

## BAB II

### LANDASAN TEORI

#### 2.1 Penelitian Sebelumnya

Pada penelitian pertama, metode C4.5 dipakai untuk melakukan penelitian tentang kepuasan klien Perum DAMRI. Hasil pengujian metode C4.5 pada 90 sampel data pelanggan untuk memprediksi kepuasan pelanggan Perum DAMRI menunjukkan bahwa metode C4.5 cukup akurat, dengan tingkat akurasi 93 persen (Rohman, 2015). Untuk membuktikan banyaknya pelanggan yang puas dan tidak puas terhadap persentase kepuasan pelanggan Perum DAMRI, peneliti membuat kuesioner dengan jumlah kuesioner yang telah ditentukan sebanyak 150 kuesioner. Jumlah responden yang senang dan tidak puas dari kuesioner yang terdaftar dimasukkan ke dalam rumus algoritma C4.5. Peneliti menyelesaikan tiga kali percobaan pada data pelanggan Bus Perum DAMRI dengan total data latih dan data test yang lain. Pada pengujian yang pertama data latih 40 persen dan data testing 60 persen, pengujian yang kedua data latih 60 persen dan data testing 40 persen, dan pengujian yang ketiga data latih 80 persen dan data testing 20 persen.

Untuk penelitian kedua, data seluler dimaksudkan untuk digunakan untuk menerapkan versi *classification decision tree* untuk memprediksi loyalitas pelanggan dan menentukan variabel yang paling berdampak. Klien data seluler adalah fokus dari evaluasi ini. Data utama dari kuesioner, yang disampaikan dalam bentuk pertanyaan. Teknik yang dipakai untuk prediksi menerapkan *method* C4.5. Untuk mencari root node dan node lainnya dilakukan perhitungan dengan menggunakan metode C4.5, dimana atribut loyalitas pelanggan data seluler adalah atribut tujuan dan karakteristik lainnya adalah atribut sumber. Teknik algoritma C4.5 menghasilkan 97,5 persen temuan klasifikasi yang akurat, menunjukkan bahwa metode C4.5 memadai untuk menentukan besarnya loyalitas pelanggan data seluler (Santoso, 2011).

Selain itu, dalam studi ketiga, pengaruh loyalitas konsumen pada atribut produk diperiksa. Tujuannya untuk mengukur minat konsumen terhadap berbagai produk Indomie dan Sedaap. Selain itu, pengaruh loyalitas konsumen terhadap kualitas barang Indomie dan Sedaap akan diselidiki. Atribut yang digunakan untuk membantu penelitian ini meliputi kemasan, harga, rasa, keragaman, dan distribusi. Mahasiswa Fakultas Ekonomi Universitas Sanata Dharma Yogyakarta mengikuti penelitian ini. Berdasarkan hasil polling, 2 orang sangat setia pada mie Sedaap, 14 orang tidak, 14 orang setia, dan 5 orang sangat setia pada mie Sedaap. Dan hasil studi regresi mie Sedaap menunjukkan bahwa loyalitas pelanggan dengan kualitas produk berdampak pada loyalitas pelanggan,  $F_{hitung} > F_{tabel}$ ,  $= 27,633 > 2,545$ .

## **2.2 Decision Support System**

Berdasarkan Turban (2005:30) “*Decision Support System* (DSS) atau Sistem Pendukung Keputusan (SPK) sebagai sistem interaktif yang membantu pengambilan keputusan dengan mengimplementasikan data dan model keputusan untuk memecahkan masalah semi terstruktur dan tidak terstruktur”.

## **2.3 Prediksi**

Prediksi didasarkan pada bagaimana prosedur akan mengekstrak pola tertentu dari data. Pola ini terbentuk dari banyak variabel data. Misalnya, jika Anda menjumpai sebuah pola, Anda dapat menggunakannya untuk memprediksi nilai atau jenis variabel lain yang tidak diketahui. Fitur ini dapat membuat prediksi yang benar menjadi lebih mudah dan lebih bermanfaat bagi semua orang (Jagoanhosting, 2020).

## **2.4 Data Mining**

Data mining juga dapat diartikan sebagai rangkaian kegiatan untuk menemukan pola-pola menarik dalam jumlah data yang besar, kemudian menyimpan data tersebut dalam database, data warehouse atau penyimpanan data. Ada beberapa ilmu yang mendukung teknik

data mining, antara lain analisis data, pemrosesan sinyal, dan jaringan saraf pengenalan pola (Irfiani & Indriyani, 2017).

Data mining memiliki sejumlah teknik bergantung aktivitas yang dapat dilakukan, masing-masing dengan algoritmanya sendiri. Teknik data mining diklasifikasikan menjadi enam jenis, yaitu *Description*, *estimation*, *prediction*, *classification*, *clustering*, dan *association* (Chapman P, 2000).

## **2.5 Klasifikasi**

Klasifikasi adalah komponen dari algoritma data mining; merupakan algoritma yang menggunakan data dengan target (*class/label*) berupa nilai kategorikal/nominal. Proses klasifikasi, menurut Gorunescu, didasarkan pada 4 komponen mendasar, yaitu *Class*, *Predictor*, *Training dataset*, dan *Testing dataset* (Gorunescu F, 2011).

## **2.6 Decision Tree Algoritma C4.5**

Salah satu metode *classification* yang terkuat dan sangat baik adalah pohon keputusan. Sistem Pohon Keputusan mengalihkan banyak data menjadi *Decision Tree* yang mencerminkan aturan (*rules*), memungkinkan manusia untuk dengan mudah memahami *rules*. *Decision Tree* dapat digunakan dalam menganalisis data dan mengungkap hubungan tersembunyi antara berbagai variable input dan variable tujuan (*class*) (Meilina, 2015).

Algoritma C4.5 adalah contoh dari algoritma pembelajaran mesin. Mesin (komputer) juga akan menyediakan kumpulan data untuk dipelajari, yang disebut sebagai learning dataset, dalam proses ini. Dataset uji akan digunakan untuk menganalisis data baru berdasarkan kesimpulan yang dibuat dari hasil pelatihan. Karena metode C4.5 dapat digunakan untuk klasifikasi, dan hasil dari penanganan dataset pengujian adalah mengkategorikan data ke dalam kelas umum, tahapan algoritma C4.5 berikut digunakan untuk membuat pohon keputusan (Kusrini dan Emha Taufiq Luthfi, 2009):

- Tentukan atribut yang akan menjadi *root*(akar).
- Tetapkan *branch* untuk setiap nilai.
- Tetapkan kelas untuk setiap cabang.
- Ulangi proses ini untuk setiap *branch* hingga semua masalah di *branch* tersebut memiliki *class* yang sama.

Atribut yang akan dijadikan akar dipilih berdasarkan nilai *gain* terbesar di antara atribut yang ada. Sedangkan untuk mendapatkan nilai *gain* terbesar, hitung *entropy* dari seluruh *value* pada atribut. Setelah menentukan nilai *entropy* pada sampel data, atribut paling penting bakal menjadi ukuran untuk mengklasifikasikan data.

Dalam algoritma C4.5, rumus berikut digunakan untuk menghitung *entropy* :

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

Keterangan:

- S menunjukkan Himpunan (dataset) kasus
- k mewakili jumlah partisi S
- $P_i$  adalah probabilitas yang didapat dari Sum (Ya) atau Sum (Tidak) dibagi total

kasus

Sesudah menghitung *entropy* untuk total kasus, telusuri setiap atribut dan nilainya masing-masing. Langkah selanjutnya adalah menghitung *Gain*, yang memiliki rumus sebagai berikut:

$$\text{Gain} (A) = \text{Entropi} (S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times \text{Entropi}(S_i)$$

## 2.7 Confusion Matrix

Untuk mengukur kinerja akurasi memakai *Confusion Matrix*, mengandung empat istilah yang mewakili hasil dari proses prediksi, yaitu *True Positive (TP)*, *False Positive (FP)*, *False Negative (FN)*, *True Negative (TN)* (Sokolova M, 2009).

**Tabel 2.1 Confusion Matrix**

Data Class	Prediksi Kelas A	Prediksi Kelas B
Kelas A	TP	FN
Kelas B	FP	TN

Keterangan :

*TP* : *True Positive*, *TP* menunjukkan nilai data Kelas A yang diprediksi benar yaitu Kelas A

*FN* : *False Negative*, *FN* menunjukkan nilai data Kelas A yang diprediksi benar yaitu Kelas B

*FP* : *False Positive*, *FP* menunjukkan nilai data Kelas B yang diprediksi benar yaitu Kelas A

*TN* : *True Negative*, *TN* menunjukkan nilai data Kelas B yang diprediksi benar yaitu Kelas B

Angka akurasi menunjukkan seberapa baik sistem dapat memprediksi data. Dengan kata lain, angka akurasi adalah prediksi yang benar dari data dibagi dengan jumlah total data. Menurut (Sokolova M, 2009), rumus akurasi adalah sebagai berikut :

$$\text{Akurasi} = \frac{\sum_{i=1}^l \frac{TP_j + TP_j}{(TP_j + TN_j + FP_j + FN_j)}}{l} \times 100\%$$

Keterangan

$TP_j$  : *True Positive* untuk kelas j

$FN_j$  : *False Negative* untuk kelas j

$FP_j$  : *False Positive* untuk kelas j

$TN_j$  : *True Negative* untuk kelas j

## 2.8 Java

Java adalah bahasa pemrograman yang berjalan di banyak platform (*Multi platform*). *Multi platform* berarti aplikasi atau *software* yang dibuat dalam bahasa pemrograman Java dapat dijalankan di komputer mana pun yang mendukung bahasa pemrograman Java tanpa memerlukan perubahan kode. Bahasa pemrograman Java memiliki kelebihan sebagai berikut:

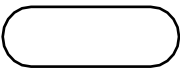

- Pemrograman berbasis objek atau *OOP (Object Oriented Programming)*.
- *Library* yang lengkap.


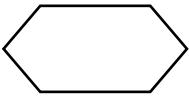
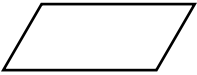
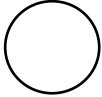
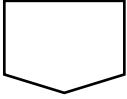

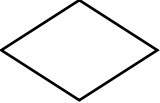


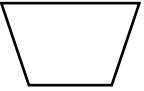
## 2.9 Flowchart

*Flowchart* ialah penggambaran visual dari tahapan suatu program dan urutan prosedur untuk membantu dalam pemecahan masalah (Indrajani, 2015).

Simbol-simbol yang dipakai *flowchart* bisa dilihat di dalam table 2.2 berikut:

**Tabel 2. 1 Komponen Pada *Flowchart***

No	Symbol	Description
1.		Permulaan ( <i>start</i> ) dan akhir ( <i>end</i> )
2.		Komputer melakukan proses/kegiatan.

3.		Aktualisasi bagian prosedur (sub-proses).  Langkah yang di informasikan disini belum spesifik maka dilain tempat akan dijelaskan.
4.		Awalan dipakai sebagai memberi nilai awalan suatu besaran.
5.		Mengutarakan masuk dan keluar ( <i>input/output</i> ).
6.		Simbol yang posisinya jauh atau sulit, dikaitkanlah garis di halaman yang sama
7.		Mengutarakan penyambung ke halaman lainnya.
8.		Masukkan yang dibutuhkan kertas, atau keluaran yang harus dicetak dikertas.
9.		Memilih proses atau keputusan
10.		Mengutarakan media pengarsipan drum magnetik.
11.		Mengutarakan <i>input/output</i> menerapkan disket.
12.		Kegiatan atau proses manual yang tidak akan dilakukan komputer.

## **2.10 Tools Design UML**

### **1.10.1 Power Designer**

PowerDesigner adalah alat lunak yang menyelaraskan bisnis dan teknologi informasi menggunakan metodologi berbasis model. Alat ini adalah alat pemodelan perusahaan yang membantu implementasi Arsitektur Perusahaan dan menyediakan lingkungan manajemen metadata yang kuat untuk pengembangan aplikasi.

PowerDesigner adalah gabungan dari teknik pemodelan konvensional seperti UML, BPMN, dan pemodelan data.

### **1.10.2 Draw.io**

Draw.io adalah situs web yang didedikasikan untuk membuat diagram online. Hanya browser yang mendukung HTML 5 yang memungkinkan pengguna untuk menggunakan semua fitur di situs web ini.

## **2.11 Unified Modeling Language (UML)**

*UML* merupakan sekumpulan perlengkapan untuk menjalankan abstraksi tentang suatu system yang akan dibuat. *UML* dapat juga digunakan untuk mendokumentasi dan menspesifikasi system *UML* dapat membantu pengembangan sistem aplikasi dengan menyajikan beragam model diagram, seperti yang juga dipakai dalam perangkat lunak ini, di antaranya:

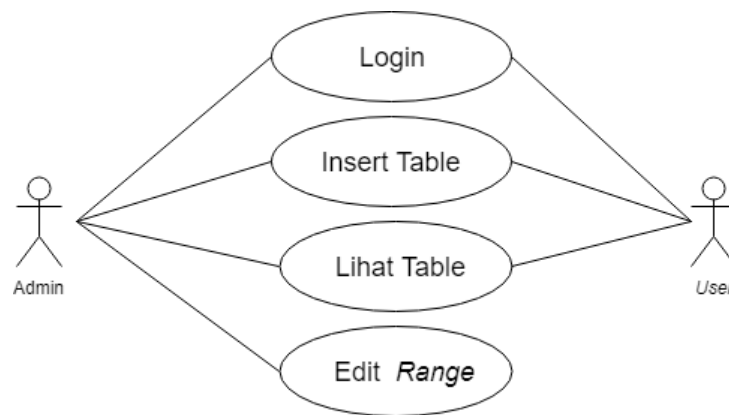
- *Use Case Diagram*
- *Activity Diagram*

### **2.11.1 Use Case Diagram**

Gambaran pada kaitan antar sistem dan aktor ialah *Use case diagram*. *Use case diagram* berperan untuk menggambarkan jenis ikatan antara aktor bersama sistemnya sendiri yang






sebenarnya menggunakan penjelasan tentang dengan cara apa sistem dipakai. (Whitten, J.L., 2004)



**Gambar 2.1.1 Contoh *Diagram Model Use Case***




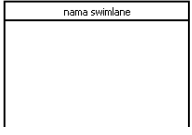

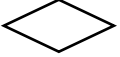
**Tabel 2. 2 Simbol-simbol *Use Case Diagram***

Symbol	Symbol Name	Description
	<i>Use Case</i>	Simbol tugas dari sistem, akhirnya aktor sistem mengerti tentang fungsi sistem yang akan dibangun
 Actor	Actor	Apa saja yang berhubungan dengan suatu sistem untuk perpindahan informasi.
	<i>Relationship/</i> hubungan	Interaksi aktor dengan <i>use case</i> dimana terjadi koneksi.

### 2.11.2 Activity Diagram

Activity Diagram memberi penyelidikan kapasitas buat menampilkan operasi dalam kerangka data. Garis besar tindakan dapat digunakan untuk menunjukan proses kerja, kasus menggunakan individu, atau alasan pilihan yang terkandung dalam strategi individu. Garis besar gerakan juga memberikan cara untuk menangani proses tampilan yang sama.

**Tabel 2. 3 Simbol-simbol Activity Diagram**

No	Symbol	Description
1.		Mengawali suatu aktivitas dalam sistem.
2.		Akhir dari suatu proses aktivitas.
3.		Mengaitkan antar lambang yang sesuai dengan arah sistem.
4.		Menampakkan pelaku dari diagram aktivitas yang dibuat
5.		Penanda aktivitas yang ada di dalam sistem.
6.		Percabangan atau <i>Decision (if-else)</i>

### 2.12 Database

Basis data, atau *Database* ialah kelompok data yang disusun agar dapat mudah dikelola, berdasarkan ketentuan terkait tertentu. Pengguna bisa dengan cepat mengakses data atau

informasi yang berguna, menyimpan informasi, dan menghapus informasi dengan manajemen ini.

Basis data, dalam arti lain ialah sistem yang terbentuk dari kelompok file, tabel, atau arsip yang dihubungkan dan disimpan di berbagai media elektronik.

### **2.13 MySQL**

My Structure Query Language (MySQL) merupakan server database atau kumpulan data yang “amat populer dan sering dipakai untuk mengembangkan aplikasi web yang membutuhkan database selaku sumber dan pengelolaan data” berdasarkan Arief (2011). Mysql adalah sistem manajemen database gratis yang berjalan pada SQL (Structured Query Language). MySQL biasanya dijalankan di berbagai sistem, termasuk Windows, Linux, dan lainnya.

### **2.14 XAMPP**

XAMPP berasal dari huruf X, A, M, P, P yang memiliki artinya masing-masing, X berarti program dapat dijalankan pada berbagai Sistem Operasi atau *Operating System* (OS). A bermakna Apache yang melambangkan aplikasi web server. M bermakna Mysql yang melambangkan *database* server. P pertama yang bermakna dari bahasa pemrograman PHP. P kedua yang bermakna dari bahasa pemrograman Perl. XAMPP adalah sebuah aplikasi gratis (*free software*) yang menunjang banyak SO (*System Operation*) yang merupakan kelompok dari beberapa program lain. XAMPP melayani beberapa tujuan, termasuk melayani sebagai server (localhost) yang mencakup Apache HTTP Server, database Mysql, dan terjemahan bahasa PHP dan Perl.

### **2.15 Neatbeans IDE**

Netbeans adalah Java IDE (Integrated Development Environment) yang berjalan di Swing dan didasarkan pada bahasa pemrograman Java. Swing adalah teknologi yang

memungkinkan pengembang membuat program desktop yang berjalan di berbagai platform, termasuk Windows, Mac OS, Linux, dan Solaris.

Integrated Development Environment (IDE) adalah sebuah pemrograman atau pengembangan sistem yang dibangun ke dalam sebuah perangkat lunak (software). Netbeans mencakup antarmuka pengguna grafis (GUI), editor kode atau editor teks, kompiler, dan debugger. Netbeans akan lebih mudah digunakan untuk programmer dan pengembang sebagai akibat dari ini. Netbeans tidak hanya mendukung bahasa pemrograman Java, tetapi juga memungkinkan Anda untuk merancang dan mengembangkan aplikasi dalam C, C++, dan bahkan bahasa dinamis seperti PHP, JavaScript, Groovy, dan Ruby.

## **2.16 Java**

Sun Microsystems pertama kali menerbitkan Java pada tahun 1995 sebagai bahasa pemrograman dan platform komputasi. Java adalah bahasa pemrograman yang mendukung kekuatan utilitas, game, dan aplikasi komersial. Java diinstal pada sekitar 850 juta komputer pribadi secara global, serta miliaran perangkat lain seperti ponsel dan pesawat televisi.

Salah satu properti Java adalah portabilitas, yang berarti bahwa program Java harus berjalan dengan cara yang sama pada perangkat keras atau platform sistem operasi apa pun. Mengkompilasi kode bahasa java ke dalam bytecode java menyelesaikan ini. Untuk menjalankan program Java atau applet Java, pengguna sering menggunakan Java Runtime Environment (JRE) yang diinstal pada PC mereka sendiri atau Java Runtime Environment (JRE) melalui browser web.

Java Development Kit (JDK) diperlukan untuk pembuatan dan pengembangan aplikasi berbasis Java, dan pemilik lisensi JDK saat ini adalah Oracle Corporation, yang membeli Sun Microsystems pada awal 2010.