

BAB II

LANDASAN TEORI

2.1 *Churn*

Menurut Hong dalam Yulianti (2016), *churn* pelanggan adalah kecenderungan pelanggan untuk meninggalkan penyedia layanan, atau beralihnya pelanggan dari satu penyedia layanan ke yang lainnya. Sementara itu, menurut Lazarov et al dalam Suryana (2014) *churn* berasal dari kata “*change*” dan “*turn*” yang berarti tidak berlanjutnya kontrak.

Churn harus di klasifikasi secara akurat melalui segmentasi karena dapat menentukan strategi dan promosi yang tepat. Tujuan utama dari klasifikasi *churn* adalah untuk memperoleh informasi dari pelanggan yang berpotensi *churn* tinggi dengan menggunakan pemodelan klasifikasi (Syamala dalam Nana, 2019). Identifikasi pelanggan *churn* dapat membantu perusahaan dalam mempertahankan pelanggannya dengan menganalisa sekumpulan data untuk menemukan pola dan kecenderungannya sehingga potensi kerugian perusahaan dapat dihindari (Umayaparvathi dalam Yulianti, 2018).

2.1.1 Tipe *Churn*

Ada tiga tipe dari *churn* yaitu:

1. Active/Deliberate

Pelanggan memutuskan untuk berhenti berlangganan dan beralih ke penyedia layanan lainnya/jasa lainnya.

2. *Rotational/Incidental*

Pelanggan memutuskan untuk berhenti berlangganan dan tidak beralih ke penyedia layanan lainnya/jasa pengiriman lainnya.

3. *Passive/Non Voluntary*

Tipe *Passive/Non Voluntary* adalah keputusan yang dilakukan oleh perusahaan.

2.1.2 Sebab & Alasan Churn

Menurut Keaveney dalam E. Parmawati (2008) (*Customer Switching Behavior in Service Industries*) dalam sebuah studi akademis sekitar 500 insiden pengalihan layanan, 8 motif terbesar yang melatari keputusan beralih (berhenti berlangganan) telah teridentifikasi. Delapan motif terbesar tersebut yaitu:

1. Masalah Tarif/Harga (*Pricing*)

Alasan tarif yang terlalu mahal, tarif mengalami kenaikan, serta tarif yang tidak wajar.

2. Masalah Ketidaknyamanan dan Ketidaksenangan Pelayanan (*Inconvenience*)

Ditunjukkan dengan ketidaknyamanan pada tempat dan waktu sehingga terlalu lama dan harus menunggu.

3. Kegagalan Inti Layanan (*Core Service Failure*)

Adanya kekeliruan dan kesalahan dalam pelayanan yang menyebabkan kerugian/masalah bagi pelanggan.

4. Kegagalan dalam Pemberian Pelayanan (*Service Encounter Failure*)

Disebabkan pelayanan yang tidak memerdulikan pelanggan, tidak sopan dan tanggap, petugas lalai dalam tugasnya.

5. Tanggapan Terhadap Kegagalan Pelayanan (*Response to Service Failure*)

Muncul tanggapan negatif dari pelanggan tentang pelayanan, tidak ada tanggapan, dan petugas melayani dengan enggan.

6. Persaingan (*Competition*)

Pelanggan menemukan layanan lain yang lebih baik.

7. Faktor Etik (*Ethical Problems*)

Menggunakan teknik pelayanan dengan kecurangan dan memaksa.

8. Faktor Peralihan Diluar Inisiatif (*Involuntary Switching*)

Pelanggan pindah atau penyedia layanan tutup.

2.2 RFM (Recency, Frequency, Monetary)

RFM merupakan metode yang digunakan untuk menganalisis nilai dan karakteristik pelanggan, umumnya digunakan oleh perusahaan retail dan industri pelayanan jasa berdasarkan catatan transaksi (Fader dalam Wikipedia, 2019). Model *RFM* telah banyak diterapkan dalam beberapa bidang terutama dalam dunia pemasaran sehingga dapat secara efektif mengidentifikasi pelanggan yang berharga yang digunakan sebagai penentuan strategi pemasaran yang efektif. Model *RFM* sering digunakan untuk proses segmentasi pelanggan.

Menurut K. Tsiptsis dan Chorianopoulos dalam Taqwim (2019), metode *RFM* memiliki penjelasan sebagai berikut :

1. *Recency*

Recency merupakan variabel untuk mengukur nilai pelanggan berdasarkan rentang waktu (tanggal, bulan, tahun) transaksi terakhir pelanggan sampai saat ini. Semakin kecil rentang waktu maka nilai *R* semakin besar.

2. *Frequency*

Frequency merupakan variabel untuk mengukur jumlah transaksi yang dilakukan pelanggan dalam satu periode. Semakin banyak jumlah transaksi yang dilakukan maka nilai *F* semakin besar.

3. *Monetary*

Monetary merupakan variabel untuk menghitung total besaran uang yang dikeluarkan pelanggan dalam satu periode. Semakin banyak jumlah besaran uang yang dikeluarkan pelanggan maka nilai *M* semakin besar.

Menurut Cheng dan Chen dalam Aulia Dewi (2018), semakin besar nilai *R* dan *F* maka kemungkinan pelanggan akan melakukan transaksi kembali dengan perusahaan tersebut. Semakin besar nilai *M* maka kecenderungan pelanggan dalam memberikan respon kepada produk dan layanan perusahaan semakin besar.

2.3 Segmentasi Pelanggan

Menurut Aulia Dewi, dkk (2018) segmentasi merupakan proses untuk mengetahui karakteristik pelanggan dengan kesamaan tertentu sehingga

mempermudah penggalan informasi terkait pelanggan yang *churn* dengan loyal. Sedangkan menurut Rully Agus, dkk (2015) segmentasi pelanggan adalah salah satu teknik dalam mengidentifikasi pelanggan berdasarkan catatan transaksi kemudian mempelajari pola dari setiap kelas pelanggan.

Segmentasi adalah proses membagi pelanggan menjadi beberapa kelas dengan beberapa karakteristik loyalitas pelanggan berdasarkan nilai *RFM* seperti yang ditunjukkan pada tabel 2.1 (K. Tsipsis dalam Beta, 2018):

Tabel 2.1 Karakteristik Pelanggan Berdasarkan Nilai *RFM*

Kelas Pelanggan	Karakteristik
<i>Superstar</i>	a. Pelanggan dengan loyaliti yang tinggi b. Mempunyai nilai <i>R, F, M</i> paling tinggi
<i>Golden</i>	a. Mempunyai nilai <i>M</i> tertinggi ke – 2 b. Mempunyai nilai <i>F</i> yang tinggi c. Mempunyai nilai <i>R</i> rata-rata
<i>Occational</i>	a. Mempunyai nilai <i>M</i> terendah ke - 2 setelah <i>dormant</i> b. Mempunyai nilai <i>F</i> paling rendah c. Mempunyai nilai <i>R</i> paling tinggi
<i>Everyday</i>	a. Memiliki <i>nilai F</i> rata-rata b. Mempunyai nilai <i>R</i> yang rendah c. Mempunyai nilai <i>M</i> sedang sampai dengan rendah
<i>Dormant</i>	a. Mempunyai nilai <i>R, F, M</i> paling rendah b. Pelanggan dengan <i>churn</i> paling tinggi

2.4 *Naïve Bayes*

Naïve Bayes merupakan salah satu metode *machine learning* dengan menggunakan metode probabilitas. Probabilitas keanggotaan dapat dihitung pada proses pengklasifikasian didasarkan pada teorema bayes (Gayatri dalam Yulianti,

2016). Kelebihan penggunaan *Naïve Bayes* adalah metode ini hanya membutuhkan jumlah data pelatihan (*data training*) yang kecil dalam proses pengklasifikasian (Rezdy dalam Hairul Umam, 2017).

Persamaan metode *Naïve Bayes* adalah sebagai berikut (Sartika, 2017).

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad \dots (1)$$

Keterangan :

X : Data dengan kelas yang belum diketahui

H : Hipotesis data X merupakan suatu kelas spesifik

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X

P(H) : Probabilitas hipotesis H

P(X|H) : Probabilitas hipotesis X berdasarkan kondisi H

P(X) : Probabilitas X

Nilai P (X) dapat diabaikan karena nilainya adalah konstan untuk semua kelas. Sedangkan untuk mengurangi perhitungan dan evaluasi P(X|H), dibuat asumsi tinggi (naif) tentang independensi, mengansumsikan bahwa nilai – nilai atribut saling bebas (independen) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu persamaan (2).

$$\begin{aligned} P(X|H) &= \prod_{k=1}^n P(X_k|H) \\ &= P(X_1|H) \times P(X_2|H) \times \dots \times P(n|H) \quad \dots (2) \end{aligned}$$

2.5 *RapidMiner*

RapidMiner adalah *platform* perangkat lunak yang menyediakan prosedur *data mining* untuk pengolahan data, *machine learning*, model prediktif, *text mining*, dan analisis prediktif (*RapidMiner*). *RapidMiner* memiliki operator yang fleksibel untuk *input* dan *output* data dalam berbagai format *file*, berisi skema pembelajaran untuk *classification*, *regression* and *clustering* (Naik dan Samant dalam Ardiansah, 2019). Pada *RapidMiner* prosedur *data mining* di proses tersusun atas operator-operator yang *nestable*, dideskripsikan dengan XML dan dibuat dengan GUI, lalu penyajiannya dituliskan dalam bahasa pemrograman Java (Nandrabertus, 2015).

2.6 *Confusion Matrix*

Confusion Matrix merupakan suatu metode yang digunakan untuk melakukan perhitungan akurasi serta mengukur kinerja model dari *data mining*. Metode *Confusion Matrix* berguna dalam menganalisa seberapa baik pengklasifikasian dapat mengenali setiap kelas yang berbeda (Han dalam Yulianti, 2016). *Confusion Matrix* memberikan penilaian kinerja model klasifikasi berdasarkan jumlah objek yang diprediksi dengan benar dan salah (Gorunescu dalam Maulana, 2019). Tabel klasifikasi adalah tabel yang terdiri dari kelas asli dan kelas prediksi, tabel ini digunakan dengan tujuan untuk mengukur kinerja suatu model klasifikasi. Berikut ini merupakan tabel klasifikasi *confusion matrix* (Chawla NV dalam Hanifa 2018):

Tabel 2.2 Tabel *Confusion Matrix*

		Kelas Asli	
		<i>Churn</i>	<i>Loyal</i>
Kelas Prediksi	<i>Churn</i>	<i>TP (True Positive)</i>	<i>FN (False Negative)</i>
	<i>Loyal</i>	<i>FP (False Positive)</i>	<i>TN (True Negative)</i>

- *TP (True Positive)* adalah jumlah pelanggan dengan status *churn*, ketika diprediksi hasil prediksi menunjukkan *churn*.
- *FP (False Positive)* adalah jumlah pelanggan dengan status *churn*, ketika diprediksi hasil prediksi menunjukkan loyal.
- *FN (False Negative)* jumlah pelanggan dengan status loyal, ketika diprediksi hasil prediksi menunjukkan *churn*.
- *TN (True Negative)* jumlah pelanggan dengan status loyal, ketika diprediksi hasil prediksi menunjukkan loyal.

Hasil evaluasi dari metode *confusion matrix* menghasilkan nilai *Accuracy*, *Precision*, *Recall*, serta *Classification Error* dengan penjelasannya sebagai berikut:

1. *Accuracy*

Accuracy adalah persentase kedekatan antara nilai data yang diprediksi secara benar (positif dan negatif) dengan keseluruhan data setelah dilakukan pengujian pada hasil klasifikasi (Resika, 2019).

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)}$$

2. *Precision*

Precision dihitung untuk mengevaluasi seberapa baik ketepatan model dapat memprediksi suatu kelas. *Precision* didapatkan dengan menghitung perbandingan antara jumlah data benar positif yang diprediksi dibagi dengan jumlah total keseluruhan hasil yang diprediksi positif (Emzir dalam Adiwijaya, 2017).

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

3. *Recall*

Recall didapatkan dengan menghitung perbandingan antara jumlah data benar positif yang diprediksi dengan benar dibagi dengan jumlah total keseluruhan hasil yang diprediksi benar positif (Emzir dalam Adiwijaya, 2017).

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

4. *Classification Error*

Classification Error adalah perkiraan tingkat kesalahan klasifikasi untuk mengetahui ukuran dari perkiraan kesalahan sebenarnya (Han dalam Ardiansah, 2019).

$$Classification Error = \frac{(FP + FN)}{(TP + FP + FN + TN)}$$

2.7 CRISP-DM

CRISP-DM (CRoss-Industry Standard Process for Data Mining) merupakan suatu konsorsium perusahaan yang didirikan oleh Komisi Eropa pada tahun 1996 dan telah ditetapkan sebagai proses standar dalam *data mining* yang dapat diaplikasikan di berbagai sektor industri. Menurut Pete Chapman dkk (2000) dalam bukunya yang berjudul *CRISP-DM 1.0*, ada 6 (enam) tahapan siklus hidup pengembangan *data mining* yang dijabarkan sebagai berikut:

1. *Business Understanding*

Tahap pertama adalah memahami tujuan dan kebutuhan dari sudut pandang bisnis, kemudian menerjemahkan pengetahuan ini ke dalam pendefinisian masalah pada *data mining*. Selanjutnya akan ditentukan rencana dan strategi untuk mencapai tujuan tersebut.

2. *Data Understanding*

Tahap kedua dimulai dengan pengumpulan data yang kemudian akan dilanjutkan dengan proses untuk mendapatkan pemahaman yang mendalam tentang data, mengidentifikasi masalah kualitas data, atau untuk mendeteksi adanya bagian yang menarik dari data yang dapat digunakan untuk hipotesa untuk informasi yang tersembunyi.

3. *Data Preparation*

Tahap ini meliputi semua kegiatan untuk membangun *dataset* akhir (data yang akan diproses pada tahap pemodelan) dari data primer. Tahap ini dapat diulang beberapa kali. Pada tahap ini juga mencakup pemilihan

tabel dan atribut data, termasuk proses *data cleaning* dan *data selection* untuk kemudian dijadikan masukan dalam tahap pemodelan.

4. *Data Transformation*

Pada tahap ini data yang sudah di seleksi di transformasikan ke dalam bentuk yang lebih sesuai dan terukur dengan sehingga dapat digunakan sebagai proses klasifikasi. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Proses ini disebut sebagai normalisasi data.

5. *Modelling*

Tahapan ini dilakukan pemilihan dan penerapan teknik pemodelan dan beberapa parameternya akan disesuaikan untuk mendapatkan nilai yang optimal. Secara khusus, ada beberapa teknik berbeda yang dapat diterapkan untuk masalah *data mining* yang sama.

6. *Evaluation*

Tahap terakhir akan dilakukan evaluasi terhadap keefektifan dan kualitas model sebelum digunakan dan menentukan apakah model dapat mencapai tujuan yang ditetapkan pada fase awal (*Business Understanding*).

2.8 Gambaran Umum Tempat Studi Kasus

2.8.1 Pengenalan ESL Express

ESL Express merupakan salah satu perusahaan jasa pengiriman ekspres dengan jaringan kantor cabang. Awalnya ESL Express merupakan

salah satu divisi dari PT. Eka Sari Lorena Transport “Lorena” yang didirikan pada tahun 1970 oleh Bapak G.T. Soerbakti, yaitu perusahaan yang bergerak di bisnis transportasi darat yang juga melayani jasa pengiriman barang melalui bus guna memaksimalkan armada “LORENA”. Dengan semakin meningkatnya jasa pengiriman barang maka pada tanggal 26 Desember 1995 didirikan PT. Eka Sari Lorena “ESL Express” dan terus melakukan pengembangan produk layanannya melalui udara dan laut ke seluruh wilayah Indonesia.

ESL Express mempunyai jaringan terbesar dan terluas yang berpusat di Kelapa Gading, DKI Jakarta dan mempunyai 77 kantor cabang di 40 kota, 242 kantor perwakilan, dan 505 kantor agen di 138 kota dengan wilayah jangkauan ke 300 kota/kabupaten di 34 provinsi yang tersebar di berbagai wilayah Indonesia.

ESL Express bekerja sama dengan PT. Lintasarta untuk pembuatan *website* dan Odiysys untuk pembuatan aplikasi dan SAP. Penetapan tarif pengiriman dalam ESL Express berdasarkan ukuran dan berat barang. Transaksi biasanya dilakukan dengan *cash* (tunai) dan *retail corporate* dibayar melalui *invoice* (kredit). ESL Express mempunyai visi dan misi yaitu menjadi perusahaan jasa pengiriman ekspres terbesar di Indonesia, yang mengangkut barang melalui darat, udara dan laut dengan sistem terintegrasi yang memberikan nilai tambah kepada pelanggan kami dengan menyediakan distribusi yang handal dan efisien serta pelayanan yang dapat dipercaya.

2.8.2 Kondisi Data Transaksi di ESL Express

Melalui perizinan oleh pihak ESL Express Cabang Tasikmalaya didapatkan data yang digunakan untuk proses klasifikasi *churn* merupakan data transaksi ESL Express pada tanggal 1 Januari 2017 – 2 Oktober 2018. Data yang didapatkan berjumlah 3420 data dalam format *file* Microsoft Excel. Ketentuan dari data yang diberikan hanya untuk periode 2 tahun sampai dengan Oktober 2018 dikarenakan pihak ESL Express khususnya cabang Tasikmalaya hanya mengizinkan data yang digunakan pada penelitian berupa data transaksi maksimal 2 tahun ke belakang sesuai SOP ESL Express dalam kebijakan perlindungan privasi perusahaan. Tabel 2.3 merupakan beberapa *query* data transaksi ESL Express pada bulan Januari 2017 yang terdiri dari:

Tabel 2.3 Data Pelanggan ESL Express Bulan Januari 2017

No.	No. eCN	Tanggal	Kota Asal	Kota Tujuan	Layanan	Tipe Transaksi	Pelanggan	Koli	Kilo	Biaya Kirim + Surcharge	Total
1	2204157000026	01/01/2017 13.20	TSY	PEKANBARU	ROAD	CASH-CASH	VERRA	1	13.00	169,000.00	145,086.50
2	2204157000042	02/01/2017 00.00	TSY	BANDUNG	ROAD	CASH-CASH	KAMALUDIN	2	18.00	167,400.00	126,805.50
3	2204157000067	02/01/2017 10.45	TSY	PURWAKARTA	TRUCKING	CASH-CASH	YUDI	1	750.00	3,100,000.00	2,726,350.00
4	2204157000083	02/01/2017 13.58	TSY	JAKARTA	COMAIL	COMAIL	CAPEM TASIKMALAYA - NAGARAWANGI	1	1.00	0.00	0.00
5	2204157000109	02/01/2017 14.26	TSY	JAKARTA	ROAD	CASH-CASH	YULIA INDRAWATI	1	1.00	11,400.00	11,514.00
6	2204157000182	02/01/2017 00.00	TSY	BANDUNG	ROAD	CASH-CASH	VINA	1	1.00	11,200.00	9,954.56

Tabel 2.3 Data Pelanggan ESL Express Bulan Januari 2017 (Lanjutan 1)

7	2204157000208	02/01/2017 17.28	TSY	BANDUNG	ROAD	CASH-CASH	RENITA KURNIASARI	1	1.00	11,400.00	11,514.00
8	2204157000216	03/01/2017 12.38	TSY	JAKARTA	ROAD	CASH-CASH	SURYA ELEKTRONIK	1	2.00	18,800.00	17,089.20
9	2204157000224	03/01/2017 13.21	TSY	BANDUNG	ROAD	CASH-CASH	INTAN	1	1.00	11,200.00	11,312.00
10	2204157000232	03/01/2017 13.23	TSY	JAKARTA	ROAD	CASH-CASH	BP R IRAWAN	1	1.00	11,400.00	11,514.00

Tabel 2.4 Data Pelanggan ESL Express Bulan Januari 2018

No.	No. eCN	Tanggal	Kota Asal	Kota Tujuan	Layanan	Tipe Transaksi	Pelanggan	Koli	Kilo	Biaya Kirim + Surchage	Total
1	2204157000026	01/01/2017 13.20	TSY	PEKANBARU	ROAD	CASH-CASH	VERRA	1	13.00	169,000.00	145,086.50
2	2204157000042	02/01/2017 00.00	TSY	BANDUNG	ROAD	CASH-CASH	KAMALUDIN	2	18.00	167,400.00	126,805.50
3	2204157000067	02/01/2017 10.45	TSY	PURWAKARTA	TRUCKING	CASH-CASH	YUDI	1	750.00	3,100,000.00	2,726,350.00
4	2204157000083	02/01/2017 13.58	TSY	JAKARTA	COMAIL	COMAIL	CAPEM TASIKMALAYA - NAGARAWANGI	1	1.00	0.00	0.00
5	2204157000109	02/01/2017 14.26	TSY	JAKARTA	ROAD	CASH-CASH	YULIA INDRAWATI	1	1.00	11,400.00	11,514.00
6	2204157000182	02/01/2017 00.00	TSY	BANDUNG	ROAD	CASH-CASH	VINA	1	1.00	11,200.00	9,954.56
7	2204157000208	02/01/2017 17.28	TSY	BANDUNG	ROAD	CASH-CASH	RENITA KURNIASARI	1	1.00	11,400.00	11,514.00
8	2204157000216	03/01/2017 12.38	TSY	JAKARTA	ROAD	CASH-CASH	SURYA ELEKTRONIK	1	2.00	18,800.00	17,089.20
9	2204157000224	03/01/2017 13.21	TSY	BANDUNG	ROAD	CASH-CASH	INTAN	1	1.00	11,200.00	11,312.00
10	2204157000232	03/01/2017 13.23	TSY	JAKARTA	ROAD	CASH-CASH	BP R IRAWAN	1	1.00	11,400.00	11,514.00

2.9 Penelitian Terkait

Pada bagian penelitian terkait, ada beberapa jurnal yang diangkat sebagai referensi dalam memperkaya bahan kajian pada penelitian dan sebagai acuan untuk penelitian selanjutnya. Berikut merupakan penelitian-penelitian berupa beberapa jurnal yang berkaitan dengan penelitian yang akan dilakukan, diantaranya akan dijelaskan pada tabel 2.5 sebagai berikut:

Tabel 2.5 Rangkuman Penelitian Terkait

No	Judul	Penulis	Tahun	Metodologi	Hasil Penelitian
1.	Analisa Komparasi Algoritma <i>Decision Tree</i> C4.5 dan <i>Naïve Bayes</i> untuk Prediksi <i>Churn</i> Berdasarkan Kelas Pelanggan Retail	Ni Wayan Wardani dan Ni Kadek Ariasih	2019	Mengenali karakter pelanggan dan menghasilkan kelas pelanggan menggunakan model <i>RFM</i> (<i>Recency, Frequency, Monetary</i>) serta membangun model prediksi menggunakan algoritma <i>Decision Tree</i> C4.5 dan <i>Naïve Bayes</i> .	Kelas pelanggan UD Mawar Sari yang dihasilkan dari model <i>RFM</i> adalah <i>Dormant, Everyday, Golden, dan Superstar</i> . Di semua kelas pelanggan kinerja Algoritma <i>Naïve Bayes</i> lebih baik daripada Algoritma <i>Decision Tree</i> C4.5 dengan <i>Recall</i> 95,92%, <i>Precision</i> 84,15%, dan <i>Accuracy</i> 83,49% dan kelas pelanggan yang memiliki potensi <i>churn</i> tinggi adalah <i>Dormant B, Dormant E, dan Dormant F</i> .
2.	Analisis Segmentasi Pelanggan dengan <i>RFM</i> Model pada PT.	Wiratama Ahsani Taqwim, Nanang Yudi	2019	Proses segmentasi untuk mengetahui karakteristik pelanggan menggunakan model <i>RFM. Fuzzy C-Means</i> dengan metode	Pada penelitian ini didapatkan <i>cluster 3</i> merupakan hasil terbaik. Hasil <i>cluster</i> divisualisasikan dengan <i>dashboard</i> dengan beberapa grafik yang memuat hasil segmentasi pelanggan berdasarkan nilai <i>RFM</i> dari

Tabel 2.5 Rangkuman Penelitian Terkait (Lanjutan 1)

No	Judul	Penulis	Tahun	Metodologi	Hasil Penelitian
	Arthamas Citra Mandiri Menggunakan Metode <i>Fuzzy C-Means Clustering</i>	Setiawan, Fitra A. Bachtiar		<i>Elbow</i> digunakan untuk menentukan jumlah <i>cluster</i> serta metode validasi yang digunakan adalah <i>Partition Coefficient</i> dan <i>Euclidean Distance</i> .	transaksi yang dilakukan pelanggan PT. Arthamas Citra Mandiri. Visualisasi <i>dashboard</i> diberikan kepada perusahaan dan dilakukan <i>usability testing</i> untuk mengetahui efektivitas dari visualisasi <i>dashboard</i> yang dibuat. Hasil dari <i>usability testing</i> yaitu 77,5. Hal ini menunjukkan bahwa <i>dashboard</i> masuk ke dalam kategori <i>acceptable</i> atau bisa diterima oleh perusahaan.
3.	Prediksi <i>Customer Churn</i> Berbasis <i>Adaptive Neuro Fuzzy Inference System</i>	Yayak Kartika Sari, Kusri, Ferry Wahyu Wibowo	2018	Menggunakan teknik <i>data mining</i> dengan algoritma ANFIS (Gabungan antara jaringan syaraf tiruan dengan <i>fuzzy inference system</i>).	Berdasarkan penelitian yang telah dilakukan, dengan menggunakan metode ANFIS dapat membantu perusahaan XYZ untuk mengetahui prediksi pelanggan yang melakukan <i>churn</i> . Dengan mengetahui prediksi pelanggan yang melakukan <i>churn</i> , maka akan memberikan keuntungan untuk manajemen, khususnya divisi <i>Customer Relationship Management (CRM)</i> , dapat menciptakan promosi yang bertujuan untuk meningkatkan loyalitas pelanggan dan meningkatkan strategi penjualan untuk mendapatkan pelanggan. Pelatihan dan pengujian ANFIS untuk semua

Tabel 2.5 Rangkuman Penelitian Terkait (Lanjutan 2)

No	Judul	Penulis	Tahun	Metodologi	Hasil Penelitian
					produk menghasilkan perbandingan nilai <i>error</i> rata-rata pelatihan sebesar 8,316 %.
4.	Prediksi <i>Customer Churn</i> dengan Algoritma <i>Decision Tree C4.5</i> Berdasarkan Segmentasi Pelanggan untuk Mempertahankan Pelanggan pada Perusahaan Retail	Ni Wayan Wardani, Gede Rasben Dantes, Gede Indrawan	2018	Segmentasi Pelanggan : Model <i>RFM (Recency, Frequency, Monetary)</i> , Model Prediksi : Algoritma <i>Decision Tree C4.5</i> .	Penerapan model prediksi algoritma <i>Decision Tree C4.5</i> memperoleh hasil kinerja, yaitu: kelas <i>Dormant</i> : <i>Recall</i> 97.51%, <i>Precision</i> 75.18%, <i>Accuracy</i> 76.18%. Kelas <i>Everyday</i> : <i>Recall</i> 100%, <i>Precision</i> 99.04%, <i>Accuracy</i> 99.04%. Kelas <i>Golden</i> : <i>Recall</i> 100%, <i>Precision</i> 98.84%, <i>Accuracy</i> 98.84%. Kelas <i>Superstar</i> : <i>Recall</i> 96.15%, <i>Precision</i> 99.43%, <i>Accuracy</i> 95.63%. Dari hasil evaluasi dengan <i>confusion matrix</i> dapat disimpulkan bahwa kelas pelanggan <i>dormant</i> adalah kelas pelanggan yang berpotensi <i>churn</i> .
5.	Analisis Segmentasi Pelanggan Menggunakan Kombinasi <i>RFM Model</i> dan Teknik <i>Clustering</i>	Beta Estri Adiana, Indah Soesanti, Adhistya Erna Permanasari	2018	Menggunakan metode <i>CRISP-DM</i> serta algoritma <i>K-Means</i> digunakan untuk pembentukan klaster. Pembentukan klaster didasarkan pada analisa <i>RFM (Recency, Frequency, dan Monetary)</i> . <i>Davies</i>	Hasilnya kelompok pelanggan yang terbentuk ada 3 kelompok dengan kelompok pertama berjumlah 30 pelanggan masuk dalam kategori <i>typical customer</i> , kelompok kedua terdapat 8 pelanggan yang masuk dalam kategori <i>superstar customer</i> dan kelompok ketiga berjumlah 89 pelanggan pada kategori <i>dormant customer</i> . Aplikasi model <i>RFM</i> dan

Tabel 2.5 Rangkuman Penelitian Terkait (Lanjutan 3)

No	Judul	Penulis	Tahun	Metodologi	Hasil Penelitian
				<i>Bouldin Index (DBI)</i> digunakan untuk menemukan jumlah <i>cluster</i> (k) yang optimal.	teknik <i>clustering</i> khususnya algoritma <i>K-Means</i> , dapat membantu mengidentifikasi pelanggan dengan nilai tinggi dan berpotensi memberi keuntungan serta dapat membantu dalam merancang strategi pemasaran yang tepat untuk ketiga <i>cluster</i> yang terbentuk.
6.	Metode <i>Data Mining</i> untuk Prediksi <i>Churn</i> Pelanggan	Yulianti	2018	Mengukur kinerja pengklasifikasi menggunakan <i>k-fold cross validation</i> .	<i>Dataset</i> yang diperoleh diterapkan pada model yang diusulkan menggunakan <i>software RapidMiner Studio 8.0.1 64 bit</i> . Berdasarkan hasil pengukuran menunjukkan bahwa <i>Neural Network (NN)</i> merupakan model terbaik dengan akurasi 90,55%. Sedangkan algoritma <i>Linear Discriminant Analysis (LDA)</i> tidak dapat digunakan untuk memprediksi <i>churn</i> pelanggan karena nilainya gagal mengklasifikasikan karena memiliki akurasi kurang dari 60%.
7.	<i>Data Mining</i> dengan Metode Klasifikasi <i>Naïve Bayes</i> untuk	Eka Miranda dan Julisar	2018	Mengklasifikasikan pelanggan berdasarkan tabel transaksi dengan pendekatan <i>Knowledge Discovery From Data (KDD)</i> dan metode	Penelitian ini mengklasifikasikan pelanggan menjadi dua kelas yaitu kelas pelanggan potensial dan pelanggan tidak potensial dengan menggunakan atribut prediksi klasifikasi terdiri atas Pekerjaan, Jenis Bayar,

Tabel 2.5 Rangkuman Penelitian Terkait (Lanjutan 4)

No	Judul	Penulis	Tahun	Metodologi	Hasil Penelitian
	Mengklasifikasikan Pelanggan			<i>data mining Naïve Bayes Classifier.</i>	Tenor, dan Usia. Hasil dari penelitian menunjukkan bahwa <i>Naïve Bayes Classifier</i> telah dapat mengklasifikasikan pelanggan menjadi dua kelas yaitu kelas pelanggan potensial dan pelanggan tidak potensial dengan nilai akurasi masing-masing sebagai berikut: <i>Sensitivity 97%, Specificity 99,8%, Precision 99,8%, Recall 97%, Accuracy 97%, Error Rate 3%</i> .
8.	Prediksi <i>Churn</i> dan Segmentasi Pelanggan TV Berlangganan (Studi Kasus Transvision Jawa Barat)	Nana Suryana	2017	Memprediksi potensi <i>churn</i> dan segmentasinya pada pelanggan menggunakan metode <i>Decision Tree</i> dan <i>Clustering</i> .	Metode <i>Decision Tree</i> dengan pemodelan menggunakan algoritma C4.5 menghasilkan tingkat akurasi 90,89%. Kemudian pelanggan yang mempunyai potensi <i>churn</i> tinggi dilakukan <i>clustering</i> menggunakan algoritma <i>K-Means</i> dan dari 5956 total pelanggan dapat merekomendasikan 5792 pelanggan yang mendapatkan penawaran program retensi.
9.	Model Prediksi Pola Loyalitas Pelanggan Telekomunikasi Menggunakan <i>Naïve Bayes</i>	Stefanus Santosa dan Roy Yuliantara	2017	Menggunakan algoritma <i>Naïve Bayes</i> berbasis <i>Particle Swarm Optimization</i> .	Dalam penelitian ini dilakukan pengujian model akurasi prediksi dengan menggunakan algoritma <i>Naïve Bayes</i> berbasis <i>Particle Swarm Optimization</i> guna mengetahui loyalitas pelanggan pada perusahaan

Tabel 2.5 Rangkuman Penelitian Terkait (Lanjutan 5)

No	Judul	Penulis	Tahun	Metodologi	Hasil Penelitian
	dengan Optimasi <i>Particle Swarm Optimization</i>				telekomunikasi di Columbia. Berdasarkan pengujian tingkat akurasi dan <i>AUC</i> menggunakan algoritma Naïve Bayes dan pemilihan fitur <i>Particle Swarm Optimization</i> , menunjukkan hasil bahwa akurasi yang dihasilkan sebesar 98,54 %, dan dengan menggunakan kurva <i>ROC</i> menghasilkan nilai <i>AUC</i> sebesar 0,99. Waktu yang digunakan untuk memproses pengujian ini membutuhkan waktu lebih kurang 2 jam 45 menit dengan 3333 <i>record</i> yang digunakan.
10.	Analisis <i>Churn Prediction</i> pada Data Pelanggan PT. Telekomunikasi dengan <i>Logistic Regression</i> dan <i>Underbagging</i>	Tesha Tasmalaila Hanifa,	2017	Metode <i>underbagging</i> untuk <i>imbalance data</i> yang dikombinasikan dengan metode klasifikasi menggunakan <i>Logistic Regression (LR)</i> .	Melakukan penanganan <i>imbalance data</i> dengan <i>underbagging</i> bisa meningkatkan akurasi karena data pada teknik <i>bagging</i> dilakukan <i>sampling</i> dengan atau tanpa pengembalian, sehingga distribusi data dari tiap data <i>bagging</i> berbeda. Hasil pengujian model klasifikasi <i>logistic regression</i> tanpa penanganan <i>imbalance data</i> mencapai nilai <i>f1-measure</i> 20,577%. Hal ini menunjukkan bahwa harus dilakukan adanya penanganan <i>imbalance data</i> . Setelah dilakukan <i>imbalance</i>

Tabel 2.5 Rangkuman Penelitian Terkait (Lanjutan 6)

No	Judul	Penulis	Tahun	Metodologi	Hasil Penelitian
					<i>data</i> maka <i>f1-measure</i> meningkat menjadi 85,531%.
11.	<i>Clustering</i> Loyalitas Pelanggan dengan Metode <i>RFM</i> (<i>Recency, Frequency, Monetary</i>) dan <i>Fuzzy C-Means</i>	Sudriyanto	2017	Mencari pola dengan pemilihan atribut menggunakan metode <i>RFM</i> (<i>Recency, Frequency, Monetary</i>) untuk mengklaster pelanggan yang lebih efisien dan akurat dengan algoritma <i>Fuzzy C-Means</i> untuk memilih pelanggan yang potensial dan loyal.	Dari hasil pengukuran validitas menggunakan <i>Partition Coefficient Index (PCI)</i> dan Xie dan Beni (XBI) 5 <i>cluster</i> dengan pangkat dua dan tiga dengan nilai <i>PCI</i> pangkat dua 0,8156, pangkat tiga 0,5860, untuk nilai XBI pangkat dua 0,0069, pangkat tiga 0,0632 dimana nilai <i>PCI</i> yang mendekati angka satu mempunyai arti kualitas <i>cluster</i> yang didapat semakin baik, sedangkan untuk XBI yang semakin kecil mempunyai arti kualitas hasil pengelompokan yang semakin baik. Dari hasil pengukuran validitas menggunakan <i>PCI</i> dan XBI, nilai akurasi <i>FCM</i> dengan pangkat dua lebih baik dibandingkan dengan nilai akurasi <i>Fuzzy C-Means</i> dengan pangkat tiga.
12.	Analisa <i>Customer Churn</i> pada Perusahaan <i>Internet Service Provider xyz</i> Menggunakan	Ali Reza Yudhistira	2017	Menggunakan metode <i>Neural Network</i> yang akan digabungkan dengan model <i>Backpropagation</i> dengan fungsi <i>sigmoid biner</i> .	Faktor yang memiliki pengaruh yang signifikan pada <i>customer</i> yang melakukan <i>churn</i> adalah hasil survei 1, survei 2, dan survei 3 dengan tingkat akurasi dalam memprediksi sebesar 99.99%. Untuk

Tabel 2.5 Rangkuman Penelitian Terkait (Lanjutan 7)

No	Judul	Penulis	Tahun	Metodologi	Hasil Penelitian
	<i>Backpropagation Neural Network</i>				mengurangi tingkat <i>churn</i> , maka penentuan strategi <i>CRM</i> adalah dengan meningkatkan <i>customer experience</i> dari pelanggan.
13.	Analisis Perbandingan Algoritma C4.5 dan Algoritma <i>Naïve Bayes</i> untuk Prediksi Kelulusan Mahasiswa (Studi Kasus : Prodi Teknik Informatika Universitas Muhammadiyah Jember)	M. Hairul Umam, Victor Wahanggara, Triawan Adi Cahyanto, M.Kom., Lutfi Ali Muharom, S.Si., M.Si.	2017	Membandingkan dua metode yaitu C4.5 dan <i>Naïve Bayes</i> untuk memprediksi kelulusan mahasiswa.	Berdasarkan hasil pengujian dengan mengukur kinerja kedua metode tersebut menggunakan metode pengujian <i>Confusion Matrix</i> dalam tool <i>Weka</i> diketahui bahwa C4.5 memiliki nilai akurasi yang lebih baik yakni sebesar 100%, dengan Akurasi 100%, Presisi 100%, dan <i>Recall</i> 100%, sedangkan <i>Naïve Bayes</i> memiliki nilai Akurasi 89%, Presisi 94%, dan <i>Recall</i> 100%. Kemudian berdasarkan perbandingan kinerja kedua metode tersebut, metode dengan pencapaian nilai akurasi terbaik akan presentasikan untuk pengujian prediksi kelulusan mahasiswa angkatan 2013. Sehingga di dapat hasil prediksi semakin meningkat pada angkatan 2013 dari evaluasi tahun-tahun sebelumnya. Dengan jumlah lulus 144 mahasiswa dengan total mahasiswa angkatan 2013 sebanyak 263 mahasiswa, hasil ini dilihat dibandingkan dengan data historis

Tabel 2.5 Rangkuman Penelitian Terkait (Lanjutan 8)

No	Judul	Penulis	Tahun	Metodologi	Hasil Penelitian
					angkatan 2010, 2011, dan 2012.
14.	Penerapan <i>Resampling</i> dan <i>Bagging</i> Berbasis <i>Naive Bayes</i> untuk Prediksi <i>Churn</i> Pelanggan	Yulianti	2016	Meningkatkan kinerja <i>Naive Bayes</i> dengan menyeimbangkan data latih menggunakan <i>random oversampling</i> dan menggabungkan dengan teknik <i>ensemble</i> menggunakan algoritma <i>Bagging</i> pada model prediksi <i>churn</i> pelanggan.	Model terbaik dari model yang diusulkan dalam mengenali kelas minoritas adalah model <i>RUS+BG+NB</i> dengan akurasi secara keseluruhan menurun sebesar 0,55%, tetapi <i>AUC</i> meningkat sebesar 1,18% bila dibandingkan dengan kinerja model lainnya yang tidak dapat meningkatkan kinerja model <i>Naive Bayes</i> .
15.	Prediksi <i>Customer Churn</i> Menggunakan Algoritma <i>Fuzzy Iterative Ichotomiser 3</i>	Herawati, Meyrina. Imam Mukhlash., dan Inu L. Wibowo	2016	Metode <i>Fuzzy Decision Tree</i> Menggunakan Algoritma <i>Iterative Ichotomiser 3 (ID3)</i> .	Membangun perangkat lunak untuk prediksi <i>customer churn</i> . Untuk proses utama yang dilakukan, yaitu proses pembentukan aturan dan <i>tree</i> dengan menggunakan metode <i>Fuzzy ID3</i> terhadap <i>training data</i> dan proses pengujian dengan menggunakan metode <i>Fuzzy</i> menentukan kelas klasifikasi. Dari berbagai uji coba didapat nilai akurasi terbaik 87.

Tabel 2.5 Rangkuman Penelitian Terkait (Lanjutan 9)

No	Judul	Penulis	Tahun	Metodologi	Hasil Penelitian
16.	Komparasi Akurasi Algoritma C4.5 dan <i>Naïve Bayes</i> untuk Prediksi Pendonor Darah Potensial dengan Dataset <i>RFMTC</i>	Wahyu Eko Susanto, Candra Agustina	2016	Melakukan komparasi algoritma C4.5 dengan algoritma <i>Naïve Bayes</i> yang diterapkan pada data diri pendonor dan data transaksi donor darah dengan dataset <i>RFMTC</i> .	Dari hasil pengujian dengan mengukur kinerja kedua algoritma ketika dilakukan Curva <i>ROC</i> , diketahui bahwa algoritma C4.5 secara keseluruhan nilai <i>AUC</i> memiliki klasifikasi <i>ROC</i> yang lebih baik dari algoritma <i>Naïve Bayes</i> sebesar 0.945 dengan klasifikasi <i>Excellent Classification</i> . Sehingga algoritma C4.5 dalam penelitian ini lebih baik daripada algoritma untuk diterapkan pada prediksi pendonor darah potensial.
17.	Klasifikasi Pelanggan pada <i>Customer Churn Prediction</i> Menggunakan <i>Decision Tree</i>	Nurzahputra, Aldi. Afifah Ratna Safitri., dan Much Aziz Muslim	2016	Algoritma yang digunakan untuk membangun <i>Decision Tree</i> adalah algoritma C4.5.	Berdasarkan hasil dan pembahasan, dengan menggunakan metode <i>Decision Tree</i> dan algoritma C4.5 dapat diprediksi klasifikasi pelanggan ke dalam pelanggan loyal atau pelanggan <i>churn</i> . Dalam mengklasifikasi, klasifikasi loyal jika <i>confidence</i> loyal lebih besar daripada <i>confidence churn</i> sedangkan klasifikasi <i>churn</i> jika <i>confidence churn</i> lebih besar daripada <i>confidence</i> loyal. Dari hasil klasifikasi, menunjukkan 61 pelanggan loyal dan 35 pelanggan <i>churn</i> , kemudian dapat diakumulasikan total pelanggan loyal

Tabel 2.5 Rangkuman Penelitian Terkait (Lanjutan 10)

No	Judul	Penulis	Tahun	Metodologi	Hasil Penelitian
					adalah 636 pelanggan dan pelanggan <i>churn</i> 357 pelanggan.
18.	Prediksi <i>Churn</i> dan Segmentasi Pelanggan Menggunakan <i>Backpropagation Neural Network</i> Berbasis <i>Evolution Strategies</i>	Junta Zeniarja, Ardytha Luthfiarta	2015	Teknik <i>data mining Backpropagation Neural Network (BPNN) in hybrid</i> dengan <i>Strategi Evolution (ES)</i> untuk atribut bobot. Validasi model dilakukan dengan menggunakan validasi Palang <i>10-Fold</i> dan evaluasi pengukuran dilakukan dengan menggunakan matriks kebingungan dan Area bawah <i>ROC Curve (AUC)</i> .	Model yang diusulkan dari <i>ES-BPNN</i> menghasilkan akurasi yang lebih baik dengan 96,30% dibandingkan dengan <i>Basic BPNN</i> dengan 95,32%. Tapi waktu komputasi untuk <i>ES-BPNN</i> sangat panjang yaitu 5 jam 24 menit 6 detik, dibandingkan dengan <i>Basic BPNN</i> yang hanya membutuhkan waktu 3 menit 52 detik. Meskipun <i>ES-BPNN</i> mencapai hasil yang lebih baik, tetapi waktu eksekusinya lebih lama. Kita bisa menggunakan <i>ES-BPNN</i> untuk memprediksi klasifikasi secara <i>offline</i> yang hasilnya membutuhkan akurasi tinggi.
19.	<i>IG-KNN</i> untuk Prediksi <i>Customer Churn</i> Telekomunikasi	Arifin, Muhammad	2015	Menggunakan <i>Information Gain</i> untuk pemilihan fitur pada tahap <i>preprocessing</i> dan menerapkan algoritma <i>K-NN</i> sebagai proses pelatihan untuk	Berdasarkan hasil eksperimen, mulai tahap awal hingga evaluasi, dapat ditarik kesimpulan bahwa model prediksi <i>customer churn</i> menggunakan metode <i>K-NN</i> dengan pengurangan fitur menggunakan <i>information gain</i> cukup akurat dibandingkan dengan tanpa menggunakan fitur seleksi dimana dengan

Tabel 2.5 Rangkuman Penelitian Terkait (Lanjutan 11)

No	Judul	Penulis	Tahun	Metodologi	Hasil Penelitian
				memprediksi <i>customer churn</i> telekomunikasi.	nilai k yang berbeda-beda algoritma <i>IG-KNN</i> tetap menunjukkan hasil yang paling baik. Peningkatan akurasi dari k1 sampai dengan k11 sebesar 1,7%.

Perbedaan penelitian ini dengan penelitian sebelumnya terletak pada metode, segmentasi kelas pelanggan, serta proses persiapan data yang digunakan. Metode yang digunakan penelitian sebelumnya menggunakan berbagai metode klasifikasi dan prediksi sedangkan penelitian ini menggunakan metode *Naïve Bayes*. Penelitian yang membahas tentang *churn* sebelumnya hanya mengklasifikasikan pelanggan ke dalam kelas *churn* dan loyal saja, dan dilakukan prediksi tetapi tidak ada nya bentuk prediksi sesungguhnya sedangkan penelitian ini mengklasifikasikan setiap pelanggan berdasarkan kelas pelanggan yang sudah di segmentasikan menggunakan metode *RFM* lalu dilakukan proses klasifikasi menggunakan algoritma *Naïve Bayes*. Tahap persiapan pada penelitian ini tidak hanya proses *cleaning* dan *selection* data saja melainkan tahap transformasi yang dilakukan dalam 3 tahap dalam melakukan segmentasi.