warehousing*. New Age International (P) Limited, 2007.
- [14] R. Govindaraju, T. Simatupang, and T. A. Samadhi, "Perancangan Sistem Prediksi Churn Pelanggan," *Tek. Inform.*, vol. 9, no. 1, pp. 33–42, 2008.
- [15] V. L. M. Oliviera, "Analytical Customer Relationship Management in Retailing Supported by Data Mining Techniques," University of Porto, 2012.
- [16] V. L. M. Oliviera, "Predicting Partial Customer Churn using Markov for Discrimination for Modeling First Purchase Sequence," University of Porto, 2012.
- [17] W. Buckinx and D. Van Den Poel, "Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting," *Eur. J. Oper. Res.*, vol. 164, no. 1, pp. 252–268, 2005.
- [18] Y. Liu and Y. Zhuang, "Research Model of Churn Prediction Based on Customer Segmentation and Misclassification Cost in the Context of Big Data," *J. Comput. Commun.*, vol. 3, no. 3, pp. 87–93, 2015.