

Prediksi Pelanggan Loyal Menggunakan Metode Naïve Bayes Berdasarkan Segmentasi Pelanggan dengan Pemodelan RFM

Ni Wayan Wardani¹, Diah Juniari Arnidya², I Nyoman Agus Suarya Putra³, Ni Made Mila Rosa Desmayani⁴, Putu Gede Surya Cipta Nugraha⁵, Eddy Hartono⁶, Gede Surya Mahendra⁷

^{1,2,3,4,5,6}Teknik Informatika, Institut Bisnis dan Teknologi Indonesia
Jalan Tukad Pakerisan No 97, Denpasar, Bali

⁷Sistem Informasi, Universitas Pendidikan ganesha
Jalan Udayana No 11, Singaraja, Bali

*Pos-el: niwayan.wardani@instiki.ac.id

ABSTRAK

Persaingan pasar yang semakin ketat telah mendorong perusahaan untuk mengupayakan bisnis yang lebih baik dan perusahaan diharapkan mengubah cara pandang dari pendekatan berorientasi produk menjadi pendekatan yang berfokus terhadap loyalitas pelanggan. Pelanggan merupakan aset yang paling berharga dalam menjalankan bisnis. Oleh karena itu, pelanggan yang loyal perlu diprediksi dengan akurat atau untuk mengetahui tingkat akurasi agar dapat membantu dalam proses pengambilan keputusan terkait masalah tersebut. Dengan menerapkan data mining untuk menemukan pola yang tersembunyi dalam data besar yang digunakan. Implementasi data mining untuk memprediksi pelanggan loyal menggunakan metode Naïve Bayes berdasarkan segmentasi pelanggan dengan pemodelan RFM dapat dilakukan dengan baik dengan memanfaatkan data transaksi penjualan. Proses segmentasi pelanggan diperlukan untuk mengelompokkan pelanggan yang memiliki kesamaan karakteristik dengan menggunakan model RFM maka menghasilkan sepuluh segment. Naïve Bayes digunakan untuk membentuk model prediksi pelanggan loyal dan hasil evaluasi dari model yang telah diuji yakni akurasi sebesar 97,27%, *precision* sebesar 100% dan *recall* sebesar 96,98%.

Kata-Kata Kunci: prediksi, pelanggan loyal, RFM, Naïve Bayes

ABSTRACT

Increasingly fierce market competition has encouraged companies to strive for better business, and companies are expected to change their perspective from a product-oriented approach to a customer loyalty-focused approach. Customers are the most valuable asset in running a business. Therefore, loyal customers need to be accurately predicted or to know the level of accuracy in order to help in the decision-making process related to the problem by applying data mining to find patterns hidden in the big data used. The implementation of data mining to predict loyal customers using the Naïve Bayes method based on customer segmentation with RFM modeling can be done well by utilizing sales transaction data. The customer segmentation process is needed to group customers who have similar characteristics by using the RFM model and then produce ten segments. Naïve Bayes is used to forming loyal customer prediction models and the evaluation

results of the models that have been tested, namely accuracy of 97.27%, precision of 100%, and recall of 96.98%.

Key Words: prediction, loyal customer, RFM, Naïve Bayes

1. PENDAHULUAN

E-commerce menjadi peluang dan tantangan yang harus dimanfaatkan oleh peritel di Indonesia untuk mempertahankan dan meningkatkan bisnis di masa pandemi Covid-19 (Kompas.id, 2021). Industri ritel dapat mengumpulkan data-data berharga tentang pelanggannya. Data tersebut dapat digunakan sebagai sumber wawasan untuk membantu perusahaan memperoleh keunggulan kompetitif. Bahkan dapat untuk meningkatkan penjualan dan mengurangi biaya. Pelanggan merupakan aset yang paling berharga dalam menjalankan bisnis. Sehingga perusahaan harus memiliki cara untuk mengembangkan aset tersebut.

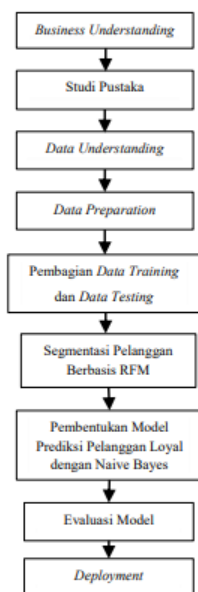
Persaingan pasar yang semakin ketat telah mendorong perusahaan untuk mengupayakan bisnis yang lebih baik dan perusahaan diharapkan mengubah cara pandang dari pendekatan berorientasi produk menjadi pendekatan yang berfokus terhadap loyalitas pelanggan (Nugroho & Saputro, 2020). Loyalitas pelanggan mutlak diperlukan agar suatu perusahaan dapat bertahan dan mampu bersaing dengan perusahaan lain. Sebagaimana yang sudah dijelaskan oleh Kiplagat dan Oyugi (2016), berpendapat bahwa mempertahankan pelanggan lebih murah daripada memperoleh pelanggan baru. Begitu pula, dikatakan oleh Nugroho dan Saputro (2016), mengklaim bahwa mempertahankan semua pelanggan yang ada umumnya akan lebih menguntungkan daripada pergantian pelanggan karena biaya untuk mendapatkan pelanggan baru bisa lima kali lipat dari biaya mempertahankan pelanggan yang sudah ada. Oleh karena itu, pelanggan yang loyal perlu diprediksi

dengan akurat atau untuk mengetahui tingkat akurasinya agar dapat membantu dalam proses pengambilan keputusan terkait masalah tersebut.

Dengan menerapkan data mining untuk menemukan pola yang tersembunyi dalam data besar yang digunakan (Sudriyanto, 2017). Pada penelitian ini menggunakan dataset transaksi penjualan dengan menggunakan metode Naïve Bayes berdasarkan segmentasi pelanggan dengan pemodelan RFM. Model RFM merupakan model berbasis perilaku digunakan untuk menganalisis perilaku pelanggan dan memprediksi loyalitas pelanggan berdasarkan perilaku *database* (Chrisnanto & Kianingsih, 2019). Terdapat tiga parameter dalam model RFM yaitu *recency*, *frequency* dan *monetary*. Salah satu keuntungan dari model RFM antara lain mudah dipahami, efektif dalam mengidentifikasi pelanggan yang berharga. Lalu kelemahannya hanya dapat menggunakan sejumlah variabel seleksi yang terbatas (Wei dkk., 2020). Dari model tersebut menghasilkan pengelompokan pelanggan dengan karakteristik pelanggan berbasis model RFM yang berupa segmen pelanggan. Untuk membangun model prediksi pelanggan loyal menggunakan metode Naïve Bayes yang merupakan sebuah metode penggolongan berdasarkan probabilitas sederhana. Keuntungan dari metode tersebut mudah dipahami dan tidak memerlukan banyak data untuk pelatihan data (Widianto, 2019).

2. METODE PENELITIAN

Penelitian ini dilakukan pada sebuah *e-commerce* berbasis UK. Dalam penelitian ini akan dilakukan beberapa tahapan seperti gambar 1.



Gambar 1
Tahapan Penelitian

Business Understanding

Tujuan dari penelitian untuk mengukur nilai akurasi mengenali pola terhadap perilaku karakteristik pelanggan untuk memprediksi pelanggan yang loyal serta melakukan uji perbandingan pada jumlah dataset yang digunakan untuk mendapatkan hasil akurasi yang terbaik.

Studi Pustaka

Tahap ini meliputi pengumpulan dan pengkajian jurnal, buku, artikel dan sebagainya yang berkaitan dengan penelitian yang sedang dilakukan. Adapun studi pustaka ini juga dijadikan sebagai acuan.

Data Understanding

Pada tahap ini dilakukan pemahaman terhadap kebutuhan data terkait pencapaian tujuan bisnis. Proses pengumpulan data, identifikasi kualitas data, dan melakukan pengecekan data. Sumber dataset transaksi penjualan diperoleh dari sebuah *website* yaitu <https://www.kaggle.com/mashlyn/online-retail-ii-uci>. Jumlah dataset yang diperoleh sebanyak 1.048.575 *record* selama dua tahun terhitung sejak tanggal

12 Januari 2009 hingga 10 Desember 2011.

Data Preparation

Pada tahap ini data yang tersimpan akan dipersiapkan untuk memudahkan proses penambangan (*mining*). Berikut beberapa proses yang dilakukan pemilihan atribut data (*data selection*) yang sesuai dengan model RFM dapat dilihat pada Tabel 1. Selanjutnya memastikan kualitas data yang telah dipilih dengan membersihkan data (*data preprocessing*). Pada tahap ini menghadapi masalah terkait data yang noise, data yang duplikat dan *missing value*. Tahap berikutnya *data transformation* yang melakukan proses perubahan format file untuk kebutuhan dalam memasukkan dataset ke dalam aplikasi RapidMiner yaitu file .csv menjadi .xlsx.

Tabel 1
Hasil Pemilihan Atribut

Nama Atribut	Keterangan
<i>CustomerID</i>	Kode pelanggan
<i>InvoiceDate</i>	Tanggal transaksi penjualan
<i>Invoice</i>	Nomor unik transaksi penjualan
<i>TotalPrice</i>	Jumlah total uang yang dibayar oleh pelanggan dalam satu kali transaksi

Pembagian Data Training dan Data Testing

Untuk menghitung nilai akurasi dari penggunaan jumlah *record* data yang berbeda. Uji perbandingan berdasarkan masing-masing jumlah *record* data yang dilakukan dengan split data menggunakan rasio sebesar 80% sebagai *data training* dan 20% sebagai *data testing* bisa dilihat pada Tabel 2. Alasan menggunakan rasio tersebut karena semakin besar data training maka dapat mewakili kumpulan data secara

keseluruhan dengan karakteristik berbeda (Wicaksono dkk., 2021). Pembagian data tersebut secara *random* dengan menggunakan excel.

Tabel 2
Rasio Perbandingan Data

Rasio (%)	Data Training	Data Testing	Total
80:20	4.689	1.172	5.861
80:20	80	20	100

Segmentasi Pelanggan dengan Pemodelan RFM

Segmentasi pelanggan diperlukan untuk mengelompokkan pelanggan dengan karakteristik yang sama (Adiana dkk., 2018). Terdapat segmentasi pelanggan dibagi menjadi sepuluh karakteristik sesuai dengan nilai RFM pada tabel 3 (Sutresno dkk., 2018; Yildirim, 2022). Saat membuat model RFM, skor 1 hingga 5 diterapkan untuk setiap pelanggan dalam *range* nilai *recency*, *frequency* dan *monetary*. Skor 5 yaitu nilai tertinggi dan skor 1 yaitu nilai terendah. Skor akhir dihitung sebagai kombinasi dari setiap skor atribut. Penilaian tersebut memfasilitasi proses segmentasi pelanggan. Nilai R, F dan M dibagi menjadi lima bagian dengan nilai 5, 4, 3, 2, dan 1.

Tabel 3
Segmen Pelanggan berdasarkan Skor RFM

Segmen	Skor <i>Recency</i>	Skor <i>Frequency</i>	Skor <i>Monetary</i>	Karakteristik
<i>Champions</i>	5	4-5	4-5	Membeli baru-baru ini, sering membeli dan menghabiskan paling banyak uang
<i>Loyal Customers</i>	2-5	3-5	3-5	Sering menghabiskan banyak uang, mempunyai frekuensi rata-rata
<i>Potential Loyalist</i>	3-5	1-3	1-3	Pelanggan melakukan pembelian baru-baru ini, sering melakukan transaksi. Namun menghabiskan jumlah uang yang cukup banyak
<i>Recent Customers</i>	4-5	1	1	Pembeli melakukan transaksi baru-baru ini namun tidak begitu sering
<i>Promising</i>	3-4	1	1	Pembeli baru-baru ini, tetapi belum menghabiskan banyak uang
<i>Customer Needing Attention</i>	2-3	2-3	2-3	Di atas rata-rata <i>recency</i> , <i>frequency</i> , dan <i>monetary</i> . Mungkin tidak membeli baru-baru ini
<i>About to Sleep</i>	2-3	1-2	1-2	Di bawah rata-rata <i>recency</i> , <i>frequency</i> dan <i>monetary</i>

<i>At Risk</i>	1-2	2-5	2-5	Menghabiskan banyak uang dan sering membeli, tapi sudah lama sekali
<i>Can't Lose Them</i>	1	4-5	1-5	Melakukan pembelian terbesar dan cukup sering melakukan transaksi. Tapi belum kembali untuk waktu yang lama
<i>Hibernating</i>	1-2	1-3	1-5	Pembelian terakhir sudah lama, <i>monetary</i> dan <i>frequency</i> yang rendah

Nilai *recency* dihitung berdasarkan tanggal transaksi terakhir dengan saat ini. Pelanggan dengan tanggal terakhir memiliki nilai tertinggi yaitu 5, dan pelanggan dengan tanggal transaksi terjauh di masa lalu memiliki nilai terendah yaitu 1. Serupa dengan nilai *frequency*, pelanggan yang sering bertransaksi memiliki nilai *frequency* yang tinggi yaitu 5. Sedangkan pelanggan yang memiliki nilai 1 berarti pelanggan tersebut jarang bertransaksi. Di sisi lain, pelanggan dengan total nilai transaksi terkecil mempunyai nilai *monetary* rendah yaitu nilai 1, sebaliknya pelanggan yang memiliki nilai transaksi tertinggi maka mempunyai nilai *monetary* tinggi yaitu 5. Rentang skor dan domain nilai diperoleh dari hasil bagi 5 dari nilai atribut terkecil hingga nilai atribut tertinggi pada tabel 4.

Skor akhir RFM dan pelabelan segmen diperoleh ketika menentukan skor dan domain nilai. Ada total

kombinasi 125 ($5 \times 5 \times 5$) skor akhir RFM. Pelanggan berharga atau memiliki tingkat loyalitas tinggi mendapatkan skor akhir 555 dan pelanggan yang memiliki tingkat loyalnya rendah menerima skor akhir 111.

Pembentukan Model Prediksi Pelanggan Loyal dengan Naïve Bayes

Model yang digunakan dalam penelitian ini untuk membangun model prediksi pelanggan loyal adalah algoritma Naïve Bayes. Data hasil model RFM masih berupa numerik, namun pemodelan Naïve Bayes memerlukan data kategorikal, sehingga perlu dilakukan konversi data RFM menjadi kategorikal. Tabel 5 menampilkan klasifikasi atribut yang telah dikonversi.

Sebelumnya telah dilakukan pengelohan data agar mendapatkan nilai loyal maupun tidak loyal dari skor akhir RFM dapat dilihat pada tabel 6.

Tabel 4
Skor dan Domain Nilai RFM

Atribut	Skor	Domain Nilai
<i>Recency</i>	5	$0 \text{ hari} \leq R \leq 23 \text{ hari}$
	4	$24 \text{ hari} \leq R \leq 72 \text{ hari}$
	3	$73 \text{ hari} \leq R \leq 204 \text{ hari}$
	2	$205 \text{ hari} \leq R \leq 415 \text{ hari}$
	1	$R > 415 \text{ hari}$
<i>Frequency</i>	1	$0 \leq F \leq 16$
	2	$17 \leq F \leq 36$
	3	$37 \leq F \leq 74$
	4	$75 \leq F \leq 171$
	5	$F > 171$
<i>Monetary</i>	1	$\$ 0 < M \leq \$ 362$
	2	$\$ 362 < M \leq \$ 787$
	3	$\$ 787 < M \leq \$ 1625$
	4	$\$ 1625 < M \leq \$ 3860$
	5	$M > \$ 3860$

Tabel 5
Klasifikasi Atribut dengan Variabel Kategorikal

Atribut	Skor	Domain Nilai	Variabel Kategorikal
Recency	5	$0 \text{ hari} \leq R \leq 23 \text{ hari}$	Baru Saja
	4	$24 \text{ hari} \leq R \leq 72 \text{ hari}$	Baru
	3	$73 \text{ hari} \leq R \leq 204 \text{ hari}$	Cukup Lama
	2	$205 \text{ hari} \leq R \leq 415 \text{ hari}$	Lama
	1	$R > 415 \text{ hari}$	Sangat Lama
Frequency	1	$0 \leq F \leq 16$	Sangat Jarang
	2	$17 \leq F \leq 36$	Jarang
	3	$37 \leq F \leq 74$	Cukup Sering
	4	$75 \leq F \leq 171$	Sering
	5	$F > 171$	Sangat Sering
Monetary	1	$\$ 0 < M \leq \$ 362$	Sangat Sedikit
	2	$\$ 362 < M \leq \$ 787$	Sedikit
	3	$\$ 787 < M \leq \$ 1625$	Cukup Banyak
	4	$\$ 1625 < M \leq \$ 3860$	Banyak
	5	$M > \$ 3860$	Sangat Banyak

Untuk membedakan antara loyal maupun tidak loyal harus menggunakan nilai kuartil pertama (Q1) sebagai dasar untuk memprediksi pelanggan loyal. Untuk memperoleh nilai kuartil pertama (Q1) didapatkan dari skor akhir RFM (Novendri dkk., 2021). Jadi hasil analisis RFM adalah kolom berlabel yang digunakan untuk memprediksi pelanggan loyal.

Tabel 6
Deskripsi Label

Nilai	Deskripsi
Loyal	Status yang diberikan kepada pelanggan dengan skor akhir RFM > 221
Tidak Loyal	Status yang diberikan kepada pelanggan dengan skor akhir RFM ≤ 221

Evaluasi Model

Untuk melihat tingkat *performance* dari pola yang dihasilkan oleh algoritma. Ukuran akurasi dari algoritma Naive

Bayes adalah *confusion matrix*. Ada empat nilai dalam matriks sebagai acuan dalam perhitungan meliputi :

- 1) TP (*True Positive*) : datanya positif yang diprediksi positif.
- 2) FN (*False Negative*) : datanya positif yang diprediksi sebagai negatif.
- 3) FP (*False Positive*) : datanya negatif namun diprediksi sebagai positif.
- 4) TN (*True Negative*) : datanya negatif yang diprediksi sebagai negatif.

Tabel 7
Konsep Confusion Matrix

Kelas Prediksi	Nilai Aktual	
	Positif	Negatif
Positif	TP (<i>True Positive</i>)	FP (<i>False Positive</i>)
Negatif	FN (<i>False Negative</i>)	TN (<i>True Negative</i>)

Proses evaluasi ini menghasilkan nilai *accuracy*, *precision*, dan *recall*. *Accuracy* adalah persentase perbandingan data yang diprediksi benar dengan data keseluruhan. *Precision* adalah tingkat

keakuratan antara data prediksi benar positif yang diminta dengan hasil prediksi yang diberikan oleh model. Sementara *recall* adalah keberhasilan model dalam menemukan kembali sebuah informasi. Berikut rumusnya meliputi :

- 1) $Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$
- 2) $Precision = \frac{TP}{TP+FP}$
- 3) $Recall = \frac{TP}{TP+FN}$

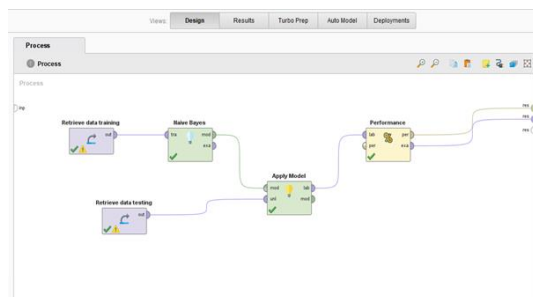
Deployment

Menjelaskan rencana pembuatan laporan yang akan dibuat dan memberikan visualisasi laporan.

3. HASIL DAN PEMBAHASAN

Proses Pembentukan Model Naïve Bayes

Tahap ini membentuk model Naïve Bayes pada RapidMiner. Pada pemodelan tersebut proses training melibatkan penggunaan data untuk melatih model (Navie Bayes) menggunakan 4.689 *record* yang terhubung ke blok model Naïve Bayes untuk mengetahui distribusi data. Kemudian pada proses *testing* menggunakan *data training* sekaligus 1.172 *record* yang terhubung ke garis penghubung pada blok *apply model* dan *performance* sebagai penampil informasi hasil pengujian data.



Gambar 2
Model Naive Bayes pada RapidMiner

Evaluasi Model

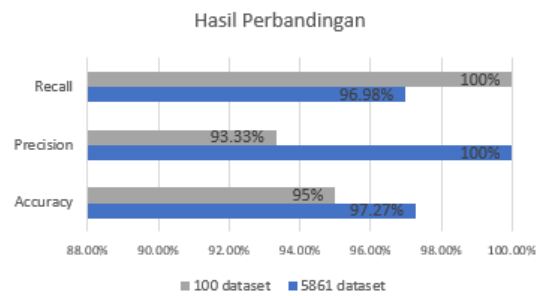
Model akan dievaluasi hasil prediksinya dengan *confusion matrix* maka akan menghasilkan data yang dapat dilihat dari tabel 8.

Tabel 8
Confusion Matrix pada Model

		<i>Actual</i>	
		<i>Loyal</i>	Tidak Loyal
<i>Predicted</i>	Loyal	1029	0
	Tidak Loyal	32	111

Berdasarkan tabel diatas, maka dapat disimpulkan bahwa :

- 1) Hasil prediksi terhadap pelanggan loyal yang benar (TP) adalah sebanyak 1029 prediksi, sedangkan hasil prediksi yang salah (FP) memiliki jumlah sebanyak 0.
- 2) Hasil prediksi terhadap pelanggan tidak loyal yang benar (TN) adalah sebanyak 111 prediksi, sedangkan hasil prediksi yang salah (FN) memiliki jumlah sebanyak 32.



Gambar 3
Hasil Perbandingan

Berdasarkan grafik diatas terlihat bahwa dari perbandingan jumlah dataset dengan algoritma Naïve Bayes, akurasi cenderung lebih tinggi saat menggunakan jumlah dataset yang lebih banyak ataupun dengan menggunakan dataset yang jumlahnya 100 dataset juga menghasilkan akurasi yang baik.

Analisis berdasarkan Visualisasi

Pada gambar 4, gambar 5 dan 6 dapat dilihat pertumbuhan pelanggan ternyata memiliki pola yang cukup bagus dari setiap periodenya. Dari grafik tersebut terlihat tingkatan pelanggan pada tahun

2011 memiliki jumlah pelanggan yang cukup banyak setiap bulannya dibandingkan dengan periode lainnya.



Gambar 4
Customers periode 2019



Gambar 5
Customers periode 2010

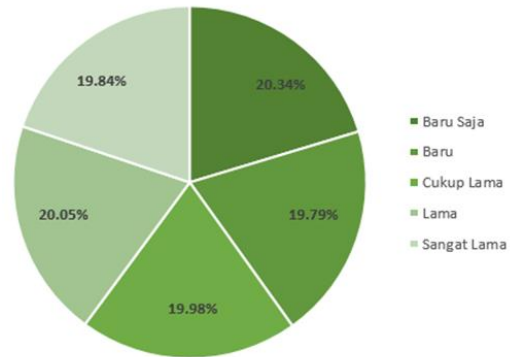


Gambar 6
Customers periode 2011

Analisis Karakteristik Data berdasarkan RFM

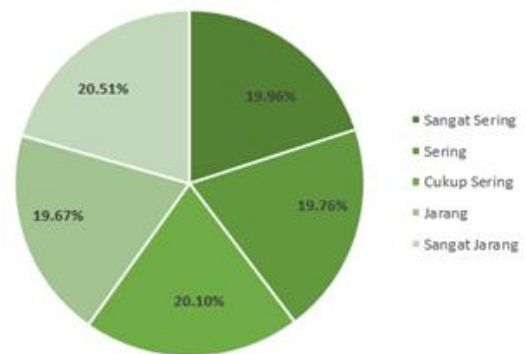
Gambar 7 menjelaskan rentang waktu terakhir bertransaksi pelanggan loyal maupun tidak loyal, pelanggan yang baru saja melakukan transaksi selama 0-23 hari sebanyak 20,34% (1.192 pelanggan).

Sedangkan pelanggan yang baru melakukan transaksi dalam kurun waktu 24-72 hari sebesar 19,79% (1.160 pelanggan).



Gambar 7
Karakteristik Data berdasarkan Recency

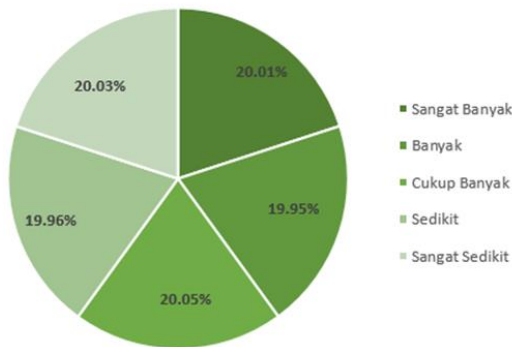
Gambar 8, menampilkan tingkat keseringan pelanggan loyal maupun tidak loyal bahwa sebanyak 20,51% (1.202 pelanggan) pelanggan ini sangat jarang bertransaksi antara 0-16 kali bertransaksi, sedangkan persentase rendah dimiliki oleh pelanggan yang jarang melakukan transaksi sebesar 19,67% (1.153 pelanggan) antara 17-36 kali bertransaksi.



Gambar 8
Karakteristik Data berdasarkan Frequency

Pada gambar 9, dilihat dari jumlah total pengeluaran pelanggan loyal maupun tidak loyal, mayoritas pelanggan yang cukup banyak dengan jumlah total pengeluarannya sekitar \$788-\$1.625

yaitu 20,05% (1.175 pelanggan). Kemudian pelanggan yang memiliki persentase rendah sebesar 19,95%, itu artinya terdapat 1.169 pelanggan yang banyak mengeluarkan uangnya saat bertransaksi sekitar \$1.626- \$3.860.



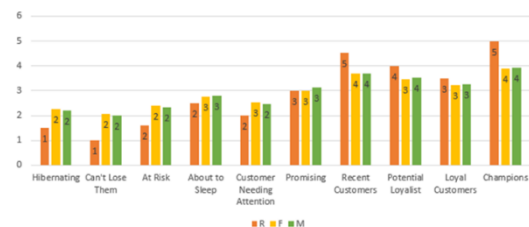
Gambar 9
Karakteristik Data berdasarkan Monetary

Analisis Segmentasi Pelanggan berdasarkan RFM

Pada gambar 10, menampilkan hasil klasifikasi dalam bentuk tampilan visual. Dilihat bahwa dalam pelanggan loyal maupun tidak loyal pada segmen Champions memiliki nilai rata-rata *recency*, *frequency* dan *monetary* lebih unggul dibandingkan dengan segmen lainnya diikuti oleh segmen *Recent*

menggunakan metode Naïve Bayes berdasarkan segmentasi pelanggan dengan pemodelan RFM dapat dilakukan dengan baik dengan memanfaatkan data transaksi penjualan. Dari 9 variabel dilakukan seleksi atribut sehingga menghasilkan 4 atribut yang digunakan yaitu: *CustomerID*, *InvoiceDate*, *Invoice*, dan *TotalPrice*. Dari hasil pengujian akurasi dengan menerapkan metode Naïve Bayes dalam memprediksi pelanggan loyal diperoleh *accuracy* sebesar 97,27%, *precision* sebesar 100% dan *recall* sebesar 96,98%. Dengan jumlah 1.729 pelanggan, segmen *Loyal Customers* memiliki pelanggan terbanyak selama 2 tahun dengan karakteristik sering menghabiskan uang dan

Customers. Pelanggan pada segmen tersebut merupakan pelanggan yang baru saja melakukan transaksi, pelanggan yang sering bertransaksi dengan jumlah uang yang dikeluarkan banyak. Sementara itu, segmen *Can't Lose Them* merupakan segmen dengan nilai rata-rata RFM terendah. Pada pelanggan tersebut memiliki nilai rata-rata *recency* dengan skor 1, artinya waktu transaksi terakhir dilakukan sangat lama. Namun demikian pelanggan tersebut dengan nilai rata-rata *frequency* dengan skor 2, yaitu pelanggan yang jarang melakukan transaksi dan jumlah uang yang dikeluarkan sedikit.



Gambar 10
Customer Segmentation by RFM

4. SIMPULAN

Berdasarkan dari hasil penelitian dapat ditarik beberapa kesimpulan, dengan implementasi data mining untuk memprediksi pelanggan loyal mempunyai frekuensi rata-rata. Sedangkan segmen dengan jumlah pelanggan paling sedikit adalah segmen *At Risk* dengan pelanggan yang karakteristiknya menghabiskan banyak uang dan sering membeli, tetapi sudah lama sekali.

Saran dari hasil pengujian yang telah dilakukan adalah peneliti telah membahas penggunaan metode Naïve Bayes berdasarkan segmentasi pelanggan dengan pemodelan RFM dalam penelitian memprediksi pelanggan loyal, diharapkan dapat dikembangkan lebih lanjut dengan metode klasifikasi lainnya seperti metode C4.5, K-Nearest Neighbor, dan lain-lain. Lalu disarankan untuk menambahkan atribut atau variabel

yang mempengaruhi hasil segmentasi pelanggan dan prediksi yang lebih akurat, serta mengembangkan penelitian dan mempertajam analisis.

DAFTAR PUSTAKA

- Adiana, B. E., Soesanti, I., & Permanasari, A. E. (2018). Analisis Segmentasi Pelanggan Menggunakan Kombinasi Rfm Model Dan Teknik Clustering. *Jurnal Terapan Teknologi Informasi*, 2(1), 23–32. <https://doi.org/10.21460/jutei.2018.21.76>
- Bhatia, P. 2019. Data Mining and Data Warehousing: Principles and Pratical Techniques. *Studies in Computational Intelligence*, 47. https://doi.org/10.1007/978-3-540-48399-1_10.
- Chrisnanto, Y. H., & Kanianingsih, A. (2019). Pengelompokan Ekuitas Pelanggan Berbasis Recency Frequency Monetary (Rfm) Menggunakan K-Means Clustering. *Seminar Nasional Teknologi Informasi Dan Komunikasi 2019 (SENTIKA 2019)*, 2019(Sentika), 13–14. <https://fti.uajy.ac.id/sentika/publikasi/makalah/2019/1.pdf>
- Drajana, I. C. 2019. "Prediksi Loyalitas Pelanggan Indihome Dengan Metode K-Nearest Neighbor". *Simtek : jurnal sistem informasi dan teknik komputer*, 4(2), 100–103. <https://doi.org/10.51876/simtek.v4i2.56>.
- Feblian, D., dan Daihani, D. U. 2017. "Implementasi Model Crisp-Dm Untuk Menentukan Sales Pipeline Pada Pt X". *Jurnal Teknik Industri*, 6(1), 1–12. <https://doi.org/10.25105/jti.v6i1.1526>.
- Ginantra, N. L. W. S. R., & Wardani, N. W. (2019). IMPLEMENTASI METODA NAÏVE BAYES DAN VECTOR SPACE MODEL DALAM DETEKSI KESAMAAN ARTIKEL JURNAL BERBAHASA INDONESIA. *Jurnal Infomedia: Teknik Informatika, Multimedia & Jaringan*, 4(2), 94-100.
- Harahap, M., Lubis, Y., dkk. 2019. "Data Science bidang Pemasaran: Analisis Prilaku Pelanggan", 1(1), 21-32.
- Kompas.id. (2021, November 11). Sektor Ritel Dorong Transformasi Digital - Kompas.id. *Grahadyarini, BM Lukita*. <https://www.kompas.id/baca/ekonomi/2021/11/11/sektor-ritel-dorong-transformasi-digital>
- Novendri, R., Andreswari, R., Pratiwi, O. N., Industri, F. R., & Telkom, U. (2021). *Implementasi Data Mining Untuk Memprediksi Customer Churn Menggunakan Algoritma Naive Bayes Implementation of Data Mining To Predict Customer Churns Using*. 8(2), 2762–2773.
- Nugraha, P. G. S. C., Dantes, G. R., & Aryanto, K. Y. E. (2017). Implementasi metode c4. 5 dan naive bayes berbasis adaboost untuk memprediksi kelayakan pemberian kredit. *International Journal of Natural Science and Engineering*, 1(2), 65-71.
- Nugroho, D. W. P. S., & Saputro, P. A. (2020). "UPAYA – UPAYA MEMPERTAHAKAN LOYALITAS PELANGGAN PADA PT. PERTAMINA LUBRICANTS REGION IV SEMARANG" (Pada Produk Fastron). *Angewandte Chemie International Edition*, 6(11), 951–952., 18(3), 1–22.
- Ramayu, I. M. S., Susanto, F., & Mahendra, G. S. (2022, March). PENERAPAN DATA MINING DENGAN ALGORITMA C4. 5 DALAM PEMESANAN OBAT GUNA MENINGKATKAN KEUNTUNGAN APOTEK.

- In *SENADA (Seminar Nasional Manajemen, Desain Dan Aplikasi Bisnis Teknologi)* (Vol. 5, pp. 237-245).
- Sudriyanto. (2017). *CLUSTERING LOYALITAS PELANGGAN DENGAN METODE RFM (RECENCY, FREQUENCY, MONETARY) DAN FUZZY C-MEANS*. 815–822.
- Sutresno, S. A., Iriani, A., & Sedyono, E. (2018). Metode K-Means Clustering dengan Atribut RFM untuk Mempertahankan Pelanggan. *Jurnal Teknik Informatika Dan Sistem Informasi*, 4, 433. <http://dx.doi.org/10.28932/jutisi.v4i3.878>
- Utami, Yohana Tri., Shofiana D. A., dkk. 2020. " Penerapan Algoritma C4.5 Untuk Prediksi Churn Rate Pengguna Jasa Telekomunikasi", 8(2), 69-76. Waryadi, W. 2018. "Penerapan Data Minig Dengan Algoritma Naive Bayes Untuk Menganalisa Pelanggan Aktif Dalam Penjualan Produk Di PT Youm Kwang Indonesia".
- Wardani, N. W. (2020). *Penerapan Data Mining Dalam Analytic CRM*. Yayasan Kita Menulis.
- Wardani, N. W., Dantes, G. R., & Indrawan, G. (2018). Prediksi customer churn dengan algoritma decision tree C4. 5 berdasarkan segmentasi pelanggan untuk mempertahankan pelanggan pada perusahaan retail. *Jurnal RESISTOR (Rekayasa Sistem Komputer)*, 1(1), 16-24.
- Wardani, N. W., & Hartono, E. (2020). IDENTIFIKASI KARAKTER PELANGGAN RETAIL BERPOTENSI CHURN DENGAN METODA RECENCY, FREQUENCY DAN MONETARY. *Jurnal Teknologi Informasi dan Komputer*, 6(3).
- Wardani, N. W., & Ariasih, N. K. (2019). Analisa komparasi algoritma decision tree C4. 5 dan naïve bayes untuk prediksi churn berdasarkan kelas pelanggan retail. *International Journal of Natural Science and Engineering*, 3(3), 103-112.
- Wei, J. T., Lin, S. Y., Yang, Y. Z., & Wu, H. H. (2020). Using a combination of RFM model and cluster analysis to analyze customers' values of a veterinary hospital. *IAENG International Journal of Computer Science*, 47(3), 442–448.
- Wicaksono, A., Anita, A., & Padilah, T. N. (2021). Uji Performa Teknik Klasifikasi untuk Memprediksi Customer Churn. *Bianglala Informatika*, 9(1), 37–45. <https://doi.org/10.31294/bi.v9i1.9992>
- Widiyanto, M. H. (2019). Algoritma Naive Bayes. *Binus University*. <https://binus.ac.id/bandung/2019/12/algorithm-naive-bayes/>