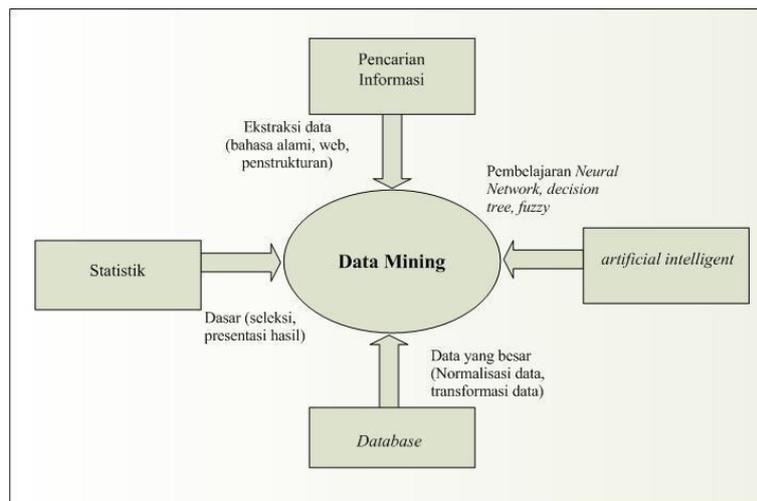


## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Data Mining

Perkembangan teknologi informasi yang pesat ditambah dengan dukungan database, mengakibatkan semakin membanjirnya data yang ada baik dalam perusahaan maupun instansi. Data-data yang sangat banyak tersebut bisa diolah lebih lanjut menjadi suatu informasi yang berguna, misalnya digunakan sebagai dasar pengambilan keputusan dalam perusahaan (Kristanto, 2013). Data mining adalah suatu proses menemukan hubungan yang berarti, pola dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. Data mining adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*). (Widayu, 2017) Proses data mining dapat dilihat pada Gambar 2.1.



Gambar 2.1 Bidang Ilmu Data mining(Widayu, 2017)

## **2.2. *Customer Relationship Management*(CRM)**

Manajemen hubungan pelanggan adalah kegiatan keseluruhan perusahaan untuk memperoleh dan mempertahankan pelanggan. *Customer Relationship Management* (CRM) mengenali pelanggan sebagai inti dari bisnis dan bahwa keberhasilan perusahaan tergantung pada hubungan dengan pelanggan yang dikelola secara efektif. CRM berfokus pada pembangunan hubungan pelanggan jangka panjang dan berkelanjutan, yang menambah nilai bagi pelanggan dan perusahaan. Secara umum CRM adalah pendekatan yang mengenali pelanggan sebagai inti bisnis dan bahwa keberhasilan perusahaan tergantung pada efektivitas pengelolaan hubungan dengan pelanggan. Dengan kata lain CRM adalah strategibisnis untuk memilih dan mengelola pelanggan agar dapat mengoptimalkan nilai jangka panjang. CRM membutuhkan filosofi bisnis dan budaya yang berpusat pada pelanggan untuk mendukung proses pemasaran, penjualan dan pelayanan yang efektif(Imasari, 2011).

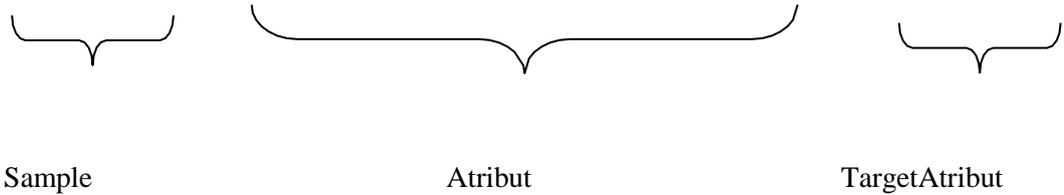
## **2.3. *Pohon Keputusan* (*Decision tree*)**

*Decision tree* adalah sebuahstruktur yang dapat digunakan untuk membagi kumpulan-kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan masing-masing rangkaian pembagi, anggota himpunan hasil menjadi mirip satu dengan yang lainnya. Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan *record*. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon. Misalkan untuk menentukan main tenis, kriteria yang diperhatikan adalah cuaca, angin dan temperatur. Salah satu atribut merupakan atribut yang menyatakan data solusi per item data yang disebut target atribut. Atribut memiliki nilai-nilai yang dinamakan dengan *instance*.

Misalkan atribut cuaca mempunyai *instance* berupa cerah, berawan dan hujan, konsep data dalam *decision tree* dapat dilihat pada Tabel 2.1.

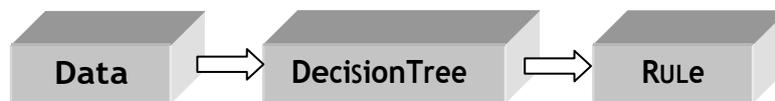
Tabel 2.1 Tabel Proses *Decision tree* (Mardi, 2017)

Nama	Cuaca	Angin	Temperatur	Main
Faris	Cerah	Keras	Panas	Tidak
Totok	Cerah	Lambat	Panas	Iya
Dinda	Berawan	Keras	Sedang	Tidak
Fendik	Cerah	Lambat	Dingin	Iya



Proses pada pohon keputusan yaitu dengan mengubah bentuk data (tabel) menjadi model pohon dan menyederhanakan *rule* (Mardi, 2017).

Pohon keputusan adalah salah satu metode belajar yang sangat populer dan banyak digunakan secara praktis. Metode ini merupakan metode yang berusaha menemukan fungsi-fungsi pendekatan yang bernilai diskrit. Adapun yang menjadi konsep pohon keputusan adalah mengubah data yang ada pada tabel keputusan menjadi sebuah pohon keputusan dan aturan-aturan keputusan (*rule-rule*). Berikut adalah konsep gambaran pohon keputusan dapat dilihat pada Gambar 2.2.



Gambar 2.2 Konsep *decision tree* (Rosela, 2019)

Manfaat utama dari penggunaan *decision tree* adalah kemampuannya untuk *break down* proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan. *Decision tree* juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan

sebuah variabel target. *Decision tree* memadukan antara eksplorasi data dan pemodelan, sehingga sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain (Rosela, 2019).

#### 2.4. Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Sedangkan pohon keputusan kumpulan-kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan (Utari, 2015)

Secara umum untuk membangun Algoritma C4.5 adalah sebagai berikut:

- a. Hal yang pertama dilakukan adalah menyiapkan dataset yang akan digunakan sebagai bahan pengujian.
- b. Setelah itu mencari nilai *entropy* dan *information gain* total yang diperoleh dari jumlah seluruh data, dengan menggunakan persamaan seperti dibawah ini :
  1. Menghitung nilai *entropy*. *Entropy* merupakan jumlah bit yang dibutuhkan untuk menyatakan suatu kelas. Nilai *entropy* dihitung dengan persamaan (1) yang ditulis sebagai berikut:
 
$$Entropy(S) = \sum_i^n -p_i \cdot \log_2(p_i) \dots\dots\dots (1)$$
 perhitungan nilai *entropy* dari atribut yang ada di dataset bisa dilakukan dengan persamaan (1)
  2. Menghitung nilai *gain*. *Gain* merupakan salah satu langkah pemilihan atribut yang digunakan untuk memilih tes atribut setiap simpul pada pohon keputusan atau dengan kata lain *gain* merupakan tingkat pengaruh suatu atribut terhadap keputusan atau ukuran efektifitas suatu variabel dalam

mengklasifikasikan data. *Gain* dihitung dengan persamaan (2) yang ditulis sebagai berikut:

$$Gain(S, A) = Entropy(S) - \sum_i^n 1 - \frac{s_i}{s} * Entropy(S) \dots\dots\dots(2)$$

perhitungan nilai *gain* dari atribut yang ada di dataset bisa dilakukan dengan persamaan (2)

- c. Selanjutnya adalah membuat pohon keputusan dengan melihat nilai *gain* tertinggi dari semua atribut yang ada, setelah itu nilai *gain* ratio tertinggi dari atribut tadi akan dijadikan sebagai akar dalam pohon keputusan.
- d. Kemudian membuat cabang dari masing-masing nilai atribut yang tersisa sesuai dengan jumlah nilai variabel *gain* ratio tertinggi menggunakan persamaan (1) dan (2).
- e. Tahap terakhir adalah mengulangi proses pada setiap cabang hingga hingga semua kasus pada cabang sampai memiliki kelas yang sama dan sampai tidak bisa dilakukan sebuah proses lagi.

### 2.5. Kepuasan Pelanggan

kepuasan pelanggan merupakan respon pelanggan terhadap perbandingan antara kinerja dan harapan sesuai dengan evaluasi ketidaksesuaian setelah pelanggan memakai suatu produk. Kepuasan ini akan dirasakan oleh pelanggan apabila mereka telah mengkonsumsi suatu produk atau jasa. Jika pelanggan menyukai produk yang mereka konsumsi maka pelanggan sudah merasakan kepuasan, sebaliknya apabila produk yang mereka konsumsi tidak sesuai dengan kebutuhan mereka dan berpindah membeli produk lain maka pelanggan tidak merasakan kepuasan (Endrawati, 2018). Memuaskan kebutuhan pelanggan adalah keinginan setiap perusahaan. Selain faktor penting bagi kelangsungan hidup perusahaan, memuaskan kebutuhan pelanggan dapat meningkatkan keunggulan

dalam persaingan. Pelanggan yang puas terhadap produk dan jasa pelayanan cenderung untuk membeli kembali produk dan menggunakan kembali jasa pada saat kebutuhan yang sama muncul kembali dikemudian hari. Hal ini berarti kepuasan merupakan faktor kunci bagi konsumen dalam melakukan pembelian ulang yang merupakan porsi terbesar dari volume penjualan perusahaan. Faktor-faktor pendorong kepuasan kepada pelanggan (Maulana, 2019).

- a. Kualitas Produk Pelanggan akan merasa puas setelah membeli dan menggunakan produk tersebut yang memiliki kualitas produk baik.
- b. Harga. Biasanya harga murah adalah sumber kepuasan yang penting. Akan tetapi biasanya faktor harga bukan menjadi jaminan suatu produk memiliki kualitas yang baik.
- c. Kualitas Jasa Pelanggan merasa puas apabila mereka memperoleh jasa yang baik atau sesuai dengan yang diharapkan dari pegawai maupun karyawan perusahaan.
- d. *Emotional factor* Kepuasannya bukan karena kualitas produk, tetapi harga diri atau nilai sosial yang menjadikan pelanggan puas terhadap merek produk tertentu.
- e. Biaya atau kemudahan untuk mendapatkan produk atau jasa kenyamanan dan efisien dalam mendapatkan suatu produk atau jasa serta mudah mendapatkan jasa produk memberikan nilai tersendiri bagi kepuasan pelanggan.

## **2.6. Studi Pustaka**

Beberapa penelitian telah banyak dilakukan dengan menggunakan teknik data mining untuk menggali berbagai informasi tentang kepuasan pelanggan dan penelitian lain yang berkaitan dengan Algoritma C4.5 dan Naïve bayes, sebagai salah satu dari penerapan metode penelitian yang akan dilakukan. Adapun beberapa penelitian tersebut disajikan pada Tabel 2.2.

Tabel 2.2 Studi Pustaka

Penulis	Judul Penelitian	Metode	Pokok Bahasan	Akurasi
(Wibowo, 2019)	Komparasi Algoritma <i>Naive Bayes</i> dan <i>Decision tree</i> Untuk Memprediksi Lama Studi Mahasiswa	Algoritma C4.5 dan <i>Naive bayes</i>	Penelitian ini bertujuan memprediksi lama studi mahasiswa lulus tepat waktu atau terlambat lulus. Menggunakan 100 data sample data yang digunakan yaitu: gender, status mahasiswa, nilai, dan beasiswa. Dari komparasi menggunakan Algoritma <i>Naive Bayes</i> dan <i>Decision tree</i> dapat diambil kesimpulan bahwa metode <i>Decision Tree</i> memiliki persentase keakuratan yang lebih tinggi dibandingkan <i>Naive Bayes</i> sehingga dapat dikatakan metode <i>decision Tree</i> lebih akurat dan lebih detail daripada <i>Naive Bayes</i> .	Metode Algoritma C4.5 <i>decision Tree</i> mendapatkan hasil dengan tingkat akurasi 55% dan perhitungan menggunakan metode <i>Naive Bayes</i> mendapatkan tingkat akurasi sebesar 30%
(Galih, 2019)	Data Mining di Bidang Pendidikan untuk Analisa Prediksi Kinerja Mahasiswa dengan Komparasi 2 Model Klasifikasi pada STMIK Jabar	Algoritma C4.5 <i>decision Tree</i> dan <i>Naive Bayes Classifier</i> (NBC)	Melakukan penelitian dengan menggunakan algoritma <i>Naive Bayes Classifier</i> dan Algoritma C4.5 <i>decision Tree</i> untuk melakukan perbandingan algoritma mana yang lebih baik sehingga menjadi landasan dan masukan untuk Labkom STMIK Jabar agar diterapkan pada pengembangan sistem informasi akademik cerdas kedepannya yang dapat membantu lebih cepat dalam membuat keputusan bagi pihak kampus terhadap penanganan mahasiswa yang terindikasi	Berdasarkan pengujian Algoritma C4.5 <i>decision Tree</i> dan <i>Naive Bayes</i> tersebut dengan menggunakan ratio data training dapat mempengaruhi hasil dari nilai akurasi masing-masing, adapun nilai akurasi terbaik pada model algoritma <i>Naive Bayes Classifier</i> (NBC) adalah 86,83% dengan ratio data training 80%, sedangkan pada model algoritma C4.5 adalah 88,10
			terlambat dalam masa studinya. Teknik yang digunakan dalam pengujiannya menggunakan teknik <i>split validation</i> adalah salah satu operator di dalam <i>Rapidminer</i> yang memiliki fungsi dalam melakukan validasi sederhana secara acak kemudian membagi dataset menjadi 2 bagian yaitu data	% dengan ratio data training 90%.

			<i>training</i> dan data <i>testing</i> dalam mengevaluasi model. Dapat disimpulkan nilai akurasi terbaik algoritma C4.5.	
(Iskandar, 2020)	Komparasi akurasi metode k-nearest neighbor dan C4.5 untuk <i>Customer Relationship Management</i> pada perusahaan pembiayaan	Algoritma C4.5 dan Algoritma KNN	Melakukan komparasi akurasi metode K-Nearest Neighbor dan Algoritma C4.5 untuk <i>Customer Relationship Management</i> pada perusahaan pembiayaan dalam melakukan klasifikasi loyalitas pelanggandan jumlah pengujian data 122 sample. Pembuatan aplikasi komparasi untuk klasifikasi pelanggan menggunakan teknik algoritma C4.5 menghasilkan model klasifikasi terbaik dengan.	dilakukan pengujian perbandingan ditemukan akurasi Algoritma C4.5 sebesar 93.10% dan nilai akurasi Algoritma KNN sebesar 90.52% dengan selisih akurasi 2,58%.
<b>Penulis</b>	<b>Judul Penelitian</b>	<b>Metode</b>	<b>Pokok Bahasan</b>	<b>Akurasi</b>
(Gupitha, 2018)	Penentuan Ekstra Kurikuler siswa Smk Terpadu Lampang Subang dengan Algoritma C 4.5	Algoritma C 4.5	Penelitian akan menggunakan pendekatan algoritma C4.5 untuk menentukan ekstra kurikuler yang akan dipilih oleh siswa sesuai dengan keinginan, kepribadian, minat, dan bakat. Penelitian ini menggunakan <i>Split Validation</i> sebagai validasi data <i>training</i> maupun <i>testing</i> yang digunakan serta tingkat konsisten data yang digunakan. Data yang diuji yaitu sebanyak 20 data. Hasil evaluasi dan validasi dengan <i>Split Validation</i> menggunakan <i>training</i> 70% dan data <i>testing</i> 30%,	Menunjukkan tingkat akurasi pada Algoritma C4.5 sebesar 81,94%.
(Febriyanto, 2018)	Implementasi Algoritma C4.5 Untuk Klasifikasi Tingkat Kepuasan Pembeli	Algoritma C 4.5	Klasifikasi dengan konsep data mining dengan menggunakan 100 data yang kemudian diambil sampel sebanyak 15 data konsumen dengan menggunakan variabel pelayanan, fasilitas,	Penelitian ini menghasilkan akurasi 91%, dengan nilai presisi pada prediksi puas sebesar 66.67% dan nilai presisi pada prediksi tidak puas

Tabel 2.2 Lanjutan Studi Pustaka

	<i>Online Shop</i>		<p>produk untuk melakukan perhitungan Algoritma C4.5 mampu mengklasifikasi tingkat kepuasan pelanggan pada <i>online shop</i>. Pohon keputusan yang dihasilkan dengan teknik klasifikasi Algoritma C4,5 dari variabel yang memiliki <i>gain</i> tertinggi yaitu pelayanan, sehingga pemilik <i>onlineshop</i> dapat menjadikan acuan untuk memperbaiki kualitas <i>online shop</i>.</p>	<p>sebesar 33.3%,</p>
--	--------------------	--	---	-----------------------

Dari beberapa penelitian pada Tabel 2.2 mengenai data mining terutama bagaimana sebuah Algoritma C4.5 dan *Naïve Bayes* bisa digunakan dalam mengukur hasil kepuasan pelanggan, dengan aturan asosiasi sebuah data bisa diolah sebagai sebuah informasi serta dapat bermanfaat untuk perusahaan. Dari beberapa penelitian yang dilakukan sebelumnya akan digunakan sebagai acuan bagi penulis untuk melakukan penelitian yang berkaitan tentang kepuasan pelanggan yaitu penelitian (Iskandar, 2020) Melakukan komparasi akurasi metode K-Nearest Neighbor dan Algoritma C4.5 untuk *Customer Relationship Management* pada perusahaan pembiayaan dalam melakukan klasifikasi loyalitas pelanggan hasil dari komparasi ditemukan akurasi Algoritma C4.5 sebesar 93.10% dan nilai akurasi Algoritma KKN sebesar 90.52% dengan selisih akurasi 2,58%, menggunakan pengujian data 122 sample. Selanjutnya penelitian yang dilakukan oleh (Gupitha, 2018) penentuan ekstra kurikuler siswa smk terpadu lampang subang dengan Algoritma C4.5 menggunakan 20 data hasil evaluasi dan validasi dengan *confusion matrix* dengan membagi data *training* 70% dan data *testing* 30%, menunjukkan tingkat akurasi pada Algoritma C4.5 sebesar 81,94%. Sementara pada penelitian yang dilakukan oleh penulis dalam pengujian dilakukan menggunakan teknik *Split Validation* dengan membagi data *training* dan data *testing* pada data set, kemudian membandingkan Algoritma C4.5 dan *Naïve Bayes* untuk mengetahui tingkat akurasi data dengan menggunakan dua algoritma tersebut untuk mengukur hasil kepuasan pelanggan pada CV. Jatimas Furniture.