

BAB II

TEORI PENUNJANG

Bab ini menjelaskan tentang dasar teori yang digunakan dalam menyusun tugas akhir ini, yaitu tentang gambaran secara umum perusahaan XYZ, pengertian *data mining*, *naïve bayes classifer* dan komponen penunjang lainnya.

2.1 Tinjauan Tempat Penelitian

PT. XYZ berdiri pada tanggal 01 September 2015 di desa Karanganyar kecamatan Kandanghaur kabupaten Indramayu dengan status dealer H123. PT. XYZ berdiri atas keinginan owner ibu Hj. Nurlaela Azis yang ingin mempunyai dealer cabang di wilayah Indramayu yaitu daerah asal beliau yang di mana pada waktu itu sudah mempunyai dealer di wilayah Cimalaka kabupaten Sumedang. Sehingga atas kerja kerasnya untuk dapat membangun dealer di daerahnya akhirnya berdirilah dealer resmi honda dengan nama PT. XYZ dan alamat dealer berpindah mulai tanggal 1 Januari 2017 ke alamat Jl. Raya Muntur No. 126 Losarang Indramayu, sekarang ini rata-rata penjualan 220 unit setiap bulannya.

2.1.1 Visi dan Misi Perusahaan

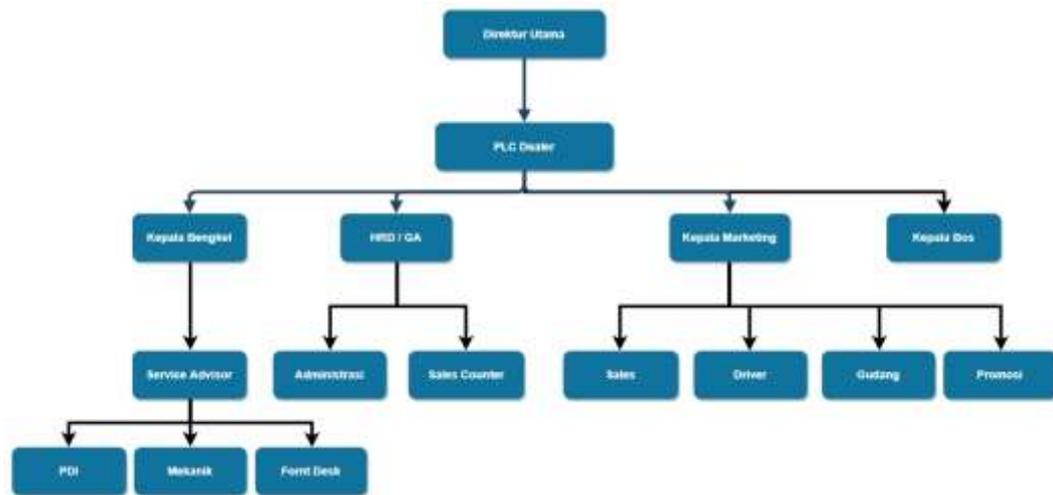
Visi dan misi PT. XYZ adalah sebagai berikut:

Visi: “Menjadi dealer yang maju, handal dan terpercaya serta mengedepankan kepuasan konsumen”.

Misi: “Menjadi dealer pilihan masyarakat Indramayu”.

2.1.2 Struktur Organisasi Perusahaan

PT. XYZ dipimpin oleh seorang Direktur yang dibantu oleh PLC Dealer serta jajaran lainnya yang bekerja di dalam divisi masing-masing. Berikut ini adalah gambaran dari struktur organisasi PT. XYZ.



Gambar 2.1 Struktur Organisasi PT. XYZ

Pada gambar 2.1 ditampilkan bagian-bagian dalam struktur organisasi di dalam PT. XYZ.

2.1.3 Sumber Data

Sumber data yang digunakan dalam penelitian ini diperoleh dari PT. XYZ yang merupakan dealer resmi PT. Astra Honda Motor (AHM) di Indramayu. Sumber data ini terdiri dari 7000 data hasil penjualan onderdil sepeda motor.

GRUP	Kode	Nama	Satuan	Qty	Nilai Transaksi
BUSI	31916KRM841	SPARK PLUG CPR8EA-9 (NG)	PCS	2	26.000
	31926KRM842	SPARK PLUG CPR8EA-9 (NG)	PCS	7	120.750
	31928MFFD01	SPARK PLUG U24EPR9 (DS)	PCS	2	26.000
	9806657723	SPARK PLUG U22FSU (DS)	PCS	1	12.400
	9806786871	SPARK PLUG U20EPR9 (DS)	PCS	3	39.000
	HGP	06401K15900	DRIVE CHAIN KIT	PCS	1
06401K18900		DRIVE CHAIN KIT	PCS	1	247.000
06401KPH881		DRIVE CHAIN KIT	PCS	1	145.000
06401K1VB900		DRIVE CHAIN KIT	PCS	1	148.200
06401KYE900		DRIVE CHAIN KIT	PCS	1	252.700
06435KPP901		PAD SET RR	PCS	2	86.000
06435KSP001		PAD SET RR	PCS	4	223.300
06455KPP901		PAD SET FR	PCS	3	125.400
06455KVBT01		PAD SET FR	PCS	1	51.000
06455KY911		PAD SET FR	PCS	1	45.500
06455KYE901		PAD SET FR	PCS	3	139.500
082322MSKB.N1		SCOOTER GEAR OIL (120ML) IDE	PCS	-35	418.800
08294M9928YN9		SCOOTER GEAR OIL (120ML) FED	PCS	7	91.000
08CLAH50500		HONDA MIC GENUINE COOLANT	PCS	14	245.000
11100KVBB31		CRANKCASE COMP R	PCS	1	400.000
11103KVB930		11103KVB001	PCS	2	22.000
11393KQH901		GASKET R COVER	PCS	1	22.000
12191GGC900		GASKET CYLINDER	PCS	1	6.175
12209GB4682		SEAL VLV. STEM	PCS	4	23.400

Gambar 2.2 Sumber data PT. XYZ

Gambar 2.2 menunjukkan cuplikan dari hasil penjualan onderdil sepeda motor PT. XYZ pada bulan februari tahun 2017.

2.2 Data Mining

Dalam buku “*Decision Support System and Intelligent Systems*” karangan dari Efraim Turban menjelaskan bahwa *data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar[9].



Gambar 2.3 Data mining merupakan hasil irisan dari berbagai ilmu disiplin

Proses data *mining* sendiri dikelompokkan menjadi beberapa kelompok berdasarkan tugasnya masing-masing, yaitu[10], [11]:

1. Deskripsi

Merupakan suatu proses di mana mencari suatu karakteristik penting dari kumpulan data untuk menggambarkan pola dan kecenderungannya yang terdapat pada data tersebut.

2. Prediksi

Mencari pola dari data yang dimiliki dengan menggunakan beberapa variabel agar dapat memprediksi variable lainnya yang tidak diketahui nilai atau jenisnya.

3. Klasifikasi

Merupakan suatu proses untuk menemukan model untuk menggambarkan class atau konsep dari suatu data. Dengan adanya proses ini diharapkan dapat mengetahui kecenderungan data pada masa yang akan datang.

4. Clustering

Merupakan kumpulan dari *record* yang memiliki kemiripan maupun ketidakmiripan antara satu dengan yang lainnya.

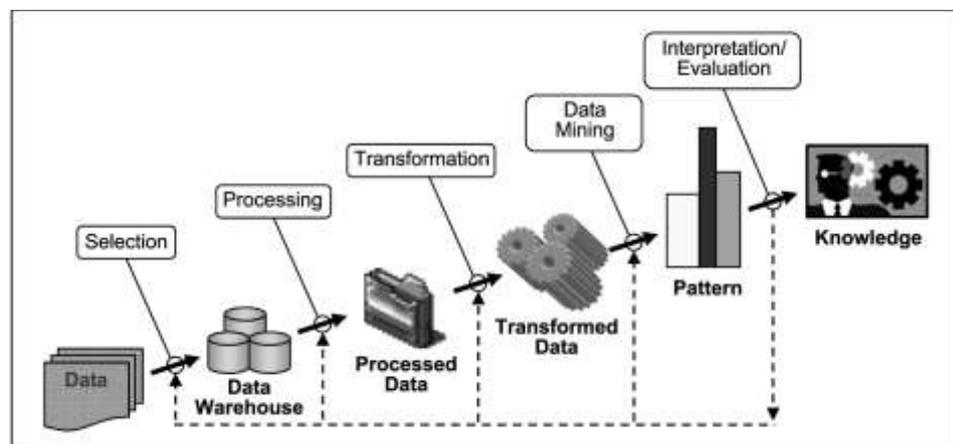
5. Asosiasi

Proses yang digunakan untuk menemukan suatu hubungan yang terdapat pada nilai atribut dari data yang terkumpul.

6. Estimasi

Estimasi hampir sama dengan klasifikasi hanya saja berbeda dibagian *variable* targetnya karena estimasi lebih kearah numerik.

Data mining memiliki beberapa tahapan dalam proses kerjanya yang digambarkan dalam ilustrasi berikut[12]–[14].



Gambar 2.4 Tahapan dalam data mining

1. Data selection

Seleksi data dari sekumpulan data operasional perlu dilakukan sebelum melakukan tahap pengumpulan atau penggalian informasi dalam *knowledge discovery in databases* (KDD). Data hasil seleksi yang digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. Pre-Processing/cleaning

Sebelum proses *data mining* dilakukan terlebih dahulu lakukan proses *cleaning* pada data yang menjadi focus KDD. Proses ini mencakup membuang duplikasi data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data.

3. Transformation

Koding adalah salah satu proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining.

4. Data mining

Proses yang telah berhasil akan menghasilkan pola atau informasi dari data yang terpilih dengan menggunakan teknik tertentu.

5. Interpretation/Evaluation

Pola yang dihasilkan dari proses *data mining* akan menghasilkan data yang dapat dimengerti oleh orang lain, bisa dalam bentuk grafik ataupun bentuk angka.

2.3 Naïve Bayes Classifier

Sebuah metode klasifikasi yang menggunakan metode seperti statistic dan probabilitas, metode ini pertama kali dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes. Dengan menggunakan metode ini dapat memprediksi peluang yang akan terjadi di masa depan berdasarkan dari pengalaman yang telah dipelajari sebelumnya. Ciri yang paling mudah dilihat dari metode ini adalah adanya asumsi yang sangat kuat dari masing-masing kondisi yang ada [6], [15], [16].

Menggunakan metode ini memiliki beberapa keuntungan jika dibandingkan dengan metode lainnya, seperti metode ini hanya membutuhkan jumlah data pelatihan yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Metode *naïve bayes classifier* sendiri memiliki beberapa tahapan yang harus dilakukan seperti pada gambar 2.5 [8], [17], [18]:

```

INPUT: DataSet
OUTPUT: TrainingSet, TestSet
1 BEGIN
2 {
3   SET trainingSetSize = DataSet.Count / 2
4   SET testSetSize = DataSet.Count - trainingSetSize
5   SET TrainingSet = DataSet.Take(trainingSetSize)
6   SET TestSet = DataSet.Skip(trainingSetSize).Take(testSetSize)
7 }
8 END

```

Gambar 2.5 Pseudocode Training Set Naive Bayes Classifier

Pada gambar 2.5 ditampilkan tahapan-tahapan untuk *pseudocode training set* NBC dimulai dari tahapan *input* sampai tahapan *output*.

2.3.1 Teorema Bayes

Teorema Bayes atau yang biasa dikenal dengan kaidah Bayes, merupakan salah satu bagian terpenting dalam penerapan probabilitas bersyarat. Teori ini sendiri pertama kali diperkenalkan dan dikembangkan oleh Thomas bayes pada tahun 1702-1763[19]. Kaidah Bayes sendiri merupakan kaidah yang berfungsi untuk memperbaiki suatu probabilitas dengan cara memanfaatkan informasi tambahan. Teori lain mengatakan bahwa teori Bayes merupakan kesimpulan statistik yang membuktikan atau pengamatan yang digunakan untuk memperbarui suatu probabilitas yang mungkin benar[20], [21].

2.3.2 HMAP (Hypothesis Maximum Appropri Probability)

HMAP merupakan hasil dari hipotesa yang diambil berdasarkan nilai probabilitas berdasarkan kondisi prior yang telah diketahui. HMAP sendiri adalah penyederhanaan dari metode bayes yang disebut dengan Naïve Bayes[22], [23].

$$\begin{aligned}
 P(S|X) &= \operatorname{argmax}_{x \in X} \frac{P(Y|X)P(X)}{P(X)} \\
 &= \operatorname{argmax}_{x \in X} P(Y|X)P(X)
 \end{aligned}
 \tag{2.1}$$

Berikut ini adalah contoh kasus sederhana yang diselesaikan dengan HMAP atau *Naïve Bayes*[22], [23]:

1. Dari hasil sebuah survey yang dilakukan oleh sebuah lembaga kesehatan menyatakan bahwa 30% penduduk dunia menderita sakit paru-paru. Dari 90% penduduk yang sakit paruparu ini 60% adalah perokok dan dari penduduk yang tidak menderita sakit paru-paru 20% perokok.[5]

Dari fakta di atas bisa didefinisikan menjadi $X = \text{sakit paru-paru}$ dan $Y = \text{perokok}$ menggunakan rumus 2.1.

Maka:

$$P(X) = 0.9, P(\sim X) = 0.1$$

$$P(Y|X) = 0.6, P(\sim Y|X) = 0.4$$

$$P(Y|\sim X) = 0.2, P(\sim Y|\sim X) = 0.8$$

Dengan menggunakan metode bayes dapat dihitung:

$$P(\{Y\}|X) = P(Y|X) \cdot P(X) = (0.6) \cdot (0.9) = 0.54$$

$$P(\{Y\}|\sim X) = P(Y|\sim X) \cdot P(\sim X) = (0.2) \cdot (0.1) = 0.02$$

Bila diketahui seseorang adalah perokok maka dia menderita sakit paru-paru karena $P(\{Y\}|X)$ lebih besar dari pada $P(\{Y\}|\sim X)$. HMAP atau *naïve bayes* dapat diartikan mencari nilai probabilitas terbesar dari semua *instance* pada *attribute* target atau dari semua kemungkinan keputusan mengenai kasus yang terjadi.

2.4 UML

UML merupakan bahasa yang menspesifikasi, memvisualisasi, membangun dan mendokumentasikan *artifacts* (bagian dari informasi yang digunakan untuk dihasilkan oleh proses pembuatan perangkat lunak, *artifact* tersebut berupa model, deskripsi atau perangkat lunak) dari sistem perangkat lunak, seperti pada pemodelan bisnis dan sistem non perangkat lunak lainnya[24]. UML juga adalah

bahasa pemodelan yang menggunakan konsep orientasi object (OOP). UML menyediakan notasi-notasi yang membantu dalam memodelkan sistem dari berbagai prespektif[24].



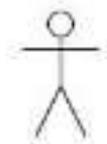
Gambar 2.6 Logo UML

2.5 Use Case Diagram

Use Case Diagram adalah gambaran graphical dari beberapa atau semua aktor, use case, dan interaksi diantaranya yang memperkenalkan suatu sistem. Use case diagram tidak menjelaskan secara detail tentang penggunaan use case tetapi hanya memberi gambaran singkat tentang hubungan antara use case, actor dan sistem[25].

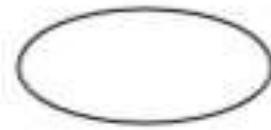
Element-element pada Use Case Diagram[25]:

1. Aktor: Mempresentasikan seseorang atau sesuatu (seperti perangkat, sistem lain) yang berinteraksi dengan sistem. Aktor hanya berinteraksi dengan use case tetapi tidak memiliki control atas use case.



Gambar 2.7 Simbol Aktor

2. Use Case: Adalah gambaran fungsionalitas dari suatu sistem, sehingga costumer atau pengguna sistem paham dan mengerti mengenai kegunaan sistem yang akan dibangun.



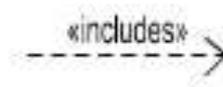
Gambar 2.8 Simbol Use Case

3. Association: Berfungsi untuk menghubungkan link antar element. [7]



Gambar 2.9 Simbol Association

4. <<Include>>: Yaitu kelakuan yang harus terpenuhi agar sebuah event dapat terjadi, dimana pada kondisi ini sebuah use case adalah bagian dari use case lainnya.



Gambar 2.10 Simbol Include

2.6 Sequence Diagram

Sequence Diagram adalah salah satu dari diagram-diagram yang ada pada UML, sequence diagram ini adalah diagram yang menggambarkan kolaborasi yang dinamis antara sejumlah objek. Fungsinya untuk menunjukkan rangkaian pesan yang dikirim antara objek juga interaksi antara objek[26].

Berikut adalah komponen-komponen yang ada pada sequence diagram[26]:

1. Objek: Adalah komponen berbentuk kotak yang mewakili sebuah *class* atau objek. Mereka mendemonstrasikan bagaimana sebuah objek berperilaku pada sebuah sistem.
2. Activation Boxes: Adalah komponen yang terbentuk persegi panjang yang menggambarkan waktu yang diperlukan sebuah objek untuk

menyelesaikan tugas. Lebih lama waktu yang diperlukan, maka *activation boxes* akan lebih panjang.

3. Aktor: adalah komponen yang berbentuk *stick figure*. Komponen yang mewakili seseorang pengguna yang berinteraksi dengan sistem.
4. Lifeline: adalah komponen yang berbentuk garis putus-putus. Lifeline biasanya memuat kotak yang berisi nama dari sebuah objek. Berfungsi menggambarkan aktifitas dari objek.

2.7 Class Diagram

Sebuah spesifikasi yang jika *diinstansiasi* akan menghasilkan sebuah objek dan merupakan inti dari pengembangan dan desain berorientasi objek. *Class* menggambarkan keasaan atribut/property) suatu sistem, sekaligus menawarkan layanan untuk memanipulasi keadaan tersebut[27].

Sebuah *class* memiliki tiga area pokok[27]:

1. Nama: Merupakan nama dari sebuah kelas.
2. Atribut: Merupakan property dari sebuah kelas. Atribut melambangkan batas nilai yang mungkin ada pada objek dari *class*.
3. Operasi: adalah sesuatu yang bisa dilakukan oleh sebuah *class* atau yang dapat dilakukan oleh *class* lain terhadap sebuah *class*.

2.8 Collaboration Diagram

Collaboration diagram merupakan salah satu bentuk alternatif dalam menggambarkan suatu skenario dari suatu sistem. *Collaboration diagram* juga dapat menggambarkan interaksi yang terjadi antar objek seperti pada peran masing-masing objek tersebut[27].

2.9 Entity Relationship Diagram (ERD)

ERD merupakan suatu model untuk menjelaskan hubungan antar data dalam basis data berdasarkan objek-objek data yang sering kita jumpai dalam aktifitas pengembangan sistem oleh sistem analisis[28].

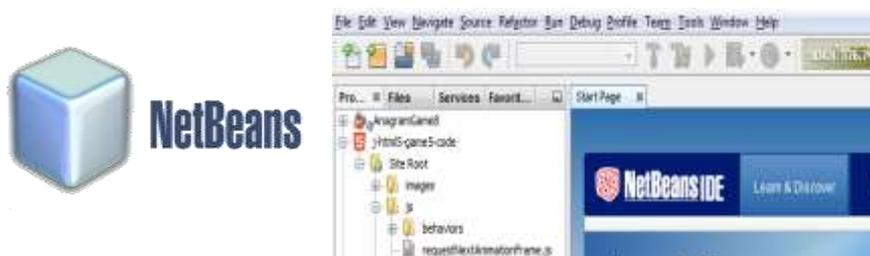
Sedangkan menurut Mata-Toledo dan Cushman (2007:139) Mendefinisikan “Entity Relationship Diagram (ERD) merupakan representasi grafis dari logika database dengan menyertakan deskripsi detail mengenai seluruh entitas (entity), hubungan (relationship), dan batasan (constraint)” [28].

2.10 Software Pendukung

Software pendukung merupakan *tools* yang digunakan dalam membangun sistem atau aplikasi yang akan dibangun. *Software* yang digunakan adalah NetBeans IDE, XAMPP, Microsoft Excel.

2.10.1 NetBeans IDE

NetBeans merupakan salah satu IDE yang dikembangkan dengan menggunakan bahasa java. NetBeans sendiri mempunyai lingkup pemrograman yang sudah terintegrasi dalam suatu perangkat lunak yang didalamnya menyediakan pembangunan pemrograman GUI, *text editor*, *compiler* dan *interpreter*. NetBeans adalah sebuah perangkat lunak *open source* sehingga dapat digunakan secara gratis untuk keperluan komersial maupun nonkomersial yang didukung oleh *Sun Microsystem*[29].



Gambar 2.11 NetBeans IDE

2.10.2 XAMPP

MySQL adalah sebuah program database server yang mampu menerima dan mengirimkan datanya dengan sangat cepat, multi user serta menggunakan perintah standar SQL (Structured Query Language). MySQL juga telah mendukung bahasa

pemrograman berfitur API seperti Java sehingga memudahkan para programmer java untuk berkoneksi dengan menggunakan MySQL[30].



Gambar 2.12 XAMPP

2.10.3 Microsoft Excel

Pengertian Microsoft Excel adalah sebuah program atau aplikasi yang merupakan bagian dari paket instalasi Microsoft Office, berfungsi untuk mengolah angka menggunakan spreadsheet yang terdiri dari baris dan kolom untuk mengeksekusi perintah. Microsoft Excel telah menjadi software pengolah data / angka terbaik di dunia, selain itu Microsoft Excel telah didistribusikan secara multi-platform. Microsoft Excel tidak hanya tersedia dalam platform Windows, Microsoft Excel juga tersedia di MacOS, Android dan Apple[31].

Microsoft Excel secara fundamental menggunakan spreadsheet untuk manajemen data serta melakukan fungsi-fungsi Excel yang lebih dikenal dengan formula Excel. Excel merupakan program spreadsheet elektronik. Spreadsheet adalah kumpulan dari Sel yang terdiri atas baris dan kolom tempat anda memasukkan angka pada Microsoft Excel. Jumlah Sel Microsoft Excel 2016 terdiri dari 1.048.576 Baris dan 16.384 Kolom atau 17.179.869.184 Sel[31].