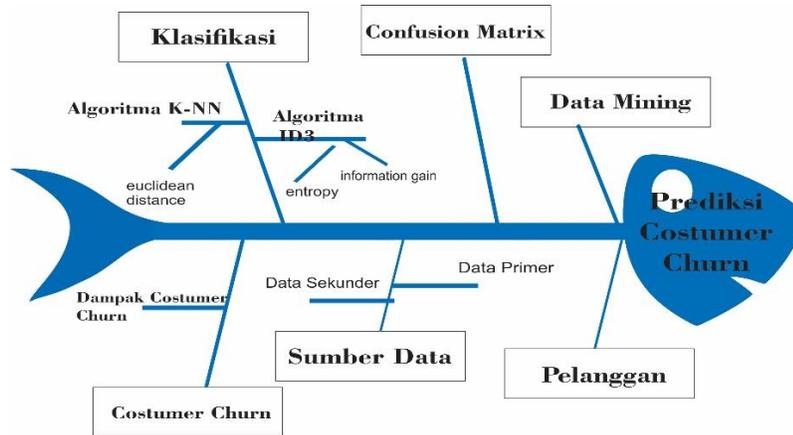


## BAB II

### LANDASAN TEORI



Gambar 2.1 Diagram *Fishbone* Penelitian

#### 2.1 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Turban, dkk. 2005). Definisi umum dari data mining itu sendiri adalah proses pencarian pola-pola yang tersembunyi (*hidden patern*) berupa pengetahuan (*knowledge*) yang tidak diketahui sebelumnya dari suatu sekumpulan data yang mana data tersebut dapat berada di dalam *database*, *data warehouse*, atau media penyimpanan informasi yang lain.

Berdasarkan definisi-definisi yang telah disampaikan, hal penting yang terkait dengan data mining adalah :

1. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan data mining adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

## **2.2 Klasifikasi**

Klasifikasi dapat didefinisikan sebagai proses untuk menyatakan suatu objek data sebagai salah satu kategori (kelas) yang telah didefinisikan sebelumnya (Zaki et al, 2013).

### **2.2.1 Algoritma *Decision Tree* (ID3)**

*Decision Tree* merupakan salah satu metode klasifikasi yang populer dan banyak digunakan secara praktis. Salah satu metode *Decision Tree* yang sangat populer adalah *Iterative Dichotomizer version 3 (ID3)*. Metode ID3 merupakan model klasifikasi yang berupa *decision tree* (pohon keputusan) secara top-down (dari atas ke bawah). Caranya dengan mengevaluasi semua atribut menggunakan suatu ukuran *statistic*, biasanya berupa *information gain*, untuk mengukur efektifitas suatu atribut dalam mengklasifikasikan himpunan sampel

data (Suyanto, 2017:134). Algoritma ini merupakan *decision tree learning* yang melakukan pencarian secara rakus (*greedy*) sehingga belum tentu optimal. Pada (Mitchell 1997), algoritma ID3 diimplementasikan menggunakan sebuah fungsi rekursif (memanggil dirinya sendiri).

**a. Entropy**

*Entropy* merupakan keberbedaan atau keberagaman. Dalam data mining, *entropy* didefinisikan sebagai suatu parameter untuk mengukur heterogenitas (keberagaman) dalam suatu himpunan data. Semakin heterogen suatu himpunan data, semakin besar pula nilai *entropi*-nya. Semakin matematis, *entropy* dirumuskan sebagai berikut (Suyanto, 2017:134):

$$Entropy (S) = \sum_i^c - P_i \log_2 P_i \dots \dots \dots (2. 1)$$

**b. Information Gain**

*Information Gain* merupakan perolehan informasi. *Information Gain* di definisikan sebagai ukuran efektivitas suatu atribut dalam mengklasifikasi data. Secara matematis, *informatin gain* dari atribut A, dituliskan sebagai berikut (Suyanto, 2017:136):

$$Gain (S, A) \equiv Entropy (S) - \sum_{A=1}^c \frac{|S_i|}{|S|} \times Entropy (S_i) \dots \dots \dots (2. 2)$$

Dimana:

$C$  = Jumlah nilai yang terdapat pada atribut target (jumlah kelas)

$P_i$  = Rasio antar jumlah sampel di kelas  $i$  dengan jumlah sampel pada himpunan data

$A$  = Atribut

$C$  = Nilai yang mungkin untuk atribut  $A$

$|S_i|$  = Jumlah sampel untuk nilai  $i$

$|S|$  = Jumlah seluruh sampel data

$Entropy(S_i)$  = *Entropy* untuk sampel- sampel yang memiliki nilai  $i$

### c. Kekuatan dan Kelemahan *Decision Tree*

Kekuatan *Decision Tree*, yaitu (Suyanto, 2017:159):

1. Mempunyai konsep yang jelas dan mudah dipahami. Pohon keputusan yang dihasilkan dari proses pelatihan dapat menjelaskan bagaimana model klasifikasi data bekerja.
2. Mudah diimplementasikan dengan menggunakan algoritma rekursif.

Kelemahan *Decision Tree* (Suyanto, 2017:159):

1. Sulit diimplementasikan untuk himpunan data yang sangat besar dengan ribuan atribut dan milyaran objek data.

2. Mudah mengalami *overfit* karena proses pelatihan *greedy* tidak menjamin dihasilkannya pohon keputusan yang optimum.

### 2.2.2 Algoritma K-NN (*K-Nearest Neighbour*)

*K-Nearest Neighbor* (KNN) adalah metode melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Metode ini bertujuan untuk mengklasifikasikan objek baru berdasarkan atribut dan *training sample*. Diberikan suatu titik *query*, selanjutnya akan ditemukan sejumlah K objek atau titik *training* yang paling dekat dengan titik *query*.

Sebelum melakukan perhitungan dengan metode *K-Nearest Neighbor*, terlebih dahulu harus menentukan data latih dan data uji. Kemudian akan dilakukan proses perhitungan untuk mencari jarak menggunakan *Euclidean*. Setelah itu, akan dilakukan tahapan perhitungan dengan metode KNN seperti pada Gambar 2.1.

Algoritma *K-Nearest Neighbor* (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Teknik ini sangat sederhana dan mudah diimplementasikan. Mirip dengan teknik *clustering*, yaitu mengelompokkan suatu data baru berdasarkan jarak data baru itu ke beberapa data/ tetangga terdekat. Pertama sebelum mencari jarak data ke tetangga adalah menentukan nilai K tetangga (*neighbor*). Lalu, untuk mendefinisikan jarak antara dua titik yaitu titik pada data training dan titik

pada data *testing*, maka digunakan rumus *Euclidean* dengan persamaan 2.3, sebagai berikut:

$$\text{dist}(x, y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \dots\dots\dots(2.3)$$

Keterangan :

Dist (x,y) = jarak *Euclidean*

X = data 1

Y = data 2

i = fitur ke

n = jumlah fitur

**a. Kekuatan dan kelemahan K-NN**

“Algoritma *k-nearest neighbor* (k-NN atau K-NN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut” (Suyanto, 2017:211).

Nilai *k* yang terbaik untuk algoritme ini tergantung pada data; secara umumnya, nilai *k* yang tinggi akan mengurangi efek *noise* pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi lebih kabur. Nilai *k* yang bagus dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan *cross-validation*. Kasus khusus di mana klasifikasi diprediksikan berdasarkan data pembelajaran yang paling dekat (dengan kata lain, *k* = 1) disebut algoritme *nearest neighbor* (Lestari 2014).

Kelebihan K-NN (*K-Nearest Neighbor*):

1. Mempunyai konsep yang jelas dan mudah dipahami.
2. Tangguh terhadap training data yang memiliki banyak noise.

Kelemahan K-NN (*K-Nearest Neighbor*):

1. K-NN perlu menentukan nilai dari parameter k (jumlah dari tetangga terdekat).
2. Training berdasarkan jarak tidak jelas mengenai jenis jarak apa yang harus digunakan.
3. Atribut mana yang harus digunakan untuk mendapatkan hasil terbaik.
4. Biaya komputasi cukup tinggi karena diperlukan perhitungan jarak dari tiap query instance pada keseluruhan training sample.

### **2.3 Costumer Churn**

*Costumer Churn* adalah pelanggan yang telah memutuskan untuk meninggalkan penyedia layanan, produk, atau bahkan perusahaan dan bergeser ke pesaing lain. Perilaku *costumer churn* memiliki dampak tertentu pada kinerja perusahaan, yaitu sebagai berikut: (Amin et al, 2017)

1. Dampak negatif pada kinerja keseluruhan perusahaan.
2. Penyebab potensial untuk penjualan rendah karena pelanggan baru / jangka pendek pembeli lebih sedikit.
3. Membantu pesaing untuk mendapatkan pelanggan yang tidak puas dengan promosi bisnis perusahaan.
4. Menyebabkan kerugian pendapatan.

5. Memberikan dampak negatif pada pelanggan jangka panjang.
6. Meningkatkan ketidakpastian yang mengurangi rasio kemungkinan pelanggan baru.
7. Menarik pelanggan baru lebih mahal daripada mempertahankan yang sudah ada.
8. Risiko terhadap citra perusahaan di pasar yang kompetitif dengan hilangnya basis pelanggan.

Dengan kata lain, *Customer Churn* merupakan perpindahan pelanggan dari perusahaan satu ke perusahaan lain, yang akan mempengaruhi perusahaan dalam mendapatkan keuntungan penjualan.

#### **2.4 Pelanggan (*Customer*)**

Pelanggan adalah seorang individu atau kelompok yang membeli produk fisik atau jasa dengan mempertimbangkan berbagai macam faktor seperti harga, kualitas, tempat, pelayanan dan lain sebagainya berdasarkan keputusan mereka sendiri (Greenberg, 2010). Pelanggan adalah masyarakat yang pada umumnya membutuhkan barang dan jasa yang berpotensi melakukan pembelian (Yoeti, 2006).

Dengan kata lain, bisa diartikan bahwa pelanggan adalah instansi, lembaga, organisasi ataupun orang yang membeli produk atau jasa perusahaan secara rutin atau berkesinambungan karena produk atau jasa yang dibelinya merasa sangat

bermanfaat. Dapat dikatakan juga pelanggan yaitu orang-orang yang memberikan perhatian penuh terhadap produk atau jasa yang dihasilkan oleh perusahaan.

## **2.5 Jenis dan Sumber Data**

### **2.5.1 Data Primer**

Data primer merupakan data yang dikumpulkan sendiri oleh perorangan atau suatu organisasi secara langsung dari objek yang diteliti dan untuk kepentingan studi yang bersangkutan yang dapat berupa wawancara langsung dengan pimpinan atau bagian yang menangani langsung permasalahan di lapangan atau dapat berupa observasi kegiatan sehari-hari suatu objek yang diteliti. Hasilnya dapat berbentuk data yang dibutuhkan atau data yang diharapkan sesuai dengan penelitian yang dilakukan, seperti data transaksi pelanggan yang ada di Radity Craft dalam periode tertentu yang nantinya akan digunakan sebagai bahan untuk analisis prediksi *customer churn*.

### **2.5.2 Data Sekunder**

Data sekunder adalah data yang diperoleh atau dikumpulkan dan disatukan oleh studi-studi sebelumnya atau yang diterbitkan oleh berbagai instansi lain. Biasanya sumber tidak langsung berupa data dokumentasi dan arsip-arsip resmi. Data sekunder didapatkan dari pustaka-pustaka yang relevan dengan penelitian yang dilakukan, yaitu seperti literature dari jurnal “Prediksi *Customer Churn* Dengan Algoritma *Decision Tree C4.5* Berdasarkan Segmentasi Pelanggan Pada Perusahaan Retail” oleh Ni Wayan Wardani, Gede

Rasben Dantes, Gede Indrawan dan literature lain seperti jurnal "Segmentasi Pelanggan Menggunakan Algoritma K-Means Untuk *Customer Relationship Management* (CRM) Pada Hijab Miulan " oleh Gita Febrina Wulandari.

## 2.6 Confusion Matrix

*Confusion matrix* adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Rumus ini melakukan perhitungan dengan 4 keluaran, yaitu: *recall*, *precision*, *accuracy* dan *error rate*. Namun yang digunakan dalam penelitian ini hanya menghitung tingkat *accuracy* saja. Adapun perhitungan tingkat akurasi pada confusion matriks (Hanjer dan Kammer M):

$$Accuracy = \frac{\text{jumlah nilai benar}}{\text{jumlah data keseluruhan}} \times 100 \% \dots\dots\dots(2.4)$$

## 2.7 Literatur Review

Pada penyusunan tugas akhir ini, ada beberapa referensi dan jurnal penelitian terkait yang dilakukan sebelumnya. Berikut ini ulasan singkat mengenai referensi dan jurnal penelitian sebelumnya pada tabel 2.1

**Tabel 2.1** *State Of The Art* (Lanjutan)

<b>No.</b>	<b>Nama Pengarang</b>	<b>Tahun</b>	<b>Judul</b>	<b>Isi Ringkasan</b>	<b>Hasil</b>
1.	Sri Mulyati, Yulianti, dan Aries Saifudin	2018	Penerapan Resampling Dan Adaboost Untuk Penanganan Masalah Ketidakseimbangan Kelas Berbasis Naïve Bayes Pada Prediksi Churn Pelanggan	Kemudahan pelanggan untuk berpindah ke pesaing merupakan perhatian utama bagi bagian CRM, karena untuk mendapatkan pelanggan baru membutuhkan biaya yang jauh lebih mahal daripada mempertahankan pelanggan yang sudah ada. Untuk mengambil tindakan yang tepat dalam mempertahankan pelanggan harus mengetahui kecenderungan pelanggan apakah akan mengalami churn atau tidak. Prediksi dilakukan dengan menggunakan model data mining, diterapkan teknik resampling dan teknik ensemble AdaBoost untuk memperbaiki kinerja pengklasifikasi sedangkan untuk mengukur kinerja model digunakan software RapidMiner	Hasil penelitian menunjukkan bahwa model integrasi random oversampling, AdaBoost, dan Naïve Bayes memiliki kinerja yang lebih baik karena memiliki nilai AUC (Area Under the ROC (Receiver OperatingCharacteristic) Curve) yang lebih baik. Kata

**Tabel 2.1** *State Of The Art* (Lanjutan)

No.	Nama Pengarang	Tahun	Judul	Isi Ringkasan	Hasil
2.	S Arifin, dan F Samopa.	2018	Analysis Of Churn Rate Significantly Factors In Telecommunication Industry Using Support Vector Machines Method Analysis Of Churn Rate Significantly Factors In Telecommunication Industry Using Support Vector Machines Method	Mengetahui faktor-faktor yang mempengaruhi tingkat churn secara signifikan di perusahaan telekomunikasi melalui penelitian data tagihan dan profil pelanggan.	Hasil penelitian menunjukkan bahwa ada tiga variabel yang mempengaruhi tingkat churn secara signifikan yaitu penggunaan suara, penggunaan data dan isi ulang dengan persentase kinerja kurang dari 5% dari total kinerja.
3.	Ni Wayan Wardani, Gede Rasben Dantes, Gede Indrawan	2018	Prediksi Customer Churn Dengan Algoritma Decision Tree C4.5 Berdasarkan Segmentasi Pelanggan Pada Perusahaan Retail	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	Divisi CRM UD. Mawar Sari perlu melakukan program mempertahankan pelanggan (customer retention).

**Tabel 2.1** *State Of The Art* (Lanjutan)

No.	Nama Pengarang	Tahun	Judul	Isi Ringkasan	Hasil
				karakteristik semua pelanggan dan mampu mempertahankan pelanggan yang telah ada agar tidak menghentikan pembelian dan berpindah ke perusahaan retail pesaing (churn).	
4.	Yayak Kartika Sari , Kusrini , Ferry Wahyu Wibowo	2018	Prediksi Customer Churn Berbasis Adaptive Neuro Fuzzy Inference System	Customer Churn adalah pelanggan yang berhenti berlangganan dan pindah pada perusahaan lain, karena berbagai faktor. Customer churn merupakan masalah yang sangat penting yang harus dihadapi oleh perusahaan karena berhentinya pelanggan akan berdampak pada retensi perusahaan. Oleh sebab itu, dibuatkan sistem prediksi customer churn untuk mengetahui tingkat pelanggan yang churn, apabila customer churn dapat diketahui terlebih dahulu, maka akan menguntungkan bagi pihak CRM untuk mengatur strategi-strategi	Berdasarkan penelitian yang telah dilakukan, dengan menggunakan metode ANFIS dapat membantu perusahaan XYZ untuk mengetahui prediksi pelanggan yang melakukan churn. Dengan mengetahui prediksi pelanggan yang melakukan churn, maka akan memberikan keuntungan untuk manajemen, khususnya divisi Customer Relationship Management (CRM), dapat menciptakan promosi yang bertujuan untuk meningkatkan loyalitas pelanggan dan

**Tabel 2.1** *State Of The Art* (Lanjutan)

No.	Nama Pengarang	Tahun	Judul	Isi Ringkasan	Hasil
				mencegah pelanggan yang melakukan churn.	meningkatkan strategi penjualan untuk mendapatkan pelanggan baru dan melakukan retensi pelanggan.
5.	Tesha Tasmalaila Hanifa, Adiwijaya, Said Al-Faraby	2017	Analisis Churn Prediction Pada Data Pelanggan PT. Telekomunikasi Dengan Logistic Regression Dan Underbagging	Kebutuhan akan informasi dan komunikasi bertambah. Persaingan untuk mendapatkan pendapatan antar perusahaan telekomunikasi menimbulkan adanya Churn. Churn adalah pindahnya pelanggan dari satu provider ke provider lainnya. Perusahaan lebih memilih untuk mempertahankan customer, karena dibutuhkan biaya yang lebih sedikit daripada menambah customer yang baru. Namun dalam permasalahan ini churn memiliki data yang tidak seimbang dan ekstrim dibanding dengan non-churn, sehingga perlu adanya penanganan pada distribusi kelas mayoritas (non-churn) dan minoritas (churn).	Hasil pengujian model klasifikasi logistic regression tanpa penanganan imbalance data mencapai nilai f1- measure 20,577% Hal ini menunjukkan bahwa harus dilakukan adanya penanganan imbalance data. Setelah dilakukan imbalance data maka f1-measure meningkat menjadi 85,531%. Penggunaan estimasi koefisien nilai atau isi dari setiap atribut (variabel predictor) ini adalah yang memberikan pengaruh pada hasil performansi klasifikasi untuk memprediksi churn menggunakan logistic regression.

**Tabel 2.1** *State Of The Art* (Lanjutan)

No.	Nama Pengarang	Tahun	Judul	Isi Ringkasan	Hasil
6.	Muhammad Faruq Mujaddid, Adiwijaya, Said Al-Faraby	2017	Analisis Churn Prediction Menggunakan Metode Logistic Regression Dan Smote (Synthetic Minority Over-Sampling Technique) Pada Perusahaan Telekomunikasi Churn	Persaingan antara perusahaan telekomunikasi seluler pada masa ini adalah dengan mempertahankan pelanggan. Pelanggan menjadi salah satu faktor utama dalam kesuksesan perusahaan telekomunikasi seluler. Pelanggan dapat memilih sesuai dengan keinginan dan kebutuhan, hal ini menjadi faktor utama pemicu terjadinya churn. Churn prediction adalah metode yang digunakan untuk memprediksi pelanggan yang kemungkinan churn dan pelanggan yang tetap bertahan pada suatu perusahaan tertentu	Penelitian yang dilakukan menggunakan metode logistic regression dan penanganan imbalance data dengan SMOTE memiliki hasil performansi dengan tingkat akurasi sebesar 92,4% dan f1-measure sebesar 31,27%
7.	Sudriyanto	2017	Clustering Loyalitas Pelanggan Dengan Metode Rfm (Recency, Frequency, Monetary ) Dan Fuzzy C-Means	Semakin maju dan berkembangnya dunia usaha, menciptakan persaingan yang luar biasa ketata. Persaingan yang ada mengharuskan pemilik usaha untuk selalu dapat memahami sebuah pasar yang terus	Dengan menggunakan metode RFM Recency, Frequency dan Monetary untuk pemilihan atribut dan Algoritma Fuzzy C-Means untuk proses pengklasteran lebih objektif

**Tabel 2.1** *State Of The Art* (Lanjutan)

No.	Nama Pengarang	Tahun	Judul	Isi Ringkasan	Hasil
				berkembang. Banyak faktor yang mempengaruhi berkembangnya usaha, salah satunya pelanggan. Untuk membangun strategi yang handal dalam dunia usaha, harus menemukan cara untuk menarik dan mengembangkan aset manusia yang tepat dan menjaga mereka	dalam meneliti loyalitas pelanggan serta mempermudah pemilik usaha sebagai dasar pengambilan keputusan dalam pemberian reward kepada pelanggan agar lebih tepat sasaran.
8.	Amri Danades, Devie Pratama, Dian Anggraini, Diny Anggriani	2017	Comparison Of Accuracy Level K-Nearest Neighbor Algorithm And Support Vector Machine Algorithm In Classification Water Quality Status	Air diklasifikasikan menjadi empat status kualitas air, yang kondisinya bagus, tercemar ringan, tercemar sedang, dan sangat tercemar. Status klasifikasi kualitas air sangat penting untuk mengetahui penggunaan dan penanganan yang tepat. Keakuratan dalam klasifikasi status kualitas sangat penting, sehingga kedua algoritma klasifikasi K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) digunakan.	Penelitian ini membahas algoritma perbandingan KNN dan SVM dalam klasifikasi status kualitas air, perbandingan dilakukan untuk mengetahui nilai bahwa algoritma tersebut memiliki akurasi tertinggi dalam penentuan status klasifikasi kualitas air, pengujian algoritma KNN dan SVM menggunakan 10 kali lipat Validasi Silang.

**Tabel 2.1** *State Of The Art* (Lanjutan)

No.	Nama Pengarang	Tahun	Judul	Isi Ringkasan	Hasil
9.	Adnan Amina, Sajid Anwara, Awais Adnana, Muhammad Nawaz, Khalid Alawfi, Amir Hussain, Kaizhu Huang	2017	Customer Churn Prediction In Telecommunication Sector Using Rough Set Approach	Customer Churn adalah masalah kritis dan menantang yang mempengaruhi bisnis dan industri, khususnya, sektor telekomunikasi yang berkembang pesat dan sangat kompetitif. Sangat menarik bagi para peneliti akademis dan praktisi industri, tertarik untuk memperkirakan perilaku pelanggan untuk membedakan churn dari pelanggan non-churn. Motivasi utama adalah kebutuhan yang sangat besar dari bisnis untuk mempertahankan pelanggan yang sudah ada, ditambah dengan biaya tinggi yang terkait dengan memperoleh pelanggan baru.	analisis tingkat atribut dapat membuka jalan untuk mengembangkan kebijakan retensi pelanggan yang sukses yang dapat membentuk bagian tak terpisahkan dari pengambilan keputusan strategis dan proses perencanaan di sektor telekomunikasi.
10.	Sri Mulyati, Yulianti, dan Aries Saifudin	2017	Penerapan Resampling Dan Adaboost Untuk Penanganan Masalah Ketidakseimbangan Kelas Berbasis Naïve	Banyaknya operator seluler mendorong persaingan usaha yang sangat ketat. Kemudahan pelanggan untuk berpindah ke pesaing merupakan perhatian utama bagi bagian CRM (Customer Relationship	Pada penelitian ini akan diterapkan teknik resampling dan teknik ensemble AdaBoost untuk memperbaiki kinerja pengklasifikasi sedangkan untuk mengukur kinerja model

**Tabel 2.1** *State Of The Art* (Lanjutan)

No.	Nama Pengarang	Tahun	Judul	Isi Ringkasan	Hasil
			Bayes Pada Prediksi Churn Pelanggan	Management), karena untuk mendapatkan pelanggan baru membutuhkan biaya yang jauh lebih mahal daripada mempertahankan pelanggan yang sudah ada.	digunakan software RapidMiner. Hasil penelitian menunjukkan bahwa model integrasi random oversampling, AdaBoost, dan Naïve Bayes memiliki kinerja yang lebih baik karena memiliki nilai AUC (Area Under the ROC (Receiver Operating Characteristic) Curve) yang lebih baik.
1	Devi Yunita	2017	Perbandingan Algoritma K-Nearest Neighbor Dan Decision Tree Untuk Penentuan Risiko Kredit Kepemilikan Mobil	Perbandingan Algoritma K-Nearest Neighbor Dan Decision Tree untuk Risiko Kredit Kepemilikan Mobil adalah sarana agar orang atau perusahaan dapat meminjam modal atau uang dan membayarnya dalam tempo yang sudah ditentukan. Agar kredit yang diberikan sesuai tujuan atau sasaran, yaitu aman, maka perlu dilakukan analisis kredit	Hasil penelitian dengan menggunakan aplikasi Rapid Miner menunjukan bahwa Algoritma K-Nearest Neighbor (K-NN) memiliki nilai akurasi yang lebih baik
11	Meyrina Herawati, Imam	2016	Prediksi Customer Churn Menggunakan	Setiap industri telekomunikasi sering menawarkan layanan yang menggi-	Penelitian ini bertujuan untuk melakukan pengem- bangan

**Tabel 2.1** *State Of The Art* (Lanjutan)

No.	Nama Pengarang	Tahun	Judul	Isi Ringkasan	Hasil
	Mukhlash, Inu L. Wibowo		Algoritma Fuzzy Iterative Dichotomiser 3	urkan. Hal ini bertujuan untuk menarik pelanggan sebanyak mungkin agar pendapatan yang masuk semakin optimal. Akan tetapi hal-hal seperti ini yang menyebabkan fenomena churn terjadi, dimana pelanggan memutuskan berhenti berlangganan dan berpindah dari satu provider ke provider lain- nya.	model churn prediction (classification) dengan menggunakan fuzzy decision tree. Fuzzy decision tree merupakan salah satu metode klasifikasi berbasis pohon keputusan dengan memasukkan konsep himpunan fuzzy dalam proses pembentukannya. Setelah tahap implementasi, hasil dari penelitian ini berupa model pohon keputusan prediksi customer churn.
11	Andri Wijaya and Abba Suganda Girsang	2016	The Use Of Data Mining For Prediction Of Customer Loyalty	analisis loyalitas pelanggan menggunakan tiga metode penambangan data: C4.5, Naif Bayes, dan Algoritma Neighbor Terdekat dan data empiris dunia nyata. Data berisi sepuluh atribut terkait dengan loyalitas pelanggan dan diperoleh dari a	Hasilnya menyarankan algoritma C4.5 menghasilkan akurasi klasifikasi tertinggi di pesanan 81% diikuti oleh metode Naif Bayes 76% dan Tetangga Terdekat 55%. Selain itu, angka

**Tabel 2.1** *State Of The Art* (Lanjutan)

No.	Nama Pengarang	Tahun	Judul	Isi Ringkasan	Hasil
				perusahaan multimedia nasional di Indonesia.	evaluasi juga menunjukkan bahwa proporsi 80% adalah optimal untuk set pelatihan.
14	Mustakim dan Giantika Oktaviani .	2016	Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa	Predikat prestasi mahasiswa diperoleh dari hasil sebuah prediksi. Proses prediksi dilakukan dengan menggunakan metode K-Nearest Neighbor (KNN). Atribut yang digunakan dalam proses prediksi adalah Jenis Kelamin, Jenis Tinggal, Umur, Jumlah Satuan Kredit Semester (SKS), dan Jumlah Nilai Mutu (NM), sehingga dengan menerapkan algoritma KNN dapat dilakukan sebuah prediksi berdasarkan kedekatan dari histori data lama (training) dengan data baru (testing).	Hasil dari perhitungan algoritma KNN diimplementasikan terhadap sebuah Early Warning System (EWS). Output dari sistem yang dibangun dapat dijadikan sebagai acuan bagi Mahasiswa untuk meningkatkan prestasi dan predikat perkuliahan dimasa yang akan datang.
15	Muhammad Arifin	2015	IG Dan K-NN Untuk Prediksi Customer Churn Telekomunikasi	IG-KNN merupakan gabungan dari algoritma pemilihan fitur information gain dengan algoritma klasifikasi KNN, kedua algoritma ini diharapkan dapat meningkatkan	Berdasarkan hasil penelitian ini prediksi customer churn telekomunikasi dengan menggunakan IG-KNN menunjukkan akurasi yang lebih

**Tabel 2.1** *State Of The Art* (Lanjutan)

No.	Nama Pengarang	Tahun	Judul	Isi Ringkasan	Hasil
				akurasi dalam memprediksi customer churn telekomunikasi. Prediksi customer churn telekomunikasi merupakan kebutuhan yang sangat penting bagi kelangsungan hidup perusahaan telekomunikasi, dimana dengan banyaknya pelanggan yang meninggalkan perusahaan maka perusahaan berpeluang untuk merugi.	baik meski dengan nilai k yang berbeda- beda bila dibandingkan dengan prediksi customer churn telekomunikasi dengan menggunakan KNN tanpa fitur seleksi Information Gain
16	Md.Mahbubur Rahman, Samsuddin Ahmed, Md. Hossain Shuvo	2014	Nearest Neighbor Classifier Method For Making Loan Decision In Commercial Bank	Kegagalan dan keberhasilan sektor perbankan tergantung pada kemampuan untuk melakukan evaluasi risiko kredit yang tepat. Evaluasi risiko kredit dari setiap aplikasi kredit potensial tetap menjadi tantangan bagi bank di seluruh dunia hingga saat ini. Jaringan saraf tiruan memainkan peran yang luar biasa di bidang keuangan untuk membuat keputusan yang kritis, penuh teka-teki, dan sensitif, yang terkadang tidak mungkin bagi manusia. Seperti	Merancang Jaringan Saraf Tiruan (Neural Network) yang dapat memfasilitasi petugas pinjaman untuk membuat keputusan yang tepat untuk memberikan pinjaman kepada klien yang tepat.

**Tabel 2.1** *State Of The Art* (Lanjutan)

No.	Nama Pengarang	Tahun	Judul	Isi Ringkasan	Hasil
				keputusan penting lainnya dalam keuangan, keputusan pemberian sanksi pinjaman kepada pelanggan juga merupakan masalah yang membingungkan.	
17	Özden Gür Ali dan Umur Arıtürk	2014	Dynamic Churn Prediction Framework With More Effective Use Of Rare Event Data: The Case Of Private Banking	Prediksi churn spesifik waktu dapat membantu bisnis untuk mengalokasikan upaya retensi sepanjang waktu, serta pelanggan, dan mengidentifikasi pemicu awal dan indikator churn pelanggan. Kami mengusulkan kerangka prediksi churn dinamis untuk menghasilkan data pelatihan dari catatan pelanggan, dan memanfaatkannya untuk memprediksi churn pelanggan dalam berbagai horizon menggunakan pengklasifikasi standar.	membandingkan analisis survival yang digunakan terutama sebagai alat deskriptif. Metode yang diusulkan mengungguli analisis survival dalam hal akurasi prediksi untuk semua waktu, dengan variabilitas yang jauh lebih rendah.
18	Mei Lestari	2014	Penerapan Algoritma Klasifikasi Nearest Neighbor (K-NN)	Data WHO menyatakan bahwa sebanyak 7,3 juta penduduk dunia meninggal dikarenakan penyakit	Hasil penelitian diperoleh algoritma K-NN dapat digunakan

**Tabel 2.1** *State Of The Art* (Lanjutan)

No.	Nama Pengarang	Tahun	Judul	Isi Ringkasan	Hasil
			Untuk Mendeteksi Penyakit Jantung	jantung. Meskipun penyakit jantung merupakan penyakit yang tidak menular, penyakit ini merupakan jenis penyakit yang mematikan nomor satu di dunia. Penerapan Algoritma Klasifikasi Nearest Neighbor (K-NN) diharapkan dapat mengatasi masalah efektifitas dan akurasi dalam mendeteksi penyakit jantung.	dan diterapkan untuk mendeteksi penyakit jantung.
19	Ricky Imanuel Ndaumanu, Kusri, M. Rudyanto Arief	2014	Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa Dengan Metode K-Nearest Neighbor	Kebutuhan akan analisis mengenai prediksi tingkat pengunduran diri mahasiswa di STIKOM UYELINDO Kupang, menjadi alasan dilakukannya penelitian terhadap prediksi tingkat pengunduran diri mahasiswa. Menentukan prediksi tingkat pengunduran diri mahasiswa dalam jumlah besar tidak mungkin dilakukan secara manual karena membutuhkan waktu yang cukup lama. Untuk itu dibutuhkan sebuah algoritma yang dapat mengkategorisasikan prediksi	Hasil pengujian menggunakan prototype sistem pendukung keputusan dan dibandingkan secara manual yang menggunakan metode Algoritma K-Nearest Neighbor dengan 4 variabel yaitu IPK, Pekerjaan orang tua, jurusan dan semester.

**Tabel 2.1** *State Of The Art* (Lanjutan)

No.	Nama Pengarang	Tahun	Judul	Isi Ringkasan	Hasil
				tingkat pengunduran diri mahasiswa secara otomatis menggunakan komputer.	
20	Gita Febrina Wulandari	2014	Segmentasi Pelanggan Menggunakan Algoritma K-Means Untuk Customer Relationship Management (Crm) Pada Hijab Miulan	Persaingan yang ketat di bidang sejenis mendorong HIJAB MIULAN untuk mengelola pelanggannya secara maksimal. Upaya untuk meningkatkan loyalitas pelanggan adalah dengan mengelompokkan pelanggan dan menerapkan strategi pemasaran yang sesuai untuk setiap kelompok. Untuk melakukan pengembangan perusahaan, maka dibutuhkan strategi pengembangan pelanggan, adapun untuk melakukan strategi pengembangan pelanggan dapat dilakukan dengan pendekatan data mining meliputi pengelompokan pelanggan dengan metode clustering. Metode yang digunakan adalah CRISP-DM dengan melalui tahapan business understanding, data	Upaya untuk meningkatkan loyalitas pelanggan adalah dengan mengelompokkan pelanggan dan menerapkan strategi pemasaran yang sesuai untuk setiap kelompok. Untuk melakukan pengembangan perusahaan, maka dibutuhkan strategi pengembangan pelanggan,

**Tabel 2.1** *State Of The Art* (Lanjutan)

<b>No.</b>	<b>Nama Pengarang</b>	<b>Tahun</b>	<b>Judul</b>	<b>Isi Ringkasan</b>	<b>Hasil</b>
				understanding, data preparation, modeling, evaluasi dan development. Algoritma yang digunakan untuk pembentukan kluster adalah algoritma K-Means, dengan tools Rapidminer 5.3.	

Penelitian yang akan dilakukan memiliki keterkaitan dengan penelitian sebelumnya yaitu mengenai masalah *costumer churn*. Penelitian ini bertujuan untuk mengetahui dan memprediksi pelanggan yang melakukan *churn*. *Costumer churn* merupakan masalah yang harus dihadapi oleh perusahaan karena berhentinya pelanggan akan merugikan perusahaan (Muhammad. 2015).

Seperti pada penelitian sebelumnya yang dilakukan oleh Meyrina, dkk (2016) dalam penelitiannya menggunakan algoritma *Decision Tree* (ID3) untuk memprediksi *costumer churn* didapatkan nilai akurasi sebesar 87%. Kemudian, diperkuat dalam penelitian Muhammad (2014) dalam memprediksi *costumer churn* dengan menggunakan algoritma K-NN menghasilkan tingkat akurasi sebesar 88% pada nilai K 5 keatas.

Penelitian yang dilakukan oleh Wardani, dkk (2018) dalam penelitiannya menggunakan algoritma *Decision Tree* (ID3) untuk memprediksi *costumer churn* berdasarkan segmentasi pelanggan dengan menggunakan model RFM sehingga mendapatkan kelas pelanggan. Kemudian, kelas pelanggan *dormant* yang berpotensi untuk *churn* sebesar 97.51%.

Berdasarkan penelitian yang dilakukan oleh Meyrina, dkk (2016), Muhammad (2014) dan Wardani, dkk(2018) membuktikan bahwa *costumer churn* dengan menggunakan algoritma ID3 dan K-NN memiliki tingkat akurasi yang relevan dan

dapat membantu perusahaan untuk menentukan indikasi pelanggan yang akan melakukan *churn*.

Perbedaan penelitian ini dengan sebelumnya adalah penelitian ini menggunakan dua algoritma ID3 dan K-NN untuk memprediksi *costumer churn* berdasarkan segmentasi pelanggan menggunakan model RFM yang akan membentuk kelas pelanggan dalam perusahaan kerajinan tangan Radity Craft. Kemudian, membandingkan algoritma ID3 dan K-NN untuk mengetahui tingkat akurasi data dengan menggunakan dua algoritma tersebut untuk memprediksi *costumer churn*.