



Analisa Komparasi Algoritma *Decision Tree C4.5* dan *Naïve Bayes* untuk Prediksi *Churn* Berdasarkan Kelas Pelanggan Retail

Ni Wayan Wardani^{1,*}, Ni Kadek Ariasih²

¹² STMIK STIKOM Indonesia

ARTICLE INFO

Article history:

Received 19 July 2019
Received in revised form
16 August 2019
Accepted 19 September
2019
Available online 30 October
2019

Kata Kunci:

Prediksi *Churn*, RFM, C4.5,
Naïve Bayes

Keywords:

Churn prediction, RFM, C4.5,
Naïve Bayes

ABSTRAK

Pelanggan adalah salah satu aset utama bagi perusahaan ritel. Perusahaan harus dapat mengenali bagaimana karakter pelanggan mereka sehingga mereka dapat mempertahankan pelanggan yang sudah ada agar tidak berhenti membeli dan pindah ke perusahaan ritel yang bersaing (*churn*). Salah satu model yang tepat untuk mengenali karakter pelanggan adalah model RFM (Recency, Frekuensi, Moneter). Model RFM mampu menghasilkan kelas pelanggan dan di setiap kelas pelanggan dapat dianalisis atau diprediksi dengan konsep data mining apakah pelanggan tetap sebagai pelanggan atau *churn*. Data yang digunakan berasal dari data pelanggan dan data penjualan di UD. Mawar Sari. Kelas pelanggan UD Mawar Sari yang dihasilkan dari model RFM adalah Dormant, Everyday, Golden dan Superstar. Konsep data mining dengan membangun model prediksi dalam penelitian ini menggunakan algoritma *Decision Tree C4.5* dan *Naïve Bayes*. Di semua kelas pelanggan kinerja Algoritma *Naïve Bayes* lebih baik daripada Algoritma *Decision Tree C4.5* dengan Recall 95,92%, Precision 84,15%, dan Accuracy 83,49% dan kelas pelanggan yang memiliki potensi *churn* tinggi adalah Dormant B, Dormant E, dan Dormant F.

ABSTRACT

Customers are one of the main assets for retail companies. Companies must be able to recognize how the character of their customers is so that they can retain existing customers so as not to stop buying and move to competing retail companies (*churn*). One of the right models for recognizing customer characters is the RFM (Recency, Frequency, Monetary) model. RFM model is able to produce customer classes and in each customer class can be analyzed or predicted with the concept of data mining whether the customer remains as a customer or *churn*. The data used comes from customer data and sales data at UD. Mawar Sari. UD Mawar Sari customer class produced from the RFM model are Dormant, Everyday, Golden and Superstar. The concept of data mining by building prediction models in this study uses the *Decision Tree C4.5* algorithm and *Naïve Bayes*. In all customer classes *Naïve Bayes* Algorithm performance better than *Decision Tree C4.5* Algorithm with Recall 95.92%, Precision 84.15%, and Accuracy 83.49% and customer classes that have high potential for *churn* are the Dormant B, Dormant E, and Dormant F.

Copyright © Universitas Pendidikan Ganesha. All rights reserved.

¹ Corresponding author.

E-mail addresses: niwayan.wardani@stiki-indonesia.co.id (Ni Wayan Wardani)

1. Pendahuluan

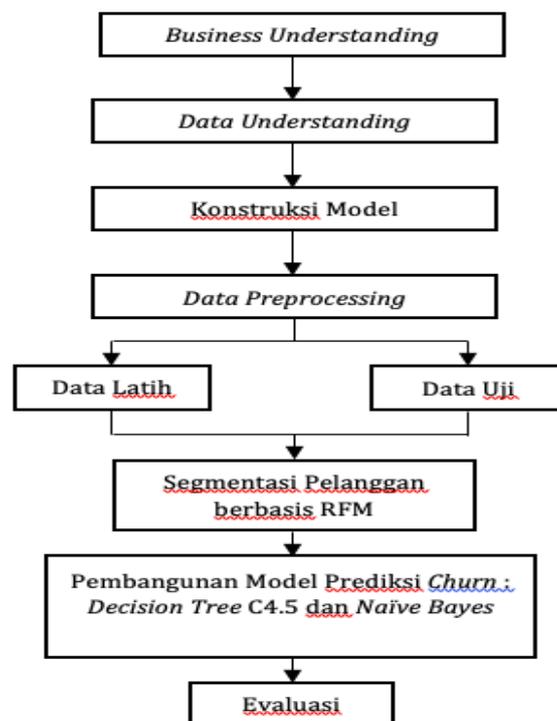
Pertumbuhan bisnis retail saat ini menunjukkan trend yang positif, tidak hanya bertumbuh di kota-kota besar tetapi juga di kota-kota kecil bahkan sampai pelosok desa. Dari pertumbuhan tersebut, tentunya memunculkan persaingan bisnis retail yang sangat ketat. Pelanggan akan dengan sangat mudah memilih toko retail yang terbaik buat mereka. Hal tersebut menjadikan perusahaan retail tidak hanya berfokus pada pengembangan produk dan layanan tetapi juga berfokus pada pelanggan dengan mampu menciptakan pengalaman personal pelanggan sehingga dapat meningkatkan loyalitas pelanggan.

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%. Untuk mempertahankan pelanggan lama, tentunya perusahaan perlu mengetahui karakter – karakter pelanggan mereka sehingga dapat mengambil langkah – langkah strategis, dengan harapan pelanggan akan semakin loyal dan tidak berpindah ke perusahaan pesaing (*churn*).

Dalam penelitian ini akan dilakukan identifikasi karakter pelanggan dan melakukan prediksi *churn* berdasarkan karakter pelanggan yang dimiliki perusahaan retail UD. Mawar Sari. Metoda yang digunakan untuk mengidentifikasi karakter pelanggan dengan menghasilkan kelas pelanggan adalah metoda RFM (*Recency, Frequency, Monetary*). Data yang digunakan untuk mendapatkan fitur *recency, frequency, dan monetary* adalah berasal dari data pelanggan dan transaksi pelanggan. Data setiap kelas pelanggan digunakan untuk pembangunan model prediksi *customer churn* menggunakan Algoritma *Decision Tree C4.5* dan Algoritma *Naïve Bayes*. Algoritma *Decision Tree C4.5* merupakan salah satu teknik klasifikasi dalam *data mining*. Kefleksibelan membuat metoda ini atraktif, khususnya karena memberikan keuntungan berupa visualiasasi keuntungan berupa visualiasasi saran dalam bentuk pohon keputusan yang membuat prosedur prediksinya dapat diamati. Algoritma *Naïve Bayes* adalah suatu klasifikasi yang bekerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas.

2. Metode

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



Gambar 1. Diagram Ikhtisar Penelitian

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 membuat prediksi *customer churn*.

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.

Konstruksi model

Konstruksi model dalam penelitian ini adalah menentukan *data window* dan *forecasting window*. *Data window* adalah kerangka waktu untuk variable 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).

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

Data Latih

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

Data Uji

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

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 sebagai berikut :

1. Superstar

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

2. Golden

Pelanggan yang mempunyai nilai *monetary* tertinggi kedua, *frequency* 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 *frequency* 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
<i>Recency</i>	1	$r > 150$ hari
	2	$61 \text{ hari} \leq r \leq 150 \text{ hari}$
	3	$0 \leq r \leq 60$ hari
<i>Frequency</i>	1	$0 < f \leq 30$
	2	$31 \leq f \leq 60$
	3	$f > 60$
<i>Monetary</i>	1	$0 < m \leq \text{Rp. } 1.500.000$
	2	$\text{Rp. } 1.500.000 < m \leq \text{Rp. } 3.500.000$
	3	$m > \text{Rp. } 3.500.000$

Pembangunan Model Prediksi *Customer Churn* dengan *Decision Tree C4.5* dan *Naïve Bayes*

Model yang digunakan untuk membangun model prediksi *customer churn* dalam penelitian ini adalah algoritma *Decision Tree C4.5* dan *Naïve Bayes*. 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 dikonversi 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
<i>Recency</i>	1	$r > 150$ hari	Lama
	2	$61 \text{ hari} \leq r \leq 150 \text{ hari}$	Agak lama
	3	$0 \leq r \leq 60$ hari	Baru saja
<i>Frequency</i>	1	$0 < f \leq 30$	Jarang
	2	$31 \leq f \leq 60$	Agak sering
	3	$f > 60$	Sering
<i>Monetary</i>	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 ($3 \times 3 \times 3$) 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			Skor Akhir RFM	Label pelanggan
	<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>		
K1	Baru saja	Jarang	Rendah	311	<i>Dormant D</i>
K2	Baru saja	Jarang	Sedang	312	<i>Dormant A</i>
K3	Baru saja	Jarang	Tinggi	313	<i>Occational A</i>
K4	Baru saja	Agak sering	Rendah	321	<i>Everyday D</i>
K5	Baru saja	Agak sering	Sedang	322	<i>Golden D</i>
K6	Baru saja	Agak sering	Tinggi	323	<i>Superstar D</i>
K7	Baru saja	Sering	Rendah	331	<i>Everyday A</i>
K8	Baru saja	Sering	Sedang	332	<i>Golden A</i>
K9	Baru saja	Sering	Tinggi	333	<i>Superstar A</i>
K10	Agak lama	Jarang	Rendah	211	<i>Dormant E</i>
K11	Agak lama	Jarang	Sedang	212	<i>Dormant B</i>
K12	Agak lama	Jarang	Tinggi	213	<i>Occational B</i>
K13	Agak lama	Agak sering	Rendah	221	<i>Everyday E</i>
K14	Agak lama	Agak sering	Sedang	222	<i>Golden E</i>
K15	Agak lama	Agak sering	Tinggi	223	<i>Superstar E</i>
K16	Agak lama	Sering	Rendah	231	<i>Everyday B</i>
K17	Agak lama	Sering	Sedang	232	<i>Golden B</i>
K18	Agak lama	Sering	Tinggi	233	<i>Superstar B</i>
K19	Lama	Jarang	Rendah	111	<i>Dormant F</i>
K20	Lama	Jarang	Sedang	112	<i>Dormant C</i>
K21	Lama	Jarang	Tinggi	113	<i>Occational C</i>
K22	Lama	Agak sering	Rendah	121	<i>Everyday F</i>
K23	Lama	Agak sering	Sedang	122	<i>Golden F</i>
K24	Lama	Agak sering	Tinggi	123	<i>Superstar F</i>
K25	Lama	Sering	Rendah	131	<i>Everyday C</i>
K26	Lama	Sering	Sedang	132	<i>Golden C</i>
K27	Lama	Sering	Tinggi	133	<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 :

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

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

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

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

Dimana :

S : himpunan kasus

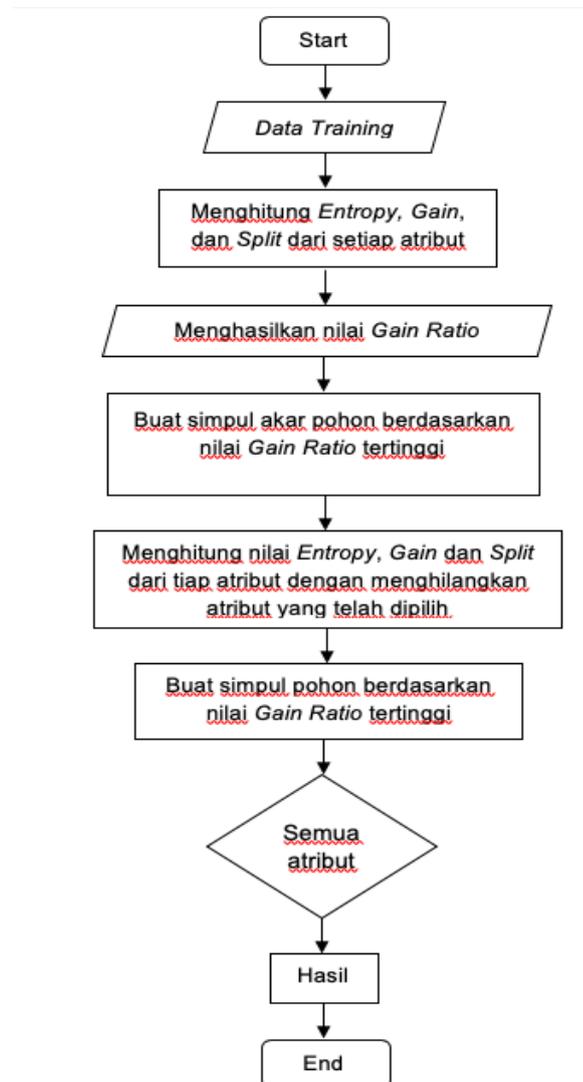
A : atribut

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*

Naïve Bayes merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan aturan *bayes* dengan asumsi ketidaktergantungan yang kuat. Selain itu *Naïve Bayes* juga dapat menganalisa variable – variabel yang paling mempengaruhinya dalam bentuk peluang (Prasetyo, 2012).

Berikut merupakan persamaan dari teorema *Bayes* :

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{p(X)}$$

Keterangan :

X : Data dengan class yang belum diketahui H : hipotesis data X merupakan suatu class spesifik

P(H|X) : Probabilitas akhir bersyarat suatu hipotesis H terjadi jika diberikan bukti X terjadi

P(H) : Probabilitas awal (*priori*) hipotesis H terjadi tanpa memandang bukti apapun

P(X|H) : Probabilitas sebuah bukti X terjadi akan mempengaruhi hipotesis H

P(X) : Probabilitas awal (*priori*) bukti X terjadi tanpa memandang hipotesis/bukti yang lain

Evaluasi

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

Tabel 5. *Confusion Matrix*

Kelas Asli	Kelas Hasil Prediksi		
	Tidak (+) Ya (-)	Tidak (+) TP (<i>True Positive</i>) FP (<i>False Positive</i>)	Ya (-) FN (<i>False Negative</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* = $TP / (TP + FN)$
2. *Precision* = $TP / (TP + FP)$
3. *Accuracy* = $(TP + TN) / (TP + FP + FN + TN)$

Keterangan :

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

3. Hasil dan pembahasan

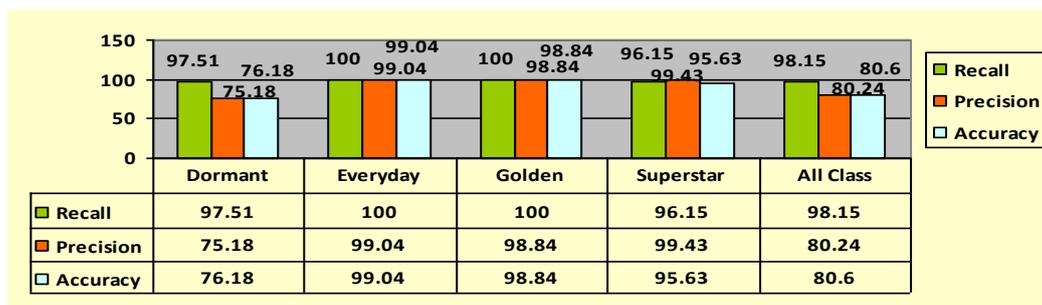
Mengidentifikasi karakter pelanggan UD.Mawar Sari melalui proses segmentasi pelanggan berbasis RFM (*Recency, Frequency, Monetary*) menghasilkan kelas pelanggan *Dormant, Everyday, Golden* dan *Superstar*. Penerapan model prediksi *customer churn* diuji coba pada 4 data yaitu :

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) dan grafik pada Gambar 3.

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%

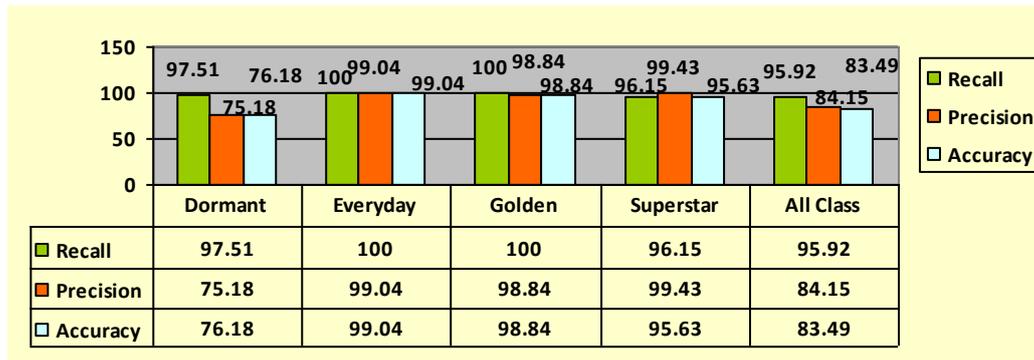


Gambar 3. Grafik Kinerja Algoritma *Decision Tree C4.5*

Berikut adalah hasil kinerja algoritma *Naïve Bayes* pada setiap kelas pelanggan (Tabel 7) dan grafik pada Gambar 4.

Tabel 7. Hasil Kinerja Algoritma *Naïve Bayes*

Kelas Pelanggan	<i>Naïve Bayes</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%



Gambar 4. Grafik Kinerja Algoritma *Naïve Bayes*

Menganalisa hasil perhitungan *recall*, *precision* dan *accuracy* dari Algoritma *Decision Tree C4.5* dan *Naïve Bayes* dapat disimpulkan bahwa kelas pelanggan UD. Mawar Sari yang berpotensi *churn* adalah kelas pelanggan *dormant*. Untuk mengetahui lebih detail bagaimana karakter pelanggan *dormant* yang berpotensi *churn*, perlu dilakukan perhitungan *recall*, *precision* dan *accuracy* pada masing-masing label pelanggan *dormant* yang dimiliki oleh UD. Mawar Sari yaitu *dormant A*, *dormant B*, *dormant D*, *dormant E* dan *dormant F*.

Penerapan model prediksi *customer churn* diuji coba pada 4 data kelas pelanggan *dormant* yaitu:

1. Data uji dengan data pelanggan kelas *Dormant A* (65 kasus)
2. Data uji dengan data pelanggan kelas *Dormant B* (6 kasus)
3. Data uji dengan data pelanggan kelas *Dormant D* (894 kasus)
4. Data uji dengan data pelanggan kelas *Dormant E* (176 kasus)
5. Data uji dengan data pelanggan kelas *Dormant F* (249 kasus)

Tabel 8. Hasil kinerja Algoritma *Decision Tree C4.5* pada kelas pelanggan *Dormant*

Kelas Pelanggan	<i>Decision Tree C4.5</i> <i>Recall</i>	<i>Precision</i>	<i>Accuracy</i>
<i>Dormant A</i>	100%	98.46%	98.46%
<i>Dormant B</i>	100%	16.67%	16.67%
<i>Dormant D</i>	100%	79.31%	79.31%
<i>Dormant E</i>	100%	26.70%	26.70%
<i>Dormant F</i>	0%	0%	84.34%

Tabel 9. Hasil kinerja Algoritma *Naïve Bayes* pada kelas pelanggan *Dormant*

Kelas Pelanggan	<i>Naïve Bayes</i> <i>Recall</i>	<i>Precision</i>	<i>Accuracy</i>
<i>Dormant A</i>	100%	98.46%	98.46%
<i>Dormant B</i>	100%	16.67%	16.67%
<i>Dormant D</i>	100%	79.31%	79.31%
<i>Dormant E</i>	100%	26.70%	26.70%
<i>Dormant F</i>	0%	0%	84.34%

4. Simpulan dan saran

Dari penjelasan di atas,, adapun simpulan yang dapat disampaikan pada penelitian ini antara lain:

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*.
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. Algoritma *Decision Tree C4.5* dan *Naïve Bayes* menghasilkan kinerja yang sama pada setiap kelas pelanggan. Hasil kinerja 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. Kinerja algoritma *Decision Tree C4.5* dan *Naïve Bayes* pada semua kelas pelanggan menghasilkan *recall, precision* dan *accuracy* yang berbeda, dimana kinerja algoritma *Naïve Bayes* lebih baik dari kinerja algoritma *Decision Tree C4.5*. Hasil kinerja *Decision Tree C4.5* yaitu *recall* 98.15%, *precision* 80.24%, dan *accuracy* 80.6% dan hasil kinerja *Naïve Bayes* yaitu *recall* 95.92%, *precision* 84.15%, dan *accuracy* 83.49%.
6. Kinerja algoritma *Decision Tree C4.5* dan *Naïve Bayes* dalam prediksi *customer churn* sangat baik pada data uji yang berisi data dengan kemiripan atau kesamaan karakteristik sedangkan pada data uji yang berisi data dengan keragaman karakteristik, kinerja algoritma *Decision Tree C4.5* dan *Naïve Bayes* dalam prediksi *customer churn* mengalami penurunan.
7. Kinerja algoritma *Naïve Bayes* lebih baik dibandingkan kinerja algoritma *Decision Tree C4.5* dalam prediksi *customer churn* yang menggunakan data uji gabungan semua kelas pelanggan.
8. 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*. Pelanggan dengan label *Dormant B, Dormant E* dan *Dormant F* adalah pelanggan yang paling berpotensi untuk *churn*.

Dari simpulan yang telah dijabarkan peneliti menyampaikan saran sebagai berikut: 1) Model segmentasi pelanggan dan prediksi dapat mempertimbangkan variable – variable lainnya seperti lamanya berlangganan data demografi pelanggan maupun data dari kompetitor seperti harga dan jenis produk yang ditawarkan oleh kompetitor. Dengan adanya variable tambahan diharapkan hasil segmentasi pelanggan dan hasil prediksi dapat lebih akurat, dan 2) Untuk penelitian berikutnya, memprediksi *customer churn* dapat menggunakan algoritma klasifikasi lainnya seperti *Artificial Neural Network, Support Vector Machine, Nearest Neighbor*, dll. Dengan harapan hasil kinerja algoritma lainnya dapat dibandingkan dengan hasil kinerja algoritma pada penelitian ini.

Daftar Rujukan

- 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.
- A. Chorianopoulos, *Effective CRM using Predictive Analytics*. Wiley, 2009.
- 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.
- E. Prasetyo, *Data mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Penerbit ANDI Yogyakarta, 2014.
- E. Prasetyo, *Data mining : Konsep dan Aplikasi Menggunakan Matlab*. Penerbit ANDI Yogyakarta, 2013.
- F. Gorunescu, *Data Mining Concept Model and Techniques*. Berlin: Springer, 2011.

- G. Klepac, *Developing Churn Models Using Data mining Techniques and Social Network Analysis*. USA: IGI Global, 2015.
- 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.
- 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.
- P. S. Venatesan, *Data Mining and Warehousing*. New Age International (P) Limited, 2007.
- R. Govindaraju, T. Simatupang, and T. A. Samadhi, "Perancangan Sistem Prediksi Churn Pelanggan," *Tek. Inform.*, vol. 9, no. 1, pp. 33–42, 2008.
- V. L. M. Oliviera, "Analytical Customer Relationship Management in Retailing Supported by Data Mining Techniques," University of Porto, 2012.
- V. L. M. Oliviera, "Predicting Partial Customer Churn using Markov for Discrimination for Modeling First Purchase Sequence," University of Porto, 2012.
- 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.
- 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.